

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA ELÉCTRICA Y ELECTRÓNICA

IDENTIFICACIÓN DE PÉRDIDAS NO TÉCNICAS DE ENERGÍA ELÉCTRICA MEDIANTE LA COMBINACIÓN DE UN CLASIFICADOR DE SVM (SUPPORT VECTOR MACHINE) Y UN ESTIMADOR DE ESTADO

**TESIS DE GRADO PREVIO A LA OBTENCIÓN DEL TÍTULO DE MAGÍSTER EN
INGENIERÍA ELÉCTRICA EN DISTRIBUCIÓN**

BRAULIO DAVID MUYULEMA MASAQUIZA

braulio.muyulema@epn.edu.ec

DIRECTOR: DR. PAÚL FABRICIO VÁSQUEZ MIRANDA

paul.vasquez@epn.edu.ec

Quito, Enero 2023

AVAL

Certifico que el presente trabajo fue desarrollado por Braulio David Muyulema Masaquiza, bajo mi supervisión.

DR. PAÚL FABRICIO VÁSQUEZ MIRANDA
DIRECTOR DEL TRABAJO DE TITULACIÓN

DECLARACIÓN DE AUTORÍA

Yo, Braulio David Muyulema Masaquiza, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentada para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedo mis derechos de propiedad intelectual correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

BRAULIO DAVID MUYULEMA
MASAQUIZA

DEDICATORIA

Este trabajo va dedicado de manera especial a Dios, a mi amada esposa, mis hijos, a mis padres, a mis hermanos, mis suegros, a mis seres queridos que ya no se encuentran conmigo en especial mi abuelita Virginia que desde el cielo me sigue bendiciendo, a la Virgen de Agua Santa, a la Virgen del Quinche, a San Isidro Labrador Patrono de mi Parroquia (Picaihua-Ambato-Tungurahua), Gracias por todas sus bendiciones y su ayuda Espiritual.

AGRADECIMIENTO

A DIOS: Por darme la salud necesaria para culminar con mis estudios profesionales.

A MI AMADA ESPOSA: Lorena. Dios y la vida nos pusieron en el mismo destino y juntos caminaremos hasta el final, gracias por su confianza por creer en mí, porque cada logro suyo o mío ahora es de los dos.

A MIS HIJOS: Pablito y Felipito Dios me dio oportunidad de ser padres de dos hermosos niños y que ahora son fuente de mi inspiración, por cada esfuerzo en la vida es para ellos.

MIS PADRES: Raúl y Clementina por brindarme y seguirme apoyando, por sus consejos y bendiciones, ya que sin su apoyo no hubiese sido posible cumplir esta meta.

A MIS HERMANOS: Darío, Paúl y Ligia gracias por su apoyo incondicional y sus consejos, por estar siempre pendiente mí en las buenas y las malas.

A MIS SOBRINOS: Julián y Thomas gracias por su cariño.

A MIS SUEGROS: Melida y Jorge Gracias por ser un soporte fundamental para mi hogar, por ayudar a cuidar y educar a mis hijos sin su ayuda no hubiese sido posible culminar este proyecto.

A MI DIRECTOR: Dr. Paul Vásquez por haber aceptado ser mi director de proyecto y transmitir sus conocimientos como profesional para la elaboración de este proyecto, mostrando su calidad como docente y persona.

A LA EMPRESA ELECTRICA RIOBAMBA S.A: Por haberme permitido desarrollar mi proyecto en sus instalaciones y todos los miembros que conforman esta gran institución.

A Luis Suntaxi: La vida me permitió elegir un hermano y así te considero gracias por estar en los momentos más felices y tristes de mi vida y por apoyarme en este proyecto académico.

A MIS AMIGOS: Rafita, Javier, Cesar y demás amigos que he conocido en el transcurso de la Maestría, Gracias

ÍNDICE DE CONTENIDO

AVAL I	
DECLARACIÓN DE AUTORÍA.....	II
DEDICATORIA	III
AGRADECIMIENTO	IV
ÍNDICE DE CONTENIDO.....	V
ÍNDICE DE FIGURAS	IX
DEFINICIONES	XII
RESUMEN	XIV
ABSTRACT	XV
1 INTRODUCCIÓN	16
1.1 Objetivo General	16
1.2 Objetivos Específicos	16
1.3 Alcance	17
1.4 Planteamiento del Problema	18
1.5 Marco Teórico	20
1.5.1 Pérdidas no técnicas en Distribución	20
1.5.2 Clasificación de pérdidas no técnicas	20
1.5.2.1 Robo.....	20
1.5.2.2 Fraude.....	20
1.5.2.3 Problemas de Facturación.....	21
1.5.2.4 Problemas de Recaudación	21
1.5.3 Métodos de Detección	21
1.5.3.1 Método Indirecto	21
1.5.3.2 Método Directo.....	22
1.5.4 Métodos que utilizan datos históricos.	22
1.5.5 Proceso de minería de datos.....	23
1.5.6 Aprendizaje automático o aprendizaje de máquina	24
1.5.6.1 Aprendizaje supervisado.	25
1.5.6.1.1 Clasificación.....	25
1.5.6.1.2 Regresión.....	26
1.5.6.2 Aprendizaje no supervisado.	27
1.5.6.2.1 Agrupamiento	27

1.5.7	Técnicas de modelado de aprendizaje automático para la detección de pérdidas no técnicas de energía.	28
1.5.7.1	Redes neuronales artificiales.	28
1.5.7.2	K vecinos más cercanos.	29
1.5.7.3	Lógica difusa (fuzzy logic)	30
1.5.7.4	Bosque de caminos óptimos	31
1.5.7.5	Máquinas de Vector de Soporte (SVM).....	32
1.5.7.5.1	Máquinas de Vector de Soporte (SVM).	34
1.5.7.5.2	Datos linealmente separables	34
1.5.7.5.3	Clasificador de svm lineal: caso separable.....	37
1.5.7.5.4	Límite de decisión de un clasificador de svm lineal.....	37
1.5.7.5.5	El margen de un clasificador de svm lineal.....	38
1.5.7.5.6	Entrenamiento clasificador lineal de svm.....	39
1.5.7.5.7	El clasificador svm lineal: caso inseparable.....	42
1.5.7.5.8	El clasificador svm no lineal	44
1.5.8	Métodos que utilizan datos obtenidos en tiempo real.....	47
1.5.9	Estimación de Estado en Sistemas.....	47
1.5.10	Modelado Trifásico para Sistemas de Distribución.....	50
1.5.10.1	Estructura estándar de los sistemas de distribución.....	50
1.5.10.1	Elemento Genérico SD	50
1.5.11	Estimación de Estados Trifásicos en Sistemas de Distribución	52
1.5.11.1	Introducción.....	52
1.5.11.2	Modelos de Medida.....	53
1.5.11.3	Observabilidad del Plan de Medición	54
1.5.11.4	Redundancia del plan de medición	54
1.5.11.5	Pseudomedidas de inyección.....	55
1.5.11.6	Restricciones de igualdad.....	55
1.5.11.7	Restricciones Estructural	56
1.5.11.8	Restricciones Operacionales.....	57
1.5.11.9	Modelado del nivel de incertidumbre de las restricciones	58
1.5.11.10	Formulación matemática del estimador de estado	59
1.5.11.11	Resolviendo el Problema de Optimización	60
1.5.11.12	Vector de ecuación y matriz jacobiana.....	63
1.5.11.13	Naturaleza trifásica de la información	63
1.5.11.14	Cálculo de Elementos de vector de ecuaciones $\mathbf{h}(\mathbf{x})$ y matriz jacobiana, $\mathbf{H}(\mathbf{x})$,	64
1.5.11.15	Módulo de tensión.....	65

1.5.11.16	Módulo de corriente al cuadrado	66
1.5.11.17	Medidor Aguas Arriba del Elemento.....	68
1.5.11.18	Matriz [B] Singular.....	69
1.5.11.19	Flujo de Potencia	69
1.5.11.20	Medidor aguas abajo del elemento	70
1.5.11.21	Inyección de potencia	71
1.5.11.22	Detección e identificación de errores graves	72
1.5.11.22.1	Pruebas geométricas de colinealidad	73
1.5.11.23	Particularidades de la estimación de estados en redes de distribución.....	73
1.5.11.23.1	Información disponible.....	73
1.5.11.23.2	Medidas en tiempo real	73
1.5.11.23.3	Medidas cuasi en tiempo real (AMI).....	74
1.5.11.23.4	Pseudo medidas	74
1.5.11.23.5	Medidas virtuales.....	75
2	METODOLOGÍA	76
2.1	Estado del arte	76
2.2	Propuesta metodológica	80
2.1.1	Etapa 1 Máquinas de Vector de Soporte (SVM)	82
2.1.2	Procesamiento de datos	82
2.1.3	Adquisición e integración de datos.....	83
2.1.4	Normalización de datos	83
2.1.5	Máquinas de vector de soporte	83
2.1.5.1	MEDIDAS DE RENDIMIENTO DEL CLASIFICADOR.....	84
2.1.5.2	EL MÉTODO HOLDOUT	86
2.1.6	Etapa 2 Estimación de estado del sistema de distribución.....	88
2.1.6.1	Solución por el método del cuadro disperso	90
2.1.7	Cálculo error grueso	91
2.1.7.1	Identificación de Pérdidas No Técnicas	92
3	APLICACIÓN PRÁCTICA DE LA METODOLOGÍA Y ANÁLISIS DE RESULTADOS	96
3.1	Introducción	96
3.2	Aplicación Máquina de Vector de Soporte.....	96
3.2.1	Procesamiento de datos	96
3.2.2	Normalización de datos.....	97
3.1.1	Resultados de las máquinas vectoriales de soporte.....	98
3.1.1.1	Optimización de parámetros.....	101

3.1.1	Aplicación de del SVM para el alimentador 2 de la subestación 14 de la Empresa Eléctrica Riobamba	104
3.2	Aplicación Estimador de estado sistema para la identificación de Pérdidas No Técnicas	105
3.2.1	Sistema de prueba IEEE de 4 nodos	105
3.2.2	Sistema de prueba de 13 nodos IEEE	111
3.2.3	Sistema de prueba de 6 nodos 15 usuarios.....	116
3.2.4	Sistema real de la Empresa Eléctrica Riobamba.....	124
4	CONCLUSIONES.....	130
4.1.1	EL CLASIFICADOR SVM.....	130
4.1.2	Estimador de estado	131
4.1.3	Combinación Clasificador SVM y Estimador de Estado	131
4.1.4	Trabajos Futuros.....	132
5	REFERENCIAS BIBLIOGRÁFICAS	133
6	ANEXOS 137	
6.1.1	Modelado Matricial Generalizado SD.....	137
6.1.1.1	Líneas de distribución	137
6.1.1.2	Modelo exacto de LD.....	137
6.1.1.3	Modelo simplificado LD.....	138
6.1.1.4	LD Bifásico y Monofásico	139
6.1.1.5	Transformadores de Distribución	140
6.1.1.6	Reguladores de voltaje	144
6.1.1.7	Modelado de carga	148
6.1.1.8	<i>Cargas Monofásicas y Bifásicas.....</i>	151
6.1.1.9	Cargas Distribuidas.....	152
6.1.1.10	Flujo de Potencia	152
6.1.1.11	Método de barrido para una red monofásica	153
6.1.1.12	Método de barrido para el modelado trifásico.....	156
6.1.1.13	Barrido hacia adelante.....	156
6.1.1.14	Actualización de corrientes de carga	157
6.1.1.15	Barrido hacia atrás.....	158
6.1.1.16	Comentarios sobre el modelado SD	159
6.1.1.17	Modelo en Por Unidad	159
6.1.1.18	Extinción de Nudos y Aglutinación de Ramas	160
6.1	Consumos de Clientes	160
	ORDEN DE EMPASTADO	1

ÍNDICE DE FIGURAS

Figura 1.1	Proceso de minería de datos [27]	23
Figura 1.2	Esquema del aprendizaje automático.....	25
Figura 1.3	Representación tarea de clasificación	26
Figura 1.4	Representación tarea de regresión	26
Figura 1.5	Representación tarea de agrupamiento o clúster.....	27
Figura 1.6	Estructura de red neuronal artificial.....	29
Figura 1.7	Ejemplo K-vecinos más próximos	30
Figura 1.8	Secuencia de entrenamiento para el clasificador OPF.	31
Figura 1.9	Hiperplano de separación óptimo y su margen asociado.....	33
Figura 1.10	Datos de entrenamiento 2-D que son linealmente separables (Han y Kamber, 2006)	35
Figura 1.11	Hiperplano de margen pequeño	35
Figura 1.12	Hiperplano de margen grande.....	36
Figura 1.13	Límite de decisión y un margen del clasificador SVM para caso separable ..	37
Figura 1.14	Límite de decisión y margen del clasificador SVM para caso inseparable	43
Figura 1.15	Ramal Genérico de la Red de Distribución [20]	51
Figura 1.16	Modelo de cuadrupolo genérico [20]	51
Figura 1.17	Esquema unifilar de un sistema de distribución de media y baja tensión, con medidas disponibles.....	73
Figura 2.1	Flujograma Método propuesto	81
Figura 2.2	Diagrama de flujo de entrenamiento [9].....	84
Figura 2.3	Evaluación del desempeño con el método de retención	87
Figura 2.4	Estimación de estado y prueba de error grueso.....	88
Figura 2.5	Proyección de λ en el espacio de columnas de WS.....	93
Figura 3.1	Gráfico de líneas para el consumo total en kWh por mes.....	97
Figura 3.2	Medidas de rendimiento obtenidas en 500 iteraciones para el clasificador SVM.....	101
Figura 3.3	Medidas de rendimiento de SVM con parámetros optimizados	103
Figura 3.4	Sistema de prueba IEEE de 4 nodos.....	106
Figura 3.5	Sistema de prueba de 13 nodos IEEE	112
Figura 3.6	Sistema de prueba de 6 nodos, 15 usuarios	117
Figura 3.7	Curva de carga residencial urbana Región Costa.....	118
Figura 3.8	Curva de carga comercial urbana Región Costa.....	119
Figura 3.9	Curva de carga industrial urbana Región Costa.....	119
Figura 3.10	Curva de carga residencial Rural Región Costa.....	120

Figura 3.11	Curva de carga comercial Rural Región Costa	120
Figura 3.12	Curva de carga Industrial Rural Región Costa	121
Figura 3.13	Centro de transformación identificado	124
Figura 3.14	Simplificación sistema real Empresa Eléctrica Riobamba.....	127
Figura 6.1	Modelo Exacto de la Línea de Distribución.....	138
Figura 6.2	Modelo Simplificado de la Línea de Distribución	138
Figura 6.3	Configuración de estrella puesta a tierra - Estrella puesta a tierra	141
Figura 6.4	Configuración Delta - Delta	142
Figura 6.5	La regulación de voltaje [20]	145
Figura 6.6	Banco de Reguladores con Conexión Delta Abierta	146
Figura 6.7	Carga conectada en estrella	149
Figura 6.8	Carga conectada en triángulo	150
Figura 6.9	Equivalente Monofásico de un Alimentador Simple.	154
Figura 6.10	Fase sale directo de borne al térmico sin pasar por el medidor.....	163
Figura 6.11	Medidor averiado	163
Figura 6.12	Fase sale directo del borne al térmico sin pasar por el medidor.....	164
Figura 6.13	Fase entra directo al térmico sin pasar por el contador de energía.....	164
Figura 6.14	Fase sale directo de la bornera al térmico sin pasar por el contador de energía.....	165
Figura 6.15	Fase sale directo de la bornera al térmico sin pasar por el contador de energía.....	165
Figura 6.16	Fase con un puente entre la entrada y la salida del medidor.....	166
Figura 6.17	Fase sale directo de la bornera al térmico sin pasar por el contador de energía.....	166
Figura 6.18	Fase con un puente entre la entrada y la salida del medidor.....	167
Figura 6.19	Centro de transformación con mayor proporción de usuarios sospechosos.	168

DEFINICIONES

Acometida: Derivación física para la conexión entre la red eléctrica, propiedad de la Distribuidora, y las instalaciones del consumidor.

Calidad: Atributos técnicos y comerciales inherentes al suministro de energía eléctrica, a los cuales las empresas eléctricas deben someterse para la prestación de este servicio público.

Carga instalada: Suma de las potencias nominales de todos los equipos eléctricos que forman parte de las instalaciones de un consumidor.

Consumidor regulado o consumidor: Persona natural o jurídica que mantiene un contrato de suministro con la empresa eléctrica de distribución y que se beneficia con la prestación del servicio público de energía eléctrica.

Consumo propio o autoconsumo: Demanda de energía eléctrica del autogenerador.

Empresa eléctrica de distribución o Distribuidora: Persona jurídica cuyo título habilitante le faculta realizar las actividades de distribución y comercialización de energía eléctrica y el servicio de alumbrado público general, dentro de su área de servicio.

Lectura o medición: Acción mediante la cual se obtiene el registro del consumo de energía eléctrica y otros parámetros relacionados, desde el equipo de medición del consumidor.

Niveles de voltaje: Se definen los siguientes niveles de voltaje:

Bajo voltaje: voltaje menor igual a 0,6 kV;

Medio voltaje: voltaje mayor a 0,6 y menor igual a 40 kV;

Alto voltaje grupo 1: voltaje mayor a 40 y menor igual a 138 kV; y,

Alto voltaje grupo 2: voltaje mayor a 138 kV.

PNT: Pérdidas No técnicas de Energía

Sistema de distribución: Comprende las líneas de subtransmisión, las subestaciones de distribución, los alimentadores primarios, los transformadores de distribución, las redes secundarias, las acometidas, el equipamiento de compensación, protección, maniobra, medición, control y comunicaciones, utilizados para la prestación del servicio de distribución de energía eléctrica.

Sistema de medición: Son los componentes necesarios para la medición o registro de energía activa, energía reactiva, demandas máximas y otros parámetros relacionados. Incluyen los equipos de medición (medidores), los transformadores de medición (cuando se apliquen), los cables de conexión, los accesorios de sujeción y protección física de los medidores y de los transformadores.

SVM: Máquina Soporte de Vectores

RESUMEN

Las pérdidas de energía eléctrica corresponden a la diferencia entre lo que el sistema nacional interconectado entrega a los consumidores y lo que en realidad se paga por el consumo y está compuesto por las pérdidas técnicas(PT) y no técnicas(PNT). Las primeras ocurren de forma natural en el sistema, debido a las propiedades físicas de los componentes que conforman el mismo. Estas pérdidas consisten principalmente en disipación de energía en líneas de transmisión/distribución, transformadores y equipos de medida. Mientras que las PNT corresponden al hurto de energía y la ineficiencia de los sistemas de control, medición, facturación y recaudación de las empresas eléctricas. En Ecuador al cierre del 2021 las PT en porcentaje alcanzaron un 6,53 %, mientras que las PNT 6,53 % aun existiendo empresa de distribución con pérdidas no técnicas que sobrepasan 12% [1]. Este proyecto propone una metodología para la identificación de pérdidas no técnicas(PNT) de energía eléctrica mediante la combinación de un clasificador basado en Máquina de Vector de Soporte (SVM) y un Estimador de Estado, donde se utiliza el algoritmo de clasificación (SVM) para lo cual se utiliza una base datos real donde se entrena y se prueba el clasificador luego se aplica a un alimentador real identificado centros de transformación con mayor sospecha de PNT al cual se le aplica el Estimador de Estado y prueba de error grueso, teniendo como resultado final una ligera mejora en el porcentaje de aciertos de usuarios identificados con fraude de energía o anomalías en su sistema de distribución.

PALABRAS CLAVE: Pérdidas no Técnicas, Máquina de Vector de Soporte, Estimador de Estado, Prueba de error grueso.

ABSTRACT

Electrical energy losses correspond to the difference between what the national interconnected system delivers to consumers and what is actually paid for consumption and is made up of technical (PT) and non-technical (PNT) losses. The former occur naturally in the system, due to the physical properties of the components that make it up. These losses consist mainly of energy dissipation in transmission/distribution lines, transformers and metering equipment. While the PNT correspond to the theft of energy and the inefficiency of the control, measurement, billing and collection systems of the electricity companies. In Ecuador at the end of 2021, the PT in percentage reached 6.53%, while the PNT 6.53%, even though there is a distribution company with non-technical losses that exceed 12% [1]. This project proposes a methodology for the identification of non-technical losses (NTP) of electrical energy through the combination of a classifier based on Support Vector Machine (SVM) and a State Estimator, where the classification algorithm (SVM) is used. for which a real database is used where the classifier is trained and tested, then it is applied to a real feeder identified transformation centers with greater suspicion of PNT to which the State Estimator and coarse error test are applied, having as The final result was a slight improvement in the percentage of correct answers for users identified as having energy fraud or anomalies in their distribution system.

KEYWORDS: Non-Technical Losses, Support Vector Machine, State Estimator, Gross Error Test.

1 INTRODUCCIÓN

Actualmente, las empresas distribuidoras han presentado grandes cambios en su proceso de distribución de energía, haciendo énfasis en la calidad de energía, lo que implica un control y mitigación de las pérdidas en los sistemas de distribución, las cuales se deben a varios factores que inciden en sus pérdidas técnicas y no técnicas (comerciales), por lo que se están tomando medidas estratégicas que se focalizan en la reducción de las mismas, mediante el desarrollo de varios proyectos que se encuentran en ejecución por parte de las empresas distribuidoras de energía. Estos proyectos han sido desarrollados de forma autónoma por parte de cada una de las empresas de distribución, dependiendo del escenario real que presentan en sus pérdidas técnicas y no técnicas. Las directrices tomadas por parte de las empresas distribuidoras son implementar infraestructuras de medición avanzada (AMI), que principalmente eliminan las lecturas manuales, y registrar efectivamente la energía consumida en el periodo de facturación. Tales proyectos han sido analizados en varias ocasiones y determinado los beneficios económicos que aportarían al sistema, los cuales son previstos para futuras implementaciones, pero debido a las políticas aplicadas y la situación económica del país, no es posible desarrollarlas inmediatamente; por tal motivo, se ha creído conveniente plantear soluciones a corto plazo para las situaciones reales de los sistemas de distribución, lo cual permite el desarrollo de este proyecto

1.1 Objetivo General

Desarrollar una metodología que contribuya a la identificación de pérdidas no técnicas de energía eléctrica mediante la combinación de un clasificador basado en Máquina de Vector de Soporte (SVM) y un Estimador de Estado para un sistema de distribución con redes aéreas.

1.2 Objetivos Específicos

- Plantear un clasificador basado en Máquinas de Vector de Soporte (SVM) capaz de identificar consumidores fraudulentos de energía eléctrica.
- Desarrollar un estimador de estado con resolución de mínimos cuadrados ponderados que sea aplicable para sistemas de distribución desbalanceadas.

- Codificar, tanto el clasificador basado en Máquinas de Vector de Soporte (SVM) como el estimador de estado con resolución de mínimos cuadrados ponderados con ayuda de la herramienta computacional Matlab de tal manera que, esta metodología pueda ser utilizada en sistemas de distribución con redes aéreas.
- Evaluar la metodología en un sistema de prueba conformado por 15 usuarios y analizar los resultados al combinar las dos metodologías.
- Aplicar la metodología a los centros de transformación de un alimentador real que contengan mayor proporción de usuarios sospechosos de fraude de energía.

1.3 Alcance

En el presente proyecto de titulación, se desarrollará una metodología para la identificación de pérdidas no técnicas de energía eléctrica mediante la combinación de un clasificador de SVM (support vector machine) y un estimador de estado. La metodología propuesta consta de dos etapas, en la etapa inicial se plantea clasificar a consumidores sospechosos y no sospechosos de fraude de energía utilizando el algoritmo SVM de una base datos. En la segunda los resultados encontrados por el SVM se les aplica la estimación de estado con resolución de mínimos cuadrados y al resultado de la estimación se le realiza el análisis de error grueso el cual permite corroborar lo indicado por el clasificador SVM en la primera etapa.

Para la activación de la estimación de estado se considerarán mediciones reales para los nodos de referencia con lo correspondiente a voltajes, corrientes, potencia activa y reactiva. Las mediciones se realizarán en campo en coordinación con la distribuidora Para la demanda de los consumidores se utilizará pseudomedidas basadas en pronósticos de carga[19]; además se considerarán voltajes y flujos de potencia activa y reactiva por nodos y ramales monofásicos y bifásicos. donde su resolución se lo hará mediante el algoritmo de mínimos cuadrados ponderados, debido a su ventajosa característica estadística y a la facilidad de su implementación computacional. [18]

Para evaluar la metodología propuesta, se aplicará a un sistema de prueba conformada para 15 usuarios contenidos en una red radial donde para el clasificador basado en Máquinas de Vector de Soporte (SVM) se utilizará una base de datos real para su fase entrenamiento y prueba, la base de datos de consumo de energía será de un periodo de 24 meses.

Para la aplicación de la metodología en sistema real de distribución se plantea clasificar a consumidores sospechosos y no sospechosos de fraude de energía utilizando el algoritmo SVM de una base datos consumo de energía de un alimentador real para luego los resultados encontrados por el SVM aplicar la estimación de estado con resolución de mínimos cuadrados pero debido a la poca redundancia que se tiene en los sistemas de distribución y para aliviar la carga computacional, los centros de transformación del alimentador con mayor proporción de consumidores sospechosos de fraude encontradas por el SVM se les aplica la estimación de estado con resolución de mínimos cuadrados.

Para el modelo de pronóstico de carga se utilizará el estudio realizado (J.A. Jardini 2000). Donde esta metodología utiliza las curvas diarias representativas de la actividad de cada consumidor y el consumo mensual de energía de los consumidores conectados [19].

Debido a que, en una línea aérea de distribución, la admitancia paralela es muy pequeña, esta se la puede despreciar de los cálculos de flujos por líneas, [20]. Razón por la cual no será considerado en el desarrollo de la metodología.

1.4 Planteamiento del Problema

Las pérdidas energía eléctrica corresponden a la diferencia entre lo que el sistema nacional interconectado entrega a los consumidores y lo que en realidad se paga por el consumo y está compuesto por las pérdidas técnicas y no técnicas. Las primeras ocurren de forma natural en el sistema, debido a las propiedades físicas de los componentes que conforman el mismo. Estas pérdidas consisten principalmente en disipación de energía en líneas de transmisión/distribución, transformadores y equipos de medida. Mientras que las pérdidas no técnicas corresponden al hurto de energía y la ineficiencia de los sistemas de control, medición, facturación y recaudación de las empresas eléctricas. En el Ecuador al cierre del año 2021 la energía disponible en los sistemas de distribución (26.450,77 GWh), el 86,94 % (22.996,15 GWh) correspondió a la demanda regulada, demanda no regulada y otras ventas (energía entregada a usuarios ubicados en las fronteras de países vecinos, servidos mediante redes de distribución); y, el 13,06 % (3.454,62 GWh) a las pérdidas de energía en distribución. Las pérdidas técnicas alcanzaron un valor de 1.726,47 GWh (6,53 %), mientras que las pérdidas no técnicas fueron 1.728,14 GWh (6,53 %). [1]

Desde el punto de vista económico, la empresa de distribución encargada de la comercializadora de energía necesita medir los consumos, para así mismo cobrar al suscriptor y remunerar su actividad de comercialización. Actualmente el sistema de medición de la mayoría de las empresas de Distribución del país requiere que de forma manual se tome la lectura del medidor, estos valores visualizados son digitalizados por otro funcionario y se almacenan en la base de datos de la empresa. Este sistema de medición es muy vulnerable a cualquier tipo de error humano, ya que emplea procesos manuales tanto para la toma de lectura del medidor como para la digitalización; además, su tiempo de respuesta depende del tiempo que tarde en llegar la información a la base de datos.

Las Empresas Eléctricas de Distribución Nacional trabajan en la implementación de nuevas acciones para la reducción de pérdidas eléctricas a fin de alcanzar la meta planteada en el PME (Plan Maestro de la Electrificación) 2016–2025 de 8,79% al final del período.[2] La continuidad de las acciones a futuro tiene prioridad en aquellas Empresa Eléctricas Distribución que aún presentan espacio de mejora en pérdidas no técnicas. En el país los índices de pérdidas deben tener la meta de bajar para poder relacionarse con países industrializados, en los cuales sus índices de pérdidas totales son menores al 6%, con pérdidas no técnicas cercanas a cero (0%).[3]

Por lo anteriormente expuesto, la estrategia de las Empresas Eléctricas de Distribución Nacional establecen una línea de acción que tienen el propósito de impulsar la reducción de pérdidas de energía y éstas se refieren a la identificación y ejecución de la inversión necesaria que coadyuve a identificar y reducir las pérdidas técnicas y no técnicas costo-eficientes mediante la modernización de la red distribución, así como la incorporación gradual de tecnologías avanzadas para administrar la demanda de energía eléctrica, de tal forma que contribuyan a abatir las pérdidas en el proceso de distribución, considerando que con ello se procurará el mejoramiento de las finanzas de la empresa de electricidad de una manera socialmente responsable. [4]

Con base a lo expuesto se plantea colaborar en la solución del problema a través del desarrollo de una metodología que permita identificar las pérdidas no técnicas en consumidores sospechosos que toman la energía eléctrica de manera ilegal del sistema de distribución o tienen alguna anomalía en su sistema de medición para posteriormente tomar medidas correctivas que permitan reducir las pérdidas no técnicas de energía eléctrica de manera considerable.

1.5 Marco Teórico

1.5.1 Pérdidas no técnicas en Distribución

Las pérdidas no técnicas es la diferencia entre las pérdidas globales y pérdidas técnicas.

Las pérdidas no técnicas, a su vez, son causadas principalmente por el robo de energía, además de otros problemas que generan una falta de ingresos, tales como:

- Errores de lectura del medidor, que pueden ser causados intencionalmente o no;
- Consumidores clandestinos, no registrados en el sistema del distribuidor;
- Fallos en la actualización de bases de datos y registros.

1.5.2 Clasificación de pérdidas no técnicas

Las pérdidas no técnicas pueden ocurrir de 4 maneras diferentes, con intensidades que varían de acuerdo con factores que van desde lo cultural a lo técnico [5].

1.5.2.1 Robo

Caracterizado por la desviación directa de energía de las redes eléctricas por el consumidor clandestino, por lo tanto, la energía utilizada por este consumidor no se tiene en cuenta. Estas conexiones ilegales generalmente se realizan en el alimentador de bajo voltaje o en el transformador de servicio, en cuyo caso las conexiones están expuestas, lo que hace posible la identificación visual [5]. En Ecuador, este tipo de irregularidad a menudo ocurre en áreas de riesgo, lo que hace que la inspección y la lucha contra el robo de energía sean más complejas, y puede haber problemas de seguridad para los técnicos del distribuidor.

1.5.2.2 Fraude

Cuando el consumidor está registrado por la distribuidora, pero realiza cambios que afectan las marcas de su medidor de energía, estos cambios pueden ser rústicos, como causar daños intencionados al medidor o el simple uso de imanes para evitar la rotación del disco por interferencia magnética, en el caso de medidores analógicos, o más sofisticado, como evitar el medidor, hacer que registre un consumo cero e incluso tener

cargas en paralelo con el medidor, lo que hace que registre un consumo menor que el real.

1.5.2.3 Problemas de Facturación

Estos casos pueden ocurrir debido a varios factores, pero consisten principalmente en errores de lectura del medidor. Estos errores de facturación pueden ser involuntarios debido al mal estado o al posicionamiento de los medidores, lo que dificulta la lectura o incluso a los problemas de administración y sistemas de la empresa que causan errores de facturación. Sin embargo, hay casos de favoritismo intencional por parte del empleado del distribuidor, por parentesco o amistad con el consumidor y, en casos donde la factura de energía es alta, problemas como el pago de sobornos a los empleados para que registren menor que el valor real [25].

La metodología propuesta en este trabajo de titulación trata de identificar a los problemas que se produce por fraude de energía

1.5.2.4 Problemas de Recaudación

Estos son los casos que corresponden al impago de facturas por parte de los consumidores. Este tipo de pérdida es conocida por la compañía, pero aún trae grandes pérdidas al distribuidor [25].

1.5.3 Métodos de Detección

Las técnicas para el cálculo de pérdidas no técnicas de energía eléctrica se pueden dividir en dos clases principales: indirecta y directa. La técnica indirecta trata de estimar las pérdidas técnicas y obtener las pérdidas no técnicas a través de la diferencia entre las pérdidas totales y las pérdidas técnicas estimadas, mientras que las técnicas directas buscan detectar directamente las pérdidas no técnicas. [5]

Los métodos directos pueden utilizar los datos históricos para definir las normas de consumo y detectar comportamientos anómalos, o el uso de datos en tiempo real para detectar la ocurrencia de pérdidas no técnicas en el sistema. [5]

1.5.3.1 Método Indirecto

La estimación de pérdidas técnicas permite, además de obtener pérdidas no técnicas por la diferencia entre pérdidas globales y pérdidas técnicas, obtener la eficiencia del sistema y detectar las necesidades de mejoras. La precisión de la estimación de pérdidas técnicas depende de la información disponible de la red y, por lo tanto, puede suceder en situaciones de:

- Alto grado de conocimiento: en este caso hay mucha información sobre la red y la carga que es posible calcular las pérdidas técnicas con buena precisión.
- Bajo nivel de conocimiento: cuando no hay mucha información, las medidas disponibles se utilizan para la estimación, generalmente subestación y algunos dispositivos de red, además de las comparaciones con sistemas similares.
- Caso híbrido: cuando existe un alto grado de conocimiento de solo una parte de la red, se utilizan otras metodologías.

1.5.3.2 Método Directo

Los métodos directos pueden usar datos históricos para definir patrones de consumo y detectar comportamientos anómalos, o usar datos en tiempo real o casi en tiempo real para detectar la ocurrencia de pérdidas no técnicas en el sistema.

Dentro de los métodos Directos podemos citar los siguientes:

1.5.4 Métodos que utilizan datos históricos.

Los métodos que usan datos históricos tienen una estructura similar, divididos en tres etapas: preparación, clasificación e investigación.

La minería de datos más conocida como Big Data ó exploración de datos es el campo de la estadística y las ciencias informáticas donde su proceso intenta descubrir patrones en grandes bases de datos, lo que ha permitido el desarrollo de diversos algoritmos que abarcan distintas técnicas de aprendizaje: supervisados y no supervisados, las cuales incluyen tareas de agrupamiento, clasificación y regresión. A partir de la información almacenada en base de datos se puede obtener patrones de consumo, que reflejan el comportamiento de los consumidores.[26] Los patrones se pueden agrupar para crear perfiles o segmentos de consumo que podrían utilizarse en el pronóstico y control de

carga, también nos permite identificar de manera automática irregularidades en el consumo y la detección de las Pérdidas no técnicas.

1.5.5 Proceso de minería de datos.

Se define como un conjunto de técnicas que permiten explorar de manera automática o semiautomática grandes bases de datos, es la fase más importante del proceso de descubrimiento de conocimiento en base de datos, más conocido como (Knowledge Discovery in Databases) el proceso comprende cuatro etapas recopilación; preparación; Data Mining; interpretación y evaluación. .[27]

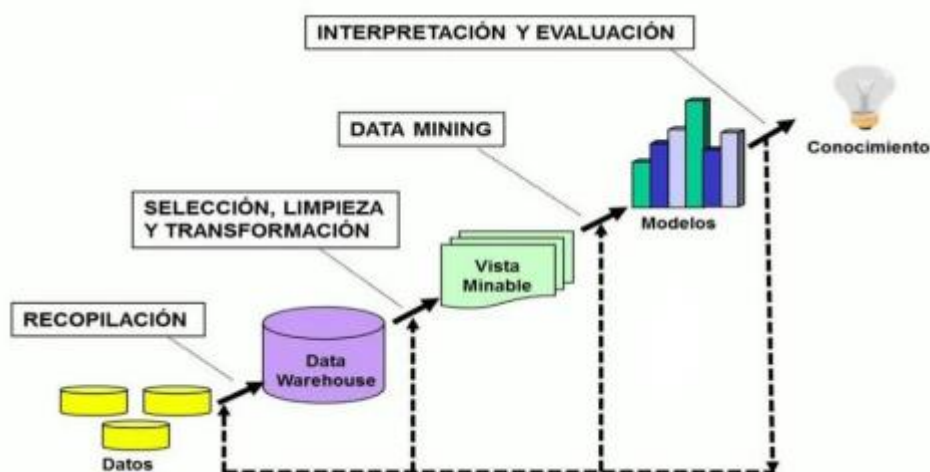


Figura 1.1 Proceso de minería de datos [27]

1. Recopilación de datos: en esta fase se toman los datos que se desean analizar procedentes de diferentes fuentes y se integran en un mismo y único repositorio de datos, denominado almacén de datos, más conocido como data warehouse. Es una tecnología diseñada especialmente para organizar grandes volúmenes de datos de procedencia generalmente estructurada. [27]

2. Preparación de los datos: toma los datos almacenados en data warehouse y los transforma en un subconjunto de datos consistente para extraer el conocimiento implícito en ellos. [27]

Esta etapa consta de varias partes:

Selección: el objetivo es seleccionar el subconjunto de datos, diferenciando las variables objetivo (variables a predecir, calcular o inferir) y las variables independientes (las cuales sirven para ayudar en el proceso de las variables objetivo), para esto se aplican técnicas de muestreo adecuadas. [27]

Limpieza de datos: se tratan todos los datos que puedan influir en un análisis inexacto y resultados incorrectos. Identificación y tratamiento de valores atípicos (outliers), datos erróneos e irrelevantes, la existencia de datos incompletos. Los valores atípicos son observaciones aisladas cuyo comportamiento se diferencia claramente del comportamiento medio del resto de las observaciones. [27]

Transformación: preparación de los datos de entrada acorde a la técnica de minería de datos que se aplicará, en esta fase se aplican técnicas de transformación como son: reducción o aumento de la dimensión, desratización o numeración, normalización de rango, escalado. [27]

3. Minería de datos: obtención de un modelo en el que están implícitos los patrones de comportamiento observados. Se distinguen entre técnicas predictivas, las cuales se utilizan para predecir el valor desconocido de uno o varios atributos para uno o varios registros y técnicas descriptivas, generan modelos que, de alguna forma, describen los datos. [27]

4. Interpretación y evaluación: consiste en evaluar la calidad de los modelos y realizar una interpretación de los mismos para obtener el conocimiento buscado, para validar los resultados se utilizan intervalos de confianza, bootstrap, análisis ROC y evaluación de modelos. [27]

1.5.6 Aprendizaje automático o aprendizaje de máquina.

El aprendizaje automático es una rama de la inteligencia artificial que se enfoca en el diseño y desarrollo de algoritmos (secuencias de instrucciones para solucionar problemas específicos), da a los ordenadores la capacidad de aprender de los datos sin ser programados explícitamente y mejorar su desempeño en el tiempo. En otras palabras, emplean métodos computacionales y estadísticas para aprender patrones a partir de ejemplos con el fin de tomar decisiones para futuros datos invisibles. Por lo tanto, se puede decir que el aprendizaje automático no sólo se relaciona con la minería de datos, reconocimiento de patrones o estadísticas, sino que también se considera como un subconjunto de las ciencias de la computación. Los algoritmos utilizados en el aprendizaje de máquina se organizan de acuerdo al tipo de datos de entradas

(etiquetados o no etiquetados) o del resultado deseado. Los algoritmos que usan datos etiquetados se conocen como aprendizaje supervisados o técnicas predictivas, y aquellos que utilizan datos no etiquetados se denominan aprendizaje no supervisado o técnicas descriptivas. En la figura 1.2 se muestra un esquema del aprendizaje automático.

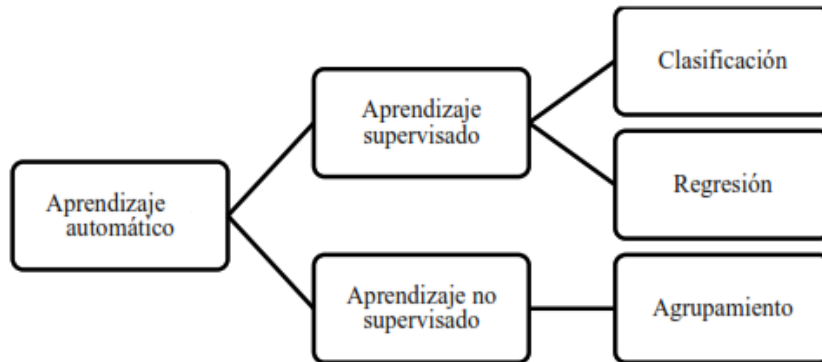


Figura 1.2 Esquema del aprendizaje automático

1.5.6.1 Aprendizaje supervisado.

El aprendizaje supervisado se relaciona con aquellos algoritmos que aprenden patrones a partir de ejemplos de entrenamientos etiquetados, los datos de entrenamientos consisten en pares de observaciones, donde una componente del par son los datos de entrada y el otro la salida o los resultados deseados. La salida puede ser un valor numérico en el caso de problemas de regresión o una etiqueta de clase en caso de problemas clasificación. El propósito de este método es predecir el valor de la función para cualquier muestra de entrada válida a partir de los datos de entrenamiento. En este tipo de aprendizaje los datos utilizados para el entrenamiento y la prueba deben guardar cierta relación con la variable de interés o dependiente para que el modelo que genere pueda asignar los ejemplos de entradas a las salidas deseadas. Sus principales tareas son:

1.5.6.1.1 Clasificación.

Consiste en predecir valores futuros o desconocidos de un atributo cualitativo para un determinado ejemplo. El atributo cuyo valor se desea predecir se denomina atributo de clase y los diferentes valores que puede tomar son las clases del problema [38]. También se conoce como reconocimiento de patrones. En el reconocimiento de patrones, la tarea de clasificación puede ser vista como un problema de dos clases

(binario) o de múltiples clases. Durante el esquema de clasificación, el algoritmo de aprendizaje toma los datos de entrenamiento y selecciona la hipótesis que mejor se ajusta a los datos. La figura 1.3 muestra la representación de un ejemplo de clasificación.

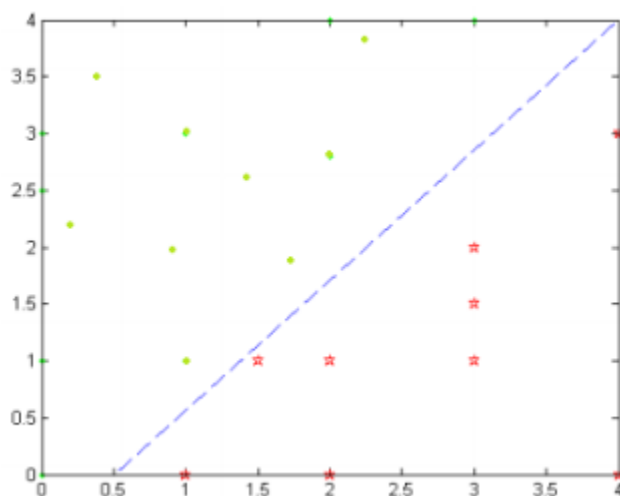


Figura 1.3 Representación tarea de clasificación

1.5.6.1.2 Regresión.

Es similar a la clasificación, la diferencia está, en que el atributo a predecir es de tipo cuantitativo. Estas técnicas buscan la obtención de una función denominada función de regresión, que ha de aproximarse en la mayor medida posible a los ejemplos del conjunto de entrenamiento. Esta función relaciona la variable a predecir (variable dependiente) con las variables utilizada para realizar la predicción (variable independiente), como se muestra en la figura 1.4.

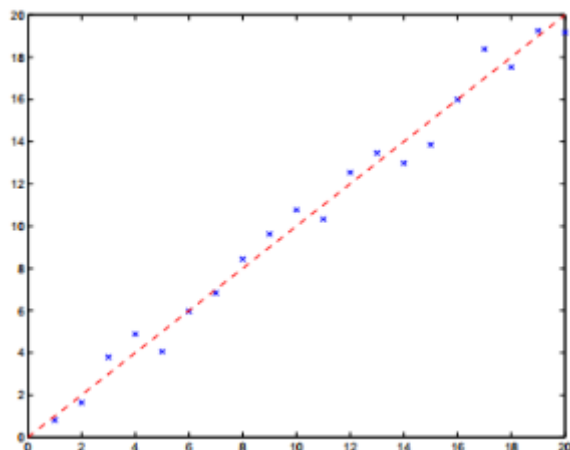


Figura 1.4 Representación tarea de regresión

1.5.6.2 Aprendizaje no supervisado.

El aprendizaje no supervisado utiliza sólo puntos de datos no etiquetados con el fin de encontrar la estructura oculta en los datos. Es decir, los únicos datos que se le proporciona son los patrones de entrada y alguna información explícita o implícita a priori. El objetivo es simplemente extraer la mayor cantidad de información que podamos de las bases de datos disponibles. En contraste con el aprendizaje no supervisado, no existen salidas objetivo explícitas o estimaciones asociados con cada entrada. Este tipo de aprendizaje es útil para tareas de agrupamientos, detección de valores atípicos y reducción de dimensionalidad. [10]

1.5.6.2.1 Agrupamiento

Es el método más popular en el aprendizaje no supervisado. Este consiste en dividir una población heterogénea de objetos en grupos homogéneos, de tal forma que los objetos de cada grupo sean similares o guarden una relación entre ellos y diferentes al resto de objetos de otros grupos [38]. La medida de similaridad está basada en los atributos que describen a los objetos. Para poder establecer los diferentes grupos de objetos similares entre sí, es necesario elegir una función de distancia y calcular con ella la distancia entre los individuos. Las medidas de distancia más empleadas son: distancia Euclídea, distancia Manhattan, distancia Minkowski. [38].

