



UNIVERSIDAD
DE PIURA

FACULTAD DE INGENIERÍA

**Propuesta metodológica para la utilización de las
herramientas de inteligencia de negocios en la gestión de
operaciones de la pequeña y mediana empresa en el
sistema eléctrico peruano**

Tesis para optar el Grado de
Doctor en Ingeniería con mención en: Automatización, Control y Optimización de
Procesos

Redy Henry Risco Ramos

Asesor(es):
Dr. Ing. Edilberto Horacio Vásquez Díaz

Piura, julio de 2022

NOMBRE DEL TRABAJO

Tesis Redy Risco (Completa) Ver2.docx

AUTOR

Redy Risco

RECUENTO DE PALABRAS

29003 Words

RECUENTO DE CARACTERES

158663 Characters

RECUENTO DE PÁGINAS

115 Pages

TAMAÑO DEL ARCHIVO

6.1MB

FECHA DE ENTREGA

May 25, 2022 4:35 PM GMT-5

FECHA DEL INFORME

May 25, 2022 4:47 PM GMT-5**● 4% de similitud general**

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para cada base de datos.

- 4% Base de datos de Internet
- Base de datos de Crossref
- 3% Base de datos de trabajos entregados
- 0% Base de datos de publicaciones
- Base de datos de contenido publicado de Crossref

● Excluir del Reporte de Similitud

- Material bibliográfico
- Material citado
- Bloques de texto excluidos manualmente
- Material citado
- Coincidencia baja (menos de 10 palabras)



Dr. Ing. Edilberto H. Vásquez Díaz

Asesor

Dedicatoria

A mis padres Manolo e Hilda, a mi hermana Kelly y a mi familia pequeña: Marina, Diego y Marcia.





Agradecimientos

El presente trabajo de investigación ha sido financiado por el Proyecto Concytec – Banco Mundial “Mejoramiento y Ampliación de los Servicios del Sistema Nacional de Ciencia Tecnología e Innovación Tecnológica” 8682-PE, a través de su unidad ejecutora ProCiencia. [contrato número 06-2018-FONDECYT/BM], con el objetivo de desarrollar investigación aplicada a la pequeña y mediana empresa del sector eléctrico nacional y regional.





Resumen

El objetivo del presente trabajo es proponer una metodología utilizable por pequeñas y medianas empresas del sector eléctrico para la mejora en la gestión de su suministro eléctrico basándose en el uso de los datos disponibles producto de sus operaciones. Para ello se realizó una búsqueda de antecedentes en la literatura científica existente, se eligió una metodología aplicable a las empresas objetivo y se probó en tres casos de estudio. Como resultado se mostró que con herramientas (hardware y software) accesibles a la pequeña y mediana empresa es posible aplicar la metodología elegida obteniéndose información y conocimiento útil para tomar decisiones operativas y/o comerciales que agreguen valor a las empresas. Basado en los resultados del trabajo se concluye que la metodología elegida es útil y vigente para las empresas objeto de estudio, que el hardware y software no son limitaciones para el desarrollo de estas soluciones, que la carencia de capacitación en el mercado laboral de este tipo de conocimiento si constituye una limitación, que la muy baja cultura en la gestión de los datos también constituye una limitación y que cuando lo que se pretende es obtener conocimiento con objetivos claros desde unos datos también definidos, como es el caso de la gestión del suministro eléctrico en una PYME, la metodología elegida constituye una minería de datos clásica y por tanto la metodología propuesta es válida y produce conocimiento útil para la toma de decisiones.



Tabla de contenido

Introducción	15
Capítulo 1 Conceptos generales y actual grado de desarrollo de las aplicaciones de inteligencia de negocios al sector eléctrico	17
1.1 Definiciones básicas	17
1.2 Aplicación de inteligencia de negocios y analítica de datos en empresas del sector eléctrico	19
1.3 Situación de la analítica de datos en las empresas del sector eléctrico del Perú	24
1.4 Limitaciones para el desarrollo de inteligencia de negocios en empresas del sector eléctrico en el país	28
1.5 Planteamiento del problema	28
1.6 Objetivos de la investigación	29
1.6.1 Objetivo general.....	29
1.6.2 Objetivos específicos	29
1.7 Justificación de la investigación.....	29
1.8 Delimitación de la investigación	32
1.8.1 Delimitación geográfica	32
1.8.2 Delimitación temporal	32
1.8.3 Delimitación temática.....	33
1.9 Limitaciones de la investigación	33
Capítulo 2 Marco teórico de la inteligencia de negocios.....	35
2.1 Grado de desarrollo de la inteligencia de negocios	35
2.2 Conceptos generales.....	37
2.2.1 Ciencia de datos	37
2.2.2 Inteligencia de negocios.....	39
2.2.3 Analítica de datos aplicada a los negocios (Business Analytics).....	40
2.2.4 Inteligencia artificial.....	41
2.2.5 Big data e internet de las cosas	46
2.2.6 Resumen	49
Capítulo 3 Marco metodológico	51
3.1 Investigación cuantitativa e investigación cualitativa	51
3.1.1 La investigación cuantitativa.....	51

