



<p>AGENCIA DE REGULACIÓN Y CONTROL DE ELECTRICIDAD</p> 	<p>Instructivo para la minería de datos de la información estadística</p>	<p>GEI-OEG-P001-I001</p>
		<p>Página 1 de 20</p>
	<p>Versión 1.0</p>	

AGENCIA DE REGULACIÓN
Y CONTROL DE ELECTRICIDAD



Instructivo para la minería de datos de la información estadística

<p>Elaborado por: Ing. Christian Junia Ing. Andrés Chiles Ing. Rodrigo Briones Srta. Diana Cajamarca Dirección Nacional de Estudios Eléctricos y Energéticos</p>	<p>Revisado por: Ing. Santiago Flores Director Nacional de Estudios Eléctricos y Energéticos</p>	<p>Aprobado por: Ing. Geovanny Pardo Coordinador Nacional de Regulación del Sector Eléctrico</p>
<p>Fecha: 15-05-2020</p>	<p>Fecha: 15-05-2020</p>	<p>Fecha: 15-05-2020</p>
<p>Firmas:</p>	<p>Firma:</p>	<p>Firma:</p>
		


	Instructivo para la minería de datos de la información estadística	<i>GEI-OEG-P001-I001</i>
		<i>Página 2 de 20</i>
	Versión 1.0	


TABLA DE CONTENIDO

TABLA DE CONTENIDO.....	2
0. CONTROL E HISTORIAL DE CAMBIOS	3
1. JUSTIFICACIÓN	4
2. OBJETIVO	4
3. ALCANCE	4
4. RESPONSABILIDAD Y AUTORIDAD	4
5. GLOSARIO DE TÉRMINOS Y ABREVIATURAS.....	5
6. INSTRUCTIVO	10
6.1 DESCRIPCIÓN DE ACTIVIDADES.....	10
6.2 DIAGRAMA DE FLUJO.....	17
6.3 INDICADORES DEL PROCESO.....	18
7. ANEXOS.....	19
7.1 FORMATO DE TABLAS Y GRÁFICOS CON LOS QUE SE MUESTRAN LOS RESULTADOS DE LA MINERÍA DE DATOS. CÓDIGO: GEI-OEG-P001-I001-R001.....	19
7.2 FORMATO DE TABLAS DE LOS RESULTADOS DE LA MINERÍA DE DATOS DEL SOFTWARE CONELEC MINERÍA DE DATOS. CÓDIGO: GEI-OEG-P001-I001-R002.....	20
7.2.1 PRIMERA PESTAÑA: ANOMALÍAS POR AGENTE	20
7.2.2 SEGUNDA PESTAÑA: DETALLE ANOMALÍAS	20

<div> </div>	<div>Instructivo para la minería de datos de la información estadística</div>	GEI-OEG-P001-I001
		Página 3 de 20
	Versión 1.0	

0. CONTROL E HISTORIAL DE CAMBIOS

Versión	Descripción del cambio	Fecha de actualización
1.0	Documento inicial	

	Instructivo para la minería de datos de la información estadística	GEI-OEG-P001-I001
		Página 4 de 20
	Versión 1.0	

1. JUSTIFICACIÓN

La Ley Orgánica del Servicio Público de Energía Eléctrica (LOSPEE), en el numeral 9 del artículo 15, establece como una de las atribuciones y deberes de la Agencia de Regulación y Control de Electricidad (ARCONEL), la referente a *“Implementar, operar y mantener el sistema único de información estadística del sector eléctrico”*.

Asimismo, la Regulación No. ARCONEL 002/19 denominada *«Sistema Único de Información Estadística del Sector Eléctrico Ecuatoriano, SISDAT 2.0»* en su capítulo V: Análisis de la Información, numeral 12.1: Minería de Datos, establece que *“El proceso de minería de datos se efectuará a partir del siguiente día laborable de la fecha límite de entrega de la información estadística y geográfica, establecida en el PAO; y, según lo que ARCONEL establezca para el efecto, en metodologías, instructivos y procedimientos”*.