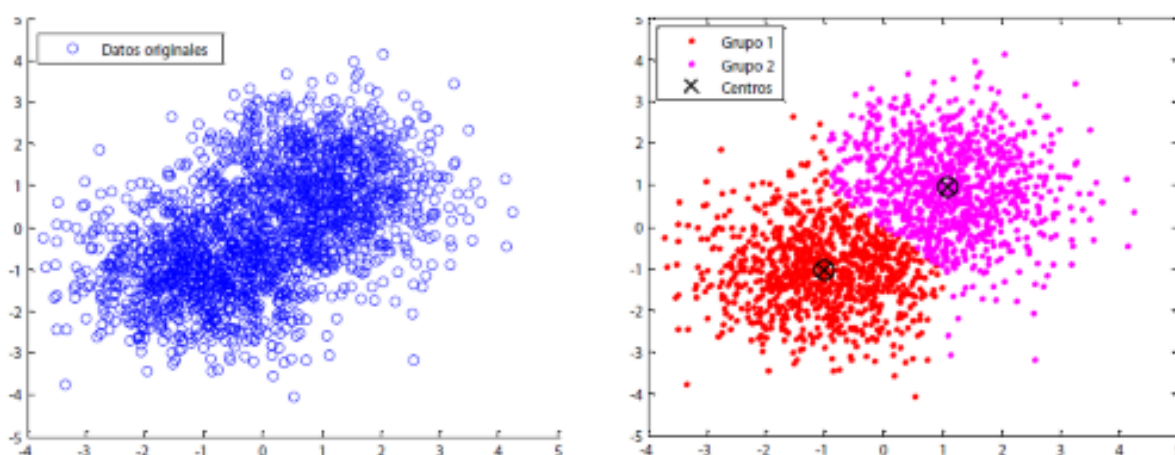


Figura 1.5 Representación tarea de agrupamiento o clúster

La figura 1.5 muestra un ejemplo que divide un conjunto de datos en dos grupos: la parte a) se presentan los datos originales sin separar y la parte b) muestra los datos ya separados en dos grupos definido por los centros.

Los algoritmos de agrupamientos se pueden dividir en métodos no jerárquicos o particionales y métodos jerárquicos. Los primeros están especialmente recomendados ante grandes tablas de datos, y además son útiles para la detección de casos atípicos. Mientras que los jerárquicos debido a la laboriosidad de los cálculos no son muy prácticos en esos casos. [38].

1.5.7 Técnicas de modelado de aprendizaje automático para la detección de pérdidas no técnicas de energía.

El aprendizaje automático también ha permitido grandes avances en el área de la ingeniería eléctrica. En los últimos años se han estado desarrollando diferentes algoritmos para la detección del robo de electricidad con el propósito de reducir las pérdidas no técnicas y a la vez, incrementar los beneficios de las compañías.

Entre los algoritmos utilizados para detectar Pérdida No Técnica de energía, que se encuentran en la literatura, son los que utilizan técnicas de inteligencia artificial para señalar posibles fraudes o defectos en los contadores de energía en los sistemas de distribución. Estos estudios han demostrado que, con la ayuda de programas informáticos con capacidad de aprendizaje, es posible aumentar la posibilidad de encontrar contadores de energía donde hay Pérdida no Técnica de Energía. Las metodologías que utilizan inteligencia artificial pueden detectar automáticamente nuevos patrones, así como buscar patrones ya conocidos por la experiencia humana, mejorando así la tasa de éxito esperada en la inspección en sitio. A continuación, se citan algunas metodologías.

1.5.7.1 Redes neuronales artificiales.

Redes neuronales artificiales esta metodología busca realizar el procesamiento de datos imitando una red neuronal natural, como el cerebro humano. Para la activación de la técnica, se requieren un historial de consumo del cliente, se comprobó que el aprendizaje que esta metodología implementada mejora significativamente la tasa de éxito en la clasificación de clientes identificados como sospechosos. Una de las limitaciones de esta técnica es que la identificación de consumidores irregulares

depende en gran medida de los datos utilizados en la formación, que se obtienen a través de las inspecciones en campo a los contadores de energía, que caracteriza el aprendizaje supervisado. Por lo tanto, las redes neuronales aprenden a identificar solo los patrones de clientes irregulares dentro del grupo de consumidores considerados fraudulentos. Esta metodología puede ser criticada por limitar el aprendizaje del reconocimiento de patrones a una base de entrenamiento que no forma una muestra estadística significativa. Una alternativa sería obtener datos de inspecciones realizadas al azar para obtener una medida estadística más significativa o emplear métodos de aprendizaje no supervisados. Los resultados obtenidos del sistema propuesto no fueron muy satisfactorios, debido a que sólo el 65.03% de los clientes preseleccionados para ser inspeccionados en campo resultaron con fraude [10]

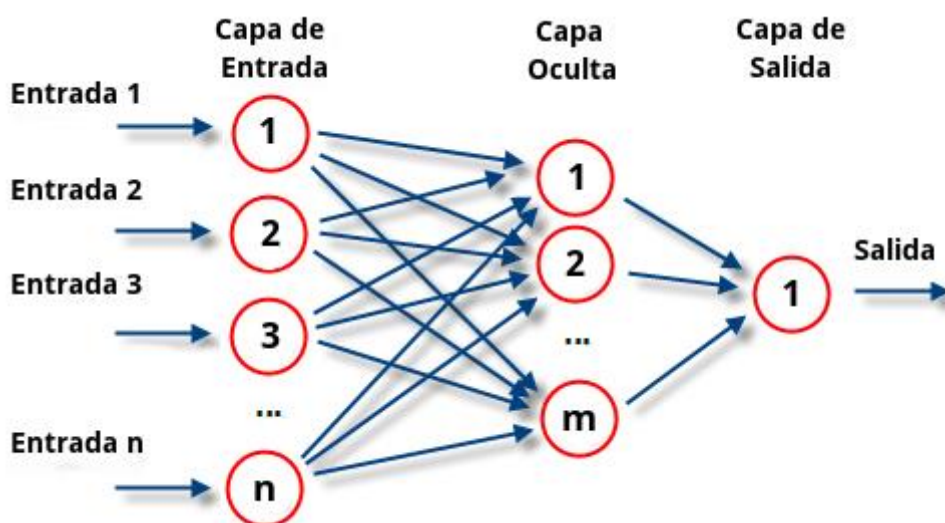


Figura 1.6 Estructura de red neuronal artificial

1.5.7.2 K vecinos más cercanos.

K- medias El k-medias (K-means) es un algoritmo que pertenece a las técnicas de agrupamiento no jerárquico o particionales, y es el más popular dentro de esta categoría. Se trata de un método en el que se parte de un número determinado de prototipos y de un conjunto de elementos a agrupar, sin etiquetar. Este algoritmo elige un punto inicial (centroide) para representar cada uno de los grupos, luego se asigna cada objeto al grupo cuyo centroide es el más próximo a él o tenga características similares. Seguidamente, el centroide de cada grupo es recalculado como la media de los puntos asignados al grupo. Este proceso que se repite hasta que alcance un número limitado

de pasos de iteración o, simplemente, que la diferencia obtenida entre los centroides de dos pasos consecutivos sea menor que un valor prefijado. Los centros se determinan con el objetivo de minimizar la suma de distancias cuadráticas euclidianas entre los patrones de entrada y el centro más cercano. Los datos usados para activar esta técnica corresponden a 5000 perfiles de clientes industriales de una empresa brasileña, donde sólo 280 de ellos presentan pérdidas no técnicas. También se aportan algunas características que consideran relevantes como: demanda facturada, demanda contratada, demanda máxima, entre otras. Los datos se dividieron en 50% para el entrenamiento y el 50% para la prueba. En los resultados obtenidos se observó que los algoritmos mejoran su desempeño cuando se incluyen ciertas características de los clientes. Este enfoque presenta algunas debilidades, una de ellas es que no toma en cuenta el desequilibrio de las clases en el conjunto de entrenamiento. El resultado final del autor indica que se obtuvo una tasa de éxito de 75% [11]

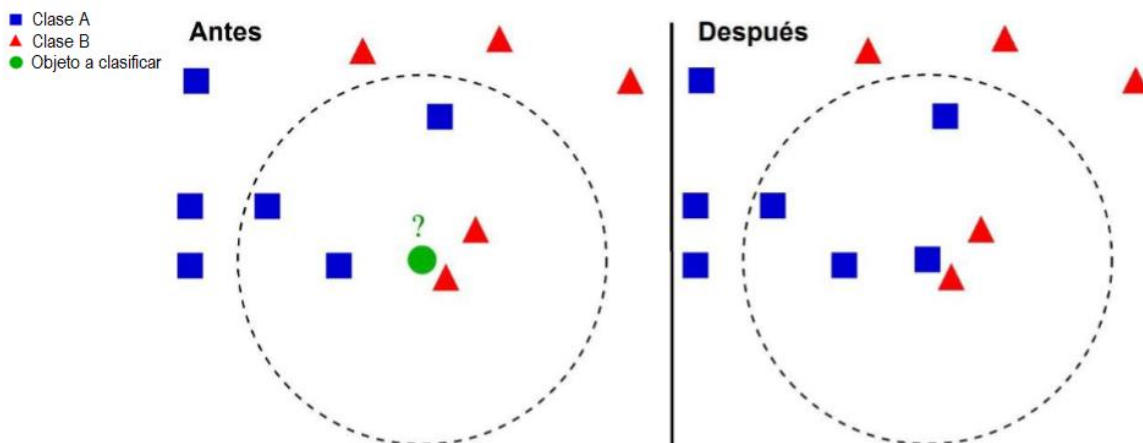


Figura 1.7 Ejemplo K-vecinos más próximos

1.5.7.3 Lógica difusa (fuzzy logic)

La aplicación de los algoritmos de lógica difusa está dentro de la tarea de agrupamiento es una de las más utilizada para determinar perfiles de carga. En la referencia [12] por ejemplo se emplea para identificar perfiles de consumo sospechosos comparándolos con perfiles de consumo regulares. Este esquema se compone de dos pasos: En el primer paso, se utiliza el algoritmo C-medias difuso para agrupar los clientes dentro de las clases correspondientes, es decir, con perfiles similares. Posteriormente, mediante una matriz de pertenencia difusa y la distancia euclídea a los centros de agrupamiento se clasifican los clientes en fraudulentos o no fraudulentos. Para desarrollar el algoritmo

se utilizaron los datos históricos de consumo de los últimos seis meses de 20,126 clientes de un área residencial y cinco atributos para el estudio (promedio de consumo, máximo consumo, desviación estándar, cantidad de inspecciones y promedio de consumo del área residencial). La tasa de éxito obtenida fue de 74.5%. [12]

1.5.7.4 Bosque de caminos óptimos

Bosque de caminos óptimos (Optimum-Path Forest - OPF), se basa en la teoría de grafos y su aplicación inicial fue el reconocimiento de patrones en imágenes, modelado como un problema de partición de grafos. El objetivo es segmentar el gráfico en dos, de modo que las muestras de cada clase de datos queden en sus respectivas particiones. El OPF tiene tres versiones: la versión con entrenamiento no supervisado, la versión con entrenamiento supervisado usando el gráfico k-NN y la versión con entrenamiento supervisado usando el gráfico completo. Un gráfico k-NN es uno en el que cada nodo está conectado por un borde solo a los k nodos de la menor distancia euclidiana, donde k es un número entero arbitrario. Un gráfico completo es aquel en el que todos los nodos están conectados entre sí por un borde, por lo que en este caso todos los nodos del gráfico son adyacentes.

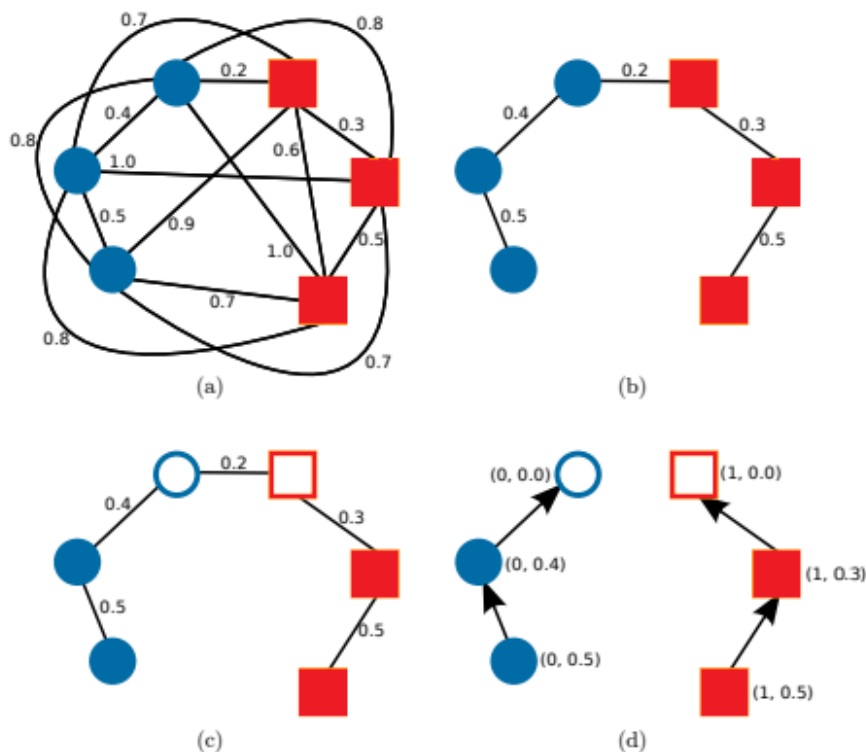


Figura 1.8 Secuencia de entrenamiento para el clasificador OPF.

(a) La gráfica completa con aristas ponderadas por disimilitud. (b) Se encuentra el árbol de expansión mínimo. (c) El siguiente paso es marcar los prototipos y asociar los costos de cada nodo. (d) Con la separación de los prototipos y costos asignados a cada nodo, tenemos un clasificador Optimum Path Forest.

En este trabajo, el problema de identificación de la Pérdidas no técnicas de Energía se trata como un problema de reconocimiento de patrones. Por lo tanto, se asume que, a partir del análisis de los datos disponibles, es posible enseñar a un clasificador a encontrar patrones de Pérdidas no técnicas de Energía de manera eficiente. Los datos para llevar a cabo la investigación fueron proporcionados por una empresa brasileña, los cuales corresponden a 736 perfiles de clientes, divididos en 116 ilegales y 620 legales. Para evaluar los modelos el conjunto de datos se dividió en 50% para la fase de entrenamiento y 50% para la fase de prueba. En los resultados registrados, el modelo obtuvo una precisión de 90.21%. En los resultados registrados se observa que la calidad y el tiempo de aprendizaje del método mejoran mucho cuando se realiza una selección óptima de características. Los resultados indican que el método propuesto mejora el rendimiento de técnicas previamente desarrolladas y es adecuado tanto para Empresa de Distribución típicas como inteligentes. Los mejores resultados se utilizan datos normalizados como entrada, con una tasa de éxito esperada de hasta 72,43% de la inspección en sitio. [13]

1.5.7.5 Máquinas de Vector de Soporte (SVM)

Máquinas de Vector de Soporte (SVM), es una técnica de aprendizaje automático que se ha utilizado con éxito en la detección y localización de Pérdidas No Técnicas de Energía principalmente porque tiene la capacidad de brindar soluciones de gran generalidad a problemas de clasificación de patrones. La clasificación es una función no lineal, cuya entrada son las características de las muestras y la salida es la clase a la que pertenecen. La formulación original de SVM trata con problemas de clasificación binaria, pero puede extenderse a problemas de clasificación con múltiples clases a un alto costo computacional. La SVM puede realizar aprendizaje supervisado o no supervisado. El método de clasificación se utilizó para entrenar y probar el clasificador SVM utilizando los datos empíricos. El autor manifiesta que se obtuvo una tasa aciertos en la identificación de usuarios fraudulentos del 76%. [9]

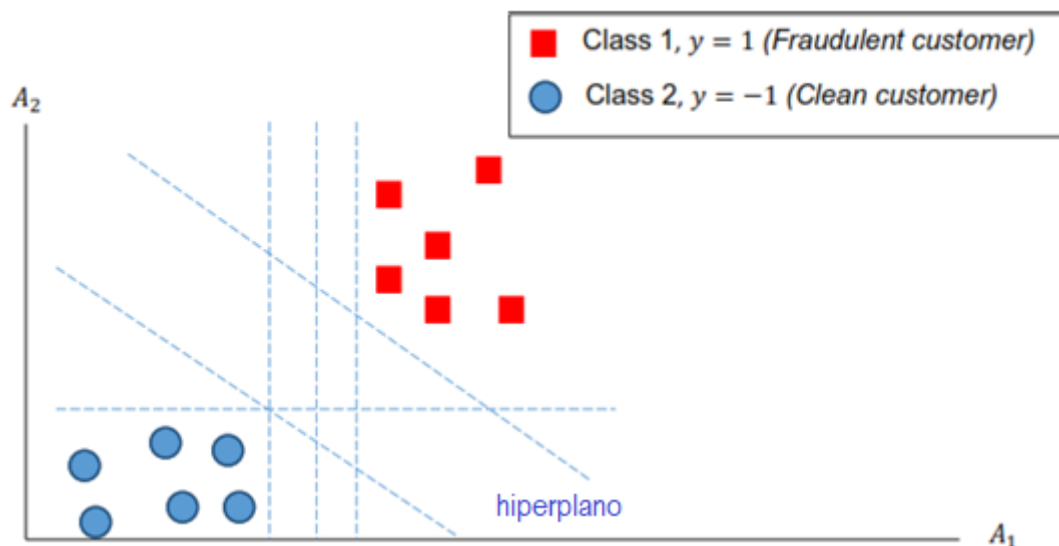


Figura 1.9 Hiperplano de separación óptimo y su margen asociado

Tabla 1.1 Resumen Técnica de modelado de aprendizaje automático

Técnica de modelado de aprendizaje automático	Tasa de aciertos de identificación de usuarios fraudulentos
Redes neuronales artificiales.	65,03%
K vecinos más cercanos.	75%
Lógica difusa (fuzzy logic)	74,50%
Bosque de caminos óptimos	72,43%
SVM	76%

La técnica de aprendizaje automático seleccionada para este trabajo de titulación es la Máquina de Vector de Soporte (SVM) por tener una tasa de aciertos para la identificación de pérdidas técnicas de un 76% ligeramente superior a las otras técnicas.

1.5.7.5.1 Máquinas de Vector de Soporte (SVM).

El método SVM se utiliza para clasificar datos lineales y no lineales. Para los datos lineales, existen dos casos: el caso en que los datos son linealmente separables y cuando son linealmente inseparables (Boser et al., 1992).

1.5.7.5.2 Datos linealmente separables

En esta sección, se discute el método SVM usando un problema binario para el caso cuando los datos son linealmente separables. Sea D un conjunto de entrenamiento dado con tuplas de entrenamiento asociadas X_i y etiquetas de clase, $y_i, i = 1, 2, 3, \dots, d$, donde D se define de la siguiente manera:

$$D = \begin{bmatrix} X_1 & y_1 \\ X_2 & y_2 \\ \vdots & \vdots \\ X_d & y_d \end{bmatrix}$$

Ecuación 1.1 Conjunto de entrenamiento

Cada y_i puede tomar uno de dos valores, -1 o 1 , es decir, $y_i \in \{-1, 1\}$, correspondientes a casos “limpios” y “fraudulentos” respectivamente. Para ayudar con la visualización, consideremos un caso en el que solo hay dos atributos de entrada A_1 y A_2 , como se muestra en la Figura 10. Los datos que se muestran en esta figura son linealmente separables ya que existe una línea recta que separa todas las tuplas de la clase 1 y la clase 2. Es evidente que hay infinitas líneas rectas que se pueden dibujar para separar las tuplas de estas dos clases. Cuando el clasificador SVM está entrenado, busca la mejor línea recta para separar los datos con un error de clasificación mínimo en tuplas. Esta línea recta se denomina límite de decisión. Si solo hubiera tres atributos de entrada, el límite de decisión sería el plano de separación. En general, en un espacio de dimensión n , donde hay atributos de entrada n , el límite de decisión es un hiperplano.

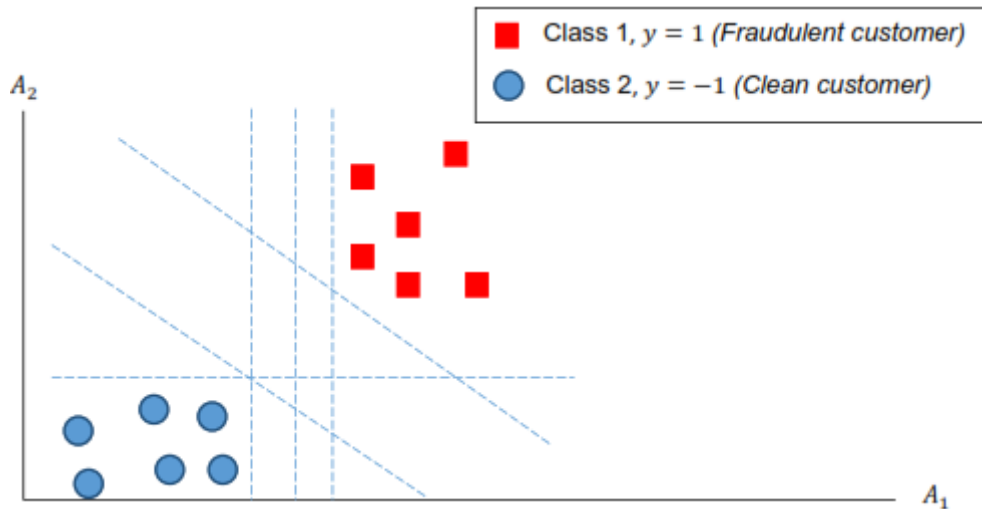


Figura 1.10 Datos de entrenamiento 2-D que son linealmente separables (Han y Kamber, 2006)

En el método SVM, el límite de decisión es el hiperplano con margen máximo. Para comprender el concepto de margen, considere las Figuras 11 y 12, que ilustran dos posibles hiperplanos y sus márgenes asociados. Ambos hiperplanos clasificaron correctamente todas las tuplas de datos, sin error de clasificación. Sin embargo, el hiperplano de la Figura 11 tiene un margen más pequeño en comparación con el hiperplano de la Figura 12.

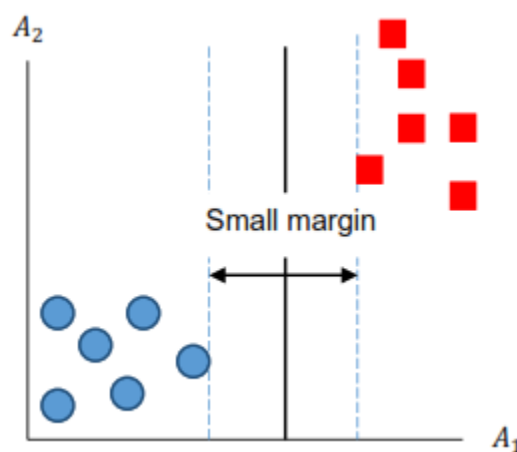


Figura 1.11 Hiperplano de margen pequeño

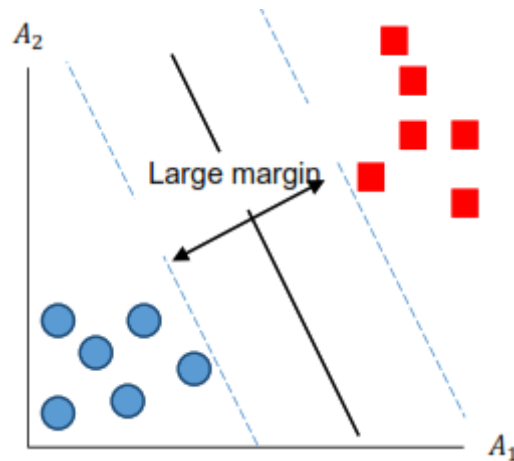


Figura 1.12 Hiperplano de margen grande

El hiperplano con el margen más grande es más preciso para clasificar tuplas de datos futuros que el que tiene el margen más pequeño. Esto se debe a que es probable que los hiperplanos con márgenes grandes tengan un error de generalización mejor que aquellos con márgenes pequeños. El error de generalización se define como una medida de la precisión con la que un algoritmo puede predecir valores de salida para datos no vistos anteriormente (Steinbach, Vipin y Tan, 2006). Es probable que los clasificadores que producen hiperplanos con un pequeño margen de error se vean afectados por el ajuste excesivo del modelo y tienden a generalizar mal en datos no vistos (Steinbach et al, 2006).

La minimización del riesgo estructural (SRM) es un principio de aprendizaje estadístico que relaciona el margen de un clasificador lineal con su error de generalización (Steinbach et al, 2006). El principio SRM proporciona un límite superior de un error de generalización R , con probabilidad $1 - \eta$ como: $R \leq R_e \varphi\left(\frac{h}{d}, \frac{\log(n)}{d}\right)$ donde φ es una función creciente monótona de la capacidad del modelo h , R_e es el error de entrenamiento y d es el número de tuplas de entrenamiento. La capacidad de un clasificador lineal es inversamente proporcional a su margen. Los clasificadores con márgenes pequeños tienen mayor capacidad y son más capaces de adaptarse a muchos conjuntos de entrenamiento. Sin embargo, de acuerdo con el principio (SRM), a medida que aumenta la capacidad del clasificador, también aumentará el límite del error de generalización. Por lo tanto, es deseable tener clasificadores lineales cuyos límites de decisión tengan márgenes máximos y límites de error de generalización más

pequeños, respectivamente. El clasificador SVM es uno de los clasificadores para los que se cumple el principio SRM (Steinbach et al, 2006).

1.5.7.5.3 Clasificador de svm lineal: caso separable

Un SVM lineal es un clasificador que busca un hiperplano lineal con el margen más grande. Por esta razón, un clasificador SVM lineal se conoce como clasificador de margen máximo. Para comprender cómo un clasificador SVM busca dicho límite, se requieren las definiciones de un límite de decisión y el margen de un clasificador lineal.

1.5.7.5.4 Límite de decisión de un clasificador de svm lineal

Considere un problema de clasificación binaria que consta de d tuplas de entrenamiento. Cada tupla se denota por, $(X_i, y_i), i = 1, 2, \dots, d$, donde $X_i = [x_{i1}, x_{i2}, \dots, x_{in}]$ corresponde a un conjunto de atributos para la i^{th} tupla. Sea $y_i \in \{-1, 1\}$ la etiqueta de clase de una tupla de entrenamiento. Para cualquier clasificador de SVM lineal, la decisión de límite se puede expresar como:

$$W \cdot X + b = 0$$

Ecuación 1.2 Limite de decisión

donde W es un vector de peso y b es un escalar, llamado sesgo, y $W \cdot X$ es un producto escalar de W y X .

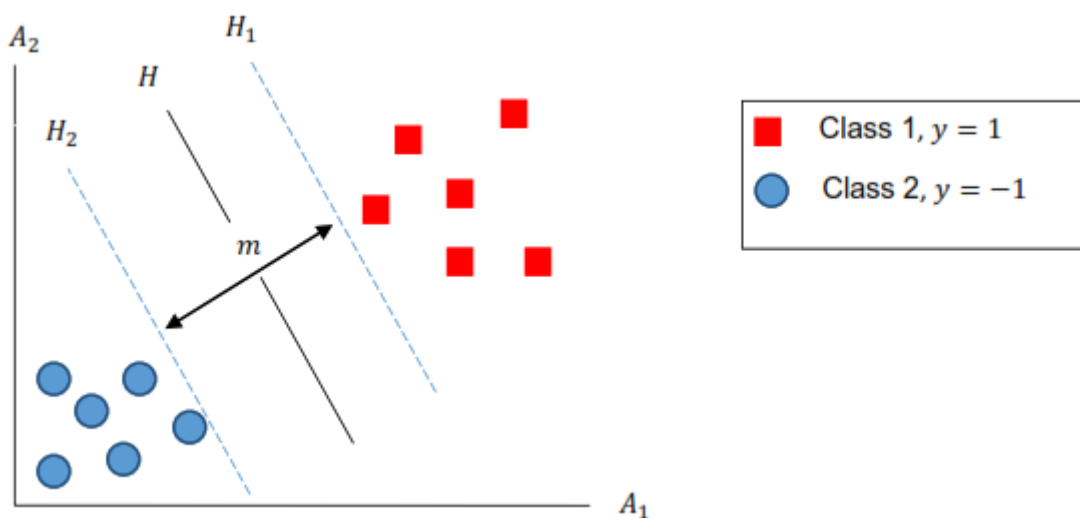


Figura 1.13 Límite de decisión y un margen del clasificador SVM para caso separable

En la Figura 1.13 se muestra un conjunto de entrenamiento bidimensional que consta de dos atributos A_1 y A_2 . Un límite de decisión que separa las tuplas de entrenamiento en sus respectivas clases, se muestra con una línea continua negra, a saber, H . Todas las tuplas ubicadas en esta línea deben satisfacer la ecuación, $W \cdot X + b = 0$. Por ejemplo, si X_a y X_b son dos tuplas ubicadas en el límite de decisión, entonces:

$$W \cdot X_a + b = 0.$$

y

$$W \cdot X_b + b = 0.$$

donde $X_a - X_b$ es un vector paralelo al límite de decisión y se dirige de X_a a X_b (Steinbach et al, 2006). Como el producto escalar $W \cdot (X_a - X_b)$ es igual a cero, la dirección de W es perpendicular al límite de decisión. Cualquier tupla de entrenamiento que pertenezca a la clase 1 satisface la siguiente ecuación

$$W \cdot X_s + b = k.$$

donde $k > 0$, y para cualquier tupla de entrenamiento que pertenezca a la clase 2, lo siguiente debe contener

$$W \cdot X_c + b = k'$$

donde $k' < 0$. Por lo tanto, la etiqueta de clase de cualquier tupla de prueba, Z , se puede predecir como:

$$y = \begin{cases} 1 & \text{si } W \cdot Z + b > 0 \\ -1 & \text{si } W \cdot Z + b < 0 \end{cases}$$

Ecuación 1.3 Etiqueta de clase

1.5.7.5.5 El margen de un clasificador de svm lineal

Los dos hiperplanos, H_1 y H_2 , que son paralelos al límite de decisión, H , en la Figura 3.5 se pueden expresar algebraicamente como

$$H_1: W \cdot X_1 + b = 1$$

$$H_2: W \cdot X_2 + b = -1$$

Ecuación 1.4 Hiperplanos

El margen del límite de decisión, m , viene dado por la distancia entre estos dos hiperplanos. Restando las dos ecuaciones, el margen del límite de decisión, m se puede calcular de la siguiente manera

$$W \cdot (X_1 - X_2) = 2$$

$$\|W\| \times m = 2$$

$$\therefore m = \frac{2}{\|W\|}$$

Ecuación 1.5 Margen del límite de decisión

1.5.7.5.6 Entrenamiento clasificador lineal de svm

La fase de entrenamiento del método SVM implica la estimación del vector de peso, W y el sesgo, b , de un conjunto de datos de entrenamiento dado. W y b se denominan parámetros del modelo SVM y se eligen de tal manera que, para cualquier tupla de entrenamiento, X_i .

$$W \cdot X_i + b \geq 1, \text{ si } y_i = 1 \text{ y } W \cdot X_i + b \leq -1 \text{ si } y_i = -1$$

Estas desigualdades se pueden combinar en una desigualdad, de la siguiente manera

$$y_i(W \cdot X_i + b) \geq 1, \quad i = 1, 2, \dots, d.$$

Para el clasificador SVM, el margen del límite de decisión debe ser el margen máximo. Encontrar este margen máximo equivale a minimizar la siguiente función objetivo:

$$f(W) = \frac{\|W\|^2}{2}$$

Ecuación 1.6 Función objetivo margen del límite de decisión

La desigualdad, $y_i(W \cdot X_i + b) \geq 1$, y la función $f(W) = \frac{\|W\|^2}{2}$ se utilizan para definir formalmente la tarea de aprendizaje involucrada en el método SVM lineal, para el caso en que los datos son linealmente separables.

Definición 1 (SVM lineal: caso separable) (Steinbach et al, 2006)

Formalmente, la tarea de aprendizaje en SVM lineal separable es el problema de optimización

$$\min_W \frac{\|W\|^2}{2}$$

Ecuación 1.7 Tarea de aprendizaje en SVM

Restricción

$$y_i(W \cdot X_i + b) \geq 1, \quad i = 1, 2, \dots, d.$$

El problema anterior es de optimización convexa, ya que la función objetivo es cuadrática y las restricciones son lineales. Este problema de optimización convexa puede resolverse mediante el método estándar del multiplicador de Lagrange. Suponga que $\lambda_i \geq 0$, para $i = 1, 2, \dots, d$. El lagrangiano para el problema de optimización viene dado por

$$L_p = \frac{1}{2} \|W\|^2 - \sum_{i=1}^d \lambda_i (y_i(W \cdot X_i + b) - 1)$$

Ecuación 1.8 Lagrangiano problema de optimización

donde los parámetros λ_i son multiplicadores de Lagrange para $i = 1, 2, \dots, d$. El método del multiplicador de Lagrange implica escribir un problema restringido como uno no restringido. El problema de la Definición 1 es equivalente al siguiente problema sin restricciones

$$\min L_p = \frac{1}{2} \|W\|^2 - \sum_{i=1}^d \lambda_i (y_i(W \cdot X_i + b) - 1)$$

Ecuación 1.9 Multiplicador de Lagrange

Para minimizar L_p , se deben tomar las derivadas con respecto a W y b , y establecerlas en cero para obtener

$$\frac{dL_p}{dW} = 0 \Rightarrow \sum_{i=1}^d \lambda_i y_i X_i$$

$$\frac{dp}{db} = 0 \Rightarrow \sum_{i=1}^d \lambda_i y_i = 0$$

Para encontrar soluciones factibles para W , b y λ_i , las restricciones de desigualdad en la Definición 1 se transforman en restricciones de igualdad. Esta transformación conduce a las siguientes condiciones de Karush-KuhnTucker (KKT) $\lambda_i \geq 0$ y $\lambda_i [y_i(W \cdot X_i + b) - 1]$

Estas condiciones KKT sugieren que los multiplicadores de Lagrange λ_i , son cero, excepto cuando $y_i(W \cdot X_i + b) = 1$

Tuplas de entrenamiento, donde $\lambda_i > 0$ se encuentran en los hiperplanos H_1 y H_2 (consulte la Figura 3.5) y se conocen como vectores de apoyo. Todas las tuplas de entrenamiento restantes, que no se encuentran ni en H_1 ni en H_2 , tienen $\lambda_i = 0$. De las ecuaciones, $W = \sum_{i=1}^d \lambda_i y_i X_i$ y $\lambda_i [y_i(W \cdot X_i + b) - 1] = 0$ es evidente que los parámetros W y b dependen únicamente de los vectores de soporte. El problema de resolver la optimización cuadrática se puede simplificar escribiendo el Lagrangiano como la función de los multiplicadores de Lagrange solamente (esto se conoce como problema dual). Sustituyendo $W = \sum_{i=1}^d \lambda_i y_i X_i$ y $\sum_{i=1}^d \lambda_i y_i = 0$ dentro $\min L_p = \frac{1}{2} \|W\|^2 - \sum_{i=1}^d \lambda_i (y_i(W \cdot X_i + b) - 1)$ produce el lagrangiano dual

$$L_p = \sum_{i=1}^d \lambda_i - \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j y_i y_j X_i \cdot X_j$$

Ecuación 1.10 Lagrangiano dual

Las principales diferencias entre los lagrangianos primarios y duales se resumen en la Tabla 1.2.

Tabla 1.2: Las principales diferencias entre lagrangianos primarios y duales

Lagrangianos primarios	Lagrangiano dual
Involucra solo multiplicadores de Lagrange y datos de entrenamiento	Involucra multiplicadores de Lagrange, datos de entrenamiento y los parámetros de límite de decisión
Problema de minimización.	Problema de maximización.

Para obtener los multiplicadores de Lagrange λ_i , el problema de optimización dual se resuelve utilizando métodos numéricos de programación cuadrática. Una vez que se conocen λ_i , los vectores de apoyo se utilizan para obtener los parámetros W y b del límite de decisión utilizando las ecuaciones, $W = \sum_{i=1}^d \lambda_i y_i X_i$ y $\lambda_i [y_i (W \cdot X_i + b) - 1] = 0$. Por lo tanto, el límite de decisión ahora se puede escribir como $\sum_{i=1}^d \lambda_i y_i X_i \cdot X + b = 0$, y cualquier tupla de prueba, Z , clasificada como

$$f(Z) = \text{sign}(W \cdot Z + b) = \text{sign} \left(\sum_{i=1}^d \lambda_i y_i x_i \cdot z + b \right)$$

Ecuación 1.11 Límite de decisión utilizando las ecuaciones

Si $f(Z) = 1$, entonces la tupla de prueba se clasifica como una clase positiva, es decir, clase 1, y si $f(Z) = -1$, la tupla de prueba se clasifica como una clase negativa, es decir, clase 2.

1.5.7.5.7 El clasificador svm lineal: caso inseparable

Los datos que se muestran en la Figura 1.14 son similares a los de la Figura 1.13. Sin embargo, dos nuevas tuplas, P y Q , se ilustran en la Figura 1.14. Este es un ejemplo de un conjunto de datos que es linealmente inseparable, ya que el hiperplano lineal con el margen máximo no puede clasificar correctamente todas las tuplas. Este hiperplano, H , clasifica erróneamente tanto P como Q con algún error de entrenamiento. Tenga en cuenta que aunque el hiperplano lineal, I , con un margen más pequeño, clasifica todas las tuplas con un error de entrenamiento cero, H sigue siendo el hiperplano preferido como límite de decisión debido a su mayor margen y su correspondiente error de generalización más pequeño. La formulación del clasificador SVM en la Sección anterior construye límites de decisión que no se acomodan a los errores de entrenamiento. Para incorporar pequeños errores de entrenamiento en el modelo, se utiliza un enfoque de

margen suave, lo que permite que el método SVM construya un límite de decisión lineal incluso cuando los datos no son separables linealmente. Por lo tanto, el problema de la Definición 1 se reformula para incorporar el caso en el que los datos son linealmente inseparables.

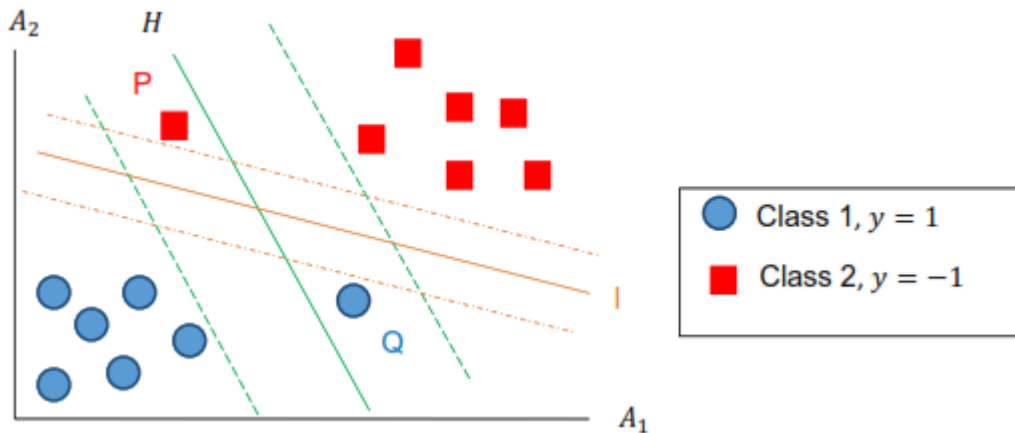


Figura 1.14 Límite de decisión y margen del clasificador SVM para caso inseparable

Para hacer esto, se introduce una variable de holgura con valor positivo, (ξ_i) , lo que da como resultado que las restricciones del problema de optimización en la Definición 1 sean

$$W \cdot X_i + b \geq 1 - \xi_i \quad \text{si } y_i = 1$$

$$W \cdot X_i + b \leq -1 + \xi_i \quad \text{si } y_i = -1$$

$$\forall_i: \xi_i > 0$$

La función objetivo también debe modificarse para penalizar un límite de decisión con valores grandes de variables de holgura. La función objetivo modificada viene dada por:

$$f(W) = \frac{\|W\|^2}{2} + C \sum_{i=1}^d \xi_i$$

Ecuación 1.12 Función objetivo valores grandes de variables de holgura

Donde C es el parámetro que representa la penalización por clasificar erróneamente las tuplas de entrenamiento. La elección del parámetro C depende del rendimiento del

modelo en el conjunto de validación. El parámetro C se denomina capacidad. El lagrangiano primario del nuevo problema de optimización se puede expresar de la siguiente manera:

$$L_p = \frac{\|W\|^2}{2} + C \sum_{i=1}^d \xi_i - \sum_{i=1}^d \lambda_i [y_i(W \cdot X_i + b) - 1 + \xi_i] - \sum_{i=1}^d \mu_i \xi_i$$

Ecuación 1.13 Lagrangiano primario del nuevo problema de optimización

Las restricciones de desigualdad luego se transforman en restricciones de igualdad utilizando condiciones KKT, de la siguiente manera:

$$\xi_i \geq 0, \lambda_i \geq 0, \mu_i \geq 0,$$

Se debe tener en cuenta que $\lambda_i \neq 0$ si las tuplas de entrenamiento se encuentran en las líneas $W \cdot X_i + b = \mp 1$ o $\xi_i > 0$. Los multiplicadores de Lagrange, μ_i son iguales a cero para cualquier tupla de entrenamiento mal clasificada. Cuando los datos son linealmente separables, las variables de holgura, λ_i , son iguales a cero, sin embargo, para los datos linealmente inseparables, las variables de holgura son positivas y menores que la capacidad, es decir, $0 \leq \lambda_i \leq C$. Los multiplicadores de Lagrange se pueden obtener resolviendo el problema dual numéricamente, utilizando técnicas de programación cuadrática. En consecuencia, estos multiplicadores se pueden utilizar para encontrar los parámetros W y b del límite de decisión. Las formulaciones de SVM explicadas en esta Sección y en la Sección 3.5.2 proporcionan una metodología para construir límites de decisión lineales que clasifican las tuplas de entrenamiento en sus respectivas clases. La siguiente sección describe un enfoque que se debe adoptar cuando se utiliza el método SVM para conjuntos de datos que tienen un límite de decisión no lineal.

1.5.7.5.8 El clasificador svm no lineal

En el caso de un clasificador SVM no lineal, el límite de decisión no es lineal. Por esta razón, los datos se transforman del espacio de coordenadas original X en un espacio transformado, $\Phi(X)$, de modo que se puede usar un límite de decisión lineal para separar tuplas de entrenamiento en el espacio transformado. El inconveniente de este enfoque es la posibilidad de sufrir la maldición de la dimensionalidad por problemas con datos de alta dimensión. Sin embargo, esto se puede evitar utilizando un método conocido como truco del kernel. El límite de decisión lineal en el espacio transformado se expresa como:

$$W \cdot \Phi(X) + b = 0$$

Ecuación 1.14 El límite de decisión lineal en el espacio transformado

Definición 2 (SVM no lineal)

La tarea de aprendizaje para una SVM no lineal se puede formalizar como el siguiente

problema de optimización: $\min_{W, b} \frac{1}{2} \|W\|^2 + C \sum_{i=1}^d \xi_i$

Sujeto a $y_i(W \cdot \Phi(X) + b) \geq 1 - \xi_i, i = 1, 2, \dots, d$.

El lagrangiano dual de este problema de optimización viene dado por:

$$L_d = \sum_{i=1}^d \lambda_i - \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j y_i y_j \Phi(X_i) \cdot \Phi(X_j)$$

Ecuación 1.15 Lagrangiano dual problema de optimización

Los multiplicadores de Lagrange, λ_i se obtienen utilizando métodos numéricos para la programación cuadrática. Los parámetros, W y b , del límite de decisión lineal en el espacio transformado se determinan luego usando las siguientes ecuaciones:

$$W = \sum_{i=1}^d \lambda_i y_i \Phi(X_i),$$

$$\lambda_i \left[y_i \left(\sum_{j=1}^d \lambda_j y_j \Phi(X_i) \cdot \Phi(X_j) + b \right) - 1 \right] = 0$$

Ecuación 1.16 Espacio transformado

donde $0 \leq \lambda_i \leq C, \forall_i$. Finalmente, cualquier entrenamiento, Z , se puede clasificar usando la siguiente ecuación:

$$f(Z) = \text{sign}(W \cdot \Phi(Z) + b) = \text{sign}\left(\sum_{i=1}^d \lambda_i y_i \Phi(X_i) \cdot \Phi(Z) + b\right).$$

Ecuación 1.17 Entrenamiento clasificador SVM

Entrenamiento clasificador SVM no lineal implica calcular el producto escalar entre los pares de vectores en el espacio transformado, es decir, $\Phi(X_i) \cdot \Phi(X_j)$. Esto puede ser computacionalmente costoso y el problema puede sufrir la maldición de la dimensionalidad. Para evitar estos problemas, se utiliza el método de truco del kernel.

