

UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO
FACULTAD DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA, INFORMÁTICA Y
MECÁNICA

ESCUELA PROFESIONAL DE INGENIERÍA INFORMÁTICA Y DE SISTEMAS



TESIS

**“APLICACIÓN DE TÉCNICAS DE MINERÍA DE DATOS PARA
IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO DE LAS
VARIABLES DE PROCESO DE GENERACIÓN Y DISTRIBUCIÓN DE
ENERGÍA ELÉCTRICA, PARA LA EMPRESA EGEMSA ”**

Para optar al título profesional de:
INGENIERO INFORMÁTICO Y DE SISTEMAS

Presentado por:
BR. CHINO ESPINOZA, HEBERT
BR. LAVILLA ALVAREZ, VANESA

Asesor:
DR. RONY VILLAFUERTE SERNA

Cusco - Perú
2019

Resumen

La revolución digital ha permitido que la captura de datos sea fácil, cada momento gran cantidad de datos son recogidos y almacenados en bases de datos a cada instante, y las herramientas tradicionales no son las adecuadas para procesarlas.

La Minería de Datos (*Data Mining*) es un conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de extraer conocimiento útil y comprensible. En el momento que se le atribuye algún significado a los datos éstos pasan a convertirse en información.

Las series temporales o series de tiempo es una colección de observaciones de una variable recogidas secuencialmente en el tiempo. La Minería de Datos Temporal (*Temporal Data Mining*, TDM), es la Minería de Datos para series temporales, que está compuesta por métodos que son capaces de caracterizar series temporales con distintas características fuera de los tradicionales.

La Empresa de Generación Eléctrica Machupicchu S.A. (EGEMSA), a través del Sistema de Supervisión, Control y Adquisición de Datos (SCADA) va generando gran cantidad de datos y estos son almacenados en su servidor histórico, éstas se almacenan por un tiempo aproximado de 3 meses y luego se eliminan, los datos relevantes se guardan en cuaderno de registro de eventos o en hojas de Excel.

El sistema SCADA recibe las señales de los distintos sensores y medidores de campo (voltímetros, capacímetros, amperímetros, medidores de caudal, sensores de temperatura), a estas señales las denominaremos Variables de proceso de generación y distribución de energía eléctrica.

La presente investigación buscó primero extraer las variables de proceso de forma segura sin vulnerar la seguridad de la red SCADA; segundo, se implementó una Base de Datos Histórica para las variables de proceso de generación y distribución; tercero, se aplicó Técnicas de Minería de Datos Temporales (TDM) y expresiones regulares para encontrar características y patrones de comportamiento; cuarto, se aplicó *clustering* (técnica descriptiva de DM) con el fin de ayudar a interpretar con ayuda de la estadística descriptiva los patrones; y por último se entregó los resultados a la empresa para la toma de decisiones frente a eventualidades y planificación de mantenimientos preventivos.

Para el trabajo de investigación, se utilizó la metodología CRISP-DM, ya que es un método muy utilizado y que nos permitió la flexibilidad de avanzar o retroceder en sus distintas fases.

Palabras clave: Patrones de comportamiento, Series de Tiempo, Minería de Datos, Minería de Datos Temporal, Variables de Generación Eléctrica, CRISP-DM.