En función de lo anterior, la Dirección Nacional de Estudios Eléctricos y Energéticos (DNEEE) de la ARCONEL elabora el *“Instructivo para la minería de datos de la información estadística”* que involucra la búsqueda y establecimiento de tendencias, reglas, patrones de los datos ingresados por los Integrantes SISDAT; orientado a la automatización de este proceso.

2. OBJETIVO

Definir y establecer las actividades y parámetros técnicos que el equipo técnico encargado de la Gestión de Información Estadística y Geográfica (GIEG) de la DNEEE debe conocer para la ejecución de la minería de datos de la información estadística reportada por los Integrantes SISDAT a la ARCONEL a través del aplicativo SISDAT.

3. ALCANCE


El presente instructivo esta comprendido desde que el Analista/Profesional encargado de la GIEG, inicia el proceso de minería de datos de manera mensual hasta la ejecución del cierre final de la información estadística.

Comprende los procesos, procedimientos, actividades y parámetros técnicos para la ejecución de la minería de datos de la información estadística, desde la etapa de captación de la información hasta la etapa de cierre final de la información.

4. RESPONSABILIDAD Y AUTORIDAD

La distribución de responsabilidades relacionadas al presente instructivo se detalla a continuación:

- Elaboración y/o modificación: Servidores públicos que conforman el equipo técnico encargado de la GIEG de la DNEEE.
- Revisión: Director Nacional de Estudios Eléctricos y Energéticos.
- Aprobación: Coordinador Nacional de Regulación del Sector Eléctrico.
- Difusión: Dirección Nacional de Estudios Eléctricos y Energéticos, con apoyo de la Dirección de Gestión Estratégica (DGE).

	Instructivo para la minería de datos de la información estadística	GEI-OEG-P001-I001
		Página 5 de 20
	Versión 1.0	

- Aplicación: Equipo técnico encargado de la GIEG de la DNEEE.

5. GLOSARIO DE TÉRMINOS Y ABREVIATURAS

Algoritmos genéticos: Son algoritmos de búsqueda y optimización basados en los mecanismos de selección natural y genética.

Análisis Multivariante: Son técnicas de análisis de datos que estudian la relación simultánea entre el comportamiento de más de dos variables.

Análisis Univariante: Son técnicas de análisis de datos que estudian el comportamiento individual de las variables.

Aplicativo SISDAT: Herramienta informática que apoya a las etapas del proceso de operación estadística del sector eléctrico ecuatoriano. El SISDAT está operando desde el año 2009 y cuenta con información histórica desde el año de 1999.


SISDAT opera en un ambiente WEB cuya dirección URL es <http://sisdat.regulacionelectrica.gob.ec>.

Árbol de decisión: Representación de una función multivariada y que fue posible utilizar en la vida práctica a partir del desarrollo de las modernas computadoras. Proveen de una herramienta de clasificación muy potente. Su uso en el manejo de datos la hace ganar en popularidad dadas las posibilidades que brinda y la facilidad con que son comprendidos sus resultados por cualquier usuario.

Base de datos: Es un almacén/repositorio que permite guardar grandes cantidades de información de forma organizada que permita fácilmente su ubicación y utilización.

Campo del formulario SISDAT: Corresponde a un parámetro del formulario SISDAT (cabecera o título de una columna).

Captación de la información: Etapa de la operación estadística en la que se ejecutan las actividades de captar (por parte de la ARCONEL) y entregar información estadística y geográfica generada (por parte de los Integrantes del SISDAT), al aplicativo SISDAT.

	Instructivo para la minería de datos de la información estadística	GEI-OEG-P001-I001
		Página 6 de 20
	Versión 1.0	

Cierre preliminar de la información: Etapa de la operación estadística en la que mensualmente se ejecutan instrucciones a nivel de base de datos (SisdatData), con la finalidad de actualizar los reportes Excel que extraen información desde el aplicativo SISDAT, esto, previo a la ejecución de los análisis y procesos inherentes a la validación y verificación de la consistencia y coherencia de la información estadística entregada por los Integrantes SISDAT.