Definición 3 (El truco del núcleo) (Steinbach et al, 2006)

El producto escalar $\Phi(X_i) \cdot \Phi(X_j)$. se puede considerar como una medida de similitud entre dos tuplas cualesquiera, (X_i) y (X_j) , en el espacio transformado. El truco del Kernel es el método utilizado para calcular similitudes en ese espacio utilizando el conjunto de atributos original. La función de similitud, K se define como:

$$K(X_i, X_j) = \Phi(X_i) \cdot \Phi(X_j).$$

Ecuación 1.18 La función de similitud

La función, K , se conoce como función del núcleo. Es necesario abordar algunos problemas cuando se utiliza el método SVM no lineal. En primer lugar, se desconoce la forma de la función de mapeo, Φ . En segundo lugar, trabajar en un espacio de características de alta dimensión podría resultar muy costoso. El truco del kernel soluciona ambos problemas. En primer lugar, la función del núcleo es la función del conjunto de atributos original, por lo que no es necesario conocer la forma de la función de mapeo, Φ . En segundo lugar, dado que los cálculos se realizaron en el espacio original, se pueden evitar los problemas relacionados con la maldición de la dimensionalidad. El teorema de Mercer proporciona condiciones bajo las cuales la función del núcleo se puede expresar como el producto escalar de dos vectores de entrada cualquiera. La función del núcleo usado en SVM no lineal debe satisfacer el teorema de Mercer.

Mercer's Theorem (Steinbach et al, 2006)

Una función del núcleo K se puede expresar como:

$$K(u, v) = \Phi(u) \cdot \Phi(v),$$

Ecuación 1.19 La función de núcleo

Si, y solo si, para cualquier función $g(x)$ tal que $\int g(x)^2 dx$ es finito, entonces

$$\int \int K(x, y)g(x)g(y)dxdy \geq 0$$

El kernel de la función de base radial (*rbf*), también conocido como kernel gaussiano, es una función del kernel basada en la distancia euclidiana (Nagi et al, 2008). En este estudio, se utilizó el kernel gaussiano para escribir el producto escalar $\Phi(X_i) \cdot \Phi(X_j)$. como función de los datos originales de la siguiente manera:

$$K(X_i, X_j) = \Phi(X_i) \cdot \Phi(X_j) = e^{-\gamma \|X_i - X_j\|^2}$$

Ecuación 1.20 Kernel gaussiano

Tenga en cuenta que el Kernel de Gauss tiene solo un parámetro, a saber, γ . Este parámetro se denomina parámetro kernel *rbf*. Otro parámetro importante en el método SVM es la capacidad C . El parámetro del kernel *rbf* y la capacidad se obtienen experimentalmente. Las medidas de desempeño se utilizan para encontrar los valores óptimos de γ y C .

En resumen, el SVM es un método de aprendizaje supervisado que se utiliza para la clasificación estadística de datos lineales y no lineales. Dado un conjunto de tuplas de entrenamiento, cada una de las cuales pertenece a una de dos clases, el algoritmo de entrenamiento de SVM crea un clasificador que asigna nuevas tuplas a una de las dos clases. El entrenamiento se realiza encontrando los parámetros del límite de decisión, que es un hiperplano con mayor margen. El límite de decisión se usa luego para clasificar tuplas con etiquetas de clase desconocidas.

1.5.8 Métodos que utilizan datos obtenidos en tiempo real.

Estos métodos se caracterizan por utilizar las mediciones obtenidas de la red para, con base en las teorías de los circuitos eléctricos, detectar pérdidas no técnicas. Podemos mencionar la onda viajera y las técnicas de estimación de estado como algunos ejemplos de estos métodos. Este trabajo utiliza la estimación de estado, que consiste en un método que utiliza las mediciones en tiempo real y casi en tiempo real disponibles en la red, para calcular el estado más probable de la red y compararlo con el observado, identificando así los nodos de consumo que tiene pérdidas no técnicas. El mayor desafío de este método es la capacidad de observación de la red, que requiere un número suficiente de medidas e información para hacer posible estimar el estado, que en las redes de distribución puede no estar disponible.

1.5.9 Estimación de Estado en Sistemas.

El proceso de estimación de estado se refiere al procedimiento para obtener fasores de voltaje y flujos de potencia en todas las barras y líneas del sistema en un momento dado. Este resultado se puede obtener directamente, a través de la medición precisa de las variables del sistema, sin embargo, este método es muy vulnerable a los problemas de medición y comunicación, además del alto costo. Por lo tanto, la estimación de estado utiliza medidas redundantes para actuar como un filtro para errores de medición y otras incertidumbres, y calcula el estado más probable de la red. Las mediciones pueden

incluir no solo las convencionales, como el módulo de voltaje y la inyección de energía, sino que también pueden contener otras, como las corrientes sincronizadas y los fasores de voltaje, obtenidos utilizando las llamadas Unidades de Medición Fasorial Sincronizadas (PMU).[27]

La información utilizada por el estimador de estado puede contener los siguientes tipos de errores o incertidumbres:

- Ruido de medición; errores inherentes a la medición. Consiste en pequeñas diferencias entre el valor medido y el valor real debido a la precisión limitada de los medidores. [27]
- Errores grueso; pueden verse como grandes diferencias entre los valores reales y medidos, pueden ocurrir por varias razones, como la existencia de fraude o defecto en los medidores. [27]
- Errores de parámetros; son errores en los datos de la red eléctrica, interfieren directamente en el cálculo del estimador de estado. [27]
- Errores de topología; Los errores asociados con la configuración de la red también comprometen la estimación correcta del estado. [27]

El estimador de estado puede facilitar una supervisión satisfactoriamente precisa y eficiente de las cantidades relacionadas con la operación, identificando el estado operativo del sistema en un momento dado, como líneas de carga y voltajes de bus. Los estimadores de estado generalmente tienen las siguientes funciones [26]:

- a) Procesador de topología: consiste en reunir todos los datos, como el estado de los interruptores y disyuntores, sobre el sistema. Esta información permite modelar el diagrama unifilar del sistema. [26]
- b) Análisis de observabilidad: determina si, en base a las medidas disponibles, es posible realizar la estimación de estado e identifica los tramos no observables si las medidas disponibles no son suficientes para realizar la estimación. [26]
- c) Solución de la estimación de estado: realiza la estimación calculando el estado, la carga de línea, los fasores de voltaje y las inyecciones de potencia activa y reactiva,

muy probablemente del sistema a partir de los datos recopilados, actuando como un filtro para las incertidumbres. [26]

- d) Identificador de error de topología y parámetro: detecta posibles errores en los parámetros o datos asociados con la topología de la red. [26]

Los estimadores de estado actúan como un filtro entre los datos medidos directamente desde el sistema y las funciones que requieren datos más precisos y confiables sobre el estado actual. Los datos obtenidos directamente del equipo instalado, como la posición del Tap del transformador y los estados del interruptor, son procesados por el procesador de topología, que genera un modelo con las barras y ramas del sistema de alimentación. Después de la construcción del modelo, el estimador analiza si el sistema es observable, si tiene secciones no observables, se puede usar el dispositivo de pseudo-medición, como se verá más adelante. Finalmente, el estimador realiza una estimación de estado e identificación de posibles errores gruesos, de topología y de parámetros, para encontrar el estado más probable del sistema, asociadas con la Detección e identificador de errores brutos de medición: detecta, identifica y elimina la existencia de errores brutos en las mediciones. Para eliminar estos errores, debe haber suficiente redundancia en los datos adquiridos para que el sistema no pierda observabilidad con la eliminación de los datos sospechosos de contener errores graves. [26]

La estimación en Sistema de Distribución puede ser muy diferente de la estimación del estado del sistema de transmisión. En comparación con las redes de transmisión, las redes de distribución difieren en las siguientes características: [27].

- Su topología es radial o de malla débil. [27].
- La carga está desequilibrada. Por lo tanto, los equivalentes de una sola fase no contienen suficientes detalles y se necesitan estimadores trifásicos. [27].
- Debido al menor número de clientes, estas redes tienen conjuntos de medidas muy limitados en comparación con los sistemas de transmisión. Debido a que el número de mediciones en tiempo real es demasiado bajo, generalmente se necesitan pronósticos de carga (también llamados pseudo-mediciones) para que el sistema sea observable. [27].
- Líneas de distribución son cortas en longitud con baja relación X/R;
- Las cargas son más difíciles de pronosticar que en las redes de transmisión, ya que el número de clientes que se transforman es menor, lo que resulta en una

aleatoriedad. Además, con la introducción de vehículos eléctricos o la respuesta a la demanda, la aleatoriedad de la carga podría ser aún mayor. [27].

1.5.10 Modelado Trifásico para Sistemas de Distribución

Con el objetivo de enriquecer y ampliar la metodología de detección e identificación de PNTs, es importante el proceso de estimación de estado utilizando un modelado adecuado a las características de los sistemas de distribución. El Modelado matricial generalizado, propuesto en Kersting(2012), permite una representación trifásica de los elementos de un SD (Sistema de Distribución), considerando impedancias mutuas, ramas bifásicas y monofásicas, presencia de cargas desequilibradas. Como es un enfoque adecuado para la representación de SD, se desarrolló un estimador de estado trifásico basado en la misma metodología. [20]

1.5.10.1 Estructura estándar de los sistemas de distribución

La distribución de energía eléctrica se caracteriza como el segmento del sector eléctrico dedicado a la operación en niveles de voltaje inferiores a los de los sistemas de transmisión, cuyo objetivo es llevar la energía eléctrica a los diferentes consumidores.

El modelado de SD por matrices generalizadas consiste en determinar un conjunto de matrices que contienen información sobre las características intrínsecas de los elementos de la red. Estas matrices, a su vez, relacionan las magnitudes eléctricas aguas abajo y aguas arriba de cada elemento, a través de ecuaciones derivadas de un modelo de cuadripolo genérico.

Se pueden deducir matrices generalizadas para cada elemento de la red a partir de estas ecuaciones genéricas, basadas en la aplicación de la teoría de circuitos. El modelado difiere ya que, a partir de la determinación de estas matrices, las ecuaciones genéricas pueden ser aplicadas a cualquier elemento del sistema, haciendo este enfoque compacto e intuitivo. [20]

1.5.10.1 Elemento Genérico SD

Considere la figura 1.15, en el que está representada cualquier rama de la SD. Esta rama se puede representar a través de un modelo cuadripolo, por matrices generalizadas, como se muestra en la figura 1.16.

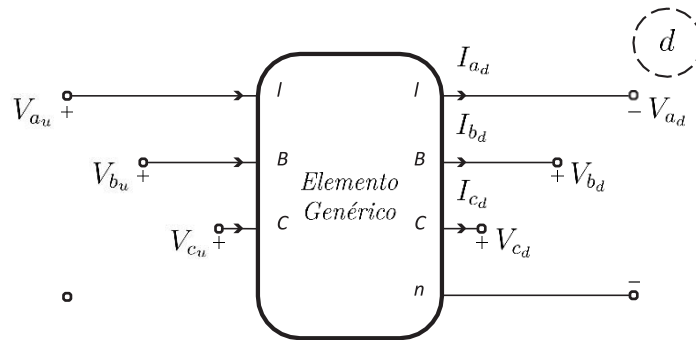


Figura 1.15 Ramal Genérico de la Red de Distribución [20]

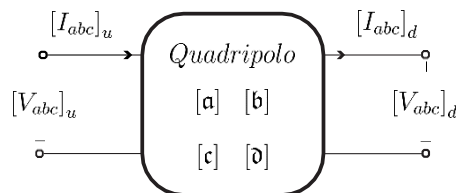


Figura 1.16 Modelo de cuadrupolo genérico [20]

En esta representación, las matrices generalizadas $[a]$, $[b]$, $[c]$ y $[d]$, de dimensión 3×3 , relacionan las tensiones de fase y las corrientes de línea de los nodos aguas arriba y aguas abajo del ramal, a través de la ecuación 1.21:

$$\begin{bmatrix} [V_{abc}]_u \\ [I_{abc}]_u \end{bmatrix} = \begin{bmatrix} [a] & [b] \\ [c] & [d] \end{bmatrix} * \begin{bmatrix} [V_{abc}]_d \\ [I_{abc}]_d \end{bmatrix}$$

Ecuación 1.21

donde $[V_{abc}]$ y $[I_{abc}]$ vienen dados por:

$$[V_{abc}] = [V_a \quad V_b \quad V_c]^T$$

Ecuación 1.22

$$[I_{abc}] = [I_a \quad I_b \quad I_c]^T$$

Ecuación 1.23

Las ecuaciones generalizadas se pueden obtener mediante el desarrollando de la Ecuación (1.21) (KERSTING,2012):

$$[V_{abc}]_u = [a] \cdot [V_{abc}]_d + [b] \cdot [I_{abc}]_d$$

Ecuación 1.24

$$[I_{abc}]_u = [c] \cdot [V_{abc}]_d + [b] \cdot [I_{abc}]_d$$

Ecuación 1.25

Las ecuaciones (1.24) y (1.25) modelan las tensiones de fase del nodo aguas arriba y las corrientes de línea inyectadas en el elemento genérico, relacionándolas con las tensiones y corrientes aguas abajo que salen del elemento, a través de sus matrices generalizadas. [20]

Es posible determinar las tensiones de fase aguas abajo a partir de las corrientes de línea aguas abajo y las tensiones de fase aguas arriba, mediante la ecuación [20]:

$$[V_{abc}]_d = [A] \cdot [V_{abc}]_u - [B] \cdot [I_{abc}]_d$$

Ecuación 1.26

donde $[A]$ y $[B]$ y son también matrices de dimensiones generalizada 3x3 que, para la mayoría de los elementos SD, puede determinarse mediante [20]:

$$[A] = [a]^{-1}$$

Ecuación 1.27

$$[B] = [a]^{-1} \cdot [b]$$

Ecuación 1.28

De las ecuaciones generalizadas (1.24), (1.25) y (1.26), es posible determinar cualquier variable eléctrica del SD.

Todo lo correspondiente al modelamiento de la red de distribución y flujo de potencia se detalla en el Anexo 6.1

1.5.11 Estimación de Estados Trifásicos en Sistemas de Distribución

1.5.11.1 Introducción

Con base a la formulación presentada, es posible desarrollar una herramienta para el monitoreo en tiempo real de los SD. Al poseer la topología de la red, los datos relacionados con los componentes del sistema y un plan de medición bien estructurado, el proceso de Estimación de Estado en Sistemas de Distribución (EESD) permite el monitoreo de voltajes complejos e inyecciones de potencia desde los nodos a lo largo del alimentador.

En esta sección se presentará la herramienta de estimación del estado trifásico, con énfasis en su uso en SD. También se discuten: aspectos relacionados con la formación

del plan de medición, formulación matemática a través de un problema de minimización, ecuaciones de las magnitudes medidas, así como la formación de la matriz jacobiana utilizada en cada iteración del estimador de estado.

1.5.11.2 Modelos de Medida

La información probabilística, a partir de medidas y pseudomedidas, puede expresarse matemáticamente por [30]:

$$z_m = z_{verd} + \eta_m$$

Ecuación 1.29

Es decir, los valores captados por los medidores, z_m , son iguales a los valores verdaderos de las magnitudes medidas, z_{verd} , contaminadas por un error aleatorio, η_m . Este error se deriva de varios factores, como el error de escala del medidor, las imprecisiones del transformador de instrumentación, la pérdida de datos debido a las conversiones analógicas/digitales, los errores de los estudios de predicción de casos, etc.

El vector de error η_m puede considerarse una variable aleatoria normal, con media cero, y matriz de covarianza R_m , es decir [31]:

$$\varepsilon\{n_m\} = 0$$

Ecuación 1.30

$$\varepsilon\{\eta_m \cdot \eta_m^T\} = R_m$$

Ecuación 1.31

Dónde $\varepsilon\{\cdot\}$ es el operador de expectativa matemática. También se considera que los errores de medida no están correlacionados, lo que implica una matriz R_m diagonal, cuyos elementos diagonales corresponden a las varianzas de sus respectivas medidas.

Del modelado SD, un conjunto de ecuaciones se puede determinar, para obtener el vector z_{verd} en función del vector de estados x :

$$h_m(x) = z_{verd}$$

Ecuación 1.32

donde x representa los ángulos y las magnitudes de las tensiones a lo largo de la SD:

$$x \triangleq \begin{bmatrix} [\delta_{abc}]_1 \\ [\delta_{abc}]_2 \\ \vdots \\ [\delta_{abc}]_{nb} \\ [E_{abc}]_1 \\ \vdots \\ [E_{abc}]_{nb} \end{bmatrix}$$

Ecuación 1.33

Entonces, reemplazando (1.32) en (1.29), es posible obtener la representación matemática de las medidas:

$$z_m = h_m(x) + n_m$$

Ecuación 1.34

Finalmente, los principales elementos diagonales de la matriz de covarianza R_m son las varianzas de medida, σ_i^2 , cuyos valores dependen de la precisión de los medidores y del valor medido.

1.5.11.3 Observabilidad del Plan de Medición

La viabilidad de la estimación del estado está directamente relacionada con la observabilidad del plan de medición del sistema. Un sistema de potencia se considera observable con respecto a un plan de medición, o conjunto de mediciones, si el estado del sistema, es decir, la magnitud y el ángulo del voltaje, se puede determinar mediante un estimador de estado [32].

Para un sistema con n variables de estado, el mínimo número de medidas e información m , una condición necesaria para que se asegure la observabilidad global del plan de medidas es:

$$m = n_e$$

Ecuación 1.35

En otras palabras, un plan de medidas con al menos n_e medidas es una condición mínima para la observabilidad, pero no garantiza esta propiedad ya que también depende de la distribución de las medidas en la red.

1.5.11.4 Redundancia del plan de medición

Para poder aplicar técnicas de análisis de errores, es necesario tener una redundancia de información y medidas en el sistema. La redundancia global (ρ) del plan de medidas viene dada por:

$$\rho = \frac{m}{n_e}$$

Ecuación 1.36

Para la redundancia necesaria del plan de medida utilizado para el análisis de errores, el valor de ρ debe ser significativamente superior a 1. En los sistemas de transporte, los niveles habituales de redundancia se encuentran entre 2,0 y 3,0, y en los sistemas de distribución, este valor normalmente no es superior a 2.0 debido a las características mencionadas.

En caso de bajo nivel de redundancia, medidas críticas y conjuntos críticos [32] puede estar presente durante el análisis del sistema. Una medida crítica contiene información crucial sobre un determinado estado del sistema. Su residual normalizado, resultante del proceso de estimación de estado, es nulo, y la eliminación de esta información del plan de medición da como resultado la pérdida de observabilidad [15]. Un conjunto crítico es un conjunto de dos o más medidas, que contiene la única información sobre un estado o conjuntos de estados dados. Es posible demostrar que los residuos normalizados de medidas pertenecientes a conjuntos críticos son idénticos en magnitud, es decir, en presencia de errores graves, es imposible discernir qué miembros de conjuntos críticos corresponden a medidas adulteradas [15].

Es interesante entender estos dos conceptos, ya que las medidas disponibles en tiempo real en los SD suelen ser escasas, lo que puede dar lugar a planes de medida con medidas y conjuntos críticos, o incluso no observables [31].

1.5.11.5 Pseudomedidas de inyección

Para enriquecer la redundancia del plan de medición, de las pseudomedidas de inyección de potencia asociadas a los autobuses con consumidores [15]. son información pronosticadores de carga que estiman el consumo horario de cada cliente SD, y que son procesados como provenientes de un medidor ficticio, dentro del proceso de estimación de estado, para obtener un plan de medición debidamente observable.

Al no medirse en tiempo real, las pseudomedidas de inyección traen mayor incertidumbre en cuanto a la calidad de su medición. Para simular esta imprecisión se utilizan valores mayores de varianza asociados a estas pseudomedidas.

1.5.11.6 Restricciones de igualdad

Las restricciones de igualdad representan condiciones estructurales y operativas a las que se somete el SD. Aunque no representan medidas en sí mismas, estas restricciones conllevan a proporcionar información relacionada con el sistema y ayudar a garantizar la

observabilidad, así como aumentar la redundancia general. Esta información se modela de forma determinista. Esto significa que las varianzas asociadas a esta información se consideran nulas, aunque en la práctica se utilizan valores muy pequeños, varios órdenes de magnitud menores que el de medidas y pseudomedidas.

1.5.11.7 Restricciones Estructural

Las restricciones estructurales representan condiciones fijas de la SD, de acuerdo con la topología y la estructura de la red. En esta trabajo, hay dos restricciones estructurales a las que el SD puede estar sujeto: ángulos de fase de los voltajes en la barra de la subestación, que se supone equilibrados; y nodos de paso o inyección nula (sin presencia de consumidores conectados).

El bus de la subestación alimenta el sistema de acuerdo con el voltaje proveniente del sistema de transmisión. Por tanto, considerando este sistema equilibrado y estableciendo referencia nula en la fase *a* de la barra de la subestación. Estas restricciones están definidas por:

$$\begin{bmatrix} \delta_a \\ \delta_b \\ \delta_c \end{bmatrix}_i - \begin{bmatrix} 0 \\ -\frac{2\pi}{3} \\ \frac{2\pi}{3} \end{bmatrix} = [\delta_{abc}]_i - \begin{bmatrix} 0 \\ -\frac{2\pi}{3} \\ \frac{2\pi}{3} \end{bmatrix} = [0]$$

Ecuación 1.37

Los nodos de paso, o inyección nula, son nodos del sistema en los que no hay conexiones a los consumidores o generadores. Por lo tanto, no hay potencia inyectada en estos puntos. Estos nodos pueden ser removidos del SD, a través de la aglutinación de elementos. Sin embargo, este proceso, por mucho que reduzca el tamaño del sistema, hace que se pierda información sobre los voltajes en esos puntos. Si no se eliminan, estos nodos se pueden representar igualando las potencias inyectadas en ellos, en función de los estados, con un vector nulo[15]:

$$\begin{cases} [P_{abc}(x)]_i = [0] \\ [Q_{abc}(x)]_i = [0] \end{cases}$$

Ecuación 1.38

Ambas restricciones descritas por Ecuaciones (1.37) y (1.38) pueden representarse colectivamente mediante un solo vector de restricciones estructurales:

$$h_s(x) = [0]$$

Ecuación 1.39

1.5.11.8 Restricciones Operacionales

Las restricciones operativas dictan las condiciones de funcionamiento del sistema. Para el caso de estudio de los SD con PNT, estas restricciones son de fundamental importancia ya que definen el objeto de estudio y análisis de este procedimiento propuesto [15].

Se conoce la topología del sistema, por lo que no es necesario modelar los interruptores automáticos. Con las derivaciones de los transformadores y los reguladores de voltaje conocidos, estos se modelan utilizando matrices generalizadas [15].

Para cumplir con el objetivo de detectar la ocurrencia de pérdidas no técnicas, se supone que el operador del sistema de distribución preselecciona un conjunto de nodos de la red n_{susp} sospechosos de estar sujetos a PNT. Esta preselección puede basarse, por ejemplo, en estudios previos que indiquen la ocurrencia de inyecciones de energía anormales en esos nodos. Para cada uno de los nodos sospechosos se definen restricciones operativas del tipo:[15]

$$h_0(x) = [0]$$

Ecuación 1.40

que, en el caso de este estudio, por la información de potencia activa/reactiva de los nodos, se reescriben de acuerdo con la ecuación:[15]

$$[P_{abc}(x)]_i - [P_{abc}^{pres}]_{carga,i} = 0, i = 1, \dots, n_{susp}$$

$$[Q_{abc}(x)]_i - [Q_{abc}^{pres}]_{carga,i} = 0, i = 1, \dots, n_{susp}$$

Ecuación 1.41

Donde $[P_{abc}^{pres}]_{carga,i}$ y $[Q_{abc}^{pres}]_{carga,i}$ son las cargas trifásicas presuntamente activas y reactivas (es decir, sin contaminación de PNT) para el nodo i . [15]

Esto significa que el operador asume que no hay PNT en los nodos preseleccionados. Sin embargo, si algunas de las cargas de barras sospechosas están realmente bajo la influencia de los PNT, los segundos términos del lado izquierdo de (1.41) debe estar dada por:[15]

$$[P_{abc}^{pres}]_{carga,i} + [PNT]_{P,i}$$

$$[Q_{abc}^{pres}]_{carga,i} + [PNT]_{Q,i}$$

Ecuación 1.42

En consecuencia, en este caso algunas de las hipótesis formuladas por el operador según son incorrectas. El problema de identificar los PNT se reduce entonces a determinar qué restricciones operativas no están suficientemente respaldadas por las medidas/pseudomedidas disponibles para el estimador de estado.[15]

1.5.11.9 Modelado del nivel de incertidumbre de las restricciones

En el proceso de estimación del estado, también es importante modelar las varianzas asociadas con cada medida o restricción. La varianza de los datos enviados al estimador de estado contiene información sobre la precisión y la calidad de la medición. En otras palabras, indica el grado de confianza que el estimador asociará con esa información, y cuánto cada medida y la restricción influye en el resultado final. [15]

La ecuación (1.31) presenta el modelado de la matriz de covarianza de la información probabilística. Como ya fue mencionado, esta información es susceptible de errores y sus variaciones son generalmente del mismo orden de magnitud. [15]

Para que el proceso EESD considere la influencia de la información determinista, es conveniente considerarla como probabilística, lo que implica asociar las varianzas con restricciones estructurales y operativas. Como se trata de información determinista, teóricamente estas varianzas son nulas. Así, las matrices de covarianza de las restricciones estructurales y operativas están dadas por:[15]

$$\begin{cases} R_s = [0] \\ R_o = [0] \end{cases}$$

Ecuación 1.43

Sin embargo, por razones prácticas, como se presentará en el uso de la Ecuación (1.43) puede causar dificultades de cálculo. Para evitar este problema se destaca la necesidad de que R_s y R_o no sean singulares.[15]

Es posible obtener matrices de covarianza invertibles, considerando información determinista como las variables aleatorias normales, de forma similar a las medidas y pseudomedidas, pero con desviaciones estándar mucho más pequeñas que las de la información probabilística, es decir: [15]

$$\sigma_s, \sigma_o \ll \sigma_m$$

Ecuación 1.44

A partir de la consideración de que la información probabilística no tiene correlación entre sí, las matrices de covarianza de las restricciones estructurales y operativas utilizadas en este trabajo se definen entonces como: [15]

$$\begin{cases} R_s = \mathcal{G} \cdot [U] \\ R_o = \mathcal{G} \cdot [U] \end{cases}$$

Ecuación 1.45

Donde \mathcal{G} es un valor muy pequeño $\mathcal{G} = 10^{-8}$ con el fin de satisfacer la relación dada por la Ecuación (1.44).

1.5.11.10 Formulación matemática del estimador de estado

El proceso de estimación de estado puede ser formulado como un problema de optimización, donde se busca obtener la mejor estimación de los estados del sistema, \hat{x} , con el fin de minimizar la suma ponderada de los cuadrados de los residuos de medición.[15]

Los residuos de medición se estipulan a partir de los vectores

Z_m y $h_m(x)$, reescritos según los estados:

$$r_m = z_m - h_m(x)$$

Ecuación 1.46

Para un SD, con el vector de estado x dado por la Ecuación (1.93), el número de estados a estimar viene dado por:

$$n_e = 6 \cdot n_b$$

Ecuación 1.47

A partir de la ecuación (1.46), es posible determinar en una matriz la función objetivo del problema de optimización.

$$J(x) = r_m^T \cdot R_m^{-1} \cdot r_m$$

Ecuación 1.48

donde los residuales son ponderados por la inversa de la matriz de covarianza de los errores de medida R_m . Esta ponderación es interesante porque da más peso a las medidas más precisas, cuyas varianzas son pequeñas, mientras relaja las medidas más propensas a errores, cuyas varianzas son más grandes. La estimación del estado puede entonces definirse formulando el problema de minimización de la función $J(x)$, sujeto a

las definiciones residuales, Ecuación (1.46), y restricciones de igualdad, Ecuaciones (1.40) y (1.42), y expresado de la siguiente manera [33]:

$$\text{Minimizar } \frac{1}{2} \cdot r_m^T R_m^{-1} \cdot r_m$$

$$\text{Sujeto a: } r_m = z_m - h_m(\hat{x})$$

$$h_s(\hat{x}) = [0]$$

$$h_o(\hat{x}) = [0]$$

Ecuación 1.49

Donde \hat{x} es el vector de los estados estimados.

1.5.11.11 Resolviendo el Problema de Optimización

Para resolver el problema de optimización con restricciones de igualdad, dado por la Ecuación (1.49), es necesario formar la función Lagrangiana correspondiente. Para ello, se asignan multiplicadores de Lagrange a las restricciones de igualdad de la Ecuación (1.49), añadiéndolos a la función objetivo [33]:

$$\mathcal{L}(\hat{x}, r, \lambda) = \frac{1}{2} \cdot r_m^T \cdot R_m^{-1} \cdot r_m - \lambda_m^T \cdot (h_m(\hat{x}) - z_m + r_m) - \lambda_s^T \cdot (h_s(\hat{x})) - \lambda_o^T \cdot (h_o(\hat{x}))$$

Ecuación 1.50

La solución óptima para la Ecuación (1.50) se obtiene satisfaciendo las condiciones de optimalidad necesarias, mediante la derivación parcial de la función Lagrangiana en relación con sus variables[33]:

$$\frac{\partial \mathcal{L}}{\partial \hat{x}} = -H_m^T(\hat{x}) \cdot \lambda_m - H_s^T(\hat{x}) \cdot \lambda_s - H_o^T(\hat{x}) \cdot \lambda_o = [0]$$

Ecuación 1.51

$$\frac{\partial \mathcal{L}}{\partial r} = R_m^{-1} r_m - \lambda_m = [0]$$

Ecuación 1.52

$$\frac{\partial \mathcal{L}}{\partial \lambda_m} = h_m(\hat{x}) - z_m + r_m = [0]$$

Ecuación 1.51

$$\frac{\partial \mathcal{L}}{\partial \lambda_s} = h_s(\hat{x}) = [0]$$

Ecuación 1.52

$$\frac{\partial \mathcal{L}}{\partial \lambda_o} = h_o(\hat{x}) = [0]$$

Ecuación 1.53

Donde:

$$\begin{cases} H_m(\hat{x}) = \frac{\partial h_m(\hat{x})}{\partial \hat{x}} \\ H_s(\hat{x}) = \frac{\partial h_s(\hat{x})}{\partial \hat{x}} \\ H_o(\hat{x}) = \frac{\partial h_o(\hat{x})}{\partial \hat{x}} \end{cases}$$

Ecuación 1.54

son las matrices jacobianas de las ecuaciones no lineales que modelan las medidas y restricciones, $h_m(\hat{x})$, $h_s(\hat{x})$ y $h_o(\hat{x})$, con respecto a las variables de estado, \hat{x} .

Para la solución de estimación de estado, es interesante linealizar este sistema de ecuaciones no lineales alrededor de un vector de estado dado \hat{x}^ℓ . Esto se hace usando la serie de Taylor, aplicando truncamiento después de la primera derivada. Así:

$$\begin{cases} h_m(\hat{x}) = h_m(\hat{x}^\ell) + H_m(\hat{x}^\ell) \cdot \Delta \hat{x} \\ h_s(\hat{x}) = h_s(\hat{x}^\ell) + H_s(\hat{x}^\ell) \cdot \Delta \hat{x} \\ h_o(\hat{x}) = h_o(\hat{x}^\ell) + H_o(\hat{x}^\ell) \cdot \Delta \hat{x} \end{cases}$$

Ecuación 1.55

Resolviendo la ecuación r_m

$$r_m = R_m \cdot \lambda_m$$

Ecuación 1.56

sustituyendo la Ecuación (1.56) en las ecuaciones (1.51), (1.52), (1.53) y (1.54), y aplicando la linealización dada por la Ecuación (1.55), se pueden reescribir las condiciones de optimalidad [31]:

$$\begin{cases} -z_m + h_m(\hat{x}^\ell) + H_m(\hat{x}^\ell) \cdot \Delta \hat{x} + R_m \cdot \lambda_m = [0] \\ h_s(\hat{x}^\ell) + H_s(\hat{x}^\ell) \cdot \Delta \hat{x} = [0] \\ h_o(\hat{x}^\ell) + H_o(\hat{x}^\ell) \cdot \Delta \hat{x} = [0] \\ H_m^T(\hat{x}^\ell) \cdot \lambda_m + H_s^T(\hat{x}^\ell) \cdot \lambda_s + H_o^T(\hat{x}^\ell) \cdot \lambda_o = [0] \end{cases}$$

Ecuación 1.57

Finalmente, para mejorar la estabilidad numérica del problema EESD, es necesario insertar las matrices de covarianza de las restricciones de igualdad, en el sistema dado por la Ecuación (1.57). Aunque las matrices R_s y R_o son teóricamente nulas, como se muestra en la Ecuación (1.43), dada por la Ecuación (1.45). Es posible demostrar que estas consideraciones conducen a la siguiente forma revisada de la Ecuación (1.57):

$$\begin{cases} -z_m + h_m(\hat{x}^\ell) + H_m(\hat{x}^\ell) \cdot \Delta\hat{x} + R_m \cdot \lambda_m = [0] \\ h_s(\hat{x}^\ell) + H_s(\hat{x}^\ell) \cdot \Delta\hat{x} + R_s \cdot \lambda_s = [0] \\ h_o(\hat{x}^\ell) + H_o(\hat{x}^\ell) \cdot \Delta\hat{x} + R_o \cdot \lambda_o = [0] \\ H_m^T(\hat{x}^\ell) \cdot \lambda_m + H_s^T(\hat{x}^\ell) \cdot \lambda_s + H_o^T(\hat{x}^\ell) \cdot \lambda_o = [0] \end{cases}$$

Ecuación 1.58

El problema dado por la Ecuación (1.58) puede ser resuelto por el método matricial aumentado de Hatchel, también conocido como Sparse Tableau a través de su representación matricial [31]:

$$\begin{bmatrix} [0] & H_m(\hat{x}^\ell)^T \\ H_m(\hat{x}^\ell) & R \end{bmatrix} \cdot \begin{bmatrix} \Delta\hat{x} \\ \lambda \end{bmatrix} = \begin{bmatrix} [0] \\ z - h(\hat{x}^\ell) \end{bmatrix}$$

Ecuación 1.59

$$H(\hat{x}^\ell) = [H_m^T(\hat{x}^\ell) \quad H_s^T(\hat{x}^\ell) \quad H_o^T(\hat{x}^\ell)]^T$$

Ecuación 1.60

$$R = \begin{bmatrix} R_m & [0] & [0] \\ [0] & R_s & [0] \\ [0] & [0] & R_o \end{bmatrix}$$

Ecuación 1.61

$$\lambda = [\lambda_m \quad \lambda_s \quad \lambda_o]^T$$

Ecuación 1.62

$$z = [z_m \quad [0] \quad [0]]^T$$

Ecuación 1.63

$$h(\hat{x}^\ell) = [h_m^T(\hat{x}^\ell) \quad h_s^T(\hat{x}^\ell) \quad h_o^T(\hat{x}^\ell)]^T$$

Ecuación 1.64

También por razones de cálculo, se acostumbra reemplazar la submatriz nula en la matriz de Tableau, en el lado izquierdo de la Ecuación (1.59), por una matriz identidad, de dimensión apropiada, multiplicada por una constante K muy pequeña. Resolviendo el problema lineal de la Ecuación (1.59), es posible actualizar la estimación del estado. Mediante [31]:

$$\hat{x}^{\ell+1} = \hat{x}^{\ell} + \Delta\hat{x}$$

Ecuación 1.65

Así, a partir de una estimación inicial \hat{x}^0 , igual al arranque plano, es decir, tensiones iguales a 1 y ángulos cero:

$$\hat{x}^0 = \begin{bmatrix} [0] \\ [1] \end{bmatrix}$$

Ecuación 1.66

la estimación de estados se puede resolver a través de un proceso iterativo, donde, en cada iteración, se aplican las Ecuaciones (1.59) y (1.65). La convergencia ocurre cuando la norma euclidiana de $\Delta\hat{x}$ es menor que una tolerancia preespecificada:

$$\|\Delta\hat{x}\|_2 \leq \epsilon$$

Ecuación 1.67

La aplicación de este procedimiento de estimación de estado es interesante para la metodología propuesta principalmente porque la solución de Tableau Esparso también proporciona los multiplicadores de Lagrange λ_m , λ_s y λ_o , relacionados con los tres tipos de información contenida en el proceso EESD. Estos vectores representan la sensibilidad de la función objetivo, en relación a sus restricciones, y se utilizan en procesos de detección de errores.[31]

1.5.11.12 Vector de ecuación y matriz jacobiana

El problema EESD consiste en resolver el sistema lineal dado por la Ecuación (1.59). Sin embargo, la información se modela a través del vector de ecuaciones $h(x)$, dado por la Ecuación (1.64), en función de las variables de estado. Tanto el vector $h(x)$ como su matriz jacobiana, $H(x)$, deben detallarse mejor para establecer el EESD.[15]

1.5.11.13 Naturaleza trifásica de la información

La ecuación de los elementos de $h(x)$ y $H(x)$, Sus matrices $[Z_{abc}]$ se expanden agregando filas y columnas nulas relativas a fases inexistentes. Este procedimiento crea

fases ficticias en los nodos conectados a estos LD. Por lo tanto, el sistema se vuelve completamente trifásico. [15]

En la Ecuación (1.33), la estructura vectorial presentada ya tiene en cuenta este efecto. El número de estados a estimar, dado por la Ecuación (1.47), también parte de este supuesto. Por tanto, es necesario comentar las medidas realizadas en los ramales bifásicos y monofásicos, así como en los nodos que son alimentados por la LD. En este escenario, las mediciones realizadas presentan los datos de medición para una o dos fases, ya que el nodo fue identificado por el algoritmo. Este procesamiento de información es importante para la determinación de $h(x)$ y $H(x)$. [15]

1.5.11.14 Cálculo de Elementos de vector de ecuaciones $h(x)$ y matriz jacobiana, $H(x)$,

Independientemente de representar medidas $h_m(x)$, o restricciones de igualdad $h_s(x)$ y $h_o(x)$, el vector de ecuaciones $h(x)$ proporciona el valor de una determinada información o medida, en función de las variables de estado x . Además de $h(x)$, su matriz jacobiana, $H(x)$, juega un papel importante, no solo por el cumplimiento de las condiciones de optimalidad, sino también por la linealización de $h(x)$, dada por la Ecuación (1.55). [15]

- Módulo de tensión;
- Módulo de corriente al cuadrado;
- Flujo de potencia activa y reactiva;
- Inyección de potencia activa y reactiva

Para aplicar la modelación propuesta, es necesario utilizar algunas de las ecuaciones generalizadas, definidas por las ecuaciones:

$$[I_{abc}]_u = [c] \cdot [V_{abc}]_d + [d] \cdot [I_{abc}]_d$$

Ecuación 1.68

$$[V_{abc}]_d = [A] \cdot [V_{abc}]_u + [B] \cdot [I_{abc}]_d$$

Ecuación 1.69

Es importante considerar la estructura del vector de estado, presentada en la Ecuación (1.33). Recordando que la dimensión de este vector está dada por la Ecuación (1.47), por lo que se puede acceder a los voltajes complejos en un nodo genérico i a través de [28]:

$$[V_{abc}]_i = \begin{bmatrix} x(3.n_b + 3.i - 2) \angle x_{(3,i-2)} \\ x(3.n_b + 3.i - 1) \angle x_{(3,i-1)} \\ x(3.n_b + 3.i) \angle x_{(3,i)} \end{bmatrix} = \begin{bmatrix} E_a \angle \delta_a \\ E_b \angle \delta_b \\ E_c \angle \delta_c \end{bmatrix}$$

Ecuación 1.70

1.5.11.15 Módulo de tensión

De toda la información adquirida por el plan de medida, las medidas de módulo de tensión son las más sencillas de formular, ya que están directamente presentes en el vector de estado. Así, una medida de tensión trifásica en el nodo i se puede expresar como:[15]

$$h_{E_i} = [E_{abc}]_i = \begin{bmatrix} x(3.n_b + 3.i - 2) \\ x(3.n_b + 3.i - 1) \\ x(3.n_b + 3.i) \end{bmatrix}$$

Ecuación 1.71

Al tratarse de una medida trifásica, el jacobiano relativo a $h_{E_i}(x)$ una matriz de dimensión $3 \times n_e$. Por lo tanto, es posible separar $H_{E_i}(x)$ en bloques de 3×3 , en relación con las fases a, b y c de cada nodo. Estos bloques, que representan derivadas parciales con respecto a variables insertadas en x , se pueden determinar por:[15]

$$\frac{\delta h_{E_i}(x)}{\delta [E_{abc}]_k} = [U] \quad , \quad \text{Para } k = i;$$

$$\frac{\delta h_{E_i}(x)}{\delta [E_{abc}]_k} = [0] \quad , \quad \text{Para } k \neq i;$$

$$\frac{\delta h_{E_i}(x)}{\delta [E_{abc}]_k} = [0] \quad , \quad \forall k,$$

Ecuación 1.72

Finalmente, la matriz jacobiana $H_{E_i}(x)$ se puede determinar a partir de estos bloques, mediante:

$$H_{E_i}(x) = \begin{bmatrix} \left[\frac{\partial h_{E_i}(x)}{\partial [\delta_{abc}]_1} \right]^T \\ \vdots \\ \left[\frac{\partial h_{E_i}(x)}{\partial [\delta_{abc}]_{nb}} \right]^T \\ \left[\frac{\partial h_{E_i}(x)}{\partial [E_{abc}]_1} \right]^T \\ \vdots \\ \left[\frac{\partial h_{E_i}(x)}{\partial [E_{abc}]_{nb}} \right]^T \end{bmatrix}^T$$

Ecuación 1.73

1.5.11.16 Módulo de corriente al cuadrado

Se requieren dos modelos diferentes para las mediciones del módulo de corriente al cuadrado, ya que los medidores se pueden colocar tanto aguas abajo como aguas arriba de un ramal determinado.[15]

Medidor aguas abajo del elemento

Para la determinación de las corrientes, considere la Ecuación (1.69). Resolviéndolo para la corriente $[I_{abc}]_d$:

$$[I_{abc}]_d = [B]^{-1} \cdot ([\mathcal{A}][V_{abc}]_u - [V_{abc}]_d)$$

Ecuación 1.74

A partir de la Ecuación (1.74), se puede determinar $h_{I_e^2}(x)$ Considere un medidor ubicado aguas abajo de cualquier ramal, que conecta los nodos i y j . La corriente de línea trifásica que lo atraviesa viene dada por[15]:

$$[I_{abc}]_{ji} = \begin{bmatrix} I_a \\ I_b \\ I_c \end{bmatrix} = [B]^{-1} \cdot ([\mathcal{A}][V_{abc}]_i - [V_{abc}]_j)$$

Ecuación 1.75

donde las tensiones $[V_{abc}]_i$ y $[V_{abc}]_j$ se puede acceder desde el vector de estado por la Ecuación (1.70).

Aplicando la operación diagonal en $[I_{abc}]_{ji}$

$$diag\{[I_{abc}]_{ji}\} = diag \left\{ \begin{bmatrix} I_a \\ I_b \\ I_c \end{bmatrix}_{ji} \right\} = \begin{bmatrix} I_a & 0 & 0 \\ 0 & I_b & 0 \\ 0 & 0 & I_c \end{bmatrix}_{ji}$$

Ecuación 1.76

el vector $h_{ji}^2(x)$ se puede determinar de la siguiente manera:

$$h_{ji}^2(x) = \text{diag}\{[I_{abc}]_{ji}\} \cdot [I_{abc}]_{ji}^*$$

Ecuación 1.77

Para determinar la matriz jacobiana, es necesario aplicar derivadas parciales, con respecto al vector de estado. De la Ecuación (1.75), se observa que la corriente de rama depende solo de los voltajes aguas abajo y aguas arriba. Así[15]:

$$\frac{\partial h_{ji}^2(x)}{\partial [E_{abc}]_k} = [0], \frac{\partial h_{ji}^2(x)}{\partial [\delta_{abc}]_k} = [0], \text{ Para } k \neq i \text{ e } k \neq j$$

Ecuación 1.78

Para determinar las demás derivadas parciales, es necesario derivar la Ecuación (1.75) para $[V_{abc}]_i$ y $[V_{abc}]_j$ Así:

$$\frac{\partial [I_{abc}]_{ij}}{\partial [E_{abc}]_i} = [B]^{-1} \cdot [A] \text{diag}\{[1\angle\delta_a \ 1\angle\delta_b \ 1\angle\delta_c]_i\}$$

$$\frac{\partial [I_{abc}]_{ij}}{\partial [E_{abc}]_i} = -[B]^{-1} \cdot \text{diag}\{[1\angle\delta_a \ 1\angle\delta_b \ 1\angle\delta_c]_j\}$$

$$\frac{\partial [I_{abc}]_{ij}}{\partial [E_{abc}]_i} = [B]^{-1} \cdot [A] \text{diag}\left\{1\angle\frac{\pi}{2}, [V_{abc}]_i\right\}$$

$$\frac{\partial [I_{abc}]_{ij}}{\partial [E_{abc}]_i} = -[B]^{-1} \cdot \text{diag}\left\{1\angle\frac{\pi}{2}, [V_{abc}]_j\right\}$$

Ecuación 1.79

Debido a que el operador diagonal $\text{diag}\{.\}$ es solo un operador lógico, y considerando la linealidad de la operación compleja conjugada, las derivadas de $h_{ji}^2(x)$ puede determinarse aplicando la regla del producto.[15]

$$\frac{\partial h_{ji}^2(x)}{\partial [E_{abc}]_i} = \text{diag}\{[I_{abc}]_{ji}\} \cdot \frac{\partial [I_{abc}]_{ij}^*}{\partial [E_{abc}]_i} + \text{diag}\{[I_{abc}]_{ji}^+\} \cdot \frac{\partial [I_{abc}]_{ji}}{\partial [E_{abc}]_i}$$

$$\frac{\partial h_{ji}^2(x)}{\partial [E_{abc}]_j} = \text{diag}\{[I_{abc}]_{ji}\} \cdot \frac{\partial [I_{abc}]_{ij}^*}{\partial [E_{abc}]_j} + \text{diag}\{[I_{abc}]_{ji}^+\} \cdot \frac{\partial [I_{abc}]_{ji}}{\partial [E_{abc}]_j}$$

$$\frac{\partial h_{ji}^2(x)}{\partial [\delta_{abc}]_i} = \text{diag}\{[I_{abc}]_{ji}\} \cdot \frac{\partial [I_{abc}]_{ij}^*}{\partial [\delta_{abc}]_i} + \text{diag}\{[I_{abc}]_{ji}^+\} \cdot \frac{\partial [I_{abc}]_{ji}}{\partial [\delta_{abc}]_i}$$

$$\frac{\partial h_{I_{ji}^2}(x)}{\partial [\delta_{abc}]_j} = \text{diag}\{[I_{abc}]_{ji}\} \cdot \frac{\partial [I_{abc}]_{ij}^*}{\partial [E_{abc}]_i} + \text{diag}\{[I_{abc}]_{ji}^+\} \cdot \frac{\partial [I_{abc}]_{ji}}{\partial [E_{abc}]_i}$$

Ecuación 1.80

Seguendo la estructura de bloques jacobiana 3x3 dada por la Ecuación (1.73), y de las Ecuaciones (1.78) y (1.80), la matriz $H_{I_{ji}^2}(x)$ se expresa por:[15]

$$H_{I_{ji}^2}(x) = \begin{bmatrix} \left[\frac{\partial h_{I_{ji}^2}(x)}{\partial [\delta_{abc}]_1} \right]^T \\ \vdots \\ \left[\frac{\partial h_{I_{ji}^2}(x)}{\partial [\delta_{abc}]_{nb}} \right]^T \\ \left[\frac{\partial h_{I_{ji}^2}(x)}{\partial [E_{abc}]_1} \right]^T \\ \vdots \\ \left[\frac{\partial h_{I_{ji}^2}(x)}{\partial [E_{abc}]_{nb}} \right]^T \end{bmatrix}^T$$

Ecuación 1.81

1.5.11.17 Medidor Aguas Arriba del Elemento

Elemento Para la medición ubicada aguas arriba del ramal se utiliza la Ecuación (1.131):

$$[I_{abc}]_{ij} = \begin{bmatrix} I_a \\ I_b \\ I_c \end{bmatrix}_{ij} = [c] \cdot [V_{abc}]_j + [d] \cdot [I_{abc}]_{ji}$$

Ecuación 1.82

Mediante el uso de la Ecuación (1.75), y la operación diagonal, se puede calcular el vector de medidas $h_{I_{ij}^2}(x)$, similar al presentado en la Ecuación (1.77):

$$h_{I_{ij}^2}(x) = \text{diag}\{[I_{abc}]_{ij}\} \cdot [I_{abc}]_{ij}^*$$

Ecuación 1.83

Similar a las derivadas presentadas en la Ecuación (1.75), la derivación parcial $h_{I_{ij}^2}(x)$, de con respecto a los voltajes $[v_{abc}]_k$, con $k \neq i$ y $k \neq j$, son nulas.

La derivada de la Ecuación (1.82), con respecto a las tensiones $[v_{abc}]_i$, y $[v_{abc}]_j$, viene dada por:

$$\frac{\partial [I_{abc}]_{ij}}{\partial [E_{abc}]_i} = \frac{\partial [I_{abc}]_{ij}}{\partial [E_{abc}]_i} \cdot [d]$$

$$\frac{\partial [I_{abc}]_{ij}}{\partial [E_{abc}]_j} = \text{diag}\{[1\angle\delta_a \quad 1\angle\delta_b \quad 1\angle\delta_c]_j\} \cdot [c] + \frac{\partial [I_{abc}]_{ij}}{\partial [E_{abc}]_j} \cdot [d]$$

$$\frac{\partial [I_{abc}]_{ij}}{\partial [\delta_{abc}]_i} = \frac{\partial [I_{abc}]_{ij}}{\partial [\delta_{abc}]_i} \cdot [d]$$

$$\frac{\partial [I_{abc}]_{ij}}{\partial [\delta_{abc}]_j} = \text{diag}\left\{ \left[1\angle\frac{\pi}{2} \cdot [V_{abc}]_j \right]_j \right\} \cdot [c] + \frac{\partial [I_{abc}]_{ij}}{\partial [\delta_{abc}]_j} \cdot [d]$$

Ecuación 1.84

donde las derivadas parciales de $[I_{abc}]_{ji}$ vienen dadas por la Ecuación (1.79).

Con la aplicación de las fórmulas dadas por las Ecuaciones (1.80) y (1.81), se puede determinar la matriz jacobiana de medidas, $H_{ij}^2(x)$, puede ser determinada.

1.5.11.18 Matriz [B] Singular

Los elementos SD tienen sus matrices [B] singulares. Incluso existe la posibilidad de que esta matriz sea nula, como en el caso del regulador de voltaje. Por lo tanto, la aplicación de la Ecuación (1.74) está comprometida, ya que es necesario invertir [B]. En la aplicación práctica, hay dos formas de resolver este problema. Para una matriz singular distinta de cero [B], se puede generalizar su inversa, utilizando la pseudoinversa de Moore Penrose [34].

$$[I_{abc}]_d = [B]^{\dagger} \cdot ([A] \cdot [V_{abc}]_u - [V_{abc}]_d)$$

Ecuación 1.85

Este procedimiento no provoca errores en el EESD, debido a las propiedades de esta operación. Para una matriz nula [B], la solución adoptada es la aglutinación de elementos en cascada. El procedimiento se puede realizar con el ramal de aguas arriba o de aguas abajo, según la posición del contador [15].

1.5.11.19 Flujo de Potencia

De manera similar a las medidas de corriente, la información relativa al flujo de potencia, tanto activa como reactiva, depende de la posición del medidor en el elemento, ya sea aguas arriba o aguas abajo de su ramal [15].

1.5.11.20 Medidor aguas abajo del elemento

Para mediciones de potencia activa y reactiva aguas abajo de un elemento, considere la rama que conecta los nodos i y j . La potencia aparente compleja incidente en el nodo j , procedente del elemento, viene dada por:

$$h_{s_{ji}}(x) = [s_{abc}]_{ji} = \text{diag}\{[V_{abc}]_j\} \cdot [I_{abc}]_{ij}^*$$

Ecuación 1.86

Los vectores de medida se pueden obtener a partir de la Ecuación (1.86), mediante las operaciones real $\Re\{\cdot\}$ e imaginario $\Im\{\cdot\}$:

$$h_{p_{ji}}(x) = \Re\{h_{s_{ji}}(x)\}$$

$$h_{q_{ji}}(x) = \Im\{h_{s_{ji}}(x)\}$$

Ecuación 1.87

Para la determinación de las matrices jacobianas se consideran las derivadas parciales de las tensiones $[V_{abc}]_j$, con respecto a las tensiones $[V_{abc}]_i$ y $[V_{abc}]_j$.

$$\frac{\partial [I_{abc}]_{ij}}{\partial [E_{abc}]_i} = [0]$$

$$\frac{\partial [V_{abc}]_{ij}}{\partial [E_{abc}]_j} = \text{diag}\{[1\angle\delta_a \quad 1\angle\delta_b \quad 1\angle\delta_c]_j\}$$

$$\frac{\partial [V_{abc}]_j}{\partial [E_{abc}]_i} = [0]$$

$$\frac{\partial [I_{abc}]_{ij}}{\partial [\delta_{abc}]_j} = \text{diag}\left\{1\angle\frac{\pi}{2} \cdot [V_{abc}]_j\right\}$$

Ecuación 1.88

Debido a las similitudes de la Ecuación (1.77) con la Ecuación (1.87), la matriz jacobiana $H_{s_{ji}}(x)$ se puede obtener utilizando la regla del producto, según la Ecuación (1.80), así como utilizando la estructura lógica que presenta la Ecuación (1.81). Por lo tanto, las matrices jacobianas de medidas de potencia activa y reactiva aguas abajo de un elemento se pueden determinar a partir de $H_{s_{ji}}(x)$, mediante [15]:

$$H_{p_{ji}}(x) = \Re\{H_{s_{ji}}(x)\}$$

$$H_{q_{ji}}(x) = \Im\{H_{s_{ji}}(x)\}$$

Ecuación 1.89

Medidor aguas arriba del elemento

Para un medidor ubicado aguas arriba de un elemento, el procedimiento para obtener los vectores de medida, $h_{p_{ji}}(x)$ y $h_{q_{ji}}(x)$, así como las matrices jacobianas, $H_{p_{ji}}(x)$, y $H_{q_{ji}}(x)$, es muy similar a los valores presentados anteriormente, con solo dos diferencias. Primero, la ecuación de $h_{s_{ji}}(x)$, viene dada por [15]:

$$h_{s_{ji}}(x) = [s_{abc}]_{ij} = \text{diag}\{[V_{abc}]_j\} \cdot [I_{abc}]_{ij}^*$$

Ecuación 1.90

Finalmente, en lugar de derivadas parciales de $[V_{abc}]_j$, se necesitan derivadas de $[V_{abc}]_i$:

$$\frac{\partial [V_{abc}]_i}{\partial [E_{abc}]_i} = \text{diag}\{[1 \angle \delta_a \quad 1 \angle \delta_b \quad 1 \angle \delta_c]_j\}$$

$$\frac{\partial [V_{abc}]_j}{\partial [E_{abc}]_i} = [0]$$

$$\frac{\partial [V_{abc}]_{ij}}{\partial [\delta_{abc}]_j} = \text{diag}\left\{1 \angle \frac{\pi}{2} \cdot [V_{abc}]_j\right\}$$

$$\frac{\partial [V_{abc}]_j}{\partial [\delta_{abc}]_i} = [0]$$

Ecuación 1.91

1.5.11.21 Inyección de potencia

La potencia aparente compleja inyectada en un nodo i es igual a la potencia suministrada por el elemento aguas arriba de i , restada de los poderes que fluyen a través de las ramas inmediatamente aguas abajo de este mismo nodo. Así, considerando k_0 el elemento aguas arriba de i , y k los elementos aguas abajo, se puede determinar el vector de ecuaciones $h_{s_i}(x)$ [15].

$$h_{s_i}(x) = [h_{s_{ii}}(x)]_{k_0} - \sum_{k \in \Omega_i} [h_{s_{ii}}(x)]_k$$

Ecuación 1.92

donde Ω_i es el conjunto de ramas aguas abajo de i . A partir de la Ecuación (1.92), se puede determinar $h_{p_i}(x)$ y $h_{q_i}(x)$:

$$h_{p_i}(x) = \Re\{h_{s_i}(x)\}$$

$$h_{Q_i}(x) = \Im\{h_{s_i}(x)\}$$

Ecuación 1.93

Debido a las propiedades de derivación, también se puede obtener la matriz jacobiana, de manera similar a la Ecuación (1.92):

$$H_{S_i}(x) = [H_{S_{ii}}(x)]_{k0} - \sum_{k \in \Omega_i} [H_{S_{ii}}(x)]_k$$

De esta manera:

$$H_{P_i}(x) = \Re\{H_{S_i}(x)\}$$

$$H_{Q_i}(x) = \Im\{H_{S_i}(x)\}$$

Ecuación 1.94

La aplicación de esta metodología al problema de detección e identificación de PNTs se presenta en la siguiente sección, junto con las principales definiciones necesarias para el análisis de los resultados posteriores.

1.5.11.22 Detección e identificación de errores graves

Una de las funciones esenciales del estimador de estado es detectar la existencia de errores brutos de medición e identificarlos y eliminarlos si es posible. Siempre existen pequeños errores de medición aleatorios debido a la precisión finita de los medidores y la pérdida de datos debido a los métodos de telecomunicaciones, sin embargo, se espera que el estimador pueda filtrar estos errores dado que hay suficiente redundancia de mediciones. Los errores graves ocurren debido a varios factores como, por ejemplo, mal funcionamiento del medidor o alteración intencionada, malas conexiones y errores en la transmisión de datos. Algunos errores graves solo pueden ser detectados mediante una verificación de plausibilidad como magnitudes de voltaje negativas, valores muy diferentes a los esperados o corriente con dirección incorrecta en la salida de subestaciones, etc. Desafortunadamente, no todos los errores graves son fácilmente detectables y, por lo tanto, el estimador debe estar equipado con métodos más avanzados para detectarlos. El manejo de errores brutos depende del método de solución utilizado para realizar la estimación. En el caso de la estimación por el método de mínimos cuadrados ponderados, la detección e identificación de errores brutos se puede realizar mediante el procesamiento de residuos de medición [15].

1.5.11.22.1 Pruebas geométricas de colinealidad

Habiendo normalizado los resultados de la prueba de multiplicadores de Lagrange, se puede aplicar la técnica de prueba de colinealidad geométrica, con el objetivo de afinar las medidas que presentan inconsistencias, permitiendo discernir qué información es realmente errónea de otras incorrectamente incluidas en el análisis preliminar.

1.5.11.23 Particularidades de la estimación de estados en redes de distribución

Los estimadores de estado se utilizan generalmente para sistemas de transmisión, sin embargo, los sistemas de distribución tienen características diferentes a los sistemas de transmisión y, por lo tanto, estas diferencias deben tenerse en cuenta. Se puede ver más información sobre la estimación en los sistemas de distribución en [9] y [11].

1.5.11.23.1 Información disponible

La Figura 1.17 representa, de forma genérica, los componentes presentes en un sistema de distribución moderno, con recursos de medición y comunicación de datos disponibles en el entorno de Red Eléctrica Inteligente (Smart Grid). En este sistema, están disponibles varios tipos de medidas, como se indica en los siguientes ítems.

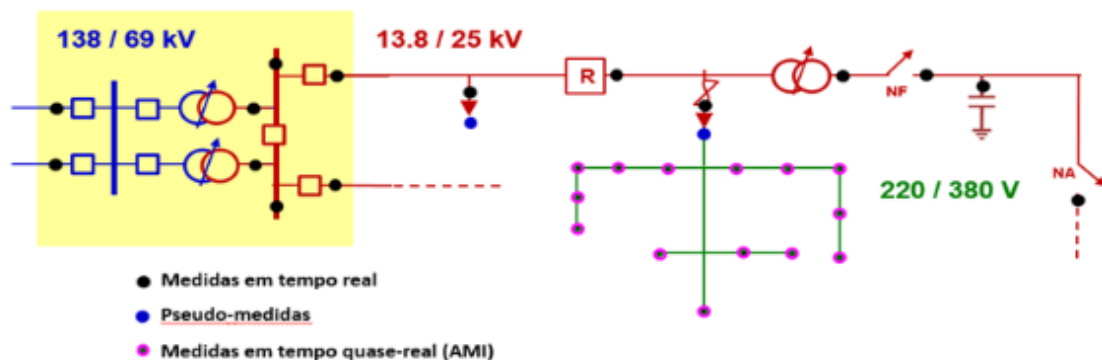


Figura 1.17 Esquema unifilar de un sistema de distribución de media y baja tensión, con medidas disponibles.

1.5.11.23.2 Medidas en tiempo real

Los sistemas de distribución modernos cuentan con centros de supervisión y control basados en el sistema SCADA (Supervisory Control and Data Acquisition). Estos centros, generalmente denominados Centros de Operación de Distribución (COD), están

conectados a Unidades Terminales Remotas (RTU) instaladas en subestaciones, que adquieren y transmiten al COD medidas analógicas de tensión, corriente, potencia activa y reactiva, además del estado (encendido / apagado) de interruptores y disyuntores.

Algunos dispositivos instalados a lo largo de alimentadores de media tensión, como reconectores, controladores de tensión, bancos de condensadores, etc., también tienen unidades para adquirir información sobre cantidades eléctricas y transmitirla al COD. La información obtenida por el sistema SCADA se recopila escaneando las RTU a intervalos de 1-2 segundos y, por lo tanto, a los efectos de trabajar en esta disertación, se pueden considerar como datos en tiempo real.

1.5.11.23.3 Medidas cuasi en tiempo real (AMI)

La implementación de contadores inteligentes (Smart Meters) representa una fuente adicional de información sobre las redes de distribución. Los contadores inteligentes se instalan individualmente en cada unidad de consumo y proporcionan información sobre el consumo / generación de potencia activa y reactiva de esa unidad. Otras magnitudes eléctricas, como voltaje, corriente, frecuencia, etc. también se puede medir, así como información de subtensión y sobretensión. Frecuencia y duración de los cortes de energía, etc. Esta información se transmite a los llamados centros de medición y se utiliza principalmente para facturación y otras aplicaciones comerciales. También se pueden enviar a los COD para su uso en el control y la supervisión de la red. Para evitar requisitos de comunicación excesivos, esta información se comprime (valores acumulativos, promedios, valores extremos, etc.) a intervalos típicos de 15 minutos y, por lo tanto, se transmite a los COD. Por tanto, la información proporcionada no puede ser considerada en tiempo real, como en los casos SCADA, y en este trabajo se hará referencia a ellos como información casi en tiempo real.

1.5.11.23.4 Pseudo medidas

En el área de interés puede haber partes inobservables, debido a pérdida temporal de medición o comunicación con los medidores, rechazo de errores graves u otras fallas inesperadas. En este caso, la estimación del estado de esta área se realizará mediante pseudomedidas. Las pseudomedidas son medidas generadas basadas en pronósticos de carga a corto plazo, pronósticos de despacho de generación, datos históricos u otros métodos de aproximación similares. Por ser medidas estimadas, pueden tener altos errores y, por tanto, se les asocian varianzas mayores que las demás, menor peso, o pueden verse obligadas a ser medidas críticas.

1.5.11.23.5 Medidas virtuales

El sistema puede tener buses sin generación ni carga, por lo tanto, la inyección de potencia es igual a cero, en este caso, aunque no exista realmente una medición de bus, esta información se puede utilizar para crear una medición libre de errores. Este tipo de medición se conoce como medición virtual [2].

2 METODOLOGÍA

2.1 Estado del arte

Los avances para la identificación de pérdidas no técnicas han sido evidentes en los últimos años, Los modelos de identificación de pérdidas no técnicas se valen de informaciones para detectar patrones o encontrar áreas donde hay grandes cantidades de pérdidas no técnicas de energía; así como también metodologías basada en mediciones en campo a los centros de transformación de la red de distribución y técnicas basadas en inteligencia artificial, por lo que es necesario investigar métodos alternativos que tengan mayor flexibilidad y se adapten fácilmente al contexto del problema.

En la referencia [6] se presenta, uno de los métodos más tradicionales de identificación de Pérdidas no Técnicas denominado Escalón de Consumo. El cual es proveniente del análisis del histórico de consumo del usuario. Cuando el consumo del usuario es reducido significativamente por algunos meses consecutivos, está caracterizado el escalón de consumo. Este es uno de los métodos tradicionales de identificación de Pérdidas no técnicas utilizados por las distribuidoras y que se basan en la experiencia de los especialistas de la empresa. La clasificación de clientes sospechosos a partir del análisis de sus datos de factura realizado por un especialista de la empresa de distribución es una práctica poco eficiente. [6] En la literatura, las experiencias de este tipo documentadas apuntan que las sospechas apuntadas por este tipo de método eran solo confirmadas en 13% a 15%. [7]

Para el método de la referencia [8] se basa en la utilización de medidores portátiles en miniatura para el procedimiento de inspección previa. Estos contadores se instalan discretamente en el ramal de conexión del usuario y registran su consumo durante algún tiempo. Luego se compara la lectura del medidor miniatura instalado y la del medidor del usuario y si hay grandes discrepancias entre las dos medidas, se realiza una inspección. En proyectos piloto, la tasa de éxito en el reconocimiento de unidades de consumo irregulares y regulares fue del 100%. La desventaja de esta metodología es la considerable cantidad que se necesita de equipos y personal técnico que realicen las mediciones a los usuarios sospechosos.

Entre los métodos utilizados para detectar Pérdida No Técnica de energía, que se encuentran en la literatura, son los que utilizan inteligencia artificial para señalar posibles fraudes o defectos en los contadores de energía en los sistemas de distribución. Estos estudios han demostrado que, con la ayuda de programas informáticos con capacidad de aprendizaje, es posible aumentar la posibilidad de encontrar contadores de energía donde

hay Pérdida no Técnica de Energía. Las metodologías que utilizan inteligencia artificial pueden detectar automáticamente nuevos patrones, así como buscar patrones ya conocidos por la experiencia humana, mejorando así la tasa de éxito esperada en la inspección en sitio. A continuación, se citan algunas metodologías.

Máquinas de Vector de Soporte (SVM), es una técnica de aprendizaje automático que se ha utilizado con éxito en la detección y localización de Pérdidas No Técnicas de Energía principalmente porque tiene la capacidad de brindar soluciones de gran generalidad a problemas de clasificación de patrones. La clasificación es una función no lineal, cuya entrada son las características de las muestras y la salida es la clase a la que pertenecen. La formulación original de SVM trata con problemas de clasificación binaria, pero puede extenderse a problemas de clasificación con múltiples clases a un alto costo computacional. La SVM puede realizar aprendizaje supervisado o no supervisado. El método de clasificación se utilizó para entrenar y probar el clasificador SVM utilizando los datos empíricos. El conjunto de datos constaba de 3 156 clientes. Este conjunto de datos se particionó aleatoriamente en un conjunto de entrenamiento y de prueba. Aproximadamente dos tercios de la base de datos formaron un conjunto de entrenamiento y el tercio restante un conjunto de prueba. El conjunto de capacitación consistió en 2 102 clientes, de los cuales aproximadamente el 78% eran casos limpios y el 22% fraudulentos. En comparación, el equipo de prueba estaba compuesto por 1052 clientes, de los cuales el 75% estaban limpios y el 25% eran fraudulentos. [9]

Redes neuronales artificiales esta metodología busca realizar el procesamiento de datos imitando una red neuronal natural, como el cerebro humano. Para la activación de la técnica, se requieren un historial de consumo del cliente, se comprobó que el aprendizaje que esta metodología implementada mejora significativamente la tasa de éxito en la clasificación de clientes identificados como sospechosos. Una de las limitaciones de esta técnica es que la identificación de consumidores irregulares depende en gran medida de los datos utilizados en la formación, que se obtienen a través de las inspecciones en campo a los contadores de energía, que caracteriza el aprendizaje supervisado. Por lo tanto, las redes neuronales aprenden a identificar solo los patrones de clientes irregulares dentro del grupo de consumidores considerados fraudulentos. Esta metodología puede ser criticada por limitar el aprendizaje del reconocimiento de patrones a una base de entrenamiento que no forma una muestra estadística significativa. Una alternativa sería obtener datos de inspecciones realizadas al azar para obtener una medida estadística más significativa o emplear métodos de aprendizaje no supervisados. Los resultados obtenidos del sistema propuesto no fueron muy satisfactorios, debido a que sólo el 65.03% de los clientes preseleccionados para ser inspeccionados en campo resultaron con fraude [10]

K- medias El k-medias (K-means) es un algoritmo que pertenece a las técnicas de agrupamiento no jerárquico o particionales, y es el más popular dentro de esta categoría. Se trata de un método en el que se parte de un número determinado de prototipos y de un conjunto de elementos a agrupar, sin etiquetar. Este algoritmo elige un punto inicial (centroide) para representar cada uno de los grupos, luego se asigna cada objeto al grupo cuyo centroide es el más próximo a él o tenga características similares. Seguidamente, el centroide de cada grupo es recalculado como la media de los puntos asignados al grupo. Este proceso que se repite hasta que alcance un número limitado de pasos de iteración o, simplemente, que la diferencia obtenida entre los centroides de dos pasos consecutivos sea menor que un valor prefijado. Los centros se determinan con el objetivo de minimizar la suma de distancias cuadráticas euclidianas entre los patrones de entrada y el centro más cercano. Los datos usados para activar esta técnica corresponden a 5000 perfiles de clientes industriales de una empresa brasileña, donde sólo 280 de ellos presentan pérdidas no técnicas. También se aportan algunas características que consideran relevantes como: demanda facturada, demanda contratada, demanda máxima, entre otras. Los datos se dividieron en 50% para el entrenamiento y el 50% para la prueba. En los resultados obtenidos se observó que los algoritmos mejoran su desempeño cuando se incluyen ciertas características de los clientes. Este enfoque presenta algunas debilidades, una de ellas es que no toma en cuenta el desequilibrio de las clases en el conjunto de entrenamiento. [11]

La aplicación de los algoritmos de lógica difusa en esta técnica la tarea de agrupamiento es una de las más utilizada para determinar perfiles de carga. En la referencia [12] por ejemplo se emplea para identificar perfiles de consumo sospechosos comparándolos con perfiles de consumo regulares. Este esquema se compone de dos pasos: En el primer paso, se utiliza el algoritmo C-medias difuso para agrupar los clientes dentro de las clases correspondientes, es decir, con perfiles similares. Posteriormente, mediante una matriz de pertenencia difusa y la distancia euclídea a los centros de agrupamiento se clasifican los clientes en fraudulentos o no fraudulentos. Para desarrollar el algoritmo se utilizaron los datos históricos de consumo de los últimos seis meses de 20,126 clientes de un área residencial y cinco atributos para el estudio (promedio de consumo, máximo consumo, desviación estándar, cantidad de inspecciones y promedio de consumo del área residencial). La tasa de éxito obtenida fue de 74.5%. [12]

Bosque de caminos óptimos (Optimum-Path Forest - OPF). En este trabajo, el problema de identificación de la Pérdidas no técnicas de Energía se trata como un problema de reconocimiento de patrones. Por lo tanto, se asume que, a partir del análisis de los datos

disponibles, es posible enseñar a un clasificador a encontrar patrones de Pérdidas no técnicas de Energía de manera eficiente. Los datos para llevar a cabo la investigación fueron proporcionados por una empresa brasileña, los cuales corresponden a 736 perfiles de clientes, divididos en 116 ilegales y 620 legales. Para evaluar los modelos el conjunto de datos se dividió en 50% para la fase de entrenamiento y 50% para la fase de prueba. En los resultados registrados, el modelo obtuvo una precisión de 90.21%. En los resultados registrados se observa que la calidad y el tiempo de aprendizaje del método mejoran mucho cuando se realiza una selección óptima de características. Los resultados indican que el método propuesto mejora el rendimiento de técnicas previamente desarrolladas y es adecuado tanto para Empresa de Distribución típicas como inteligentes. Los mejores resultados se obtienen cuando se utilizan datos normalizados como entrada, con una tasa de éxito esperada de hasta 72,43% de la inspección in situ. [13]

Un enfoque particularmente prometedor para la detección y cuantificación de pérdidas no técnicas se basa en métodos de estimación de estados. Tales métodos tienen una larga historia de éxito en sistemas de transmisión y desde la década de 90 su aplicación a redes de distribución ha sido investigada [14].

En el artículo de la referencia [15] se propone una metodología basada en la estimación del estado para un sistema de distribución y la prueba de error grueso basada en la colinealidad, capaz de no solo identificar la ocurrencia de pérdidas no técnicas, sino también de estimar su magnitud. Para ese propósito, se emplea un estimador de estado trifásico que procesa las mediciones disponibles en tiempo real y carga los valores de pronóstico.

La estimación del estado sirve como un filtro para pequeños errores que hacen que este método sea más robusto, el estimador de estado puede detectar los nodos con datos de demanda inconsistentes. La estimación de estado busca obtener la mejor estimación posible, en el sentido de los mínimos cuadrados ponderados, donde su resolución proporciona los errores en las mediciones, las mismas que son normalizadas y ayuda a tener un indicio en primera instancia de que nodos presentan inconsistencias. Sin embargo, aunque los errores normalizados suministren los datos inconsistentes en los nodos, no es posible discernir precisamente qué nodos presentan anomalías en su sistema de medición de manera efectivamente. Por lo que adicional se puede hacer un análisis de error grueso a los resultados del estimador de estado y detectar consumos fraudulentos. [15]

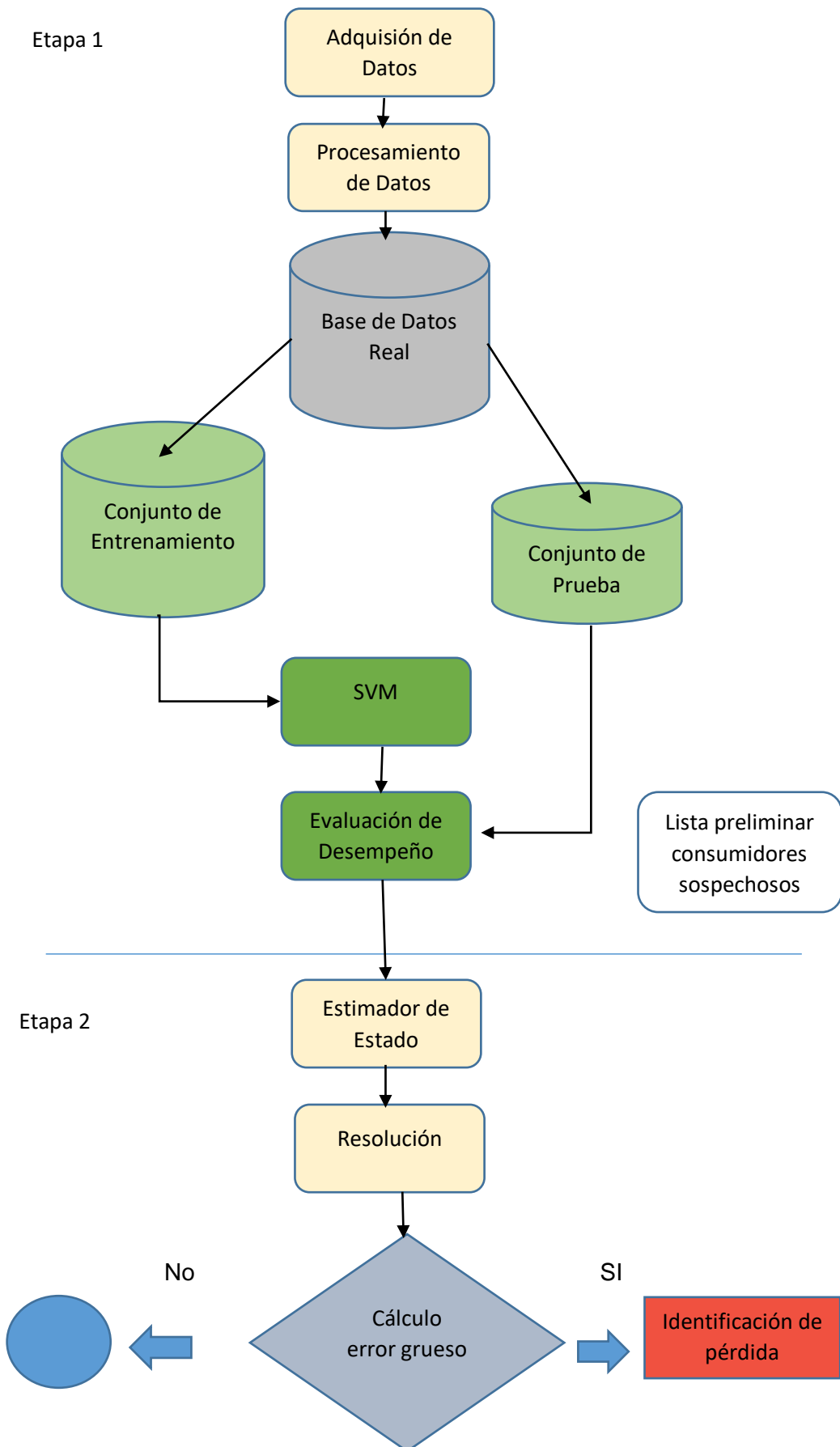
Lo que se propone desarrollar en este trabajo de titulación es la **combinación de dos metodologías para la Identificación de Pérdidas no técnicas de Energía** en los sistemas de distribución, para lo cual se combinarán un clasificador y un estimador de estado, el clasificador seleccionado es el basado en Máquinas de Vector de Soporte (SVM) debido a las ventajas que ofrece comparado con otros, como son la capacidad de manejar el problema del sobreajuste en el conjunto de entrenamiento y su metodología para minimizar el error en la clasificación [16] y un Estimador de Estado desbalanceado con un enfoque en mínimos cuadrados ponderados que se aproxima más al estado real del sistema de distribución[15].

Esta metodología propuesta pretende en su primera fase clasificar a los consumidores en sospechosos y no sospechosos de una base de datos de consumo mensual utilizando un clasificador basado en SVM. Esta información se trasladará al estimador de estado, refinando su desempeño. La tasa de éxito para encontrar consumidores con fraude es de 75% al utilizar solo el clasificador SVM. Con la metodología propuesta se pretende mejorar esta tasa de éxito, **además por la naturaleza de los sistemas de distribución se intenta aportar en el desarrollo de un estimador que permita la inclusión de ramales bifásicos y monofásicos.**

2.2 Propuesta metodológica

La metodología propuesta consiste en la unión de dos enfoques diferentes para la identificación de pérdidas no técnicas. Este método analiza la base de datos de clientes de una empresa distribuidora nacional y busca patrones considerados anómalos o correlacionados con patrones irregulares conocidos de clientes sospechosos de fraude de energía. Suponiendo que este método es capaz de indicar a la mayoría de estos clientes irregulares con una alta tasa de aciertos, esta información adicional se puede pasar al estimador de estado, refinando su desempeño y mejorar la tasa de acierto en la identificación de pérdidas no técnicas. La metodología se resume en la figura 2.1

Figura 2.1 Flujograma Método propuesto



2.1.1 Etapa 1 Máquinas de Vector de Soporte (SVM)

La clasificación de datos estadísticos es una forma de análisis de datos que se puede utilizar para extraer información útil de los datos. Dicho análisis se utiliza para construir modelos o clasificadores para describir clases de datos importantes. Por ejemplo, el clasificador se puede utilizar para clasificar una solicitud de préstamo bancario como "segura" o "arriesgada". Otro ejemplo sería asignar un correo electrónico a "spam" o "no spam", o clasificar una transacción con tarjeta de crédito como "fraudulenta" o "no fraudulenta". Hay muchas aplicaciones de clasificación de datos, incluida la detección de fraudes, el marketing objetivo, la fabricación y el diagnóstico médico, etc. En este estudio, se plantea la clasificación de datos estadísticos para analizar las pérdidas no técnicas de energía eléctrica.

2.1.2 Procesamiento de datos

La clasificación de datos es un proceso de dos pasos. El primer paso, llamado "fase de entrenamiento", implica construir un clasificador a partir de un conjunto de datos utilizando un algoritmo de clasificación. En este paso, un algoritmo de clasificación construye un clasificador de un conjunto de entrenamiento que comprende tuplas de bases de datos y sus etiquetas correspondientes. Una tupla se caracteriza por un vector n dimensional, $X = x_1, x_2, \dots, x_n$ que muestra n mediciones realizadas en la tupla de n atributos de la base de datos, A_1, A_2, \dots, A_n , respectivamente. Cada atributo representa una característica de X . Se supone que cada tupla pertenece a una clase predefinida que está determinada por otro atributo llamado "etiqueta de clase". Cada valor del atributo "etiqueta de clase" sirve como una categoría o una clase, y estos valores son discretos y desordenados. Las tuplas individuales que componen el conjunto de entrenamiento se denominan tuplas de entrenamiento. Hay dos tipos de fase de formación, a saber, "aprendizaje supervisado" y "aprendizaje no supervisado". En "aprendizaje supervisado", se conoce la etiqueta de clase de cada tupla, mientras que en "aprendizaje no supervisado", las etiquetas de clase no se conocen de antemano.[9]

En el segundo paso, se utiliza un conjunto de prueba para evaluar el rendimiento del clasificador aprendido, esto se llama la "fase de prueba". Este conjunto consta de tuplas de prueba y etiquetas de clase asociadas. Estas tuplas se seleccionan al azar del conjunto de datos generales que no se utilizaron para construir el clasificador. Es convencional designar dos tercios de los datos como conjunto de entrenamiento y el tercio restante

como conjunto de prueba. "La precisión de un clasificador en un conjunto de pruebas dado es la proporción de tuplas de conjuntos de pruebas que el clasificador clasifica correctamente". Al evaluar el rendimiento, la etiqueta de clase de cada tupla de prueba se compara con la etiqueta de clase predicha por un clasificador para esa tupla. Si la precisión de un clasificador es aceptable, entonces ese clasificador se puede utilizar para predecir tuplas de datos con etiquetas de clase desconocidas. Los datos con etiquetas desconocidas se denominan datos "no vistos anteriormente". Además de la precisión, se utilizaron la sensibilidad y la especificidad para evaluar el rendimiento de los clasificadores. La sensibilidad y la especificidad se definen en las siguientes secciones. [9]

2.1.3 Adquisición e integración de datos.

Para plantear el clasificador se tiene una base de datos real de en Excel suministrados por la Empresa Eléctrica Riobamba. Los datos disponibles consisten en 5074 clientes por un período de 24 meses. Los datos consisten en el consumo de electricidad del cliente durante todos los meses. A cada cliente se le asignará una identidad llamada ID de cliente con su respectiva etiqueta la cual consistirá que a cada cliente inspeccionado se le asigna "1", si el cliente estaba limpio, o un "0", si el cliente era fraudulento. Esta codificación se utilizó para transformar valores categóricos en valores numéricos.

2.1.4 Normalización de datos

Para la normalización de datos se decidió utilizar la normalización mínimo-máximo para normalizar los consumos promedio diarios

$$NL = \frac{L - \min(L)}{\max(L) - \min(L)}$$

Ecuación 2.1

donde L representa el consumo promedio diario de un cliente, $\min(L)$ y $\max(L)$ son los valores mínimo y máximo en el conjunto de características, respectivamente. Los datos normalizados se utilizaron luego para el análisis estadístico. [9]

2.1.5 Máquinas de vector de soporte

En este estudio, se plantea el uso del kernel gaussiano para escribir el producto escalar $\Phi(X_i) \cdot \Phi(X_j)$. como función de los datos originales de la siguiente manera:

$$K(X_i, X_j) = \Phi(X_i) \cdot \Phi(X_j) = e^{-\gamma \|X_i - X_j\|^2}$$

Ecuación 2.1

Tenga en cuenta que el kernel de Gauss tiene solo un parámetro, a saber, γ . Este parámetro se denomina parámetro kernel *rbf*. Otro parámetro importante en el método SVM es la capacidad C . El parámetro del kernel *rbf* y la capacidad se obtienen experimentalmente. Las medidas de desempeño se utilizan para encontrar los valores óptimos de γ y C . El marco de entrenamiento propuesto para la optimización de los parámetros de SVM se muestra en la Figura 2.2.

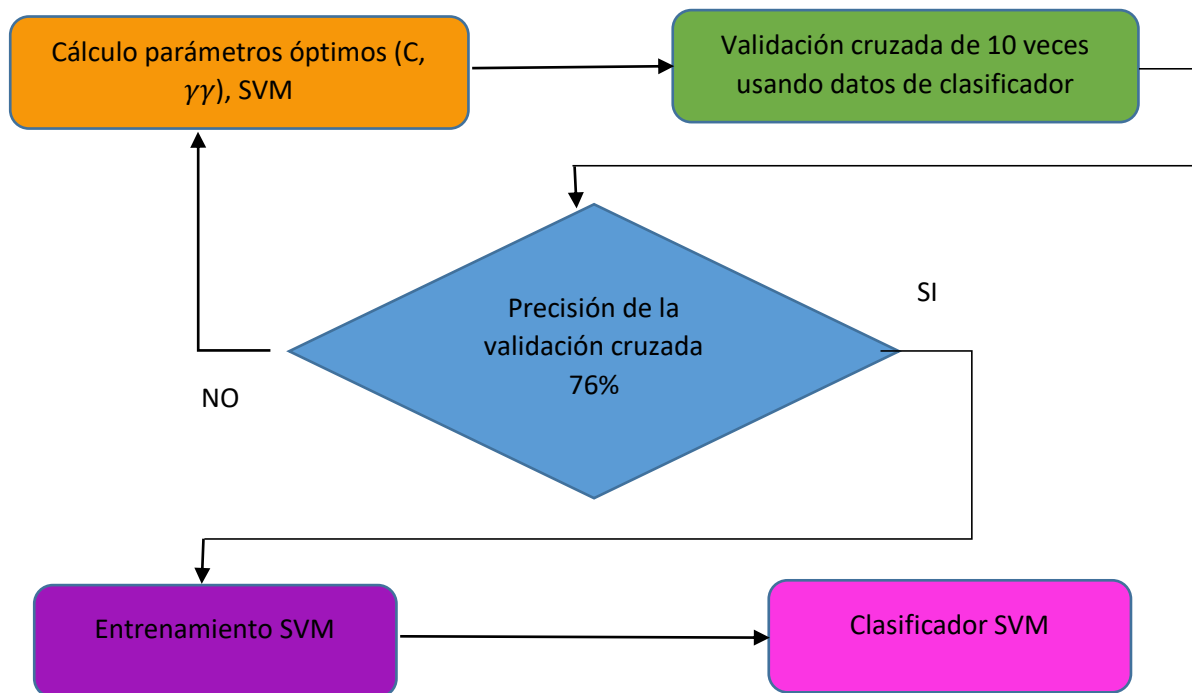


Figura 2.2 Diagrama de flujo de entrenamiento [9]

En resumen, el SVM es un método de aprendizaje supervisado que se utiliza para la clasificación estadística de datos lineales y no lineales. Dado un conjunto de tuplas de entrenamiento, cada una de las cuales pertenece a una de dos clases, el algoritmo de entrenamiento de SVM crea un clasificador que asigna nuevas tuplas a una de las dos clases. El entrenamiento se realiza encontrando los parámetros del límite de decisión, que es un hiperplano con mayor margen. El límite de decisión se usa luego para clasificar tuplas con etiquetas de clase desconocidas.[2]

2.1.5.1 MEDIDAS DE RENDIMIENTO DEL CLASIFICADOR

Después de que un clasificador ha sido entrenado, surgen las siguientes preguntas:

1. ¿Con qué precisión puede el clasificador predecir etiquetas de clase de datos no vistos anteriormente?
2. ¿Cómo se puede comparar el desempeño de diferentes clasificadores entrenados?
3. ¿Qué es el rendimiento?
4. ¿Cómo se puede estimar el desempeño?

El objetivo de esta sección es abordar las preguntas anteriores.

Usar un conjunto de datos de entrenamiento para construir un clasificador, así como usar el mismo conjunto de entrenamiento para estimar el rendimiento, puede presentar el peligro de un ajuste excesivo del modelo. De hecho, si se usa un conjunto de entrenamiento para estimar el rendimiento, entonces las estimaciones serán optimistas del verdadero rendimiento. Por esta razón, un conjunto de prueba, con tuplas etiquetadas de clase que no se usaron para entrenar el clasificador, se usa para estimar el rendimiento de un clasificador. Las siguientes son las definiciones de varias medidas de rendimiento para un problema de clasificación binaria: [35]

Tabla 2.1 Definiciones de varias medidas de rendimiento para un problema de clasificación binaria

Desempeño de medida	Definición
Precisión	La proporción de tuplas de prueba que el clasificador clasifica correctamente.
Tasa de clasificación errónea / Tasa de error	La proporción de tuplas de prueba que el clasificador clasifica incorrectamente.
Tuplas positivas	La tupla de la clase principal de interés (por ejemplo, clientes fraudulentos).
Tuplas negativas	La tupla de la otra clase (por ejemplo, Clientes limpios).
Verdaderos positivos	Tuplas positivas que están correctamente clasificadas por el clasificador.
Verdaderos negativos	Tuplas negativas que están correctamente clasificadas por el clasificador.
Falso positivo	Tupla negativa que se clasifica incorrectamente como tupla positiva.
Falso negativo	Tuplas positivas que se clasifican incorrectamente como tuplas negativas.