Cierre final de la información: Etapa de la operación estadística en la que mensualmente se ejecutan instrucciones a nivel de base de datos (SisdatData), con la finalidad de actualizar la información que se consume desde el aplicativo SISDAT, posterior a la ejecución de los análisis y procesos inherentes a la validación y verificación de la consistencia y coherencia de la información estadística entregada por los Integrantes SISDAT.

CONELEC Minería de Datos¹: Software utilizado para la ejecución de la minería de datos de la información estadística.


Coordinador SISDAT: Persona responsable de entregar la información estadística y geográfica en el SISDAT; y, que tiene a su cargo la coordinación de las tareas necesarias para cumplir con las responsabilidades establecidas en el PAO (Plan Anual de Operación Estadística) y en la Regulación correspondiente.

Dato atípico: Es una observación numéricamente distante del resto de los datos, detectado en la etapa de análisis de información de la operación estadística. Es un valor que está fuera del rango normal establecido.

Formulario SISDAT: Documento digital que se utiliza en el aplicativo SISDAT para la recolección de datos de manera estructurada. Puede ser un archivo Excel o un archivo de texto plano en formato .txt.

Los formularios SISDAT que se utilizan en la GIEG se dividen en dos grupos: infraestructura y transacciones. Los formularios de transacciones se subdividen a su vez en tres subgrupos: transacciones de generación, transmisión y distribución. Los formularios SISDAT de infraestructura son: INF-010 (Caudales), INF-020 (Cobertura Eléctrica), INF-040 (Redes Media

¹ El software no realiza el análisis de los formularios SISDAT INF-010 (Caudales), INF-090 (Personal) INF-100 (Medidores Instalados), TRA-060 (Pérdidas), TRA-130 (Facturación Clientes Regulados) y TRA-210 (Catastro de Clientes).

	Instructivo para la minería de datos de la información estadística	GEI-OEG-P001-I001
		Página 7 de 20
	Versión 1.0	

Tensión), INF-050 (Luminarias), INF-060 (Redes Secundarias), INF-070 (Acometidas), INF-080 (Medidores), INF-090 (Personal) e INF-100 (Medidores Instalados). Los formularios SISDAT de transacciones de generación son: TRA-010 (Energía Vendida), TRA-180 (Energía Producida) y TRA-190 (Balance de Producción). Los formularios SISDAT de transacciones de transmisión son: TRA-020 (SNT Energía Transportada), TRA-030 (SNT Balance de Energía), TRA-050 (SNT Potencia Línea Transmisión), TRA-090 (SNT DMax Transformadores) y TRA-140 (SNT KV Subestaciones). Los formularios SISDAT de transacciones de distribución son: TRA-040 (Balance Energía), TRA-060 (Pérdidas), TRA-080 (Facturación Clientes No Regulados), TRA-130 (Facturación Clientes Regulados), TRA-160 (Energía Comprada), y TRA-210 (Catastro de Clientes).




Grubbs' Test: También conocida como prueba residual normalizada máxima, es una prueba utilizada para detectar valores atípicos en un conjunto de datos univariado provenientes de una población normalmente distribuida.

Inconsistencia: Información estadística o geográfica inexacta y/o distorsionada, detectada en la etapa de análisis de información de la operación estadística.

Integrante SISDAT: Participantes y/o instituciones del sector eléctrico, CENACE y en general todas las entidades y organismos (públicos o privados) que deben entregar y que consumen información del aplicativo SISDAT.

K-means: Es un algoritmo de clasificación no supervisada (clusterización) que agrupa objetos en k grupos basándose en sus características. El agrupamiento se realiza minimizando la suma de distancias entre cada objeto y el centroide de su grupo o clúster. Se suele usar la distancia cuadrática.

Método bayesiano: Constituyen un método elegante y potente, basado en la probabilidad y la estadística que permite construir modelos de probabilidad conjunta manejables que representan las dependencias relevantes entre un conjunto formado por cientos de variables en aplicaciones prácticas. Los modelos resultantes permiten realizar inferencia probabilística de una manera eficiente.

  	Instructivo para la minería de datos de la información estadística	GEI-OEG-P001-I001
		Página 8 de 20
	Versión 1.0	

Minería de datos: Es el conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto.