Sensibilidad	Proporción de tuplas positivas que se clasifican correctamente.
Especificidad	Proporción de tuplas negativas que se clasifican correctamente.

La matriz de confusión es una herramienta importante que se utiliza para analizar qué tan bien un clasificador clasifica tuplas de diferentes clases C_1 y C_2 . En la tabla 2.2 se muestra una matriz de confusión para dos clases. [35]

Tabla 2.2: Una matriz de confusión para tuplas positivas y negativas

	C_1	C_2
C_1	Verdaderos positivos	Falsos negativos
C_2	Falsos positivos	Verdaderos negativos

$$\text{Sensibilidad} = \frac{t_pos}{pos}$$

Ecuación 2.2

$$\text{Especificidad} = \frac{t_neg}{neg}$$

Ecuación 2.3

$$\text{Precisión} = \text{sensibilidad} \left(\frac{pos}{pos + neg} \right) + \text{Especificidad} \left(\frac{neg}{pos + neg} \right)$$

Ecuación 2.4

Donde, t_pos es el número de tuplas positivas verdaderas correctamente clasificadas, pos es el número total de tuplas positivas t_neg el número de tuplas negativas verdaderas correctamente clasificadas y, neg el número total de tuplas negativas.[35]

2.1.5.2 EL MÉTODO HOLDOUT

El método Holdout se utilizó para calcular medidas de desempeño a partir de datos empíricos dados. En este método, un conjunto de datos dado se dividió aleatoriamente en un conjunto de entrenamiento y un conjunto de prueba. Es una convención asignar dos tercios de los datos al conjunto de entrenamiento y el tercio restante al conjunto de prueba. El conjunto de entrenamiento se usó para construir un clasificador y el conjunto

de prueba para estimar el rendimiento del clasificador. Este método se ilustra gráficamente en la Figura 2.3. Las estimaciones de sensibilidad, especificidad y tasa de error se definieron de manera similar bajo el método de submuestreo aleatorio. La ventaja de utilizar el método de submuestreo aleatorio es que se pueden calcular varias estimaciones para cada medida de rendimiento. [35] El método Holdout se repitió 500 veces y para cada caso se determinó la sensibilidad, la especificidad y las tasas de error para el clasificador SVM.

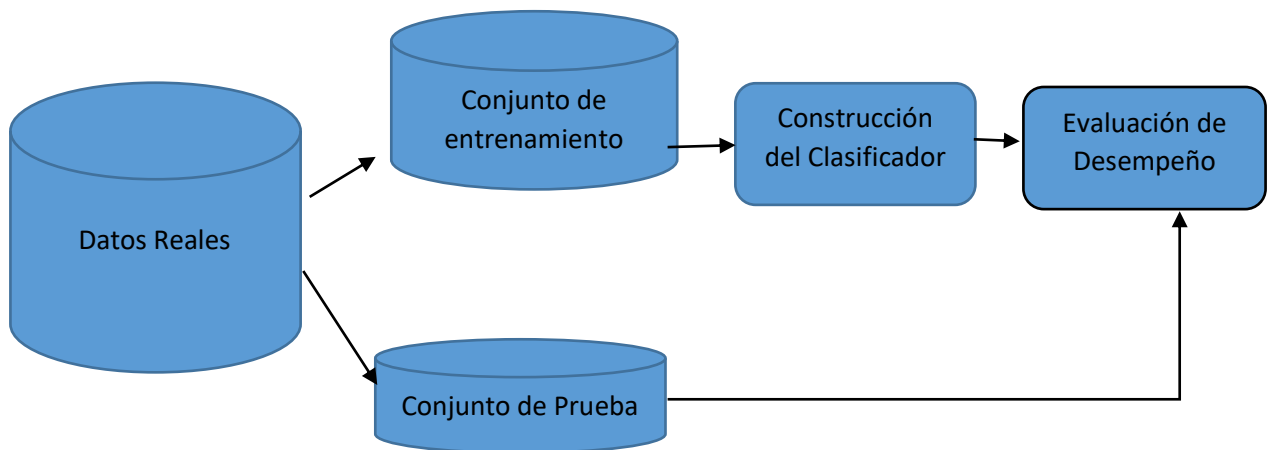


Figura 2.3 Evaluación del desempeño con el método de retención

2.1.6 Etapa 2 Estimación de estado del sistema de distribución

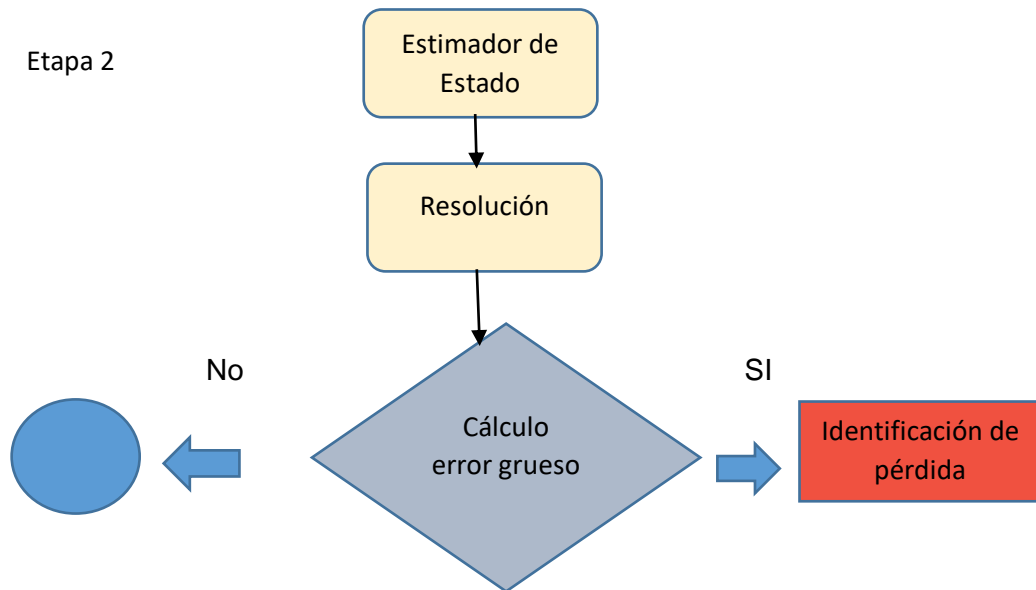


Figura 2.4 Estimación de estado y prueba de error grueso

Para la segunda etapa la hipótesis hecha por el clasificador se traslada al estimador de estado para refinar su desempeño en la identificación de pérdidas no técnicas.

La estimación de estados busca obtener las mejores estimaciones posibles, en el sentido de los mínimos cuadrados ponderados, para las tensiones complejas en los nudos de un sistema eléctrico, a partir del procesamiento de medidas adquiridas en tiempo real y de otras informaciones disponibles, tales como pronóstico de carga. Estas últimas se denominan pseudomedidas. Se admite que tanto medidas como pseudomedidas son contaminadas por ruidos. El problema de Estimación de Estados en Sistemas de Distribución (EESD) requiere que las telemidas y demás informaciones disponibles sobre el sistema monitoreado garanticen su observabilidad. Esto implica la existencia de medidas de flujos e inyecciones de potencia, corrientes y tensiones adecuadamente distribuidas sobre la red. En lo que se refiere a las telemidas, estas condiciones son más difíciles de encontrar en la práctica en sistemas de distribución que en la transmisión, por lo que es necesario complementar el plan de medición con pseudomedidas.

En el número de estados (magnitudes y ángulos de fase desconocidos de las tensiones nodales) y m el número de medidas, el problema de estimación de estados puede ser formulado como [15]:

$$\text{Minimizar } \frac{1}{2} \cdot r_m^T R_m^{-1} \cdot r_m$$

$$\text{Sujeto a: } r_m = z_m - h_m(\hat{x})$$

$$h_s(\hat{x}) = [0]$$

$$h_o(\hat{x}) = [0]$$

Ecuación 2.5

Siendo \hat{x} o vector de los estados estimado, z_m e r_m vectores $m \times 1$, de las medidas y pseudomedidas y residuos de estimación, respectivamente, e h_m o vector $m \times 1$ de las funciones no lineales que relacionan cantidades medidas y estados. Los errores de medición se modelan como variables aleatorias no correlacionadas con distribución normal cuya media es cero y cuya matriz de covariancia R_m , es diagonal. Los vectores $h_s(n_s \times 1)$ y $h_o(n_o \times 1)$ definición de restricciones de igualdad al problema de estimación de estados, a ser descritas a continuación.[15]

En este trabajo, las restricciones operativas se utilizan para la identificación de PNT. Para cumplir este objetivo, se supone que el SVM preselecciona un conjunto de N_{susp} número nodos de la red sospechosos a PNTs. Esta preselección puede basarse, por ejemplo, en estudios previos que indiquen la ocurrencia de inyecciones de potencia anormales en aquellos nodos. Para cada uno de los nudos sospechosos, no hay restricciones operativas del tipo[15]:

$$\left. \begin{array}{l} [P_{abc}(x)]_i - [P_{abc}^{pres}]_{carga,i} = 0 \\ [Q_{abc}(x)]_i - [Q_{abc}^{pres}]_{carga,i} = 0 \end{array} \right\}, i = 1, \dots, N_{susp}$$

Ecuación 2.6

Donde $[P_{abc}^{pres}]_{carga,i}$ y $[Q_{abc}^{pres}]_{carga,i} = 0$ son las cargas trifásicas activa y reactivas presumidas (es decir, sin contaminación con PNT) para el nodo i . Esto significa que el Operador parte de la hipótesis de que no existen PNT en los nodos preseleccionados. Si bien algunas de las cargas de las barras sospechosas están en

realidad bajo influencia de PNTs, los segundos términos en el lado izquierdo de (10) deberían ser dados por:

$$\begin{cases} [P_{abc}^{pres}]_{carga,i} + [PNT]_{P,i} \\ [Q_{abc}^{pres}]_{carga,i} + [PNT]_{Q,i} \end{cases}$$

Ecuación 2.7

En consecuencia, en este caso algunas de las hipótesis hechas por el Operador son incorrectas. El problema de identificación de PNTs se resume entonces en determinar cuáles restricciones operativas no son respaldadas por las medidas / pseudomedidas disponibles al estimador de estados. [15]

2.1.6.1 Solución por el método del cuadro disperso

Para resolver el problema de optimización, el problema (9), utilizamos el método Sparse Tableau, que consiste en un procedimiento iterativo en el que en cada iteración se debe resolver el siguiente sistema lineal. [15]:

$$\begin{bmatrix} [0] & H_m(\hat{x}^\ell)^T \\ H_m(\hat{x}^\ell) & R \end{bmatrix} \cdot \begin{bmatrix} [\Delta\hat{x}] \\ [\lambda] \end{bmatrix} = \begin{bmatrix} [0] \\ z - h(\hat{x}^\ell) \end{bmatrix}$$

Ecuación 2.8

Donde:

$$H(\hat{x}^\ell) = [H_m^T(\hat{x}^\ell) \quad H_s^T(\hat{x}^\ell) \quad H_o^T(\hat{x}^\ell)]^T$$

Ecuación 2.9

$$R = \begin{bmatrix} R_m & [0] & [0] \\ [0] & R_s & [0] \\ [0] & [0] & R_o \end{bmatrix}$$

Ecuación 2.10

$$\lambda = [\lambda_m \quad \lambda_s \quad \lambda_o]^T$$

Ecuación 2.11

$$z = [z_m \quad [0] \quad [0]]^T$$

Ecuación 2.11

$$h(\hat{x}^\ell) = [h_m^T(\hat{x}^\ell) \quad h_s^T(\hat{x}^\ell) \quad h_o^T(\hat{x}^\ell)]^T$$

Ecuación 2.12

Y H_m , H_s y H_o son respectivamente las matrices Jacobianas de las funciones $h_m(x)$, $h_s(x)$, y $h_o(x)$. Resolviendo el problema lineal de la ecuación (2.8), es posible actualizar la estimación de los estados. a través de:

$$\hat{x}^{l+1} = \hat{x}^l + \Delta x$$

Ecuación 2.13

Donde \hat{x}^0 es la partida plana para las tensiones nodales complejas. La convergencia ocurre cuando la norma euclidiana de Δx es menor que una tolerancia predefinida:

$$\|\Delta \hat{x}\|_2 \leq \epsilon$$

Ecuación 2.14

La aplicación de este procedimiento de estimación de estados es interesante para la metodología propuesta principalmente por el hecho de que la solución del Tableau Esparso proporciona también los multiplicadores de Lagrange λ_m , λ_s e λ_o sobre los tres tipos de información contenidos en el proceso de Estimación de Estado. En particular, los multiplicadores de Lagrange de las restricciones operativas, λ_o , son instrumentales para averiguar la corrección de las hipótesis representadas en la ecuación 2.6. [15]

2.1.7 Cálculo error grueso

Los multiplicadores de Lagrange están asociados a la sensibilidad de la función objetivo con respecto a variaciones en los valores de las restricciones de igualdad. En consecuencia, estas variables se clasificaron como instrumentos eficientes para detectar inconsistencias entre las medidas disponibles al estimador de estados y las relaciones establecidas en las restricciones. Para ello, sin embargo, los multiplicadores de Lagrange deben ser normalizados, por el hecho de estar asociados a diferentes tipos de informaciones procesadas por el estimador de estados. [36]

Considerando que todas las mediciones y restricciones son trifásicas, el i-ésimo Multiplicador de Lagrange Normalizado, relativo a la información de alguna de las fases (a, b, c), se define como [36]:

$$\lambda_i^N = \frac{\lambda_i}{\sqrt{W_{i,i}}}$$

Ecuación 2.15

donde W es la matriz de covarianza de los multiplicadores de Lagrange, definida a través de la inversión de la matriz Tableau Esparso [36]:

$$\begin{bmatrix} 0 & C^T \\ C & W \end{bmatrix} = \begin{bmatrix} 0 & H(\hat{x}^\ell)^T \\ H(\hat{x}^\ell) & R \end{bmatrix}^{-1}$$

Ecuación 2.16

Considerando que los errores de medición siguen una distribución normal y que no hay incoherencias en la información proporcionada por el plan de medición, los multiplicadores de Lagrange normalizados se comportan como variables aleatorias con distribución normal, con un promedio $\epsilon\{\lambda^N\} = 0$, y desviación estándar $\sigma = 1$. Por lo tanto, si no hay inconsistencias en los datos sometidos al estimador, y adoptando una probabilidad de falsa alarma del 1%, cada multiplicador de Lagrange i debe satisfacer la siguiente desigualdad [36]

$$\lambda_i^N \leq \lambda_t$$

Ecuación 2.17

Donde λ_t es un umbral que, en las condiciones anteriores, es igual a 3 desviaciones estándar, o sea $\lambda_t = 3\sigma = 3$.

De esta manera, si alguno de los multiplicadores de Lagrange normalizados asociados a una o más informaciones viole la desigualdad dada por la Ecuación (2.17), hay un 99% de probabilidad de ocurrencia de incongruencias en los datos o restricciones del problema de estimación.

La aplicación de este tipo de prueba permite la detección de incoherencias en el conjunto de restricciones operativas. Sin embargo, aunque los multiplicadores de Lagrange normalizados suministran indicios, no es posible discernir precisamente qué restricciones son efectivamente inconsistentes con los demás datos sometidos al estimador. Los métodos adicionales, como los basados en pruebas de colinealidad [15], se tornan, por lo tanto, necesarios.

2.1.7.1 Identificación de Pérdidas No Técnicas

Las pruebas geométricas de colinealidad se basan en una interpretación geométrica del comportamiento del vector de multiplicadores de Lagrange, λ , en presencia de errores. Para ilustrar este aspecto, se parte de la relación entre λ y los errores en los datos sometidos al estimador, η , que es dada por [36]:

$$\lambda = W \cdot \eta$$

Ecuación 2.18

La ecuación (2.18) revela que λ es una combinación lineal de las columnas de la matriz W . Por otro lado, considere la reestructuración de las columnas de la matriz de covarianza de los multiplicadores de Lagrange:

$$W = [W_S \ W_T]$$

Ecuación 2.19

Donde δ es el conjunto de informaciones sospechosas de error y τ es el conjunto de información libre de error, siendo las columnas de W_S particiones coincidentes. Suponiendo que los errores en los elementos del conjunto τ son nulos, se tiene:

$$\eta = \begin{bmatrix} \eta_S \\ 0 \end{bmatrix}$$

Ecuación 2.20

y por lo tanto, sustituyendo las Ecuaciones (2.19) y (2.20) en (2.18):

$$\lambda_S = W_S \cdot \eta_S$$

Ecuación 2.21

De acuerdo con la Ecuación (2.21), en presencia de errores groseros el vector de multiplicadores de Lagrange es una combinación de las columnas W_S sobre las informaciones sospechosas. En otras palabras λ pertenece al espacio vectorial $C(W_S)$ generado por las columnas de W_S .

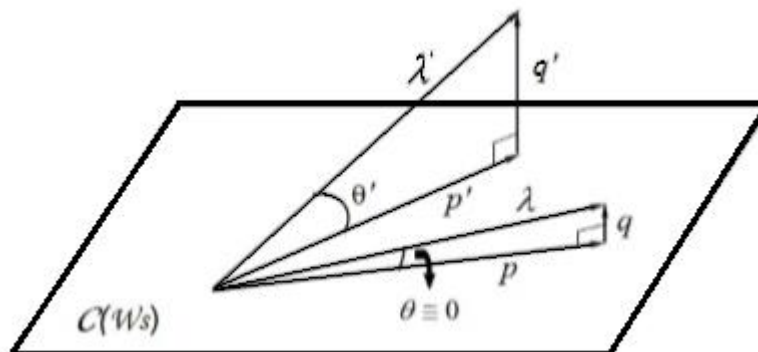


Figura 2.5 Proyección de λ en el espacio de columnas de W_S .

La Figura 2.5 resume esta característica, e ilustra cómo esta propiedad de colinealidad puede ser utilizada para la identificación de errores groseros. Si todas las restricciones

erróneas se incluyen en el conjunto sospechoso, λ teóricamente está contenido en el espacio $C(W_s)$. En la práctica, como las medidas son siempre contaminadas por ruido, esto no es rigurosamente cierto, pero el ángulo θ entre λ y $C(W_s)$ es pequeño, como se indica en la Figura 2.5. Por otra parte, si alguna restricción errónea no se ha incluido en \mathcal{S} , la ecuación (2.21) ya no se aplica, y λ se aleja significativamente de una forma $C(W_s)$. Como se muestra en la Figura 2.5, esta característica es evidente a través del surgimiento del ángulo $\theta' \gg 0$ entre λ y $C(W_s)$ es decir, permite determinar si todas las informaciones están incluidas en el conjunto sospechoso. \mathcal{S} . Se prueba que $\cos(\theta)$ puede ser calculado como:[15]

$$\cos\theta = \sqrt{\frac{\lambda_s^T \cdot W_{SS}^{-1} \cdot \lambda_s}{\lambda^T \cdot R \cdot \lambda}}$$

Ecuación 2.22

En este trabajo las cargas en las barras presumiblemente sujetas a PNTs son modeladas como restricciones operacionales, en la forma dada por la Ecuación (2.6). En el caso de que esta información sea errónea, y suponiendo que el plan de medición presenta redundancia satisfactoria, los multiplicadores de Lagrange de las restricciones afectadas asumirán valores relativamente elevados (así como esto posiblemente también ocurra con λ 's de otras restricciones, debido a efectos de dispersión de los errores). El criterio dado por la desigualdad (2.17) se utiliza entonces para definir el conjunto de restricciones sospechosas \mathcal{S} .

Como el conjunto \mathcal{S} posiblemente contendrá, además de las restricciones erróneas debido a las PNT, también restricciones asociadas a datos correctos, se hace necesario un segundo paso para depurar el conjunto \mathcal{S} .

El criterio para esta depuración hace uso de la prueba de colinealidad basada en el cálculo del $\cos(\theta)$, dado por la ecuación (2.22).

En síntesis, cada λ_i en \mathcal{S} es temporalmente retirado de este conjunto y $\cos\theta$ es recalculado. Si este valor permanece cerca de 1,0, λ_i puede ser retirado de \mathcal{S} y la restricción correspondiente se considera exenta de error. Si, sin embargo, la remoción de λ_i implica una reducción significativa de $\cos\theta$, se considera que la restricción sospechosa es de hecho errónea, siendo por lo tanto mantenida en \mathcal{S} .

La prueba para verificar si $\cos\theta$ es aproximadamente cerca de 1,0 consiste en la

comparación del coseno con un umbral dado por $1-\epsilon_\theta$ donde ϵ_θ es un valor de tolerancia entre 0,01 y 0,1.

3 Aplicación práctica de la metodología y análisis de resultados

3.1 Introducción

Para la aplicación de la metodología propuesta en este proyecto de titulación se utiliza el algoritmo de clasificación, Máquinas de vectores de soporte (SVM) a través de un lenguaje de programación en Matlab R2019a, donde se aplica a una base real de consumidores de un alimentador conformado de 3892 clientes, para luego llevar estos resultados al estimador de estado y refinar su resultado. El estimador se prueba con un sistema de 4 nodos, 13 nodos y 6 nodos, para luego finalmente ser aplicada al centro de transformador real con mayor identificación de usuarios sospechosos.

3.2 Aplicación Máquina de Vector de Soporte

3.2.1 Procesamiento de datos

Los datos se adquirieron para el desarrollo de la metodología fue proporcionada por Empresa Eléctrica Riobamba S.A. Los datos consistieron en 180200 clientes durante un período de 24 meses desde marzo de 2018 hasta febrero de 2020. A cada cliente se mantiene con su número de medidor.

El siguiente gráfico de líneas (Figura 3.1) ilustra el consumo total para cada uno de los meses registrados en la Tabla 3.1. Los enteros en el eje horizontal corresponden a los 24 meses del período que comienza en marzo de 2018 y finaliza en febrero 2020. No existe una estacionalidad aparente dentro de cada mes, ya que los datos muestran un comportamiento cíclico. con un plazo de 12 meses. No hubo una tendencia aparente en los datos durante este período.

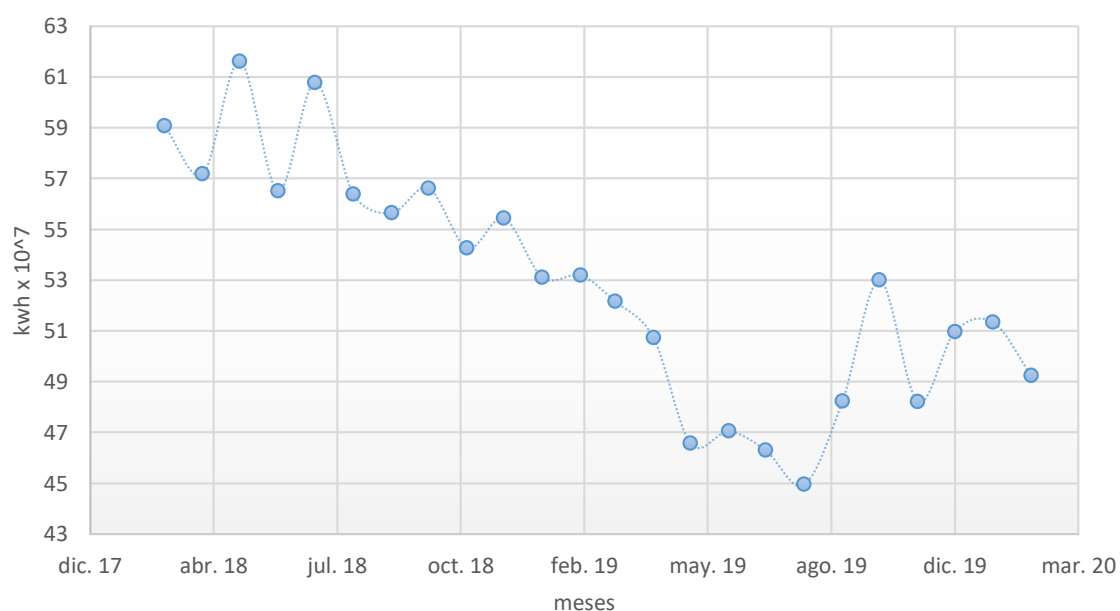


Figura 3.1 Gráfico de líneas para el consumo total en kWh por mes

Los datos se procesaron previamente antes de la evaluación estadística para corregir inconsistencias y eliminar datos ruidosos (es decir, datos con valores de atributo incorrectos), se eliminaron todos los clientes que tenían consumo cero (0 kWh) durante todo el período de 24 meses y todos los clientes que se registraron después del primer mes. Se eliminaron los duplicados para evitar la redundancia en los datos.

Se realizaron inspecciones a clientes seleccionados al azar en la Provincia de Chimborazo para identificar clientes fraudulentos y honestos. Un total de 3892 clientes fueron inspeccionados durante el período de estudio. De estos, 3161 fueron clasificados como casos limpios y 731 como fraudulentos. El estado de los contadores de electricidad de estos clientes se probó utilizando dispositivos de ingeniería eléctrica. A los clientes que cometieron fraude fueron etiquetados en la base de datos como "fraudulentos" y los clientes honestos como "limpios". Los 3893 clientes inspeccionados se extrajeron de la población de 180200 clientes. A cada cliente inspeccionado se le dio un "1", si el cliente estaba limpio, o un "0", si el cliente era fraudulento. Esta codificación se utilizó para transformar valores categóricos en valores numéricos. El método se utilizó para entrenar y probar el clasificador SVM utilizando los datos empíricos de la Tabla 3.1. Este conjunto de datos se dividió aleatoriamente en un conjunto de entrenamiento y uno de prueba. Aproximadamente dos tercios de las tuplas de datos formaron un conjunto de entrenamiento y el tercio restante un conjunto de prueba.

Tabla 3.1. Consumo real mensual de Energía

N° de medidor	mar-18	abr-18	may-18	jun-18	ene-20	feb-20	Etiqueta	Clase
547646	22	22	24	10	7	4	Fraudulento	0
2011781	161	164	167	164	0	0	Fraudulento	0
547927	0	0	0	0	78	75	Fraudulento	0
547752	105	105	105	105	116	63	Fraudulento	0
.....
113051	41	46	57	43	164	166	Limpio	1
208332	78	52	86	50	54	68	Limpio	1
117026	194	196	206	196	85	83	Limpio	1

3.2.2 Normalización de datos

Los consumos de la tabla 3.1 se normalizaron utilizando la metodología mínimo máximo de manera que todos los valores cayeron dentro del intervalo [0,1] (Nagi et al, 2008). Los consumos medios diarios se normalizaron de la siguiente manera

Tabla 3.2. Consumo normalizado mensual de Energía

Usuario	ene-17	feb-17	mar-17	abr-17	nov-18	dic-18	Etiqueta	Clase
1	0,350	0,450	0,000	0,100	0,450	0,500	Fraudulento	1
2	0,000	0,115	0,231	0,308	0,577	0,615	Fraudulento	1
3	1,000	0,000	0,650	0,664	0,714	0,721	Fraudulento	0
4	1,000	0,000	0,705	0,724	0,790	0,800	Limpio	0
5	1,000	0,000	0,340	0,359	0,427	0,437	Fraudulento	0
6	0,762	0,000	0,860	0,874	0,923	0,930	Fraudulento	0
....
37	0,150	0,000	0,725	0,752	0,849	0,862	Limpio	1
38	0,000	0,173	0,624	0,662	0,793	0,812	Limpio	1
39	0,044	0,000	0,704	0,734	0,837	0,852	Limpio	1
40	0,576	0,000	0,861	0,875	0,285	0,292	Limpio	1

3.1.1 Resultados de las máquinas vectoriales de soporte

El método de adquisición se utilizó para entrenar y probar el clasificador SVM utilizando los datos empíricos de la Tabla 3.1. El conjunto de datos constaba de 3892 clientes, 3161 de los cuales se clasificaron como casos limpios y 731 como casos limpios. Este conjunto de datos se dividió aleatoriamente en un conjunto de entrenamiento y uno de prueba. Aproximadamente dos tercios de las tuplas de datos formaron un conjunto de entrenamiento y el tercio restante un conjunto de prueba. El conjunto de entrenamiento estuvo conformado por 2595 clientes y el conjunto de prueba estaba compuesto por 1297 clientes.

Tabla 3.3 División conjunto de entrenamiento y prueba SVM

	Limpio	Fraudulento	Total
Conjunto de entrenamiento	2104	491	2595
Conjunto de Prueba	1057	240	1297
Total	3161	731	3892

El clasificador SVM se entrenó en el conjunto de entrenamiento y se estimó el rendimiento del mismo usando el conjunto de prueba. A continuación, se usó el conjunto de prueba para calcular los valores en la matriz de confusión.

La matriz de confusión correspondiente, que ilustra el número de clasificaciones correctas e incorrectas, se muestra en la Tabla 3.4. Las columnas corresponden a las etiquetas verdaderas y las filas a las etiquetas de clase previstas por el clasificador SVM. Los

números de clasificaciones correctas son los valores en la diagonal principal de la matriz de confusión, y las clasificaciones erróneas, los valores fuera de la diagonal. Idealmente, todos los valores deberían estar en la diagonal principal y todos los valores fuera de la diagonal deberían ser cero. Esto significaría que el clasificador clasifica correctamente todas las tuplas de prueba. Sin embargo, en la práctica, los valores fuera de la diagonal apenas son ceros. Así, un buen clasificador se caracteriza por tener valores fuera de la diagonal muy pequeña en comparación con los valores de la diagonal principal.

Tabla 3.4 Confusión de resultado conjunto de entrenamiento

		Etiquetas Verdaderas		
		Limpio	Fraudulento	Total
Predicción de Etiquetas	Limpio	1033	59	1092
	Fraudulento	24	181	205
	Total	1057	240	1297

A partir de la matriz de confusión anterior, las estimaciones de sensibilidad, especificidad, tasa de error se calcularon de la siguiente manera:

$$sensibilidad = \frac{1033}{1057} * 100\% = 97.72\%$$

$$especificidad = \frac{181}{240} * 100\% = 75.41\%$$

$$tasa\ de\ error = \frac{24 + 59}{1297} * 100\% = 6.4\%$$

La especificidad representó el porcentaje de casos fraudulentos identificados correctamente por el clasificador. Esta fue una medida importante, dado el objetivo de este estudio, y una estimación deseable de esta medida sería de al menos el 75.41%. El número de casos limpios siempre fue muy grande en comparación con el número de casos fraudulentos. La tasa de error fue otra medida imperativa a considerar, ya que medía el porcentaje de clientes mal clasificados. De acuerdo con la matriz de confusión de la Tabla 3.4, el clasificador SVM produjo una sensibilidad que estaba justo por encima

del 97.72%. La estimación de especificidad estuvo muy por encima del 75.41% y la tasa de error se estimó cercana al 6.4%.

Las medidas calculadas a partir de la Tabla 3.4 se obtuvieron utilizando un conjunto de entrenamiento y un conjunto de prueba. Para abordar esta pregunta, se utilizó el método de submuestreo aleatorio. En este método, el método de retención se repitió 500 veces.

Es decir, se utilizaron 500 combinaciones de conjuntos de entrenamiento y prueba para entrenar y probar el clasificador SVM, respectivamente. Se calcularon estimaciones de sensibilidad, especificidad y tasa de error para cada iteración. Los resultados de las 500 iteraciones, con los promedios asociados para cada medida, se muestran en la Tabla 5. Luego, estas medidas se trazaron en un conjunto de ejes contra el número de la iteración.

Los gráficos lineales resultantes se muestran en la Figura 3.2. Todas las estimaciones fueron relativamente constantes en diferentes conjuntos de prueba. La estimación de la sensibilidad fue de alrededor del 97 % para todos los conjuntos de prueba, mientras que la estimación de la especificidad fue constante, de aproximadamente el 75%, y la tasa de error cercana al 6 % para todos los conjuntos.

Tabla 3.5 Estimaciones de medidas de rendimiento para el clasificador SVM

Numero de Iteraciones	Sensibilidad	Especificidad	Tasa de error
1	0,972	0,768	0,065
2	0,968	0,732	0,079
3	0,958	0,779	0,075
4	0,973	0,775	0,063
:	:	:	:
499	0,967	0,746	0,072
500	0,970	0,730	0,078

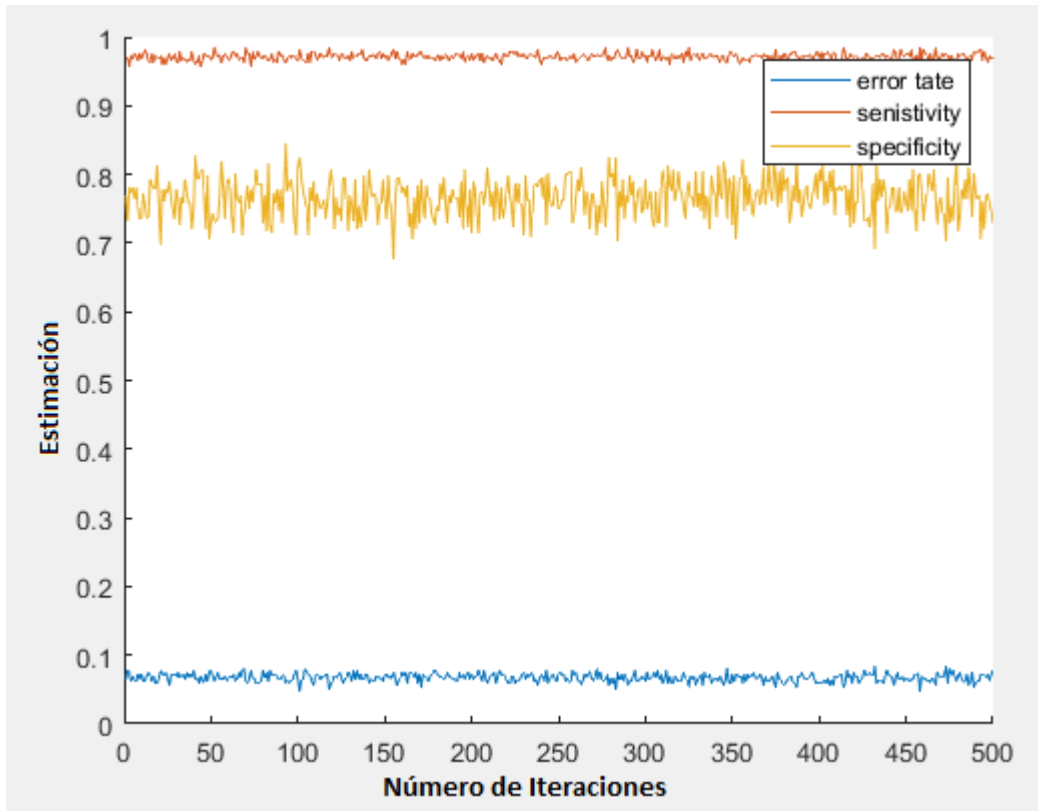


Figura 3.2 Medidas de rendimiento obtenidas en 500 iteraciones para el clasificador SVM

3.1.1.1 Optimización de parámetros

Experimentalmente se encontró que los valores óptimos de los parámetros capacidad y función de base radial (rbf) fueron, $C = 1.402$ y $\gamma = 2.963$ respectivamente. Luego, el clasificador SVM se entrenó y se probó usando los parámetros óptimos. La matriz de confusión calculada utilizando los parámetros óptimos se muestra en la Tabla 4.4. Esta matriz de confusión se usó luego para calcular las estimaciones de sensibilidad, especificidad y tasa de error.

Tabla 3.6 Confusión de resultado conjunto de entrenamiento

		Etiquetas Verdaderas		
		Limpio	Fraudulento	Total
Predicción de Etiquetas	Limpio	1032	52	1084
	Fraudulento	24	189	213
	Total	1056	241	1297

A partir de la matriz de confusión anterior, las estimaciones de sensibilidad, especificidad, tasa de error se calcularon de la siguiente manera:

$$\text{sensibilidad} = \frac{1032}{1056} * 100\% = 97.72\%$$

$$\text{especificidad} = \frac{189}{241} * 100\% = 78.42\%$$

$$\text{tasa de error} = \frac{24 + 52}{1297} * 100\% = 5.85\%$$

Tabla 3.7 Comparación de Parámetros

	Sin optimizar parámetros	Optimizando parámetros
Sensibilidad	97,73%	97,73%
Especificidad	75,42%	78,42%
Tasa de error	6,40%	5,86%

Aunque las estimaciones de que no cambiaron significativamente después de la optimización de los parámetros, la estimación de la especificidad y la tasa de error mejoró del 76.42 % al 78.42 % y de 6.40% a 5,86% respectivamente, para probar si este cambio fue significativo o se debió al azar, se utilizó el método de submuestreo aleatorio con los

parámetros óptimos para obtener 500 estimaciones diferentes para cada medida de desempeño. Estas estimaciones se muestran en la Tabla 3.8 y los gráficos de líneas correspondientes representados en la Figura 3.3.

Tabla 3.8 Estimaciones de medidas de rendimiento para el clasificador SVM después de la optimización de parámetros

Numero de Iteraciones	Sensibilidad	Especificidad	Tasa de error
1	0.9686	0.7500	0.0724
2	0.9743	0.7755	0.0632
3	0.9816	0.8106	0.0531
4	0.9723	0.7570	0.0686
:	:	:	:
499	0.9602	0.7583	0.0771
500	0.9772	0.7842	0.0585

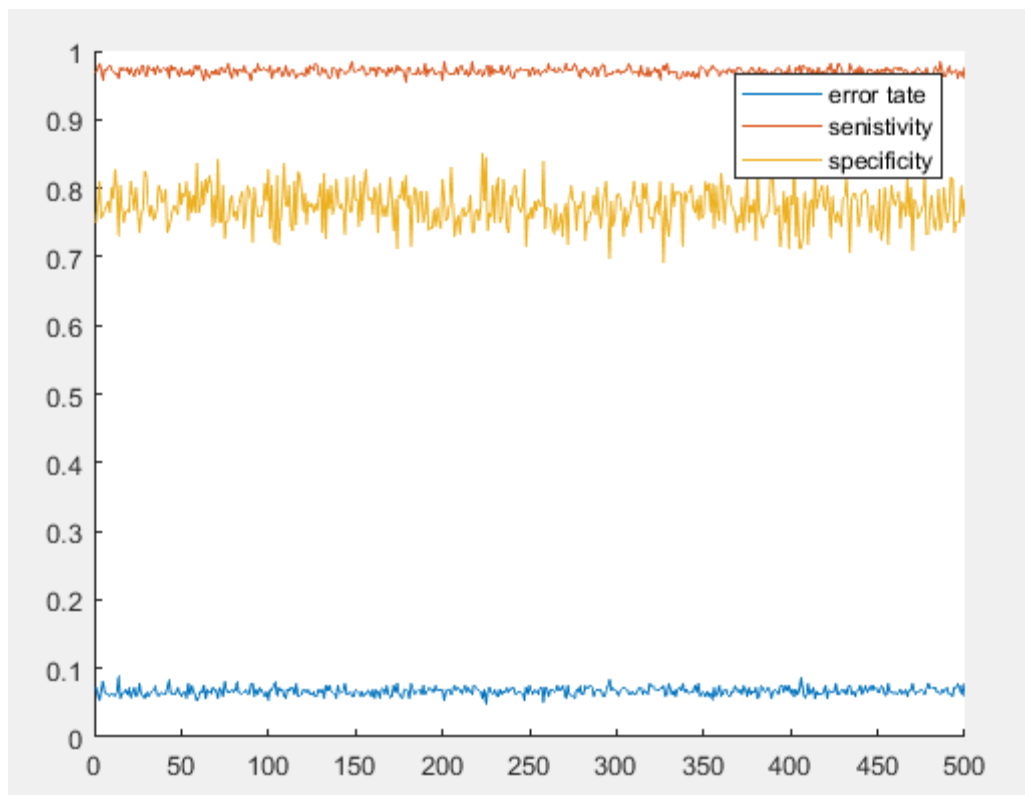


Figura 3.3 Medidas de rendimiento de SVM con parámetros optimizados

3.1.1 Aplicación de del SVM para el alimentador 2 de la subestación 14 de la Empresa Eléctrica Riobamba

El clasificador entrenado es aplicado al alimentador 2 de la subestación 14 y el alimentador 1 de la subestación 1 de la Empresa Eléctrica Riobamba, el alimentador se encuentran conectados 2874 consumidores; para lo cual se prepara la información donde al no saber la etiqueta y la clase se asume que todos los usuarios se encuentran Limpios, y su respectiva clase sería "1".

Tabla 3.9 Consumos de energía mensual alimentador 2 de la subestación 14

No de medidor	mar-18	abr-18	may-18	jun-18	:	ene-20	feb-20	Etiqueta Asumida	Clase Asumida
2283976	200	10	155	2	:	0	0	Limpio	1
2284891	200	10	155	2	:	2331	1979	Limpio	1
1329051	340	339	437	283	:	165	106	Limpio	1
1509553	2265	4379	2995	2935	:	2385	2369	Limpio	1
1435577	1752	1753	1750	2534	:	2259	1990	Limpio	1
1766377	1652	1704	1307	1451	:	754	770	Limpio	1
1446558	530	925	670	479	:	345	416	Limpio	1
:	:	:	:	:	:	:	:	:	:
1385350	107	78	173	119	73	63	85	Limpio	1
1385343	132	128	162	233	25	56	71	Limpio	1

A la base se le realiza la respectiva normalización y se aplica el calificador SVM donde se obtiene los siguientes resultados.

Tabla 3.10: Consumos de energía mensual normalizada alimentador 2 de la subestación14

no de medidor	mar-18	abr-18	:	ene-20	feb-20	Etiqueta	Clase Asumida	Predicción	Etiqueta Predicha
2283976	200	10	:	0	0	Limpio	1	0	Fraude
2284891	200	10	:	2331	1979	Limpio	1	0	Fraude
1329051	340	339	:	165	106	Limpio	1	1	Limpio
1509553	2265	4379	:	2385	2369	Limpio	1	1	Limpio
1435577	1752	1753	:	2259	1990	Limpio	1	1	Limpio
:	:	:	:	:	:	:	:	:	:
1206440	1281	2064	:	1028	1061	Limpio	1	1	Limpio
1366939	914	1488	:	450	444	Limpio	1	1	Limpio
2099174	135	252	:	304	308	Limpio	1	0	Fraude
2060440	16	9	:	4	0	Limpio	1	0	Fraude
1385350	107	78	:	63	85	Limpio	1	1	Limpio
1385343	132	128	:	56	71	Limpio	1	1	Limpio

Al aplicar la base entrena SVM al alimentador en cuestión se predice los siguientes resultados:

Tabla 3.11: Resumen de resultados Alimentador 2 de la subestación 14 y Alimentador 1 de la subestación 1

	Limpios	Fraude	Total
Consumidores Alimentador 2 de subestación 14	2039	835	2874
Consumidores Alimentador 1 de subestación 1	4443	504	4443

3.2 Aplicación Estimador de estado sistema para la identificación de Pérdidas No Técnicas

Con la hipótesis hecha por el Clarificador SVM donde se tiene identificaron usuarios con sospecha de fraude de energía, por lo que estos resultados son trasladados al estimador de estado mediante el cual se resuelve a través del algoritmo de mínimos cuadrados ponderados.

Las simulaciones se utilizan en el sistema de prueba IEEE de 4 nodos. Por su reducido tamaño que permite un mayor detalle de los resultados, luego se aplica al sistema IEEE 13, un sistema conformado de 6 nodos con 15 usuarios y finalmente en un sistema real de 17 nodos

3.2.1 Sistema de prueba IEEE de 4 nodos

Se eligió el sistema de prueba de 4 nodos IEEE (Figura 3.4) para los estudios iniciales del algoritmo propuesto destinado a detectar e identificar PNT. Este sistema de prueba consta de un alimentador longitudinal, sin ramales laterales, donde el nodo 1 corresponde a la subestación y el nodo 4 es el único nodo de carga, que se conecta radialmente a las demás, según las características descritas en la Tabla 3.12.

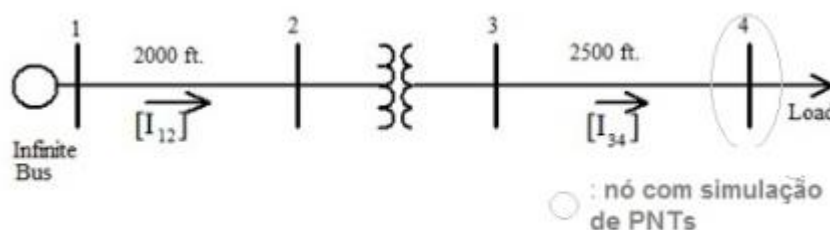


Figura 3.4 Sistema de prueba IEEE de 4 nodos

Tabla 3.12 - Nodos IEEE 4 - Información general

Datos del Sistema	
Tipo de elemento	Cantidad
Líneas trifásicas	2
Transformador estrella-estrella	1
Datos de nodo	
Tipo de nodo	Cantidad
Nodo de subestación	1
Nodos trifásicos	3
Potencia instalada	
Tipo da Carga	Carga Total (kVA)
Cargas nodales	180 + j87,2

Para aplicar el algoritmo de detección e identificación PNT al sistema de prueba de 4 nodos IEEE, se creó un plan de medición, con un índice de redundancia (ρ) de $\rho = 2.5$

Para el cálculo de redundancia se (ρ) considera la suma del número total de medidas y pseudomedidas para el plan de medición utilizado, el número de restricciones estructurales del sistema de prueba y el número de restricciones operativas, dividido por el número de variables de estado. Las características trifásicas o no de los nodos o ramas deben tenerse en cuenta para el correcto cálculo de la redundancia.

La Tabla 3.13 se presenta el plan de medición, también se discriminan las respectivas redundancias (ρ). Para las medidas de flujo de potencia activa / reactiva y módulo de corriente que se definen en las ramas, el número resaltado en negrita indica el nodo donde se realiza esta medida.

Tabla 3.13 - Nodos IEEE 4 - Plan de medición

Medidores de supervisión		
Medidor	Ubicación del medidor	Precisión (%)
Módulo de tensión	1, 4.	1
Inyección de energía activa	4	3
Inyección de potencia reactiva	4	3
Flujo de energía activa	1-2, 2-3.	3
Flujo de energía reactiva	1-2, 2-3.	3
Módulo de Corriente al Cuadrado	3-4.	3
Pseudomedidas		
Previsión de carga	Nodo consumidor	Precisión (%)

Inyección de energía activa	1, 3.	5
Inyección de potencia reactiva	3,4.	5
Restricciones estructurales		
Restricción	Nodo	Varianza (pu)
Fase de voltaje	1	10^{-8}
Inyección cero de potencia activa y reactiva	2,3	10^{-8}
Restricciones operativas		
Restricción	Posibilidad de PNTs	Varianza (pu)
Inyección de potencia activa y reactiva	4	10^{-8}

Los rangos de medición se agregan de acuerdo con los niveles de precisión especificados.

Tabla 3.14 Parámetros de Medición

Medidores de supervisión	
Medidor	Precisión (%)
Módulo de tensión	0,5
Inyección de potencia activa y reactiva	1
Flujo de potencia activa y reactiva	1
Módulo de Corriente al cuadrado	1
Pseudomedidas	
Carga	Precisión(%)
Inyección de potencia activa y reactiva	5

Antes de discutir la aplicación del algoritmo al sistema de prueba de 4 nodos IEEE, se necesitan algunas definiciones para comprender los resultados:

- La tolerancia ϵ se fijó en 10^{-6} , para los procesos iterativos de estimación de estado y flujo de potencia;
- Los parámetros λ_{lim}^0 y $\epsilon\theta$ de la prueba del coseno de la estimación de estados para la identificación de PNT se fijan inicialmente en 3,0 y 0,02, respectivamente;
- S^0 es el conjunto inicial de restricciones / medidas seleccionadas como sospechosas en la primera etapa de la prueba de colinealidad basada en pruebas $|\lambda_N| > \lambda_{lim}^0$.

- \mathcal{S}^f es el conjunto final de restricciones / medidas seleccionadas como sospechosas en la primera etapa de la prueba de colinealidad basada en pruebas $|\lambda_N| > \lambda_{lim}^f$.
- \mathcal{S} define el conjunto de restricciones/medidas seleccionadas como sospechosas por la segunda etapa de la prueba de colinealidad;

El nivel porcentual estipulado de PNT es el aumento porcentual en relación con la información asumida por el Operador como información correcta en los nodos sospechosos. Por ejemplo, si $[P_{abc}^{pres}]_{carga,i} = 180kW$ y el nivel de PNT simulados es del 30% pronto el término $[PNT]_{P,i}$ y $0,3 * 180kW = 54kW$ Por lo tanto, se aplica la Ecuación:

$$\begin{cases} [P_{abc}^{pres}]_{carga,i} + [PNT]_{P,i} \\ [Q_{abc}^{pres}]_{carga,i} + [PNT]_{Q,i} \end{cases}$$

Para sintetizar los resultados obtenidos para el sistema de prueba de 4 nodos IEEE, se caracteriza por presentar un nivel del 30% de PNT en la Fase b potencia activa y fase a potencia reactiva,

En la Tabla 3.15 y Tabla 3.16 se muestran los resultados donde los valores marcados con negrilla son los valores identificados como sospechosos.

Tabla 3.15 Resultado de Simulación con la condición $|\lambda_N| > \lambda_{lim}^0$, donde $\lambda_{lim}^0 = 3$

VALORES MEDIDOS	VALORES CALCULADOS	Denominación	H	lamb_N
1,0061	0,9983	Va	hm(x)	1,0155
1,0054	1,0037	Vb		0,3526
0,9855	0,9912	Vc		-0,7097
0,9741	0,9817	Va		-0,9851
0,9910	0,9926	Vb		-0,2036
0,9894	0,9839	Vc		0,7102
0,1893	0,1856	Pa12		0,2556
0,2447	0,2382	Pb12		2,0643
0,1774	0,1825	Pc12		-0,1206
0,1847	0,1854	Pa34		0,1104
0,2438	0,2381	Pb34		2,0400
0,1817	0,1825	Pc34		0,0259
0,1911	0,1854	Pa23		0,3190
0,2427	0,2381	Pb23		2,0035
0,1817	0,1825	Pc23		0,0234
0,1189	0,1186	Qa12		0,9690
0,0918	0,0926	Qb12		0,0790

0,0880	0,0903	Qc12	-0,0585
0,1146	0,1182	Qa23	0,8358
0,0873	0,0920	Qb23	-0,0608
0,0862	0,0900	Qc23	-0,1069
0,1219	0,1182	Qa34	1,0770
0,0917	0,0920	Qb34	0,0845
0,0957	0,0900	Qc34	0,2066
0,1808	0,1856	Pa1	-0,0221
0,2393	0,2382	Pb1	1,8926
0,1891	0,1825	Pc1	0,2649
0,0000	0,0000	Pa2	0,0000
0,0000	0,0000	Pb2	0,0000
0,0000	0,0000	Pc2	0,0000
-0,1843	-0,1831	Pa4	-0,1422
-0,2247	-0,2369	Pb4	-1,4547
-0,1785	-0,1819	Pc4	0,0485
0,1171	0,1186	Qa1	0,9066
0,0957	0,0926	Qb1	0,2085
0,0931	0,0903	Qc1	0,1115
0,0000	0,0000	Qa2	0,0000
0,0000	0,0000	Qb2	0,0000
0,0000	0,0000	Qc2	0,0000
-0,1156	-0,1157	Qa4	-0,9413
-0,0922	-0,0881	Qb4	-0,1663
-0,0876	-0,0881	Qc4	-0,0136
0,1797	0,1856	Psa1	-0,0353
0,2294	0,2382	Psb1	0,9457
0,1870	0,1825	Psc1	0,1176
0,0000	0,0000	Psa2	0,0000
0,0000	0,0000	Psb2	0,0000
0,0000	0,0000	Psc2	0,0000
-0,1815	-0,1831	Psa4	-0,0289
-0,2266	-0,2369	Psb4	-0,9088
-0,1945	-0,1819	Psc4	-0,2848
0,1157	0,1186	Qsa1	0,5165
0,0894	0,0926	Qsb1	0,0005
0,0990	0,0903	Qsc1	0,1843
0,0000	0,0000	Qsa2	0,0000
0,0000	0,0000	Qsb2	0,0000
0,0000	0,0000	Qsc2	0,0000
-0,1158	-0,1157	Qsa4	-0,5690
-0,0894	-0,0881	Qsb4	-0,0452
-0,0824	-0,0881	Qsc4	0,0949
0,0465	0,0487	la1	0,1703

0,0630	0,0649	lb1		0,7458	
0,0412	0,0422	lc1		-0,0076	
0,0454	0,0487	la3		0,1338	
0,0658	0,0649	lb3		0,8408	
0,0427	0,0422	lc3		0,0434	
0,0463	0,0487	la4		0,1667	
0,0637	0,0649	lb4		0,7710	
0,0412	0,0422	lc4		-0,0084	
0,0000	0,0000	$\delta a1$	hs(x)	0,9816	
-2,0944	-2,0944	$\delta b1$		-0,3254	
2,0944	2,0944	$\delta c1$		-0,6655	
0,0000	0,0000	Pa2		0,1171	
0,0000	0,0000	Pb2		2,4527	
0,0000	0,0000	Pc2		0,1136	
0,0000	0,0000	Pa3		0,3057	
0,0000	0,0000	Pb3		3,4601	
0,0000	0,0000	Pc3		0,1165	
0,0000	0,0000	Qa2		1,1970	
0,0000	0,0000	Pb2		0,1586	
0,0000	0,0000	Pc2		0,0833	
0,0000	0,0000	Pa3		1,7081	
0,0000	0,0000	Pb3		0,2635	
0,0000	0,0000	Pc34		0,1828	
-0,1800	-0,1831	Pga4		h0(x)	0,4236
-0,1800	-0,2369	Pgb4			4,6603
-0,1800	-0,1819	Pgc4	0,1410		
-0,0872	-0,1157	Qga4	2,3185		
-0,0872	-0,0881	Qgb4	0,4107		
-0,0872	-0,0881	Qgc4	0,0372		

Para estos casos descritos, la tabla 3.15 resume los parámetros obtenidos de las simulaciones. En la primera columna se encuentra la hipótesis realizada en la segunda se encuentra el valor calculado en la tercera se definen las denominaciones sospechosas de cada caso, la cuarta columna las restricciones operativas simuladas con errores, La quinta columna presenta los valores de los multiplicadores de Lagrange normalizados ($\lambda_{N'S}$) obtenidos para cada cantidad sospechosa de S^0 .

Tabla 3.16 Información efectivamente incorrecta no cumple la condición $|\lambda_N| > \lambda_{lim}^0$, donde $\lambda_{lim}^f = 0.7$ para coseno=0.98.

VALORES MEDIDOS	VALORES CALCULADOS	Denominación	h	lamb_N
-0,1800	-0,1818	Pga4	h0(x)	0,4236

-0,1800	-0,2295	Pgb4		4,6603
-0,1800	-0,1778	Pgc4		0,1410
-0,0872	-0,1099	Qga4		2,3185
-0,0872	-0,0875	Qgb4		0,4107
-0,0872	-0,0889	Qgc4		0,0372

Los resultados marcados con negrilla son detectados como sospechoso de fraude de energía.

Después de identificar los supuestos incorrectos sobre la carga en el nodo (4), el procedimiento propuesto aún permite estimar las PNT. La Tabla 3.17 muestra los valores de carga estimados para la restricción h_0 nodo (4), y su porcentaje en relación con la carga asumida. Los PNT estimados son aproximadamente del 30%, lo que es compatible con el nivel de PNT simulado en este caso.

La Tabla 3.17 Resultados restricción operacional porcentaje de PNT

Restricción Operacional	Hipótesis (carga asumida)	Carga Estimada	PNT estimada
Pgb4 kW	180	229,49638	27,5%
Qga4 kVAR	87,18	109,8504948	26,0%

La Tabla 3.18 Tabla Resultado estimador de estado

	modulo	ángulo	Barra
Va1	1,01	0,00	1
Vb1	0,99	-120,00	
Vc1	1,01	120,00	
Va2	1,00	-0,02	2
Vb2	0,99	-120,07	
Vc2	1,00	119,96	
Va3	1,00	-0,05	3
Vb3	0,99	-120,11	
Vc3	1,00	119,93	
Va4	0,99	-0,25	4
Vb4	0,98	-120,86	
Vc4	0,998493345	119,5128268	

3.2.2 Sistema de prueba de 13 nodos IEEE

El sistema de prueba de 13 nodos IEEE es radial, como se puede ver en la Figura 3.5. La Tabla 3.19 presenta la descripción de sus características físicas, como ramales y nodos

trifásicos, bifásicos y monofásicos, y los nodos donde se puede evaluar la presencia de PNT.

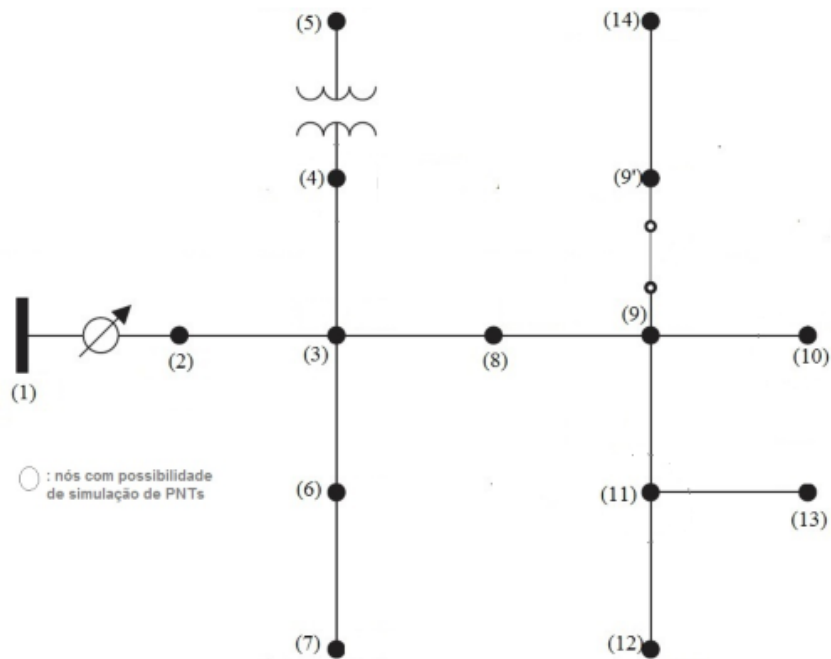


Figura 3.5 Sistema de prueba de 13 nodos IEEE

Se hicieron algunas modificaciones al sistema de prueba original. (IEEE PES, 2004) para llegar a la versión de la Figura 3.5. Ellos son: creación del nodo virtual (8), en el que la carga distribuida entre los nodos (3) y (7) se concentra; aglutinación de nódulos (9) y (9'), ya que el disyuntor entre ellos está en posición "cerrado". La presencia del regulador de voltaje entre los nodos (1) y (2) afecta los resultados debido al hecho de que su impedancia en serie y la admitancia en derivación son nulas, generando problemas numéricos en el algoritmo. Para resolver este problema, se utiliza la aglutinación.

Tabla 3.19 descripción características físicas sistema prueba de 13 nodos IEEE

Datos del sistema	
Tipo do Elemento	Cantidad
Líneas Trifásicas	6
Líneas Bifásicas	3
Líneas Monofásicas	2
Transformador Estrela-Estrela	1
Regulador de voltaje Delta abierto	1
Datos del nodo	
Tipo de nodo	Cantidad
Nodo de subestación	1
Nodos trifásicos	6

Nodos Bifásicos	3
Potencia instalada	
Tipo de carga	Carga Total (kVA)
Cargas nodales	3266+ j2102
Cargas Distribuidas	200+j116
Capacitores Shunt	-j700

Como se puede observar en la Tabla 3.19, el sistema de prueba IEEE de 13 nodos tiene una característica que lo diferencia de los sistemas de prueba IEEE 4. Este sistema de prueba es estructuralmente desequilibrado, es decir, contiene nodos y ramales bifásicos y monofásicos. Esta diferencia permite que el algoritmo propuesto sea sometido a situaciones más severas. Para ello se estipularon un plan de medición, Tabla 3.20, con redundancia = 2,231

Tabla 3.20 - Nodos IEEE 13 - Plan de medición

Medidores de supervisión		
Medidor	Ubicación del medidor	Precisión (%)
Módulo de la Tensión	1,5,9,12,13,14	1
Inyección de potencia Activa y Reactiva	3,5,9,12,13,14	3
Flujo de Potencia Activa y Reactiva	1-2,3-4,3-8,9-10,9-11,8-9	3
Módulo de la Corriente al Cuadrado	3-4, 3-8, 9-11, 2-3, 11-13	3
Pseudomedidas		
Previsión de carga	Nodo consumidor	Precisión (%)
Inyección de potencia activa	3,5,6,7,8,9,12,13,14	5
Inyección de potencia reactiva	2,5,6,7,8,9,12,13,14	5
Restricciones estructurales		
Restricción	Nodo	Varianza (pu)
Fase de voltaje	1	10^{-8}
Inyección cero de potencia activa y reactiva	3	10^{-8}
Restricciones operativas		
Restricción	Nodos con posibilidad de PNTs	Varianza (pu)
Inyección de potencia activa y reactiva	4,10,11	10^{-8}

El nivel de PNT simulados es del 30%. En la Tabla 3.16 y Tabla 3.17 se muestran los resultados donde los valores marcados con negrilla son los valores identificados como sospechoso.

Tabla 3.21 Resultado de Simulación con la condición $|\lambda_N| > \lambda_{lim}^0$, donde $\lambda_{lim}^0 = 3$

VALORES MEDIDOS	VALORES CALCULADOS	Denominación	H	λ_N
-0,200	-0,258	Pga4	h0(x)	1,470
-0,200	-0,280	Pgb4		1,993
-0,200	-0,254	Pgc4		1,333
-0,250	-0,252	Pga10		0,172
-0,250	-0,293	Pgb10		1,747
-0,250	-0,213	Pgc10		-1,660
-0,200	-0,237	Pga11		1,121
0,000	0,000	Pgb11		0,000
-0,200	-0,136	Pgc11		-2,055
-0,100	-0,094	Qga4		-0,171
-0,100	-0,119	Qgb4		0,517
-0,100	-0,108	Qgc4		0,177
-0,125	-0,080	Qga10		-1,807
-0,125	-0,194	Qgb10		2,788
-0,125	-0,108	Qgc10		-1,289
-0,100	-0,094	Qga11		-0,419
0,000	0,000	Qgb11		0,000
-0,100	-0,001	Qgc11		-2,936

Tabla 3.22 Resultado de Simulación con la condición $|\lambda_N| > \lambda_{lim}^0$, donde $\lambda_{lim}^0 = 1$ para coseno=0.98

VALORES MEDIDOS	VALORES CALCULADOS	Denominación	h	lamb_N
-0,200	-0,258	Pga4	h0(x)	1,470
-0,200	-0,280	Pgb4		1,993
-0,200	-0,254	Pgc4		1,333
-0,250	-0,252	Pga10		0,172
-0,250	-0,293	Pgb10		1,747
-0,250	-0,213	Pgc10		-1,660
-0,200	-0,237	Pga11		1,121
0,000	0,000	Pgb11		0,000
-0,200	-0,136	Pgc11		-2,055
-0,100	-0,094	Qga4		-0,171
-0,100	-0,119	Qgb4		0,517
-0,100	-0,108	Qgc4		0,177
-0,125	-0,080	Qga10		-1,807

-0,125	-0,194	Qgb10	2,788
-0,125	-0,108	Qgc10	-1,289
-0,100	-0,094	Qga11	-0,419
0,000	0,000	Qgb11	0,000
-0,100	-0,001	Qgc11	-2,936

Los resultados marcados con negrilla son detectados como sospechoso de fraude de energía

Tabla 3.23: Resultado estimador de estado

Fase	Voltaje	Ángulo	Nodo
fase a	0,995	0,000	NODO1
faseb	1,000	-120,000	
fasec	0,985	120,000	
fasea	0,995	-3,459	NODO2
faseb	1,025	-122,938	
fasec	0,996	116,651	
fasea	0,989	-3,598	NODO3
faseb	1,020	-123,101	
fasec	0,990	116,564	
fasea	0,965	-4,373	NODO4
faseb	1,003	-123,611	
fasec	0,969	116,058	
fasea	0,000	0,000	NODO5
faseb	1,016	-123,257	
fasec	0,997	116,637	
fasea	0,000	0,000	NODO6
faseb	1,015	-123,364	
fasec	0,996	116,669	
fasea	0,973	-5,348	NODO7
faseb	1,023	-123,495	
fasec	0,969	115,150	
fasea	0,949	-7,388	NODO8
faseb	1,024	-124,061	
fasec	0,944	113,840	
fasea	0,944	-7,606	NODO9
faseb	1,019	-124,330	
fasec	0,940	113,599	
fasea	0,943	-7,520	NODO10
faseb	0,000	0,000	
fasec	0,940	113,656	
fasea	0,000	0,000	NODO11

faseb	0,000	0,000	
fasec	0,938	113,497	
fasea	0,937	-7,435	NODO12
faseb	0,000	0,000	
fasec	0,000	0,000	
fasea	0,942	-7,668	NODO13
faseb	1,026	-124,212	
fasec	0,942	113,828	

La Tabla 3.24 Resultados restricción operacional porcentaje de PNT

Restricción Operacional	Hipótesis (carga asumida)	Carga Estimada	PNT estimada
Pga4 kW	200	258	-29%
Pgb4 kW	200	280	-40%
Pgc4 kW	200	254	-27%
Pgb10 kW	250	293	-17%
Pgc10 kW	250	213	15%
Pga11 kW	200	237	-19%
Pgc11 kW	200	136	32%
Qga10	125	80	36%
Qgb10	125	194	-55%
Qgc10	125	108	14%
Qgc11	100	1	99%

3.2.3 Sistema de prueba de 6 nodos 15 usuarios

Para esta parte de la aplicación se utiliza un sistema de prueba con ramales trifásicos donde las cargas a utilizar son las identificadas como sospechosas en la primera etapa del SVM. Las cargas identificadas corresponden a residenciales, comerciales e industriales los valores de las cargas indicadas extrajeron de las curvas de cargas desarrolladas en la referencia

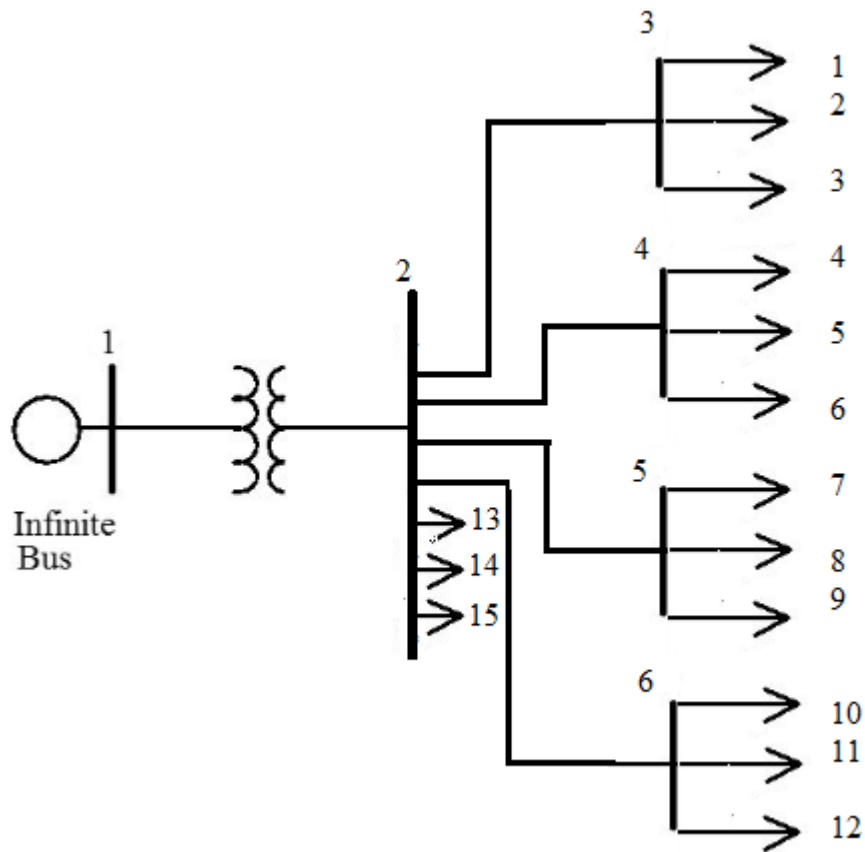


Figura 3.6 Sistema de prueba de 6 nodos, 15 usuarios

Tabla 3.25 Datos del sistema

Tipo do Elemento	Cantidad
Líneas Trifásicas	4
Transformador Estrela-Estrela	1
Datos del nodo	
Tipo de nodo	Cantidad
Nodo de subestación	1
Nodos trifásicos	6
Potencia instalada	
Tipo de carga	Carga Total (kVA)
Cargas nodales	23.1+ j7.59

Tabla 3.26 - Nodos sistema 6 nodos - Plan de medición

Medidores de supervisión		
Medidor	Ubicación del medidor	Precisión (%)
Módulo de la Tensión	1	1
Inyección de potencia Activa y Reactiva	1	3

Flujo de Potencia Activa y Reactiva	1	3
Módulo de la Corriente al Cuadrado	1	3
Pseudomedidas		
Carga	Nodo consumidor	Precisión (%)
Inyección de potencia activa	2,3,4,5,6	5
Inyección de potencia reactiva	2,3,4,5,6	5
Restricciones estructurales		
Restricción	Nodo	Varianza (pu)
Fase de voltaje	1	10^{-8}
Inyección cero de potencia activa y reactiva	1	10^{-8}
Restricciones operativas		
Restricción	Nodos con posibilidad de PNTs	Varianza (pu)
Inyección de potencia activa y reactiva	2,3,4,5,6	10^{-8}

Curva COSTA_RESIDENCIAL_URBANA

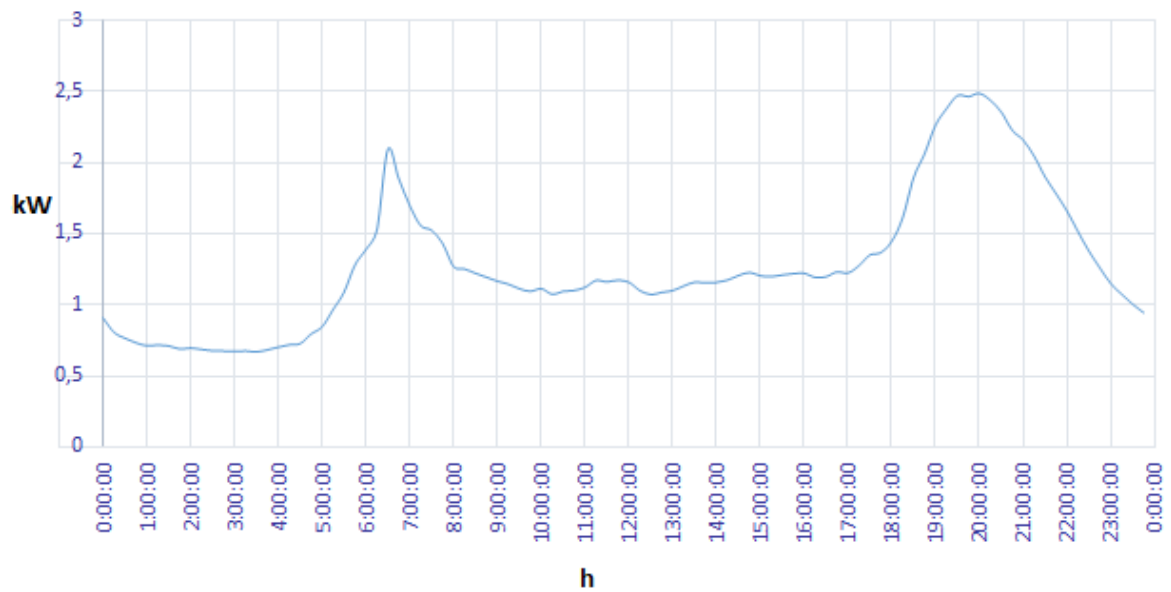


Figura 3.7 Curva de carga residencial urbana Región Costa

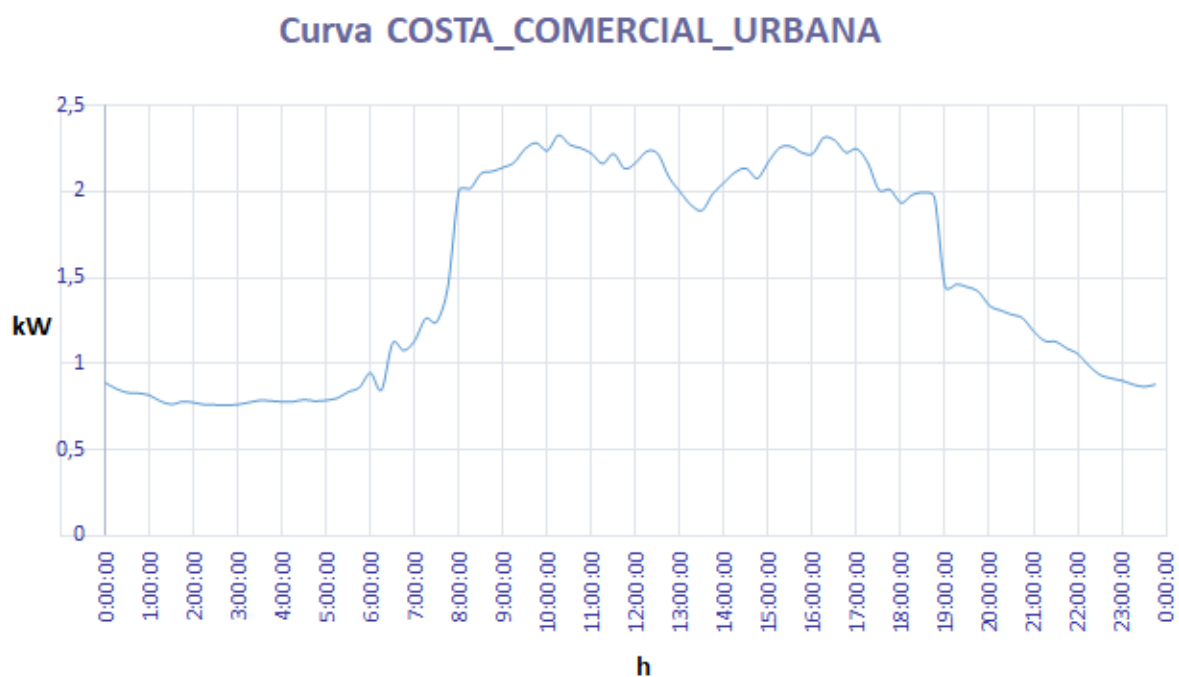


Figura 3.8 Curva de carga comercial urbana Región Costa

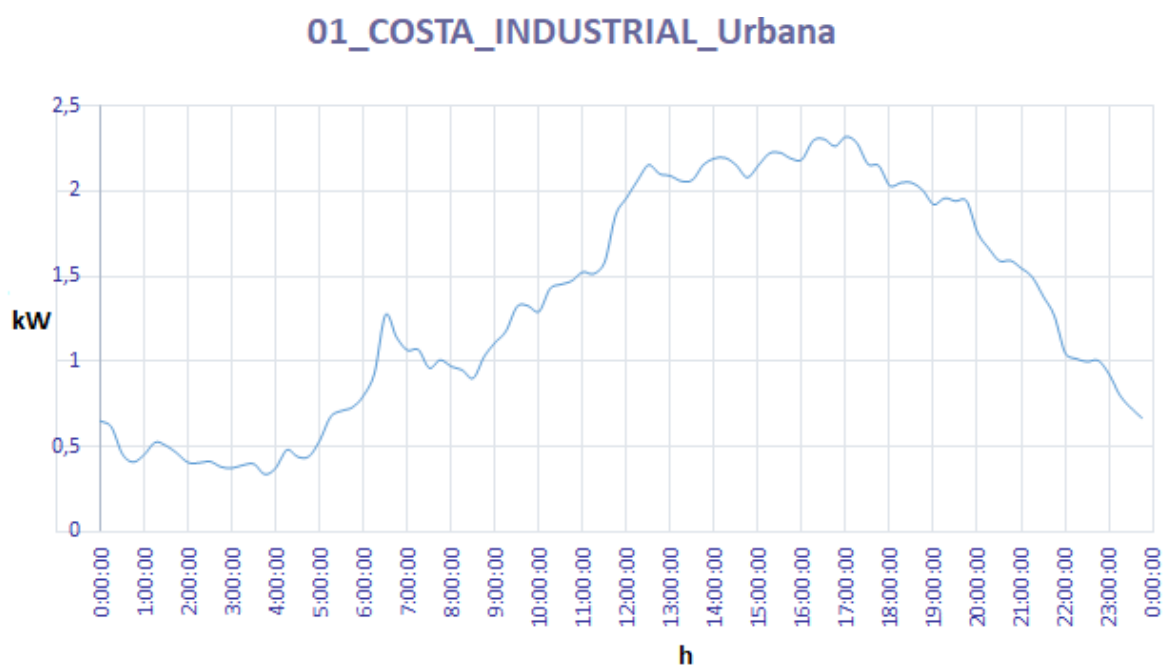


Figura 3.9 Curva de carga industrial urbana Región Costa

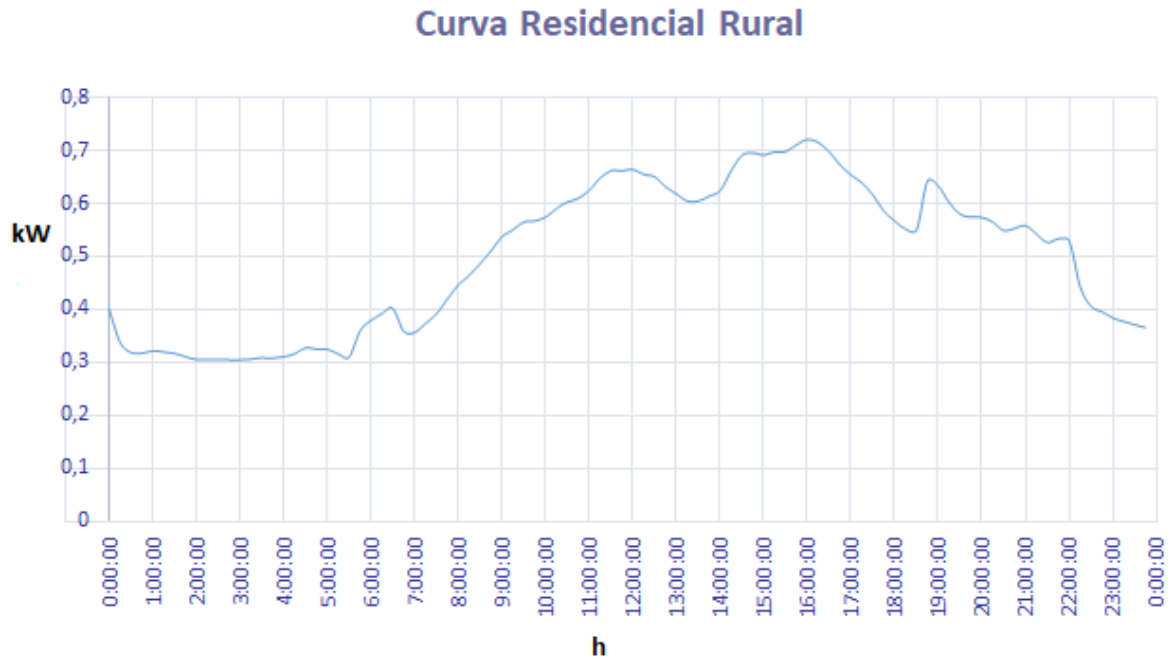


Figura 3.10 Curva de carga residencial Rural Región Costa

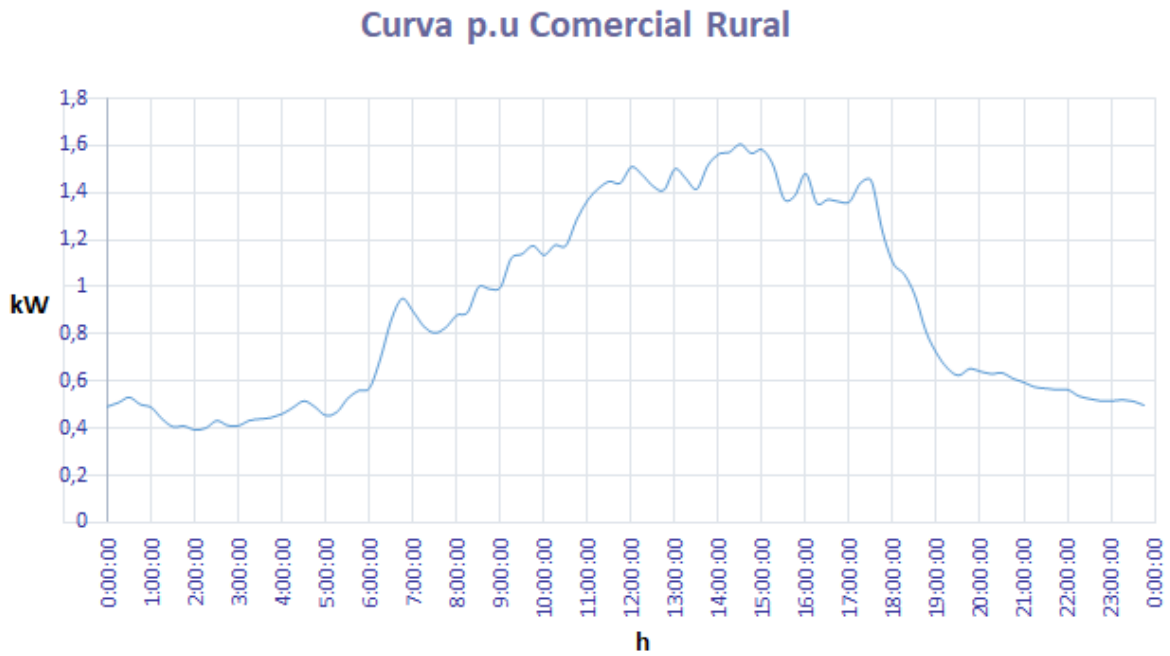


Figura 3.11 Curva de carga comercial Rural Región Costa

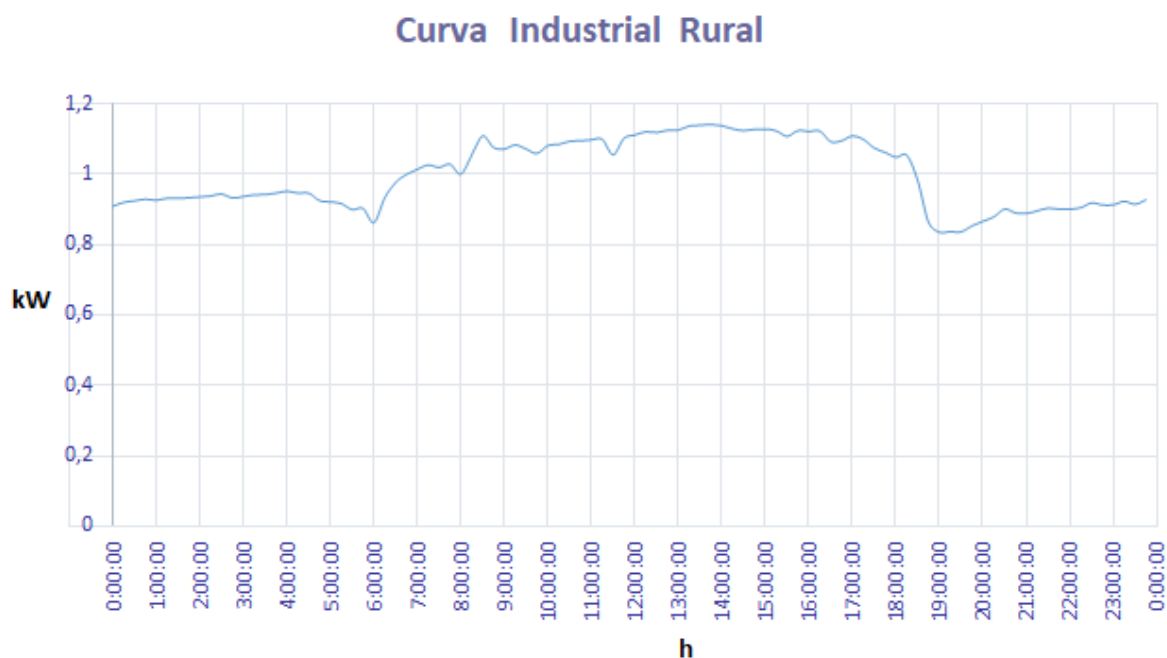


Figura 3.12 Curva de carga Industrial Rural Región Costa

Los valores de carga se extrajeron de las curvas indicadas para el horario de las 18:00 horas y la hipótesis realizada se muestra a continuación en la siguiente tabla

Tabla 3.27 Carga para los nodos

	Potencia activa Kva			Potencia Reactiva Kvar			característica
	P fase a	P fase b	P fase c	Q fase a	Q fase b	Q fase c	
NODO 2	1,99	1,99	1,99	0,654	0,654	0,654	Curva residencial urbana
NODO 3	1,94	1,94	1,94	0,638	0,638	0,638	Curva comercial urbana
NODO 4	2,05	2,05	2,05	0,674	0,674	0,674	Curva industrial urbana
NODO 5	0,58	0,58	0,58	0,191	0,191	0,191	Curva residencial rural
NODO 6	1,14	1,14	1,14	0,375	0,375	0,375	Curva comercial rural

El nivel porcentual estipulado de PNT es el aumento porcentual en relación con la información asumida por el SVM como información correcta en los nodos sospechosos. El nivel de PNT simulados se muestra en la siguiente tabla.