Participantes del sector eléctrico: Personas jurídicas dedicadas a las actividades de generación, autogeneración, transmisión, distribución y comercialización, alumbrado público general, importación y exportación de energía eléctrica, así como también las personas naturales o jurídicas que sean considerados consumidores o usuarios finales.


Plan Anual de Operación Estadística, PAO: Documento en el que se establecen las acciones y requerimientos para el funcionamiento de las etapas de la operación estadística del sector eléctrico; donde se determinan los plazos previstos, módulos de información, formatos, fecha de publicación, responsable que atenderá los requerimientos, mecanismos de captación, entre otros.

Puntuación Z: En estadística, la puntuación Z (o puntuación estándar) de una observación es el número de desviaciones estándar que hay por encima o por debajo de la media de población.

Red Neuronal: Conjunto de elementos de procesamiento de la información altamente interconectados, que son capaces de aprender con la información que se les alimenta.

Regresión lineal: En estadística la regresión lineal o ajuste lineal es un modelo matemático usado para aproximar la relación de dependencia entre una variable dependiente Y, las variables independientes X_i y un término aleatorio ε .

Reportes SISDAT: Archivos Excel alojados en la intranet que contienen tablas dinámicas que extraen información desde los formularios SISDAT. Se emplean para realizar la ejecución manual de la minería de datos de los formularios SISDAT que no pueden ser analizados con el software CONELEC Minería de Datos, es decir el INF-010 (Caudales), INF-090 (Personal), INF-100 (Medidores Instalados), TRA-020 (SNT Energía Transportada), TRA-030 (SNT Balance de Energía), TRA-050 (SNT Potencia Línea Transmisión), TRA-060 (Pérdidas), TRA-090 (SNT DMax Transformadores), TRA-130 (Facturación Clientes Regulados), TRA-140 (SNT KV Subestaciones) y TRA-210 (Catastro de Clientes).

	Instructivo para la minería de datos de la información estadística	GEI-OEG-P001-I001
		Página 9 de 20
	Versión 1.0	

Series temporales: Colección de observaciones de una variable recogidas secuencialmente en el tiempo. Estas observaciones se suelen recoger en instantes de tiempo equiespaciados. Si los datos se recogen en instantes temporales de forma continua, se debe o bien digitalizar la serie, es decir, recoger sólo los valores en instantes de tiempo equiespaciados, o bien acumular los valores sobre intervalos de tiempo.

SIGLAS

ARCONEL: Agencia de Regulación y Control de Electricidad.




DNEEE: Dirección Nacional de Estudios Eléctricos y Energéticos.

DGE: Dirección de Gestión Estratégica.

GIEG: Gestión de Información Estadística y Geográfica.

PAO: Plan Anual de Operación Estadística.

SISDAT: Sistematización de Datos del Sector Eléctrico.

<div>AGENCIA DE REGULACIÓN Y CONTROL DE ELECTRICIDAD</div> <div></div>	Instructivo para la minería de datos de la información estadística	GEI-OEG-P001-I001
		Página 10 de 20
	Versión 1.0	