Tabla 3.28 Porcentaje de variación de PNT

Potencia activa Kva			Potencia Reactiva Kvar			Incremento
P fase a	P fase b	P fase c	Q fase a	Q fase b	Q fase c	

NODO 2	2,587	2,587	2,587	0,850	0,850	0,850	30%
NODO 3	1,358	1,358	1,358	0,446	0,446	0,446	-30%
NODO 4	1,025	1,025	1,025	0,337	0,337	0,337	-50%
NODO 5	0,058	0,058	0,058	0,019	0,019	0,019	-80%
NODO 6	1,710	1,710	1,710	0,562	0,562	0,562	50%

Tabla 3.29 Resultado de Simulación con la condición $|\lambda_N| > \lambda_{lim}^0$, donde $\lambda_{lim}^0 = 3$

VALORES MEDIDOS	VALORES CALCULADOS	Denominación	H	λ_N
-1,990	-1,990	Pga2	h0(x)	-4,3727
-1,990	-1,990	Pgb2		-4,4582
-1,990	-1,990	Pgc2		-3,9179
-1,940	-1,434	Pga3		-4,3729
-1,940	-1,416	Pgb3		-4,4585
-1,940	-1,538	Pgc3		-3,9182
-2,050	-2,050	Pga4		-4,6469
-2,050	-2,050	Pgb4		-4,7589
-2,050	-2,050	Pgc4		-4,2620
-0,580	-0,059	Pgc5		-4,7516
-0,580	-0,058	Pgc5		-4,8113
-0,580	-0,058	Pgc5		-4,4477
-1,140	-1,140	Pgc6		-0,7646
-1,140	-1,140	Pgc6		-0,9173
-1,140	-1,140	Pgc6		-0,3642
-0,654	-0,654	Qga2		-1,5776
-0,654	-0,654	Qgb2		-1,8157
-0,654	-0,654	Qgc2		-1,9428
-0,637	-0,492	Qga3		-1,5776
-0,637	-0,445	Qgb3		-1,8158
-0,637	-0,420	Qgc3		-1,9428
-0,673	-0,673	Qga4		-1,7637
-0,673	-0,673	Qgb4		-2,0160
-0,673	-0,673	Qgc4		-2,1401
-0,190	-0,019	Qgc5		-
				1,703661463
-0,190	-0,019	Qgc5		-
				1,877169165
-0,190	-0,019	Qgc5	-	
			-1,9709779	
-0,374	-0,374	Qgc6	-	
			0,405206034	

-0,374	-0,374	Qgc6	- 0,698883279
-0,374	-0,374	Qgc6	- 0,807160099

Tabla 3.30 Resultado de Simulación con la condición $|\lambda_N| > \lambda_{lim}^0$, donde $\lambda_{lim}^0 = 1.3$, $\cos=0.99$

VALORES MEDIDOS	VALORES CALCULADOS	Denominación	H	λ_N
-1,990	-1,990	Pga2	h0(x)	-4,3727
-1,990	-1,990	Pgb2		-4,4582
-1,990	-1,990	Pgc2		-3,9179
-1,940	-1,434	Pga3		-4,3729
-1,940	-1,416	Pgb3		-4,4585
-1,940	-1,538	Pgc3		-3,9182
-2,050	-2,050	Pga4		-4,6469
-2,050	-2,050	Pgb4		-4,7589
-2,050	-2,050	Pgc4		-4,2620
-0,580	-0,059	Pgc5		-4,7516
-0,580	-0,058	Pgc5		-4,8113
-0,580	-0,058	Pgc5		-4,4477
-1,140	-1,140	Pgc6		-0,7646
-1,140	-1,140	Pgc6		-0,9173
-1,140	-1,140	Pgc6		-0,3642
-0,654	-0,654	Qga2		-1,5776
-0,654	-0,654	Qgb2		-1,8157
-0,654	-0,654	Qgc2		-1,9428
-0,637	-0,492	Qga3		-1,5776
-0,637	-0,445	Qgb3		-1,8158
-0,637	-0,420	Qgc3		-1,9428
-0,673	-0,673	Qga4		-1,7637
-0,673	-0,673	Qgb4		-2,0160
-0,673	-0,673	Qgc4		-2,1401
-0,190	-0,019	Qgc5		- 1,703661463
-0,190	-0,019	Qgc5		- 1,877169165
-0,190	-0,019	Qgc5	-1,9709779	
-0,374	-0,374	Qgc6	- 0,405206034	
-0,374	-0,374	Qgc6	- 0,698883279	

-0,374	-0,374	Qgc6	-
			0,807160099

Los resultados marcados con negrilla son detectados como sospechoso de fraude de energía

3.2.4 Sistema real de la Empresa Eléctrica Riobamba

Para esto se selecciona un centro de transformación que presentó mayores usuarios sospechosos según el SVM figura 3.13, donde al sistema se simplifica como se muestra en la figura 3.14, Las características se resumen a continuación

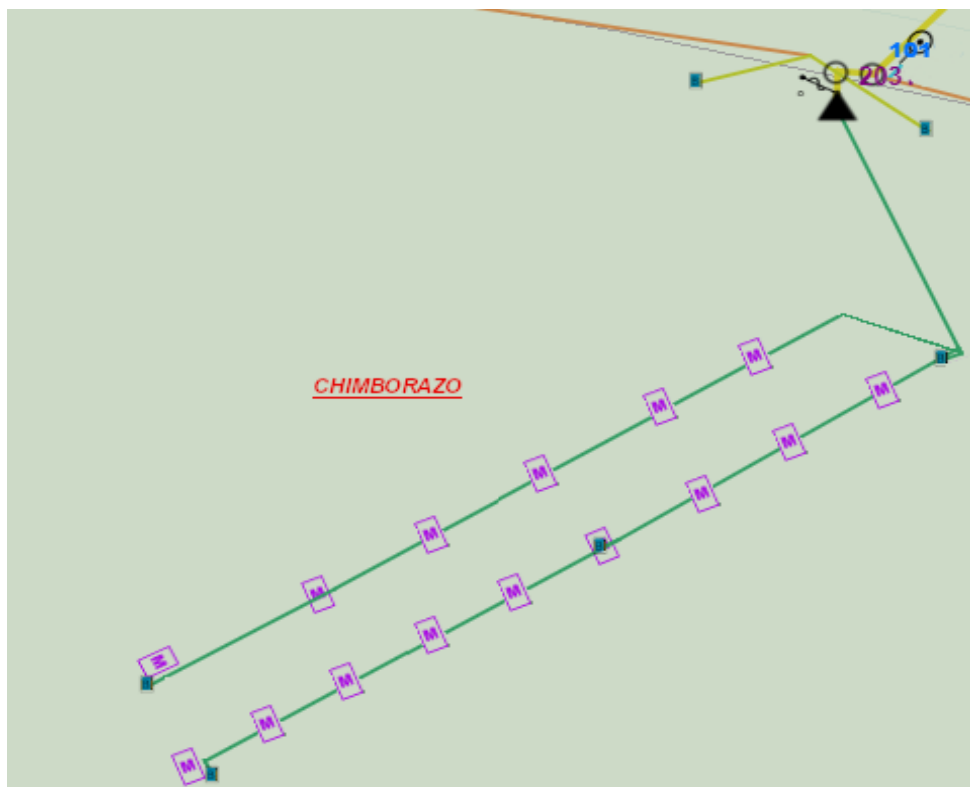


Figura 3.13 Centro de transformación identificado

Tabla 3.31: Datos del sistema

Tipo do Elemento	Cantidad
Líneas Trifásicas	8
Transformador Estrella-Estrella	1
Líneas bifásicas	4
Líneas monofásicas	4
Datos del nodo	
Tipo de nodo	Cantidad
Nodo de subestación	1
Nodos trifásicos	9
Potencia instalada	

Tipo de carga	Carga Total (kVA)
Cargas nodales	23.1+ j7.59

Se realizaron mediciones en campo para la activación del estimador de estado para el cual se realizaron mediciones en tiempo real en el nodo 1 y 2, donde se extrae los valores de 18:00 horas realiza la simulación.

Tabla 3.32 - Sistema 18 nodos - Plan de medición

Medidores de supervisión		
Medidor	Ubicación del medidor	Precisión (%)
Módulo de la Tensión	1 y 2	1
Inyección de potencia Activa y Reactiva	2	3
Flujo de Potencia Activa y Reactiva	2	3
Módulo de la Corriente al Cuadrado	2	3
Pseudomedidas		
Carga	Nodo consumidor	Precisión (%)
Inyección de potencia activa	4,5,6,7,8,9,10,11,12,13,14,15,16,17,18	5
Inyección de potencia reactiva	4,5,6,7,8,9,10,11,12,13,14,15,16,17,18	5
Restricciones estructurales		
Restricción	Nodo	Varianza (pu)
Fase de voltaje	1	10^{-8}
Inyección cero de potencia activa y reactiva	1	10^{-8}
Restricciones operativas		
Restricción	Nodos con posibilidad de PNTs	Varianza (pu)
Inyección de potencia activa y reactiva	5,7,10,11,15,17	10^{-8}

El nivel porcentual estipulado de PNT es el aumento porcentual en relación con la información asumida por el Operador como información correcta en los nodos sospechosos. El nivel de PNT simulados se muestra en la siguiente tabla.

Tabla 3.33 Porcentaje de variación de PNT

	Potencia activa Kva			Potencia Reactiva Kvar			Incremento
	P fase a	P fase b	P fase c	Q fase a	Q fase b	Q fase c	
NODO 5	0,058	0,000	0,000	0,019	0,000	0,000	-80%
NODO 7	0,058	0,058	0,058	0,019	0,019	0,019	-80%

NODO 10	0,058	0,058	0,058	0,019	0,019	0,019	-80%
NODO 11	0,058	0,058	0,000	0,019	0,019	0,000	-80%
NODO 15	0,058	0,000	0,000	0,019	0,000	0,000	-80%
NODO 17	0,058	0,058	0,058	0,019	0,019	0,019	-80%

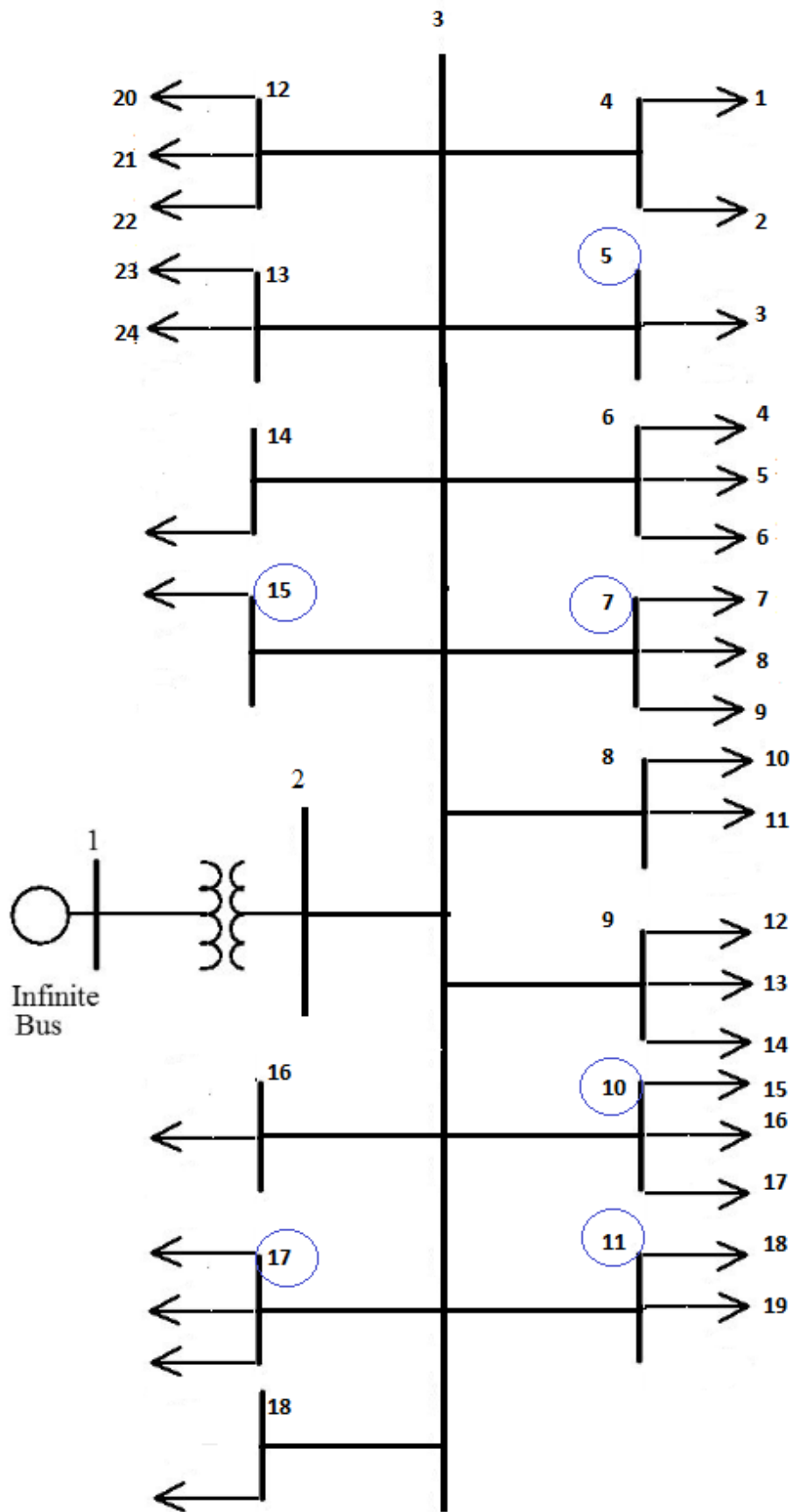


Figura 3.14 Simplificación sistema real Empresa Eléctrica Riobamba

Tabla 3.34 Resultado de Simulación con la condición $|\lambda_N| > \lambda_{lim}^0$, donde $\lambda_{lim}^0 = 3$

VALORES MEDIDOS	VALORES CALCULADOS	Denominación	H	λ_N
-0,199	-0,005072256	Pga5	h0(x)	-17,06991445
0	0	Pgb5		0
0	0	Pgc5		0
-0,199	-0,005196083	Pga7		-17,06449384
-0,199	-0,006107191	Pgb7		-10,58287181
-0,199	-0,003923014	Pgc7		-9,752734535
-0,199	-0,005234301	Pga10		-17,06296381
-0,199	-0,006312935	Pgb10		-10,57415104
-0,199	-0,00392242	Pgc10		-9,752755751
-0,199	-0,005126631	Pgc11		-17,06783573
-0,199	-0,00590651	Pgc11		-10,59166599
0	0	Pgc11		0
-0,199	-0,004915682	Pgc15		-17,10127093
0	0	Pgc15		0
0	0	Pgc15		0
-0,199	-0,005274031	Pga17		-17,06147158
-0,199	-0,006122231	Pgb17		-10,58222137
-0,199	-0,003828902	Pgc17		-9,756723473
-0,0654	-0,001978184	Qga5		-6,115127624
0	0	Qgb5		0
0	0	Qgc5		0
-0,0654	-0,002000258	Qga7		-6,113786592
-0,0654	-0,002000979	Qgb7		-4,064303128
-0,0654	-0,00203	Qgc7		-3,618800836
-0,0654	-0,002019265	Qga10		-6,112974467
-0,0654	-0,002001042	Qgb10		-4,064241722
-0,0654	-0,002028797	Qgc10		-3,618819714
-0,0654	-0,001910712	Qgc11		-6,118420025
-0,0654	-0,002022768	Qgc11		-4,063509084
0	0	Qgc11		0
-0,0654	-0,001892374	Qgc15		-6,139932519
0	0	Qgc15		0
0	0	Qgc15		0
-0,0654	-0,001989376	Qga17	-6,114260636	
-0,0654	-0,002092534	Qgb17	-4,059874111	
-0,0654	-0,002064044	Qgc17	-3,617094704	

Tabla 3.35: El nivel porcentual estipulado de PNT con un incremento del 30%

	Potencia activa Kva			Potencia Reactiva Kvar			Incremento
	P fase a	P fase b	P fase c	Q fase a	Q fase b	Q fase c	
NODO 5	2,610	0,000	0,000	0,065	0,000	0,000	+30%
NODO 7	2,610	2,610	2,610	0,065	0,065	0,065	+30%
NODO 10	2,610	2,610	2,610	0,065	0,065	0,065	+30%
NODO 11	2,610	2,610	0,000	0,065	0,065	0,000	
NODO 15	2,610	0,000	0,000	0,065	0,065	0,000	+30%
NODO 17	2,610	2,610	2,610	0,065	0,065	0,019	+30%

. Tabla 3.36 Resultado de Simulación con la condición $|\lambda_N| > \lambda_{lim}^0$, donde $\lambda_{lim}^0 = 1.2$

VALORES MEDIDOS	VALORES CALCULADOS	Denominación	H	λ_N
-0,199	-0,265652456	Pga5	h0(x)	1,031906566
0	0	Pgb5		0
0	0	Pgc5		0
-0,199	-0,262333156	Pga7		1,170120212
-0,199	-0,261025565	Pgb7		1,124219367
-0,199	-0,256603556	Pgc7		1,032433673
-0,199	-0,267716075	Pga10		2,170430597
-0,199	-0,262199606	Pgb10		1,320603493
-0,199	-0,251236552	Pgc10		1,323522221
-0,199	-0,263614132	Pgc11		2,172187468
-0,199	-0,264723637	Pgc11		1,320841027
0	0	Pgc11		0
-0,199	-0,263824075	Pgc15		2,16443921
0	0	Pgc15		0
0	0	Pgc15		0
-0,199	-0,266057182	Pga17		2,171575662
-0,199	-0,262710813	Pgb17		1,321995519
-0,199	-0,255381896	Pgc17		1,322028655

Al aplicar el estimador de estado se descarta que en el nodo 5 y 7 existe sospecha de fraude de energía

4 CONCLUSIONES

El propósito de este proyecto fue formular una metodología para ser aplicada en la identificación del uso fraudulento de la energía eléctrica mediante la combinación de dos algoritmos SVM Máquina de Soporte de Vectores y Estimador de Estado. Un modelo capaz de identificar actividades fraudulentas en el consumo de electricidad. Para lo cual se utilizó una base de consumo real donde se pudo desarrollar el clasificador SVM en su fase de entrenamiento y prueba, luego se aplicó a un alimentador real de la Empresa Eléctrica Riobamba logrando identificar usuarios sospechosos de fraude de energía, luego finalmente el centro de transformación con mayor número de usuarios sospechosos de fraude se le aplica la metodología de estimador de estado logrando descartar o reafirmar los usuarios sospechosos de fraude de energía.

4.1.1 EL CLASIFICADOR SVM

El clasificador SVM fue desarrollado con una base datos real proporcionados por la Empresa Eléctrica Riobamba para lo cual los 2/3 de la base fue utilizada para la fase de entrenamiento el 1/3 restante fue utilizada para la fase de prueba, el clasificador SVM es directamente afectado por los parámetros del núcleo rbf, γ , y la capacidad, C . En el entorno Matlab R2019a, para su diseño inicial del clasificador SVM, se entrenó y probó utilizando los parámetros predeterminados $\gamma = 1$ y $C = 1$. Luego se optimizaron los parámetros y se encontró que $C = 1.402$ y $\gamma = 2.963$ se utilizó el método de su muestreo aleatorio para entrenar y probar el clasificador SVM 500 veces. Cada vez, se calcularon estimaciones de sensibilidad, especificidad y tasa de error. Posteriormente, se calcularon los promedios para cada una de estas medidas. Este proceso se llevó a cabo tanto antes como después de la optimización de parámetros. El propósito del modelo propuesto era detectar el uso fraudulento de electricidad y, como tal, la especificidad también denominada tasa de detección de fraude, es la medida de la capacidad del modelo para detectar casos fraudulentos.

La tasa de error y la sensibilidad no se vieron afectadas por la optimización del parámetro SVM, consulte la Tabla 3.7. Por el contrario, la tasa de especificidad aumentó de aproximadamente 75.42% a 78,42% al utilizar los parámetros óptimos. Así, el clasificador SVM, diseñados con parámetros optimizados, fue utilizado en la posterior aplicación de los dos alimentadores reales de la Empresa Eléctrica Riobamba, logrando obtener una lista preliminar de usuarios sospechosos de fraude para posterior tratamiento en el estimador de estado.

4.1.2 Estimador de estado

Para la metodología de identificación pérdidas no técnicas a través de la estimación de estado se utilizó la resolución de mínimos cuadrados ponderados y prueba de error grueso basada en una interpretación geométrica de los multiplicadores de Lagrange normalizados.

Debido a que los sistemas de distribución y sus diferentes características, son de topología radial y configuraciones que varían de monofásicos, bifásicos y trifásicos, el uso del modelado trifásico para estimación de estado fue muy importante ya que posibilitó la aplicación de la metodología de identificación de PNTs, dada las dificultades impuestas por la estructura del sistema de distribución y la baja redundancia y precisión que presenta el sistema real en estudio.

4.1.3 Combinación Clasificador SVM y Estimador de Estado

Al combinar las dos metodologías en el sistema de prueba de 6 nodos donde intencionalmente se introdujo consumidores reales identificados por el SVM como sospechoso de fraude de energía en los nodos 2,3,4,5,6 al aplicar la metodología de Estimador de Estado y error grueso, se descartó al nodo 6 como libre de fraude de energía y ratificando a los demás nodos con sospecha de energía.

Para el sistema real de la Empresa Eléctrica Riobamba se aplicó la metodología propuesta correspondiente al alimentador 1 de la subestación 1, donde al aplicar el SVM se identificó un centro de transformación que corresponde a un conjunto habitacional que contiene 31 usuarios y 18 nodos, el cual presenta usuarios 13 usuarios con sospecha fraude y que están conectados a los nodos 5,7,10,11,15,17 respectivamente, al aplicar la metodología de estimador de estado y error grueso se descartan 4 usuarios de sospecha de fraude que se encuentra conectados al nodo 5 y 7, se hizo las debidas inspecciones de campo donde se pudo ratificar que efectivamente en los nodos 10,11,15 y 17 existe usuarios fraudulentos pero en el nodo 7 se descartó que está libre de fraude; además se logró ratificar que en el nodo 6 existe fraude de energía, el nodo 3,4,8,9,12,13,14,16 y 18 se constató que están libre de fraude.

Se puede concluir que al solo utilizar la metodología de SVM se podría tener un nivel de aciertos de usuarios con fraude de energía del 70% y al combinar las dos metodologías en cuestión se podría tener un nivel de aciertos del 76%. Por lo que se tendría una ligera mejora en la identificación de usuarios fraudulentos. Esto contribuirá de manera favorable en la mitigación de fraude optimizando las inspecciones en los sitios donde se procede este evento y que permitirá la reducción de PNT en el Sistema de Distribución.

4.1.4 Trabajos Futuros

En trabajos futuros se pretende a corto plazo la implantación de la metodología en la Distribuidora en cuestión y posteriormente a Empresas que contengan mayor porcentaje de pérdidas no técnicas; así como también plantear una nueva metodología de cálculo de Pérdidas de Energía con la aplicación de del estimador de estado.

5 REFERENCIAS BIBLIOGRÁFICAS

- [1] Agencia de Regulación y Control de Energía y Recursos Naturales No Renovables ESTADÍSTICA ANUAL Y MULTIANUAL DEL SECTOR ELÉCTRICO ECUATORIANO,” Quito 2021.
- [2] Ministerio de Electricidad y Energía Renovable MEER. “PME (Plan Maestro de la Electrificación 2016–2025),” Quito 2016.
- [3] S. Cañar, “Cálculo detallado en Sistemas Eléctricos de Distribución Aplicado al Alimentador Universidad,” *Departamento de Energía Eléctrica EPN*, Quito 2009,
- [4] J. Tejado, G. Duran, R. Jiménez, M. Doyle, “Incremento de la eficiencia del sector eléctrico”, *Lecciones sobre la relación de pérdidas eléctricas en Ecuador*, 2017.
- [5] T. S. D. Ferreira, F. C. L. Trindade, W. Freitas Filho, J. C. M. Vieira Jr., “Detecção de Perdas Não Técnicas na Presença dos Medidores Inteligentes”, *Anais do V Simpósio Brasileiro de Sistemas Elétricos - SBSE*, 2014.
- [6] C. Penin, “Combate, prevenção e otimização das perdas comerciais de energia elétrica”. 2008. 214p. Ph. D. thesis Universidade de São Paulo, São Paulo, 2008.
- [7] J. Hernandez, “Processo não invasivo de baixo custo para otimização da rotina de inspeção na detecção de furto de energia eléctrica”. *Revista Pesquisa e Desenvolvimento da ANEEL - P&D, [S.l.], n.5, p.47–51, Ago 2013*.
- [8] R. Cruz, F. Pérez, “Detecting Non-Technical Losses in Radial Distribution System Transformation Point through the Real Time State Estimation Method”. In: *IEEE/PES Transmission & Distribution Conference and Exposition: Latin america*, 2006,
- [9] J. Nagi, A. Mohammad, K. Yap, S. Tiong, and S. Ahmed, “Non-technical loss analysis for detection of electricity theft using support vector machines,” in *Proc. 2nd IEEE Int. Power and Energy Conf.*, 200.
- [10] I. Monedero, F. Biscarri, C. León, J. Biscarri, and R. Millán, “MIDAS: “Detection of non-technical losses in electrical consumption using neural networks and statistical techniques,” *Conf. Computational Science and Applications, Springer, Berlin/Heidelberg, Germany*, 2006.

- [11] G. Chicco y otros, “Comparisons Among Clustering Techniques for Electricity Custome Classification,” *IEEE Transactions on Power Systems* , vol. 21, n° 2, pp. 933-940, Mayo 2016.
- [12] E. Angelos y otros, “Detection and Identification of Abnormalities in Customer Consumptions in Power Distribution Systems,” *IEEE Transactions on Power Delivery*, 2011.
- [13] R. Trivizan, A. Rossoni, A. Bretas, D. Gazzana “Non-Technical Losses Identification Using Optimum Path Forest and State Estimation”, *Electrical Engineering Department – DELET Federal University of Rio Grande do Sul – UFRGS Porto Alegre, Brazil*.
- [14] A. Rossoni, “Estimação de perdas técnicas e comerciais: métodos baseados em fluxo de carga e estimador de estados,” M.S. thesis, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2014.
- [15] A. Rodrigues, A. Costa , D. Issicaba, “Identification of Non-Technical Losses in Distribution Systems via State Estimation and Geometric Tests”, *Departamento de Engenharia Elétrica, Grupo de Sistema de Potência Universidade Federal de Santa Catarina Florianópolis, SC, Brasil* 2018.
- [16] A. Jindal y otros, “Decision Tree and SVM-Based Data Analytics for Theft Detection in Smart Grid,” *IEEE Transactions On Industrial Informatics* , Junio 2016.
- [17] C. Cabrera, P. Rodas “Metodología para determinar las pérdidas no técnicas de energía en el sistema de distribución de la Empresa Eléctrica Regional CENTROSUR” M.S. tesis Universidad Politécnica Salesiana, Cuenca, jul 2016.
- [18] M. Á. G. Juárez “Estimadores de Estado en Redes de Distribución: Revisión del Estado del Arte”, *Dep. Ingeniería Eléctrica Escuela Técnica Superior de Ingeniería Universidad de Sevilla, Sevilla, 2016*.
- [19] J.A. Jardini, M.V. tahan, M.r. Gouvea, and S. Un ahn, “Daily load profiles for residential, commercial and industrial low voltage consumers”, *IEEE Trans. On Power Delivery*, Jan. 2000.
- [20] W.H. Kersting, *Distribución System Modeling and Analysis, Third Edition.2012*
- [21] A. Nizar, Z. Dong, J. Zhao, and P. Zhang, “A data mining-based NTL analysis method,” in Proc. *IEEE Power Eng. Soc. General Meeting, 2007*.
- [22] J. Nagi, K. Yap, S. Tiong, S. Ahmed, and A. Mohammad, “Detection of abnormalities and electricity theft using genetic support vector machines,” in Proc. *IEEE Tencon Region 10 Conf.*, 2008.

- [23] RAMOS, C. C. O. et al. “A new approach for non technical losses detection based on optimum-path forest”. *IEEE Transactionson Power Systems* ,2011
- [24] A. Ortiz “Estimación de estado y ubicación óptima de medidores de voltaje y corriente en redes eléctricas de distribución desbalanceadas” M.S. tesis, EPN, 2018.
- [25] I. MONEDERO. “Detection of frauds and other non-technical losses in a power utility using Pearson coefficient, Bayesian networks and decision trees”. *International Journal of Electrical Power & Energy Systems*, 2012.
- [26] Abur, A; Exposito, A. G. Power System State Estimation: Theory and Implementation. New York: Marcel Dekker, 2004.
- [27] Wenz. S. R. “Detecção de perdas no técnicas em redes de distribuição radiais usando estimación de estado” Rio de Janeiro Diciembre, 2016
- [28] COLOMBO, J. L. Monitoração dos status de fontes ativas de geração distribuída através de testes de hipóteses. Trabalho de Conclusão de Curso, Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, SC, 2013.
- [29] PAIVA, R. R. d. C. Fluxo de potência ótimo em redes de distribuição de energia com a presença de geração distribuída: um novo algoritmo para auxiliar a análise do perfil de tensão. [S.l.]: Florianópolis, SC, 2006.
- [30] SCHWEPPE, F. C.; WILDES, J. Power system static-state estimation, part i: Exact model. *IEEE Transactions on Power Apparatus and systems*, IEEE, 1970.
- [31] LOURENÇO, E. Análise de observabilidade e identificação de erros de topologia na estimação de estados generalizada. Tese de Doutorado, Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, SC, 2001.
- [32] QUINTANA, V.; SIMÕES COSTA, A.; MANDEL, A. Power system topological observability using a direct graph-theoretic approach. *Power Apparatus and Systems, IEEE Transactions on* 1982.
- [33] CLEMENTS, K.; SIMÕES COSTA, A. Topology error identification using normalized lagrange multipliers. *Power Systems, IEEE Transactions on*, IEEE, 1998.
- [34] BEN-ISRAEL, A.; GREVILLE, T. N. Generalized inverses: theory and applications. [S.l.]: Springer Science& Business Media, 2003.

- [35] S. Pazi STATISTICAL METHODS FOR THE DETECTION OF NON-TECHNICAL LOSSES: A CASE STUDY FOR THE NELSON MANDELA BAY MUNICIPALITY, Nelson Mandela Metropolitan University, Abril 2017.
- [36] K. Clements e A. Simões Costa, “Topology error identification using normalized lagrange multipliers,” IEEE Transactions on Power Systems, 1998

6 ANEXOS

6.1.1 Modelado Matricial Generalizado SD

6.1.1.1 Líneas de distribución

Para el modelado de las matrices generalizadas de las líneas de distribución (LD), se utilizaron las matrices de impedancia serie y admitancia shunt, Z_{abc} y Y_{abc} respectivamente, determinadas de acuerdo a la geometría y posicionamiento de los cables, conductores utilizados y posicionamiento de las fases, a través de las ecuaciones de Carson y la reducción de Kron [20].

De Z_{abc} y Y_{abc} , calculado como se presentarán dos conjuntos de matrices generalizadas, con base en las siguientes consideraciones: modelo de segmento de línea exacta; Modelo simplificado del segmento de línea.

Luego, se comenta el modelado de LDs bifásicos y monofásicos.

6.1.1.2 Modelo exacto de LD

El modelo exacto de la LD, similar a la representación π de las líneas, se muestra en la Figura 1.17. Para una LD trifásico, las matrices de impedancia en serie y admitancia en derivación se representan como:

$$Z_{abc} = \begin{bmatrix} Z_{aa} & Z_{ab} & Z_{ac} \\ Z_{ba} & Z_{bb} & Z_{bc} \\ Z_{ca} & Z_{cb} & Z_{cc} \end{bmatrix}$$

Ecuación 6.1 Matriz de impedancia

$$Y_{abc} = \begin{bmatrix} Y_{aa} & Y_{ab} & Y_{ac} \\ Y_{ba} & Y_{bb} & Y_{bc} \\ Y_{ca} & Y_{cb} & Y_{cc} \end{bmatrix}$$

Ecuación 6.2 Matriz de admitancia

donde los elementos de las diagonales principales representan las impedancias y admitancias propias de la LD, mientras que los elementos fuera de las diagonales principales representan las impedancias y admitancias mutuas. [20] Aplicando las ecuaciones de Kirchhoff, las matrices generalizadas de la representación exacta de la LD se pueden definir:

$$\begin{cases} [a] = [U] + \frac{1}{2} \cdot [Z_{abc}] \cdot [Y_{abc}] \\ [b] = [Z_{abc}] \\ [c] = [Y_{abc}] + \frac{1}{4} \cdot [Y_{abc}] [Z_{abc}] \cdot [Y_{abc}] \\ [d] = [U] + \frac{1}{2} \cdot [Z_{abc}] \cdot [Y_{abc}] \end{cases}$$

Ecuación 6.3

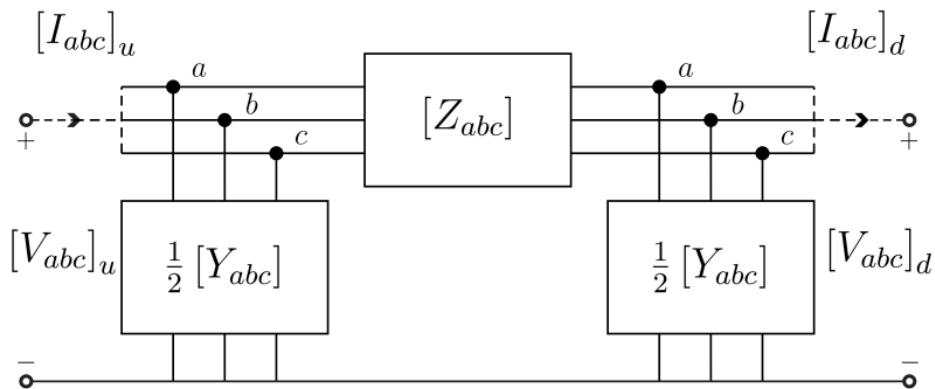


Figura 6.1 Modelo Exacto de la Línea de Distribución

donde $[U]$ representa la matriz identidad, de dimensión 3x3 Finalmente.[20]

6.1.1.3 Modelo simplificado LD

En los SD, las admitancias en derivación de las líneas aéreas tienden a asumir valores de orden de magnitud mucho más bajos que las impedancias en serie, por lo que estas magnitudes pueden despreciarse, en lo que respecta al modelado de las LD (KERSTING,2012).[20] Por lo tanto, a partir de esta simplificación, es posible representar la LD como se muestra en la Figura 6.1

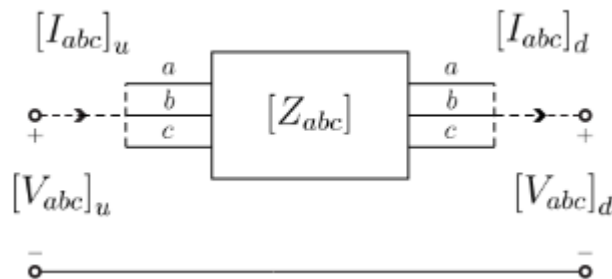


Figura 6.2 Modelo Simplificado de la Línea de Distribución

Para determinar las matrices generalizadas considerando esta hipótesis simplificadora, considerando $[Y_{abc}] = [0]$:

$$\begin{cases} [a] = [b] = [\mathcal{A}] = [U] \\ [b] = [\mathcal{B}] = [Z_{abc}] \\ [c] = [0] \end{cases}$$

Ecuación 6.4

La elección de utilizar el modelo exacto o simplificado dependerá del caso en estudio.

6.1.1.4 LD Bifásico y Monofásico

Como se discutió, el modelado de matrices generalizadas para la LD requiere la determinación de las impedancias en serie equivalentes, Z_{abc} a través de las ecuaciones de Carson y la reducción de Kron. Sin embargo, tratándose de ramales bifásicos y monofásicos, estos procedimientos darían como resultado matrices Z_{abc} de dimensiones 2x2 y 1x1 respectivamente. Así, la aplicación de la Ecuación (6.4) daría como resultado matriz generalizadas sin la dimensión 3x3 adecuada, lo que dificultaría la manipulación adicional de estas matrices para modelar estimadores de estado. [20]

El problema de la dimensión se puede resolver agregando componentes ficticios, relativos a fases inexistentes. En la práctica, este procedimiento es equivalente a agregar columnas y filas nulas a las matrices Z_{abc} de ramas bifásicas y monofásicas, expandiéndolas a dimensión 3 x3 correcta. Por lo tanto, estas ramas se pueden modelar como LD trifásicos. [20]

Como ejemplo, considere una línea bifásica con las fases *a* y *c*, y una rama monofásica, que contiene solo la fase *a*. Por lo tanto, agregando una fase *b* ficticia a la primera y las fases *b* y *c* a la última, sus matrices de impedancia en serie equivalentes se pueden reescribir como: [20]

$$Z_{abc} = \begin{bmatrix} Z_{aa} & 0 & Z_{ac} \\ 0 & 0 & 0 \\ Z_{ca} & 0 & Z_{cc} \end{bmatrix}$$

Ecuación 6.5

$$Z_{abc} = \begin{bmatrix} Z_{aa} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Ecuación 6.6

que permite el análisis de estos elementos como ramas trifásicas.

6.1.1.5 Transformadores de Distribución

Al tratarse de sistemas de distribución de energía, otro componente importante a modelar es el transformador. Para atender a los consumidores de bajo voltaje, es necesario disponer de energía en distintos niveles de voltaje, en distintos tramos de los alimentadores. Sin embargo, a diferencia de la LD, la determinación de las matrices generalizadas de estos elementos no depende solo de los aspectos constructivos de los transformadores, sino también de la forma en que se conectan.

Para ejemplificar esta diferencia, se modelan dos configuraciones de transformadores:

Estrella a tierra - Estrella a tierra;

Delta - Delta.

Configuración de estrella a tierra - Estrella a tierra

La figura 6.3 presenta un banco de transformadores en configuración estrella a tierra – estrella a tierra, conectando los nodos u y d . Considerando un banco compuesto por transformadores idénticos, las tensiones de fase de los lados primario y secundario de sus bobinas se pueden expresar a través de su relación de transformación nt :

$$[V_{abc}]_1 = \begin{bmatrix} nt & 0 & 0 \\ 0 & nt & 0 \\ 0 & 0 & nt \end{bmatrix} \cdot [V_{abc}]_2 = nt \cdot [U] \cdot [V_{abc}]_2$$

Ecuación 6.7

Del mismo modo, las corrientes aguas abajo y aguas arriba también pueden estar relacionados:

$$[I_{abc}]_u = \begin{bmatrix} 1/nt & 0 & 0 \\ 0 & 1/nt & 0 \\ 0 & 0 & 1/nt \end{bmatrix} \cdot [I_{abc}]_d = 1/nt \cdot [U] \cdot [I_{abc}]_d$$

Ecuación 6.8

Aplicando las leyes de Kirchhoff en los lados primario y secundario, para obtener $[V_{abc}]_1$ y $[V_{abc}]_2$, respectivamente, y usando la Ecuación (6.8), es posible relacionar los voltajes aguas abajo y aguas arriba a través de [20]:

$$[V_{abc}]_u = nt[U] \cdot ([V_{abc}]_d + [Z_{tabc}] \cdot [I_{abc}]_d)$$

Ecuación 6.9

donde la matriz de impedancia $[Z_{abc}]$ incluye las impedancias en serie de los devanados del transformador:

$$[Z_{tabc}] = \begin{bmatrix} z_{ta} & 0 & 0 \\ 0 & z_{tb} & 0 \\ 0 & 0 & z_{tc} \end{bmatrix}$$

Ecuación 6.10

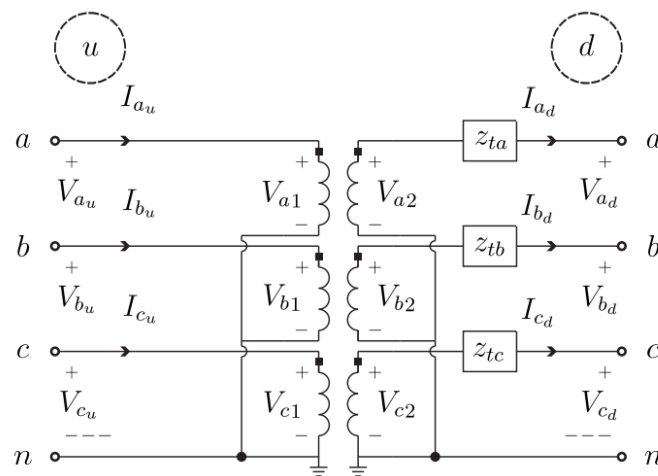


Figura 6.3 Configuración de estrella puesta a tierra - Estrella puesta a tierra

Con la manipulación de las ecuaciones (6.8) y (6.9) es posible determinar las matrices generalizadas utilizadas en las Ecuaciones (1.24) y (1.25):

$$\begin{cases} [\mathbf{a}_t] = nt \cdot [U] \\ [\mathbf{b}_t] = nt \cdot [Z_{tabc}] \\ [\mathbf{c}_t] = [0] \\ [\mathbf{d}] = \frac{1}{nt} \cdot [U] \end{cases}$$

Ecuación 6.11

Finalmente, las matrices generalizadas $[\mathcal{A}_t]$ y $[\mathcal{B}_t]$ se determinan a través de las Ecuaciones (1.27) y (1.28)

$$\begin{cases} [\mathcal{A}_t] = [\mathbf{a}_t]^{-1} = \frac{1}{nt} \cdot [U] \\ [\mathcal{B}_t] = [\mathbf{a}_t]^{-1} \cdot [\mathbf{b}_t] = [Z_{tabc}] \end{cases}$$

Ecuación 6.12

Configuración Delta – Delta

La figura 6.4 muestra un diagrama de un banco de transformadores monofásicos conectados en triángulo-triángulo, situada entre los nodos u y d.

Similar a la configuración estrella-tierra, considerando que los transformadores son idénticos, las tensiones de línea primaria y secundaria están relacionadas por:

$$[V_{Labc}]_u = \begin{bmatrix} nt & 0 & 0 \\ 0 & nt & 0 \\ 0 & 0 & nt \end{bmatrix} \cdot [V_{Labc}]_2 = nt[U] \cdot [V_{Labc}]_2$$

Ecuación 6.13

Donde:

$$[V_{Labc}] = [V_{Lab} \ V_{Lbc} \ V_{Lca}]^T$$

Ecuación 6.14

Se puede escribir una relación análoga para las corrientes de fase dentro de la configuración delta:

$$[I_{Labc}]_1 = \frac{1}{nt} \cdot [U] \cdot [I_{Labc}]_2$$

Ecuación 6.15

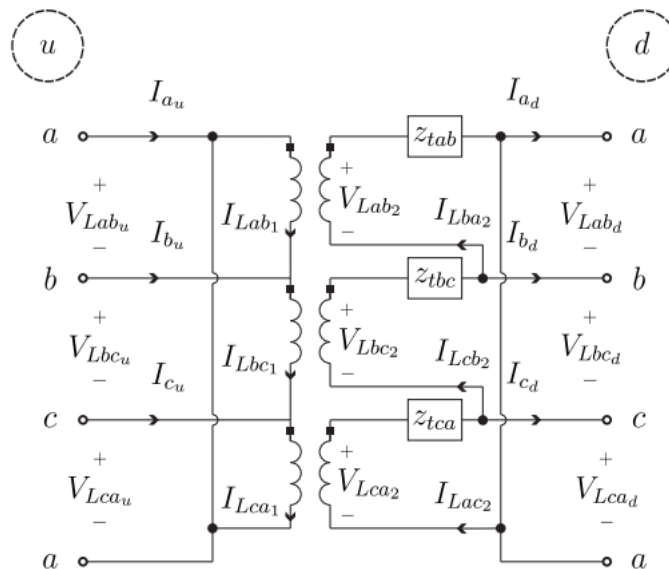


Figura 6.4 Configuración Delta - Delta

Donde:

$$[I_{Labc}]_1 = [L_{Lab1} \quad I_{Lbc1} \quad I_{Lca1}]^T$$

$$[I_{Labc}]_2 = [L_{Lab2} \quad I_{Lbc2} \quad I_{Lca2}]^T$$

Ecuación 6.16

El modelado matricial generalizado relaciona las tensiones corrientes, de fase y de línea. La ecuación (6.15) se ocupa de las corrientes de fase en la configuración delta. Sin embargo, Kersting (2012) muestra que la relación dada por la Ecuación (6.15) se puede extender a las corrientes de línea aguas abajo y aguas arriba:

$$[I_{abc}]_u = \frac{1}{nt} \cdot [U] \cdot [I_{abc}]_d$$

Ecuación 6.17

Aplicando la ley de voltajes de Kirchhoff en el lado secundario de la figura 1.20, y usando la Ecuación (6.13), puede ser determinado los voltajes de línea aguas arriba en función de los voltajes Corrientes de fase secundaria del banco de transformadores y línea aguas abajo:

$$[V_{Labc}]_u = nt \cdot [U] \cdot [V_{Labc}]_d + nt \cdot [U] \cdot [Z_{tabc}] \cdot [I_{abc}]_d$$

Ecuación 6.18

Donde

$$[Z_{tabc}] = \begin{bmatrix} Z_{tab} & 0 & 0 \\ 0 & Z_{tbc} & 0 \\ 0 & 0 & Z_{tca} \end{bmatrix}$$

Ecuación 6.19

Aplicando un procedimiento análogo, es posible determinar las tensiones de línea aguas abajo del elemento:

$$[V_{Labc}]_d = \frac{1}{nt} \cdot [U] \cdot [V_{Labc}]_u - [Z_{tabc}] \cdot [I_{abc}]_d$$

Ecuación 6.20

Las ecuaciones (6.18) y (6.20) son muy similares a las ecuaciones generalizadas (1.24) y (1.26), con la diferencia de que en el primer par las tensiones expresadas son tensiones de línea y las corrientes son corrientes de fase. Así, para obtener las matrices generalizadas, se puede determinar cómo se relacionan las corrientes y los voltajes de

fase y de línea. Para esta configuración, las corrientes de fase aguas abajo del elemento $[I_{Labc}]_d$ están relacionadas con las corrientes de línea $[I_{abc}]_d$. [20]

$$[I_{abc}]_d = [G1] \cdot [I_{abc}]_d$$

Ecuación 6.21

Siendo que:

$$[G1] = \frac{1}{z_{tab} + z_{tbc} + z_{tca}} \cdot \begin{bmatrix} z_{tca} & -z_{tbc} & 0 \\ z_{tca} & z_{tab} + z_{tca} & 0 \\ -z_{tab} - z_{tbc} & -z_{tab} & 0 \end{bmatrix}$$

Ecuación 6.22

Para el modelado generalizado de transformadores, en la referencia [28] complementa el procedimiento obtener tensiones de línea y de fase, al considerar sistemas desequilibrados, la ecuación generalizada (1.24) para la configuración delta-delta viene dada por:

$$[V_{abc}]_u = nt \cdot [F] \cdot [D] \cdot [V_{abc}]_d + nt \cdot [F] \cdot [Z_{tabc}] \cdot [G1] \cdot [I_{abc}]_d$$

Ecuación 6.23

donde las matrices [D] y [F] están dadas por Ecuaciones (6.24) y (6.25)[20]:

$$[D] = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix}$$

Ecuación 6.24

$$[F] = 1/3 \cdot \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & -1 \\ 1 & 0 & 2 \end{bmatrix}$$

Ecuación 6.25

6.1.1.6 Reguladores de voltaje

Otro componente fundamental importante para los SD son los reguladores de voltaje. Estos son autotransformadores capaces de cambiar sus tomas automáticamente, con el objetivo de mantener la magnitud del voltaje de un punto dado de SD en el nivel deseado. La posición del Tap está determinada por un circuito compensador de control. Los

reguladores estandarizados permiten la conmutación inversa, capaz de proporcionar un cambio del 10 % en el voltaje del regulador, en 32 posiciones. [28] Los reguladores de voltaje se pueden conectar en conexiones tipo "A" o tipo "B" (IEEE,1988). La figura 6.5 cuenta con la conexión más común, Tipo "B". Para un regulador Tipo "B" con 32 tomas, con una variación de $\pm 10\%$, el voltaje aumenta en $5/8\%$, o 0.00625 por:

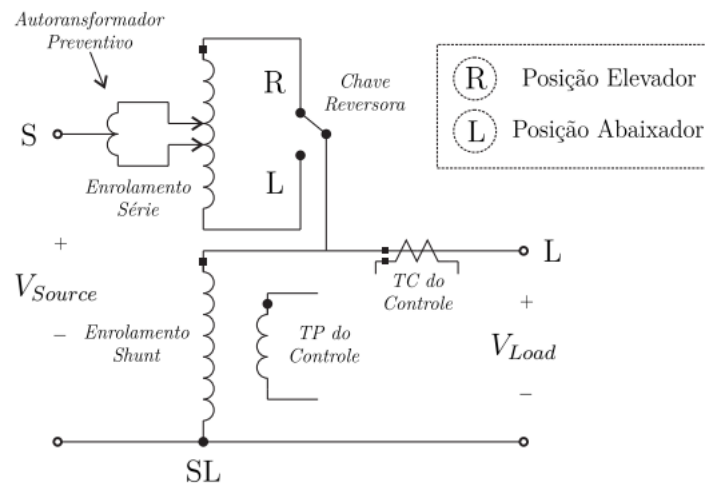


Figura 6.5 La regulación de voltaje [20]

$$a_R = 1 \mp 0,00625 * tap$$

Ecuación 6.26

donde el signo menos se aplica al regulador en la posición hacia arriba y el signo más, en la posición hacia abajo. La posición de descenso o elevación está determinada por el interruptor de inversión presente en el regulador. lado de voltaje, como se muestra en la figura 6.5. En este trabajo, los reguladores de voltaje se modelan como ideales, es decir, sus impedancias se consideran nulas. Además, se considera que las posiciones de las tomas son información conocida por el operador del sistema. A partir de estas consideraciones, se obtienen luego matrices generalizadas para bancos de reguladores de tensión Tipo "B" a partir de las siguientes conexiones:

- Conexión en estrella puesta a tierra;
- Conexión delta abierta.
- *Conexión estrella conectado a tierra*

Un banco de reguladores de voltaje conectados en estrella tiene una conexión análoga al banco de transformadores conectados en estrella con conexión a tierra - conexión en estrella con conexión a tierra, que se muestra en la figura 6.3. Cada fase tiene su propio

regulador Tipo "B", con neutro común y puesto a tierra, con relaciones de regulación de tensión aR_a , aR_b y aR_c relativas a las tomas de cada fase, y determinadas mediante la Ecuación (6.26).