6. INSTRUCTIVO

6.1 Descripción de Actividades

No	Responsable	Actividad	Descripción	Documento	Tiempo estimado de ejecución (días laborables)
1	Analista / Profesional encargado de la GIEG	Iniciar el proceso mensual de minería de datos de la información estadística	<p>El inicio del proceso de minería de datos se efectúa mensualmente a partir del siguiente día laborable de la fecha límite de entrega de la información estadística, establecida en el PAO.</p> <p>¿Existe información estadística pendiente por entregar de parte de algún Integrante SISDAT?</p> <p>SI: Se procede conforme a lo estipulado en el “Procedimiento para la Gestión de la Operación Estadística y Geográfica del Sector Eléctrico Ecuatoriano”. Con la información disponible en el aplicativo SISDAT, se continúa con la actividad Nro. 2.</p> <p>NO: Se continúa con la actividad Nro. 2.</p>	No aplica	0,5
2	Analista / Profesional encargado de la GIEG	Ejecutar el cierre preliminar de la información estadística	Se ejecutan instrucciones a nivel de base de datos (SisdatData), con la finalidad de actualizar los reportes SISDAT. Para este propósito, se utiliza la herramienta Microsoft SQL Server, en la cual se ejecuta el archivo “cierreMensual2019.sql”, ubicado en la dirección: M:\CNRSE\DNEEE\ESTADISTICAS\SQL\cierre.	No aplica	0,5
3	Analista / Profesional encargado de la GIEG	Iniciar la ejecución de la minería de datos de la información estadística	<p>La ejecución de la minería de datos de la información estadística comprende dos procesos en paralelo. El primero, que es un proceso manual, se efectúa utilizando Reportes SISDAT (para aquellos formularios SISDAT que no pueden ser analizados con el software CONELEC Minería de Datos); y, el segundo emplea el software CONELEC Minería de Datos.</p> <p>Reportes SISDAT: Se eligen y abren los reportes SISDAT para iniciar la ejecución manual de la minería</p>	No aplica	0.5


		<p>de datos de los formularios SISDAT:</p> <p>INF-010 (Caudales), INF-090 (Personal), INF-100 (Medidores Instalados), TRA-020 (SNT Energía Transportada), TRA-030 (SNT Balance de Energía), TRA-050 (SNT Potencia Línea Transmisión), TRA-060 (Pérdidas), TRA-090 (SNT DMax Transformadores), TRA-130 (Facturación Clientes Regulados), TRA-140 (SNT KV Subestaciones), y, TRA-210 (Catastro de Clientes).</p> <p>Software CONELEC Minería de Datos: Con este software se realiza la minería de datos de los formularios SISDAT:</p> <p>INF-020 (Cobertura Eléctrica), INF-040 (Redes Media Tensión), INF-050 (Luminarias), INF-060 (Redes Secundarias), INF-070 (Acometidas), INF-080 (Medidores), TRA-010 (Energía Vendida), TRA-040 (Balance Energía), TRA-080 (Facturación Clientes No Regulados), TRA-160 (Energía Comprada), TRA-180 (Energía Producida), y, TRA-190 (Balance de Producción).</p> <p>Para iniciar la ejecución de la minería de datos con este software:</p> <ul style="list-style-type: none"> - Se ejecuta el software. Se ingresa Usuario y Clave. - Se selecciona el comando “Ejecutar”, luego se selecciona “Procesos”. 		
--	--	---	--	--

4	Analista / Profesional encargado de la GIEG	<p>Elegir los parámetros necesarios para la ejecución de la minería de datos de la información estadística</p> <p>Reportes SISDAT:</p> <ul style="list-style-type: none"> - Se actualizan las tablas dinámicas de los reportes SISDAT. - Se elige el año, el mes, la empresa (integrante SISDAT), y el campo del formulario SISDAT a analizar (análisis univariante). <p>Software CONELEC Minería de Datos:</p> <ul style="list-style-type: none"> - Se elige el tipo de análisis: Univariante o Multivariante. - Para cada técnica de análisis se elige el tipo de algoritmo a emplear. Para el caso del análisis univariante los algoritmos a emplear son: GrubbTest, Puntuación Z, Regresión Lineal, Series Temporales. Redes Neuronales***, Algoritmos genéticos***. - Para el caso del análisis multivariante los algoritmos a emplear son: K-means, Árboles de decisión***, Métodos bayesianos***. - Se elige el formulario SISDAT al que se realizará el análisis: INF-020 (Cobertura Eléctrica), INF-040 (Redes Media Tensión), INF-050 (Luminarias), INF-060 (Redes Secundarias), INF-070 (Acometidas), INF-080 (Medidores), TRA-010 (Energía Vendida), 	No aplica	0.5
---	---	--	-----------	-----

  	Instructivo para la minería de datos de la información estadística	GEI-OEG-P001-I001
	Versión 1.0	Página 13 de 20

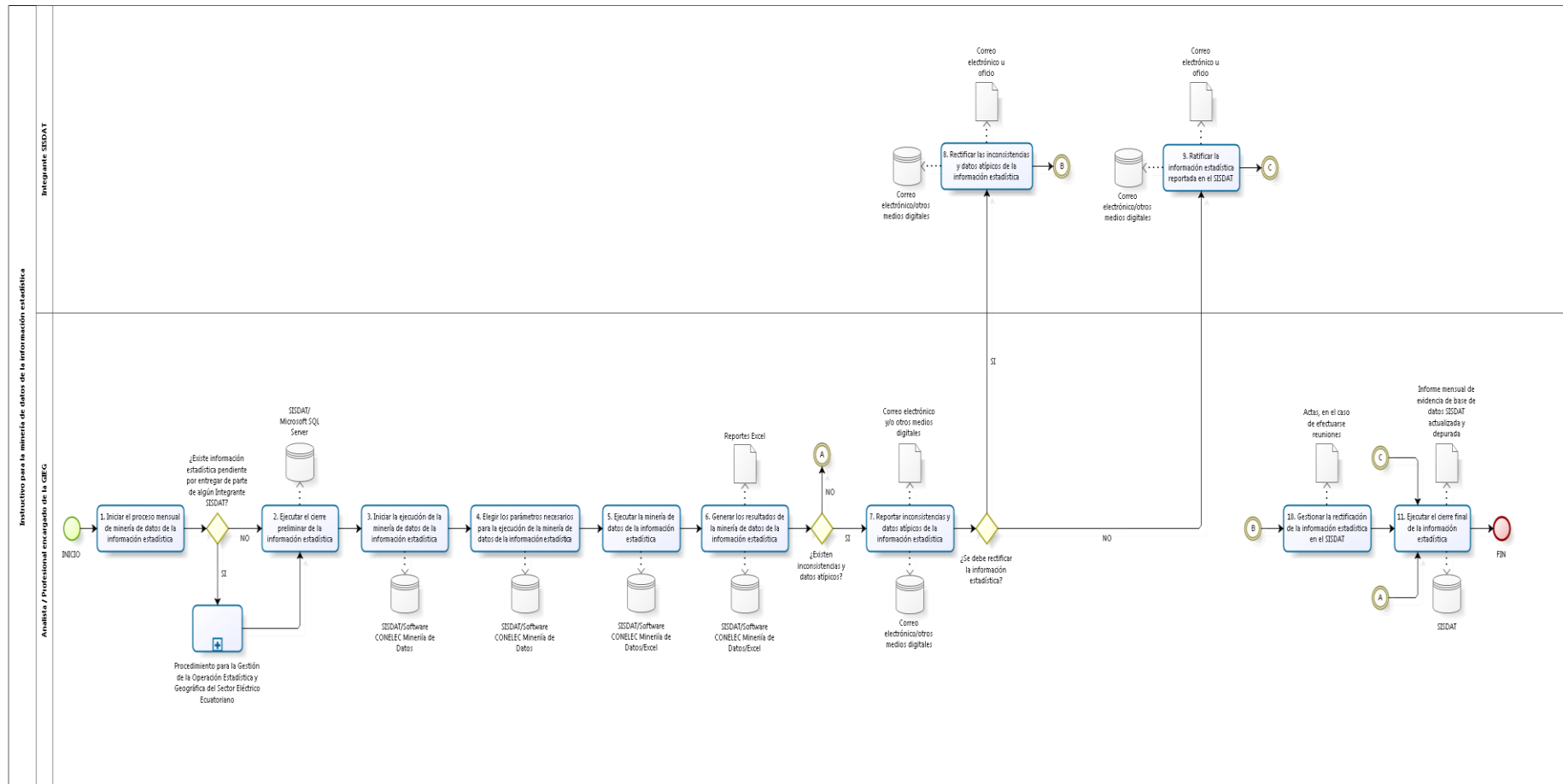
			<p>TRA-040 (Balance Energía), TRA-080 (Facturación Clientes No Regulados), TRA-160 (Energía Comprada), TRA-180 (Energía Producida), TRA-190 (Balance de Producción),</p> <p>INF-010 (Caudales)**, INF-090 (Personal)**, INF-100 (Medidores Instalados)*, TRA-020 (SNT Energía Transportada)**, TRA-030 (SNT Balance de Energía)**, TRA-050 (SNT Potencia Línea Transmisión)**, TRA-060 (Pérdidas)*, TRA-090 (SNT DMax Transformadores)**, TRA-130 (Facturación Clientes Regulados)*, TRA-140 (SNT KV Subestaciones)** y, TRA-210 (Catastro de Clientes)*.</p> <p>- Finalmente se elige el año, el mes y la empresa (integrante SISDAT).</p>		
5	Analista / Profesional encargado de la GIEG	Ejecutar la minería de datos de la información estadística	<p>Reportes SISDAT:</p> <p>- Se realiza la detección manual de inconsistencias y datos atípicos, mediante el análisis de la varianza o desviación estándar de los datos de cada campo de los formularios SISDAT que se analizan mediante Reportes SISDAT. Para este propósito se emplea las herramientas integradas de Excel como tablas dinámicas, macros, power pivot, gráficos de tendencias, entre otras.</p> <p>- Con base a los reportes SISDAT de transacciones de generación, se efectúan análisis comparativos de los datos históricos, fijando como límites la producción mensual de energía a potencia nominal y un precio medio de 40 USD¢/kWh.</p>	No aplica	1