Al tratarse de una conexión individual para cada fase, y teniendo en cuenta la idealidad de cada regulador, las matrices generalizadas de esta configuración vienen dadas por [20]:

$$\left\{ \begin{array}{l} \mathbf{a}_r = \begin{bmatrix} aR_a & 0 & 0 \\ 0 & aR_b & 0 \\ 0 & 0 & aR_c \end{bmatrix} \\ [\mathbf{b}_r] = [0] \\ [\mathbf{c}_r] = [0] \\ [\mathbf{d}_r] = \begin{bmatrix} \frac{1}{aR_a} & 0 & 0 \\ 0 & \frac{1}{aR_b} & 0 \\ 0 & 0 & \frac{1}{aR_c} \end{bmatrix} \end{array} \right.$$

Ecuación 6.27

Las matrices $[\mathcal{A}_r]$ y $[B_r]$ se obtienen aplicando las Ecuaciones (1.27) y (1.28).

Conexión Delta abierta Típica de los alimentadores conectados en triángulo de tres hilos, la configuración en triángulo abierto tiene dos reguladores de voltaje Tipo "B" conectados entre dos de las fases del sistema. A

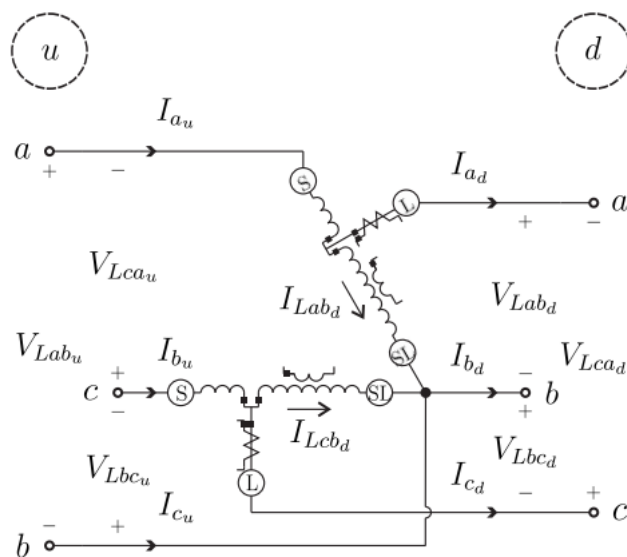


Figura 6.6 Banco de Reguladores con Conexión Delta Abierta

Debido a su diferente configuración, las tensiones de línea aguas arriba y aguas abajo están relacionadas por [20]:

$$[V_{Labc}]_u = \begin{bmatrix} aR_{ab} & 0 & 0 \\ 0 & aR_{cb} & 0 \\ -aR_{ab} & -aR_{ac} & 0 \end{bmatrix} \cdot [V_{Labc}]_d$$

Ecuación 6.28

mientras que las corrientes de línea aguas arriba y aguas abajo son dada por [20]:

$$[I_{abc}]_u = \begin{bmatrix} \frac{1}{aR_{ab}} & 0 & 0 \\ -\frac{1}{aR_{ab}} & 0 & -\frac{1}{aR_{ac}} \\ 0 & 0 & \frac{1}{aR_{cb}} \end{bmatrix} \cdot [I_{abc}]_d$$

Ecuación 6.29

Debido a que es una relación entre voltajes de línea, Ecuación (6.28) no se puede utilizar para la determinación directa de matrices generalizadas, ya que relacionan tensiones de fase. Usando ecuaciones (6.28) y (6.29), junto con las relaciones de transformación de tensiones de línea en tensiones de fase, y viceversa, descritas en [28] es posible determinar las matrices generalizadas para el banco de reguladores conectados en delta abierto:

$$\left\{ \begin{array}{l} [a_r] = [F] \cdot \begin{bmatrix} aR_{ab} & 0 & 0 \\ 0 & aR_{cb} & 0 \\ -aR_{ab} & -aR_{cb} & 0 \end{bmatrix} \cdot [D] \\ [b_r] = [0] \\ [c_r] = [0] \\ [d_r] = [F] \cdot \begin{bmatrix} \frac{1}{aR_{ab}} & 0 & 0 \\ -\frac{1}{aR_{ab}} & 0 & -\frac{1}{aR_{ac}} \\ 0 & 0 & \frac{1}{aR_{cb}} \end{bmatrix} \cdot [D] \end{array} \right.$$

Ecuación 6.30

Para las matrices generalizadas $[A_r]$ y $[B_r]$ las Ecuaciones (1.27) y (1.28) no se puede aplicar directamente, debido a la matriz $[a_r]$. Por lo tanto, el voltaje aguas abajo se puede calcular a partir del voltaje aguas arriba a través de [20]:

$$[V_{Labc}]_d = \begin{bmatrix} \frac{1}{aR_{ab}} & 0 & 0 \\ 0 & \frac{1}{aR_{cb}} & 0 \\ -\frac{1}{aR_{ab}} & -\frac{1}{aR_{cb}} & 0 \end{bmatrix} \cdot [V_{Labc}]_u$$

Ecuación 6.31

y aplicando el mismo desarrollo utilizado para obtener la Ecuación (6.30), es posible determinar las otras matrices

Generalizada:

$$\left\{ \begin{array}{l} [A_r] = [F] \cdot \begin{bmatrix} \frac{1}{aR_{ab}} & 0 & 0 \\ 0 & \frac{1}{aR_{cb}} & 0 \\ -\frac{1}{aR_{ab}} & -\frac{1}{aR_{cb}} & 0 \end{bmatrix} \cdot [D] \\ [B_r] = [0] \end{array} \right.$$

Ecuación 6.32

Las ecuaciones (6.30) y (6.32) modelan el regulador de voltaje conectado en delta abierto.

6.1.1.7 Modelado de carga

Una parte fundamental de los SD es la conexión a lo largo de los nodos de alimentación y ramal con los consumidores. con fines de estudio. En esta disertación, los consumidores son vistos como cargas en los nodos de alimentación [20].

Una carga se puede modelar como:

- Potencia Constante - donde la potencia consumida no depende de su tensión de alimentación; [20]
- Corriente constante - donde la potencia consumida es proporcional a su tensión de alimentación; [20]
- Impedancia constante: donde la potencia consumida es proporcional al cuadrado de su voltaje de suministro; [20]
- Híbrido: en el que la potencia total consumida se compone de partes pertenecientes a las categorías anteriores. [20]

Para el estudio de detección e identificación de PNTs, consideramos que todas las cargas son del tipo Potencia Constante. De esta forma, la carga total en cada nodo del sistema es fija y es independiente de su tensión de alimentación. Otro aspecto importante del modelado de carga es su conexión con el sistema. La conexión de una carga puede ser del tipo:

- Estrella o fase neutra, o
- Delta o fase-fase.

Para cada carga a lo largo del alimentador, es necesario calcular la corriente consumida en función de su voltaje en el punto de conexión y su potencia consumida.[28].

Cargas con Conexión Estrella

En esta configuración, los terminales de carga están conectados entre las fases y el neutro común. La figura 6.7 muestra el diagrama de una carga conectada en estrella.

Para cada fase de la carga, tenemos que su carga y su voltaje esta por:

$$\text{Fase } a : |S_a| \angle \theta_a = P_a + jQ_a \quad \text{y} \quad V_a = E_a \angle \delta_a$$

Ecuación 6.33

$$\text{Fase } b : |S_b| \angle \theta_b = P_b + jQ_b \quad \text{y} \quad V_b = E_b \angle \delta_b$$

Ecuación 6.34

$$\text{Fase } c : |S_c| \angle \theta_c = P_c + jQ_c \quad \text{y} \quad V_c = E_c \angle \delta_c$$

Ecuación 6.35

Debido a la suposición de que la carga es del tipo de Potencia Constante, y con base en la Figura 6.7, las corrientes de línea de cada fase

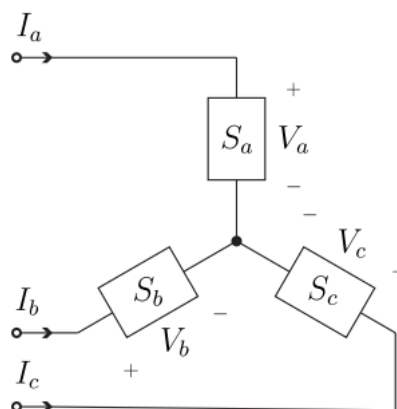


Figura 6.7 Carga conectada en estrella

donde [20]:

$$\begin{cases} I_a = \left(\frac{S_a}{V_a}\right)^* = \frac{|S_a|}{E_a} \angle(\delta_a - \theta_a) = |I_a| \angle \alpha_a \\ I_b = \left(\frac{S_b}{V_b}\right)^* = \frac{|S_b|}{E_b} \angle(\delta_b - \theta_b) = |I_b| \angle \alpha_b \\ I_c = \left(\frac{S_c}{V_c}\right)^* = \frac{|S_c|}{E_c} \angle(\delta_c - \theta_c) = |I_c| \angle \alpha_c \end{cases}$$

Ecuación 6.36

ecuación (6.36) permite el modelado de cargas conectadas en estrella.

Cargas con Conexión Delta

La Figura 6.8 muestra un diagrama de una carga conectada en delta. A diferencia de la configuración Fase-Neutro, las cargas ahora están conectadas entre las líneas de la red de distribución.

Las tensiones y potencias consumidas por esta conexión son

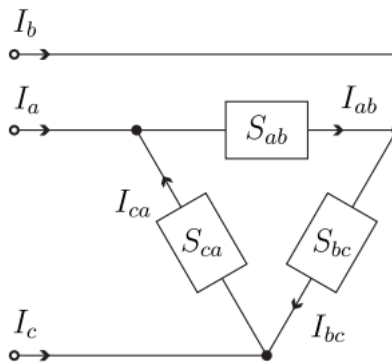


Figura 6.8 Carga conectada en triángulo

dada por:

$$\text{Fase } ab : |S_{ab}| \angle \theta_{ab} = P_{ab} + jQ_{ab} \quad \text{y} \quad V_{ab} = E_{ab} \angle \delta_{ab}$$

$$\text{Fase } bc : |S_{bc}| \angle \theta_{bc} = P_{bc} + jQ_{bc} \quad \text{y} \quad V_{bc} = E_{bc} \angle \delta_{bc}$$

$$\text{Fase } ca : |S_{ca}| \angle \theta_{ca} = P_{ca} + jQ_{ca} \quad \text{y} \quad V_{ca} = E_{ca} \angle \delta_{ca}$$

Ecuación 6.37

Las corrientes que circulan en cada fase de la carga delta, a una carga de potencia constante, vienen dadas por [20]:

$$\begin{cases} I_{ab} = \left(\frac{S_{ab}}{V_{ab}}\right)^* = \frac{|S_{ab}|}{E_{ab}} \angle(\delta_{ab} - \theta_{ab}) = |I_{ab}| \angle \alpha_{ab} \\ I_{bc} = \left(\frac{S_{bc}}{V_{bc}}\right)^* = \frac{|S_{bc}|}{E_{bc}} \angle(\delta_{bc} - \theta_{bc}) = |I_{bc}| \angle \alpha_{bc} \\ I_{ca} = \left(\frac{S_{ca}}{V_{ca}}\right)^* = \frac{|S_{ca}|}{E_{ca}} \angle(\delta_{ca} - \theta_{ca}) = |I_{ca}| \angle \alpha_{ca} \end{cases}$$

Ecuación 6.38

Debido a que es una carga conectada en delta, las corrientes en cada fase no coinciden con las corrientes de línea. La aplicación de la ley de Kirchhoff a las corrientes en cada nodo de la conexión de carga, dada en la Figura 1.24, nos permite calcular la Corrientes de línea que absorbe una carga delta [20]:

$$\begin{bmatrix} I_a \\ I_b \\ I_c \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} I_{ab} \\ I_{bc} \\ I_{ca} \end{bmatrix}$$

Ecuación 6.39

O, en forma matricial:

$$[I_{abc}] = [DI] \cdot [I_{Labc}]$$

Ecuación 6.40

Donde

$$[DI] = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$$

Ecuación 6.41

Las ecuaciones (6.38) y(6.40) modelan las corrientes de cargas conectadas en delta.

6.1.1.8 Cargas Monofásicas y Bifásicas

Principal punto de uso de la modelización [28] para esta tesis es la posibilidad de utilizarlas mismas ecuaciones (6.36), (6.38) y (6.40) para modelar cargas desequilibradas,

muy común en SD. Para cargas monofásicas y bifásicas, se pueden seguir utilizando estas mismas ecuaciones, siempre que para las fases que alimentan la carga, las corrientes se calculen normalmente. Para fases que no tienen conexión con la carga, es suficiente considerar que estas corrientes son nulas [20].

Este procedimiento permite calcular las corrientes, tanto para cargas conectadas a nodos trifásicos, como para conexiones en nodos monofásicos y bifásicos, ya que las ramas que alimentan a estos nodos se modelan como matrices generalizadas de 3×3 . [20].

6.1.1.9 Cargas Distribuidas

En el contexto de SD, es posible tener la presencia de cargas distribuidas. Estas cargas no son más que múltiples consumidores, con un perfil similar, alimentados en diferentes nodos a lo largo de una LD. Debido a que sus cargas puntuales son muy pequeñas en comparación con el consumo global del SD, una práctica común es agregar las potencias consumidas y nombrarlas como carga distribuida a lo largo de una rama. [20]

Sin embargo, aunque cada consumo individual puede ser pequeño, una carga distribuida puede presentar un valor significativo y, por lo tanto, debe modelarse. En la referencia [28] propone que, para el modelado de cargas distribuidas, se debe crear un nodo central imaginario a lo largo de la rama. Debido a que la impedancia en serie Z_{abc} de una LD es directamente proporcional a su longitud, la impedancia en serie de los LD resultantes se compone de todos los elementos de Z_{abc} divididos por 2.

La carga total distribuida se concentra entonces en este nodo ficticio y se modela según el tipo de conexión realizada por los consumidores locales, ya sea en estrella o en triángulo.

6.1.1.10 Flujo de Potencia

Como herramienta básica para el análisis de sistemas eléctricos, el flujo de potencia permite el cálculo de tensiones complejas en la red a partir de información de generación y consumo.

Los principales métodos para calcular el flujo de potencia en redes de distribución radial se dividen en tres categorías: flujo de potencia a través de Newton-Raphson, métodos basados en la matriz de impedancia nodal implícita y métodos de exploración hacia adelante y hacia atrás [29].

El cálculo mediante el método de Newton-Raphson se utiliza tradicionalmente en redes de extra y ultra alta tensión, y también en de distribución. Es un método basado en la matriz de impedancia nodal, y se debe tener cuidado para evitar un mal acondicionamiento de la matriz, en presencia de tramos de alta y muy baja impedancia.

Por otro lado, los métodos basados en la matriz de impedancia nodal implícita utilizan una formulación más adecuada para sistemas mallados, basados en la factorización de la matriz de admitancia nodal e inyecciones de corriente equivalente para resolver la red [29].

Los métodos de barrido directo e inverso son los más recomendados para los sistemas radiales. Este método consiste en realizar dos barridos, uno directo y otro inverso, donde el barrido inverso parte de los puntos más extremos y con una estimación inicial de las tensiones nodales, calcula las corrientes o caudales hacia la subestación. En base a este resultado y con el valor de tensión conocido en la subestación, se inicia el barrido directo, donde se recalculan los valores de tensión desde los nodos hasta los puntos más extremos. Este proceso se repite hasta que la diferencia en el valor del voltaje en dos iteraciones sucesivas no sea mayor que una tolerancia predeterminada [29].

En este trabajo, el flujo de potencia propuesto por Kersting (2012), se utiliza como herramienta de apoyo. Debido a las características particulares de los SD descritos [29], los métodos de flujo de potencia tradicionalmente utilizados en los sistemas de transmisión, cuando se aplican a la distribución, pueden presentar propiedades de convergencia indeseables. Por lo tanto, la solución adoptada es la aplicación de un algoritmo de exploración dedicado a los sistemas de distribución trifásicos.

El flujo de potencia de barrido asume que las cargas a lo largo de los nodos del SD, y que la subestación puede considerarse una barra infinita, que suministra energía a un voltaje específico V_{sub} dependiente de la red de transmisión y la toma del transformador de la subestación. A partir de esta información, el algoritmo recorre los alimentadores SD, hacia adelante y hacia atrás, en un proceso iterativo diferenciado que se beneficia del funcionamiento radial de estos sistemas.[29]

Para una mejor comprensión del flujo de potencia trifásico por método de barrido, primero se analiza una red monofásica. Luego, este procedimiento se amplía para soportar el modelado de matrices generalizadas, pudiendo así ser utilizado para redes trifásicas.

6.1.1.11 Método de barrido para una red monofásica

Considere el equivalente monofásico de un SD radial simple, que se muestra en la Figura.6.9, donde se conoce la tensión V_{sub} , así como las cargas S_K , donde $K = 2 \dots 5$

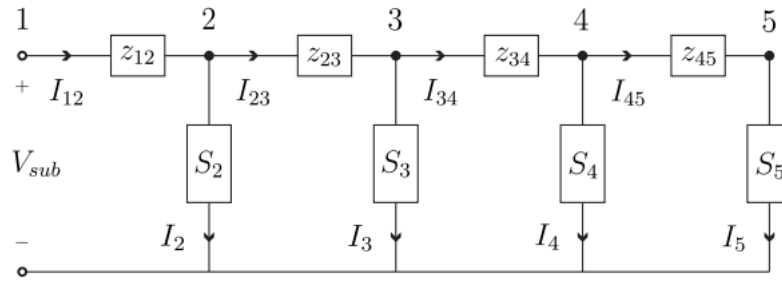


Figura 6.9 Equivalente Monofásico de un Alimentador Simple.

Para este sistema, es necesario determinar las tensiones complejas a lo largo de los nodos. Para ello, el método de barrido será utilizado. Este flujo de potencia consiste en un proceso iterativo en el cual, para cada iteración ℓ se realizan dos medias iteraciones:[20]

- Barrido hacia adelante: le permite actualizar las corrientes I_k^ℓ
- Barrido hacia atrás: le permite actualizar los voltajes V_k^ℓ

y, para la primera iteración, las corrientes en las cargas se consideran nulas,

es decir, $I_k^0 = 0$

Cada iteración del flujo de potencia comienza estableciendo el valor de voltaje en la barra de la subestación, V_1 igual al voltaje V_{sub} , especificado. Para una iteración dada ℓ , esta inicialización es impartido por:[20]

$$V_1^{\ell-1/2} = V_{sub}$$

Ecuación 6.42

Donde $V_1^{\ell-1/2}$ indica el voltaje inicial de media iteración. En el proceso de barrido avanzado, también llamado iteración de la primera mitad, los voltajes de la mitad de la iteración se calculan a lo largo del alimentador, $V_k^{\ell-1/2}$. Este procedimiento se aparta de la aplicación de la ley kirchhoff de los voltajes en cada malla de alimentación. Por ejemplo, para calcular el voltaje $V_2^{\ell-1/2}$, tenemos[20] :

$$V_2^{\ell-1/2} = V_1^{\ell-1/2} - z_{12} \cdot I_{12}^{\ell-1}$$

Ecuación 6.43

siendo que la corriente I_{12} se calcula mediante la aplicación de la ley de Kirchhoff de corrientes en el nodo 2. Este proceso continúa aplicando la teoría de circuitos para calcular las corrientes de ramas y tensiones en los nodos, desde la subestación hasta el último nodo terminal, es decir, barriendo el sistema de avance. Con las tensiones de media iteración, considerando las cargas en los nodos, es posible actualizar las corrientes consumidas en cada nodo k , por [20]:

$$I_k^\ell = \left(\frac{S_k}{V_k^{\ell-1/2}} \right)^*$$

Ecuación 6.44

procedimiento que completa la iteración de la primera mitad. En el proceso de barrido de retroceso equivalente a la iteración de la segunda mitad del flujo de potencia, a partir de las corrientes calculadas por ecuación (6.44) se actualizan las tensiones en cada nodo. Para ello, se actualizan las tensiones en los nodos terminales, actualizando a la tensión calculada por la Ecuación (6.43):

$$V_{kf}^\ell = V_k^{\ell-1/2}, \forall kf \text{ equivalente a un nodo final}$$

Ecuación 6.45

por ejemplo, en el caso de la Figura 6.9 $k_f = 5$, y $V_5^\ell = V_5^{\ell-1/2}$ La ley de tensiones de Kirchhoff se aplica de nuevo, buscando, desde los nodos terminales, actualizar las tensiones a la barra de la subestación, podemos calcular V_4^ℓ por:

$$V_4^\ell = V_5^\ell + z_{45} \cdot I_{45}^\ell$$

Ecuación 6.46

Nuevamente, las corrientes de las ramas se calculan aplicando la ley de corriente de Kirchhoff en el nodo aguas abajo de ellas. Así, la red de distribución es atravesada una vez más, solo que esta vez en un barrido hacia atrás. Al final de la segunda mitad de la iteración ℓ , la iteración está completa. La convergencia del flujo de potencia ocurre cuando la diferencia entre el voltaje calculado para la subestación, V_1^ℓ , y el voltaje especificado de la subestación, V_{sub} , es menor que una tolerancia predeterminada [20]:

$$\|V_1^\ell - V_{sub}\| \leq \epsilon$$

Ecuación 6.47

De lo contrario, $\ell = \ell + 1$, y se realiza una nueva iteración.

6.1.1.12 Método de barrido para el modelado trifásico

Es posible expandir el método de flujo de potencia de barrido al modelado trifásico mediante matrices generalizadas para determinar las tensiones trifásicas $[V_{abc}]$ a lo largo de la SD. La estructura de flujo de potencia es la misma, y cada iteración tiene dos medias iteraciones, que constituyen un barrido hacia delante y atrás. Las corrientes iniciales de cada nodo k se consideran nulas, y al comienzo de cada iteración ℓ , el voltaje de media iteración en el nodo 1 se corresponde con el voltaje especificado para la subestación:[20]

$$\begin{cases} [V_{abc}]_1^{\ell-1/2} = [V_{sub}] = [E_{sub} \angle \delta_{sub}] \\ [I_{abc}]_k^0 = [0] \end{cases}$$

Ecuación 6.48

Con los datos de la ecuación (6.48), es posible describir cada paso de una iteración dada ℓ del flujo de potencia.

6.1.1.13 Barrido hacia adelante

Los procedimientos de barrido, para el modelado por matrices generalizadas, tienen dos diferencias principales: se aplican en las extensiones, respetando su ordenamiento desde la subestación hasta las extensiones finales, y viceversa; y las actualizaciones están dadas por ecuaciones generalizadas, Ecuaciones (1.24), (1.25) y (1.26).[20]

En primer lugar, para una iteración dada ℓ , se inicializan las corrientes aguas arriba y aguas abajo de cada una de las ramas. Si es la primera iteración, $\ell = 1$, las corrientes iniciales son nulas, según la ecuación (6.48).[20]

$$\begin{cases} [I_{abc}]_u^{\ell-1/2} = [0] \\ [I_{abc}]_u^{\ell-1/2} = [0], \forall \text{ rama de SD} \end{cases}$$

Ecuación 6.49

Para las demás iteraciones, estas corrientes se actualizan por las corrientes obtenidas en la iteración anterior:

$$\begin{cases} [I_{abc}]_u^{\ell-1/2} = [I_{abc}]_u^{\ell-1} \\ [I_{abc}]_d^{\ell-1/2} = [I_{abc}]_d^{\ell-1}, \forall \text{ rama de SD} \end{cases}$$

Ecuación 6.50

Con las corrientes en las ramas inicializadas, y la tensión en la subestación determinada por la ecuación (6.48), comienza el barrido avanzado, que recorre las extensiones desde la subestación hasta los nodos terminales, en función de las ecuaciones generalizadas. Para cualquier extensión de la SD, las tensiones aguas abajo de la SD se actualizan mediante la ecuación (1.26), de la siguiente manera:

$$[Vabc]_k^{\ell-1/2} = [A]. [Vabc]_u^{\ell-1/2} - [B]. [Iabc]_d^{\ell-1/2}$$

Ecuación 6.51

La ecuación (6.51) se aplica sistemáticamente a cada extensión del sistema, comenzando desde las extensiones adyacentes a la barra de la subestación, y siguiendo la red hasta que se atraviesan todas las extensiones. Debido a la conexión radial de la SD, el procedimiento de barrido avanzado permite obtener todas las tensiones trifásicas de semi-iteración, $[Vabc]_k^{\ell-1/2}$, para cada nodo k de la SD.[20]

6.1.1.14 Actualización de corrientes de carga

Al final del proceso de análisis directo, es necesario actualizar las corrientes de carga. Para cada nodo k con cargas conectadas en estrella, la corriente consumida en cada fase se actualiza mediante la ecuación (6.36), considerando el voltaje obtenido del barrido avanzado, $[Vabc]_k^{\ell-1/2}$:[20]

$$[Iabc]_k^{\ell} = \begin{bmatrix} \frac{S_a}{V_a^{\ell-1/2}} \\ \frac{S_b}{V_b^{\ell-1/2}} \\ \frac{S_c}{V_c^{\ell-1/2}} \end{bmatrix}_k, \forall k \text{ con carga en estrella}$$

Ecuación 6.52

En el caso de las cargas conectadas en delta, el procedimiento es más elaborado. En primer lugar, se calculan las tensiones de línea para todos los nodos con carga delta, considerando el procedimiento para sistemas desequilibrados propuesto por [28]:

$$[V_{Labc}]_k^{\ell-1/2} = [D]. [Vabc]_k^{\ell-1/2} \quad \forall \text{ nodo } k \text{ con carga delta}$$

Ecuación 6.53

Las corrientes de fase de carga delta se calculan aplicando la ecuación (6.38), considerando la tensión de línea obtenida en la ecuación (6.53).

$$[I_{Labc}]_k^\ell = \begin{bmatrix} \frac{S_{ab}}{V_{ab}^{\ell-1/2}} \\ \frac{S_{bc}}{V_{bc}^{\ell-1/2}} \\ \frac{S_{ca}}{V_{ca}^{\ell-1/2}} \end{bmatrix}_k^* , \forall k \text{ con carga en estrella}$$

Ecuación 6.54

Finalmente, aplicando la ecuación (6.40) la corriente consumida en cada fase de los nodos con carga delta es finalmente actualizada:

$$[I_{abc}]_k^\ell = [DI] \cdot [I_{Labc}]_k^\ell , \forall k \text{ con carga en delta}$$

Ecuación 6.55

Con las cadenas de iteración ℓ actualizadas, se puede iniciar el proceso de barrido hacia atrás.

6.1.1.15 Barrido hacia atrás

El barrido hacia atrás para el modelado trifásico es similar al barrido hacia adelante, con respecto al seguimiento de las extensiones. A partir de las extensiones cuyos nodos kf aguas abajo son nodos terminales, se actualizan las tensiones aguas arriba, así como las corrientes en las ramas a lo largo de la SD. El proceso comienza en todas las extensiones finales del SD. Inicialmente, sus tensiones aguas abajo se actualizan, similar a la ecuación (6.45):[20]

$$[V_{abc}]_d^\ell = [V_{abc}]_k^{\ell-1/2} , \forall k \text{ ramas finales SD}$$

Ecuación 6.56

Sus corrientes aguas abajo también se actualizan, aplicando la Ley de Kirchoff a los Estados. Debido a que son extensiones de terminal, las corrientes descendentes son idénticas a las corrientes de carga en los nodos finales kf : [20]

$$[I_{abc}]_d^\ell = [I_{abc}]_{kf}^\ell , \forall k \text{ rmas finales SD}$$

Ecuación 6.57

Las tensiones aguas arriba de estas extensiones finales se actualizan de acuerdo con la Ecuación (1.24)

$$[V_{abc}]_u^\ell = [a][V_{abc}]_d^\ell + [b] \cdot [I_{abc}]_d^\ell$$

Ecuación 6.58

Las corrientes aguas arriba de los ramales finales se actualizan de manera similar, esta vez de acuerdo con la Ecuación (1.25)

$$[I_{abc}]_u^\ell = [c][V_{abc}]_d^\ell + [d] \cdot [I_{abc}]_d^\ell$$

Ecuación 6.59

El proceso continúa siguiendo las ramas inmediatamente aguas arriba de las extensiones finales. Aplicando la ley de Kirchhoff de las cadenas en todos los k nodos aguas abajo de las ramas no terminales, se obtiene:[20]

$$[I_{abc}]_d^\ell = [I_{abc}]_k^\ell + \sum_{i \in \Omega_k} ([I_{abc}]_u^\ell)_i$$

Ecuación 6.60

Donde Ω_k es el conjunto de ramas aguas abajo del nodo k . La aplicación de las ecuaciones (6.58), (6.59) y (6.60) en todas las ramas, desde las ramas finales, hasta la rama de la subestación, constituye el proceso de retroceso. Al final de la segunda mitad de la iteración, el voltaje de la subestación actualizado se compara con el voltaje especificado. $[V_{abc}]_1^\ell$ y $[V_{sub}]$. [20]

$$\|[V_{abc}]_1^\ell - [V_{sub}]\|_\infty \leq \epsilon$$

Ecuación 6.61

menor que la tolerancia especificada, el flujo de potencia converge correctamente. De lo contrario, hay un aumento en la iteración, $\ell = \ell + 1$, y el proceso de barrido se reinicia.

6.1.1.16 Comentarios sobre el modelado SD

Con el modelado presentado, es posible describir los SD con mucha precisión. La plantilla requiere algunos comentarios adicionales para esbozar soluciones a las posibles dificultades encontradas.

6.1.1.17 Modelo en Por Unidad

En los procesos de estimación de estado, así como en aplicaciones prácticas de flujo de potencia, es natural utilizar el sistema pu. Sin embargo, este tema no se trata en detalle [20], donde todos los sistemas, incluida la determinación de matrices generalizadas, se dan en unidades reales.

Sin embargo, para la normalización, muy común en los procesos de estimación de estado y análisis de error bruto, la modelización por unidades reales puede presentar problemas

computacionales. Un modelado pu, además de estar estandarizado, no implicaría dificultades innecesarias en la implementación de técnicas correctivas.[20]

6.1.1.18 Extinción de Nudos y Aglutinación de Ramas

Muchos de los elementos del SD, principalmente el transformador con conexión delta-triángulo, y los reguladores de tensión, tienen matrices generalizadas $[B]$ no invertibles. Sin embargo, como se presenta en un capítulo posterior, esta inversión es necesaria para el estimador de estado propuesto.

La solución a este problema radica en la aglutinación de elementos en cascada, descrita en[28], ya que normalmente estas ramas están en serie con las LD, por lo que las correspondientes matrices generalizadas $[B]$ tienen una inversa.

Los dos elementos se fusionan en una rama equivalente y el nodo intermedio se elimina del SD. La matriz $[B_{eq}]$ puede invertirse y, por lo tanto, usarse en el EESD, resolviendo el problema para muchos casos que se encuentran en el modelado SD.[28]

Además, la reducción de nodos implica menos pasos en el proceso de flujo de potencia, así como menos variables de estado a estimar.[28]

6.1 Consumos de Clientes

Tabla 6.1: Consumos de los clientes de electricidad en kWh

Usuario	ene-17	feb-17	mar-17	abr-17	nov-18	dic-18	Etiqueta	Clase
1	130,0	132,0	123,0	125,0	132,0	133,0	Fraudulent	1
2	120,0	123,0	126,0	128,0	135,0	136,0	Fraudulent	1
3	150,0	10,0	101,0	103,0	110,0	111,0	Fraudulent	0
4	120,0	15,0	89,0	91,0	98,0	99,0	Clean	0
5	113,0	10,0	45,0	47,0	54,0	55,0	Fraudulent	0
6	112,0	3,0	126,0	128,0	135,0	136,0	Fraudulent	0
7	113,0	100,0	90,0	92,0	99,0	100,0	Clean	1
8	123,0	130,0	54,0	56,0	63,0	64,0	Clean	1
9	5,0	124,0	18,0	20,0	27,0	28,0	Fraudulent	0
10	1,0	134,0	114,0	116,0	123,0	10,0	Fraudulent	0
11	0,0	124,0	123,0	125,0	132,0	133,0	Clean	0
12	0,1	90,0	132,0	134,0	6,6	123,0	Clean	0
13	80,0	90,0	141,0	143,0	150,0	151,0	Fraudulent	1
14	83,0	79,0	150,0	152,0	14,0	15,0	Fraudulent	1
15	175,0	134,0	114,0	116,0	123,0	124,0	Clean	1
16	123,0	145,0	114,4	116,4	123,4	124,4	Clean	1
17	98,0	100,0	114,8	116,8	123,8	124,8	Fraudulent	1
18	97,0	98,0	115,2	117,2	124,2	125,2	Fraudulent	1

19	98,0	99,0	115,6	117,6	124,6	125,6	Clean	1
20	100,0	1,0	116,0	118,0	125,0	126,0	Clean	0
21	129,0	126,0	116,4	118,4	125,4	126,4	Fraudulent	1
22	87,0	89,0	116,8	118,8	125,8	126,8	Fraudulent	1
23	97,0	100,0	117,2	119,2	126,2	127,2	Clean	1
24	85,0	0,0	117,6	119,6	126,6	127,6	Clean	1
25	88,0	97,0	118,0	120,0	127,0	128,0	Fraudulent	0
26	46,0	97,0	118,4	120,4	127,4	128,4	Fraudulent	0
27	87,0	110,0	118,8	120,8	127,8	128,8	Clean	1
28	90,0	107,0	119,2	121,2	128,2	129,2	Clean	1
29	95,0	100,0	119,6	121,6	25,0	26,0	Fraudulent	1
30	113,0	100,0	120,0	122,0	129,0	130,0	Clean	1
31	90,0	89,0	120,4	122,4	129,4	130,4	Clean	1
32	15,0	5,0	120,8	122,8	129,8	130,8	Clean	1
33	69,0	80,0	121,2	123,2	130,2	131,2	Clean	1
34	81,0	86,0	121,6	123,6	130,6	131,6	Clean	1
35	87,0	90,0	122,0	124,0	131,0	132,0	Clean	1
36	86,0	88,6	122,4	124,4	131,4	132,4	Clean	1
37	81,0	70,1	122,8	124,8	131,8	132,8	Clean	1
38	90,0	99,2	123,2	125,2	132,2	133,2	Clean	1
39	79,0	76,0	123,6	125,6	132,6	133,6	Clean	1
40	83,0	0,0	124,0	126,0	41,0	42,0	Clean	1

Tabla 6.2: Consumos de los clientes de electricidad en kWh

Usuario	ene-17	feb-17	mar-17	abr-17	nov-18	dic-18	Etiqueta	Clase
1	0,350	0,450	0,000	0,100	0,450	0,500	Fraudulent	1
2	0,000	0,115	0,231	0,308	0,577	0,615	Fraudulent	1
3	1,000	0,000	0,650	0,664	0,714	0,721	Fraudulent	0
4	1,000	0,000	0,705	0,724	0,790	0,800	Clean	0
5	1,000	0,000	0,340	0,359	0,427	0,437	Fraudulent	0
6	0,762	0,000	0,860	0,874	0,923	0,930	Fraudulent	0
7	1,000	0,435	0,000	0,087	0,391	0,435	Clean	1
8	0,908	1,000	0,000	0,026	0,118	0,132	Clean	1
9	0,000	1,000	0,109	0,126	0,185	0,193	Fraudulent	0
10	0,000	1,000	0,850	0,865	0,917	0,068	Fraudulent	0
11	0,000	0,867	0,860	0,874	0,923	0,930	Clean	0
12	0,000	0,592	0,868	0,882	0,043	0,809	Clean	0
13	0,000	0,123	0,753	0,778	0,864	0,877	Fraudulent	1
14	0,456	0,431	0,875	0,888	0,025	0,031	Fraudulent	1
15	1,000	0,328	0,000	0,033	0,148	0,164	Clean	1
16	0,281	1,000	0,000	0,065	0,294	0,327	Clean	1
17	0,000	0,054	0,457	0,511	0,701	0,728	Fraudulent	1
18	0,000	0,026	0,476	0,529	0,712	0,738	Fraudulent	1

19	0,000	0,027	0,468	0,521	0,707	0,734	Clean	1
20	0,733	0,000	0,852	0,867	0,919	0,926	Clean	0
21	0,630	0,480	0,000	0,100	0,450	0,500	Fraudulent	1
22	0,000	0,040	0,598	0,639	0,779	0,799	Fraudulent	1
23	0,000	0,075	0,502	0,552	0,726	0,751	Clean	1
24	0,618	0,000	0,855	0,869	0,920	0,927	Clean	1
25	0,000	0,180	0,600	0,640	0,780	0,800	Fraudulent	0
26	0,000	0,552	0,784	0,805	0,881	0,892	Fraudulent	0
27	0,000	0,444	0,614	0,653	0,788	0,807	Clean	1
28	0,000	0,346	0,593	0,634	0,776	0,797	Clean	1
29	0,624	0,666	0,831	0,848	0,034	0,042	Fraudulent	1
30	0,325	0,000	0,500	0,550	0,725	0,750	Clean	1
31	0,019	0,000	0,611	0,650	0,786	0,805	Clean	1
32	0,074	0,000	0,853	0,867	0,919	0,926	Clean	1
33	0,000	0,152	0,723	0,751	0,848	0,861	Clean	1
34	0,000	0,083	0,670	0,703	0,818	0,835	Clean	1
35	0,000	0,055	0,636	0,673	0,800	0,818	Clean	1
36	0,000	0,046	0,645	0,681	0,805	0,823	Clean	1
37	0,150	0,000	0,725	0,752	0,849	0,862	Clean	1
38	0,000	0,173	0,624	0,662	0,793	0,812	Clean	1
39	0,044	0,000	0,704	0,734	0,837	0,852	Clean	1
40	0,576	0,000	0,861	0,875	0,285	0,292	Clean	1

Nodo 10

Figura 6.10 Fase sale directo de borne al térmico sin pasar por el medidor

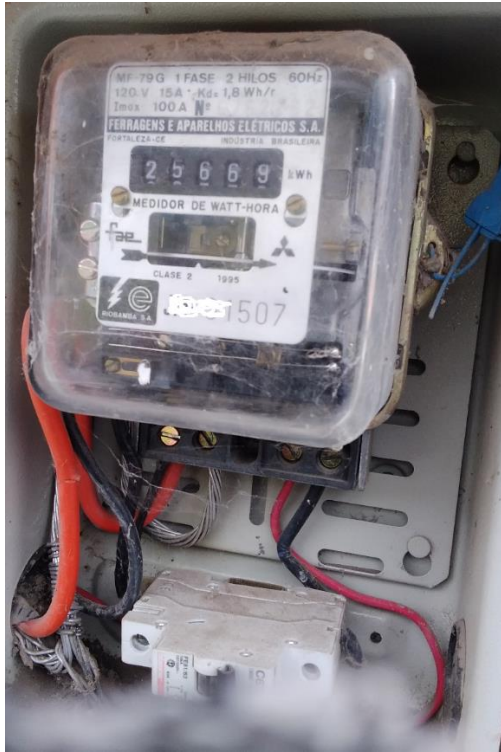


Figura 6.11 Medidor averiado



Figura 6.12 Fase sale directo del borne al térmico sin pasar por el medidor



Nodo 11

Figura 6.13 Fase entra directo al térmico sin pasar por el contador de energía



Figura 6.14 Fase sale directo de la bornera al térmico sin pasar por el contador de energía



Nodo 15

Figura 6.15 Fase sale directo de la bornera al térmico sin pasar por el contador de energía



Nodo 17

Figura 6.16 Fase con un puente entre la entrada y la salida del medidor



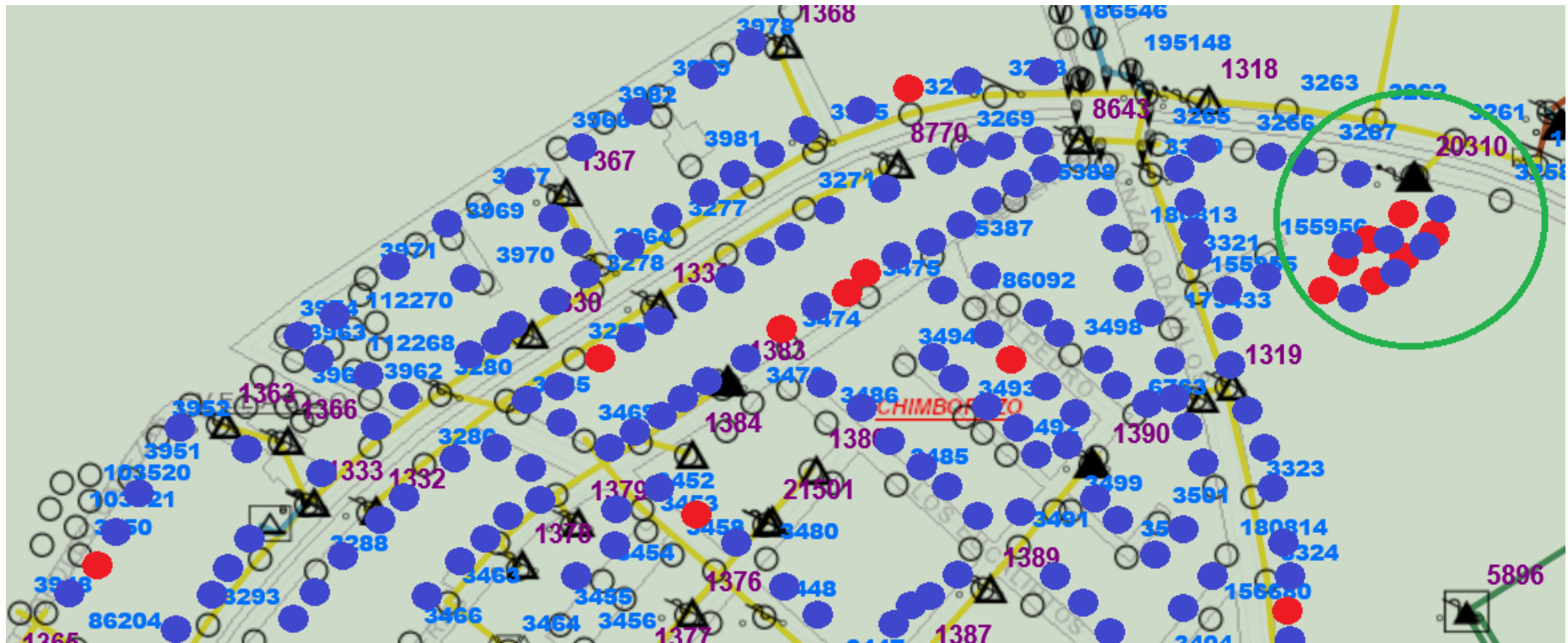
Figura 6.17 Fase sale directo de la bornera al térmico sin pasar por el contador de energía



Figura 6.18 Fase con un puente entre la entrada y la salida del medidor



Figura 6.19 Centro de transformación con mayor proporción de usuarios sospechosos



ORDEN DE EMPASTADO