			Software CONELEC Minería de Datos: - Una vez elegidos los parámetros de la actividad Nro. 4, se ejecuta el comando “Procesar”. Con la ejecución de este comando, el software realiza la detección automática de inconsistencias y datos atípicos. - Cuando se quiere procesar información que ya ha sido procesada previamente, el software presenta el siguiente mensaje: “¡Ya existen datos procesados!. Desea volver a procesarlos?”; y, se da la opción de “Si” y “No”. Si se elige “Si” el software vuelve a procesar la información. Si se elige “No”, el software permite consultar los resultados previamente procesados, mediante la ejecución del comando “Consultar”.		
6	Analista / Profesional encargado de la GIEG	Generar los resultados de la minería de datos de la información estadística	Reportes SISDAT: - Los resultados de la detección de inconsistencias y datos atípicos, se los presenta mediante tablas y gráficos realizados manualmente con las herramientas integradas de Excel (ver Anexo 7.1). Software CONELEC Minería de Datos: Los resultados de la minería de datos se los exporta en un archivo de Excel mediante la ejecución del comando “Exportar”. El archivo (ver Anexo 7.2) consta de dos pestañas: la primera presenta un resumen por empresa (Integrante SISDAT) de las anomalías encontradas en la ejecución de la minería de datos; mientras que en la segunda pestaña se encuentra el detalle de las anomalías. Este software no permite obtener resultados gráficos de la minería de datos como los obtenidos mediante los Reportes SISDAT (Ver Anexo 7.1***). Para los Reportes SISDAT, Software CONELEC Minería de Datos, y, Requerimiento: ¿Existen inconsistencias y datos atípicos? SI: Se continúa con la actividad Nro. 7. NO: Se continúa con la actividad Nro. 11.	Reportes Excel	1

	Instructivo para la minería de datos de la información estadística	GEI-OEG-P001-I001
		Página 15 de 20
Versión 1.0		

7	Analista / Profesional encargado de la GIEG	Reportar inconsistencias y datos atípicos de la información estadística	<p>Se indicará al Coordinador SISDAT las inconsistencias y datos atípicos, detectados en los análisis y procesos inherentes a la minería de datos de la información estadística; y, se solicitará la ratificación o rectificación de la misma.</p> <p>¿Se debe rectificar la información estadística?</p> <p>SI: Se continúa con la actividad Nro. 8.</p> <p>NO: Se continúa con la actividad Nro. 9.</p>	Correo electrónico y/o otros medios digitales	1
8	Integrante SISDAT	Rectificar las inconsistencias y datos atípicos de la información estadística	<p>En el caso que se deba rectificar la información, el Integrante SISDAT, en un plazo máximo de cinco (5) días laborables, deberá remitir según criterio de la ARCONEL:</p> <ul style="list-style-type: none"> - Un correo electrónico; y/o, - Un informe con los justificativos a las inconsistencias y datos atípicos detectados en la información, que deberá ser entregado mediante oficio. Se continúa con la actividad Nro. 10. 	Correo electrónico u oficio	Hasta 5 días
9	Integrante SISDAT	Ratificar la información estadística reportada en el SISDAT	En el caso que se ratifique la información, el Integrante SISDAT, en un plazo máximo de cinco (5) días laborables, notificará este particular a la ARCONEL mediante correo electrónico u oficio. Se continúa con la actividad Nro. 11.	Correo electrónico u oficio	Hasta 5 días
10	Analista / Profesional encargado de la GIEG	Gestionar la rectificación de la información estadística en el SISDAT	En el caso que se deba rectificar la información, se brindará asesoría para coordinar y efectuar la rectificación de la información por parte del Integrante SISDAT.	Actas, en el caso de efectuarse reuniones	Hasta 5 días
11	Analista / Profesional encargado de la GIEG	Ejecutar el cierre final de la información estadística	Una vez que se haya validado y verificado la información disponible en el aplicativo SISDAT, se ejecutan instrucciones a nivel de base de datos, con la finalidad de actualizar los reportes Excel que extraen información desde el aplicativo SISDAT.	Informe mensual de evidencia de base de datos SISDAT actualizada y depurada	1

6.2 Diagrama de flujo


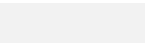



	Instructivo para la minería de datos de la información estadística	GEI-OEG-P001-I001
		Página 18 de 20
	Versión 1.0	

6.3 Indicadores del proceso

FICHA DE INDICADORES	
Nombre del indicador:	<i>Tiempo de ejecución de la minería de datos.</i>
Fórmula de cálculo:	<i>(Fecha de reporte de inconsistencias y datos atípicos de la información estadística) – (Fecha de inicio del proceso mensual de minería de datos de la información estadística).</i>
Responsable de la medición:	<i>Analista / Profesional encargado de la GIEG</i>
Frecuencia de medición:	<i>Mensual</i>
Unidad de medida:	<i>Tiempo (días)</i>
Fuente de medición:	<i>Correos electrónicos y PAO</i>
Meta:	<i>Hasta 5 días (de acuerdo a la Regulación No. ARCONEL - 002/19)</i>
<u>SIMBOLOGÍA</u>	
<i>No aplica.</i>	
<i>NOTA:</i> <i>Los días a los que se hace mención son días laborables.</i>	

7.1 Formato de tablas y gráficos con los que se muestran los resultados de la minería de datos. Código: GEI-OEG-P001-I001-R001

Nombre de la Empresa	DATOS	Año n-1	Año n												Año n+1	Tendencias	Observaciones
		Dic n-1	Enero - n	Febrero - n	Marzo - n	Abril - n	Mayo - n	Junio - n	Julio - n	Agosto - n	Septiembre - n	Octubre - n	Noviembre - n	Diciembre - n	Enero n+1		
Variable 1	SISDAT																
	Δ SISDAT																
	Δ SISDAT [%]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
Variable 2	SISDAT																
	Δ SISDAT																
	Δ SISDAT [%]																
Variable 3	SISDAT																
	Δ SISDAT																
	Δ SISDAT [%]																
Variable n	SISDAT																
	Δ SISDAT																
	Δ SISDAT [%]																

7.2 Formato de tablas de los resultados de la minería de datos del Software CONELEC Minería de Datos. Código: GEI-OEG-P001-I001-R002

7.2.1 Primera pestaña: Anomalías por Agente

Empresa	Total atípicos	Unidad
Empresa 1		celdas
Empresa 2		celdas
Empresa 3		celdas
Empresa 4		celdas

7.2.2 Segunda pestaña: Detalle Anomalías

Agente	Año	Mes	Variable 1	Variable 2	Variable 3	Variable 4	Variable 5	Variable 6	Variable 7	Variable n
Empresa 1	n	m								
Empresa 2	n	m				(Dato atípico)				
Empresa 3	n	m								
Empresa 4	n	m		(Dato atípico)					(Dato atípico)	