3.1.2	La investigación cualitativa	52
3.2	El método de estudio de casos como estrategia de investigación.....	54
3.3	Selección de los casos de estudio	57
3.4	Recopilación de datos.....	57
3.5	Análisis de los datos.....	58
3.6	Desarrollo del marco.....	58
3.7	Descripción de los casos	58
3.7.1	Caso 1: Calidad de energía y clasificación.....	58
3.7.2	Caso 2: Análisis de patrones de operación	59
3.7.3	Caso 3: Predicción de producción.....	59
Capítulo 4	Propuesta metodológica aplicada a una PYME del sector eléctrico peruano	61
4.1	Introducción.....	61
4.2	Justificación.....	65
4.3	Propuesta.....	66
4.3.1	Paso uno: Comprender el negocio.....	67
4.3.2	Paso dos: comprensión de los datos.....	68
4.3.3	Paso tres: preparar los datos	69
4.3.4	Paso cuatro: Modelado.....	69
4.3.5	Paso cinco: Evaluación	70
4.3.6	Paso seis: Implementación	70
4.4	Aplicación de la metodología a los casos de estudio	72
4.4.1	Caso 1: Calidad de energía eléctrica y clasificación.....	72
4.4.2	Caso 2: Determinación de patrones de operación normal.....	77
4.4.3	Caso 3: Predicción de producción hidráulica	93
4.4.4	Requerimientos y comentarios sobre la metodología	103
Conclusiones		105
Referencias bibliográficas		107
Apéndices		113
Apéndice A.	Exploración de la situación actual de la empresa – Focus Group.....	115
Apéndice B.	Resultados de la encuesta.....	120
Apéndice C1.	Paper caso 1 Uso de un nuevo marco de inteligencia empresarial en la gestión de la calidad del suministro eléctrico en pequeñas y medianas empresas	124
Apéndice C2.	Paper caso 2: Aplicación de técnica de minería de datos para determinar un modelo de operación normal de n turbogenerador hidráulico.....	125
Apéndice C3.	Paper Caso 3: Prediction of hidropower energy in a Peruvian company using a business intelligence framework	126

Lista de tablas

Tabla 1	Aplicaciones de data analytics por subsectores.....	31
Tabla 2	Número de artículos según criterios de búsqueda utilizados.....	32
Tabla 3	Definiciones de Business Intelligence en el tiempo	39
Tabla 4	Definiciones de Business Analytics en el tiempo	41
Tabla 5	Diferencias entre los métodos Cualitativos - Cuantitativos y estudios de caso	54
Tabla 6	Resumen de métodos analizados.....	65
Tabla 7	Etapas del Modelo CRISP-DM	66
Tabla 8	Fases de la implementación	71
Tabla 9	Caso 1 - Resultado del agrupamiento en clusters de los factores de potencia	75
Tabla 10	Caso 1 - Pasos del banco de condensadores.....	75
Tabla 11	Caso 2: Períodos de operación del Turbogenerador	79
Tabla 12	Caso 2 - Variables registradas en el sistema SCADA	80
Tabla 13	Caso 2 -Funciones Excel utilizadas para el cálculo de la distribución Chi cuadrado y distribución F	83
Tabla 14	Caso 2 - Valores calculados utilizando Excel y sus funciones, para los límites de control superior del SPE	83
Tabla 15	Caso 2 - Valores calculados utilizando Excel y sus funciones, para los límites de control superior del T2 de Hotelling	84
Tabla 16	Caso 2 - Conjuntos de muestras analizadas	84
Tabla 17	Caso 3 - Resultados numéricos de las predicciones.....	102
Tabla 18	Caso 3 - Error cuadrático medio (MSE) de las predicciones	102



Lista de figuras

Figura 1	Población mundial – Proyección	17
Figura 2	Demanda de electricidad global y mix de generación por escenario de proyección	18
Figura 3	Proyección de la demanda global del Perú en escenarios (MW).....	18
Figura 4	Modelo clásico y futuro de una red eléctrica.....	20
Figura 5	Dimensiones del análisis de datos en el sector eléctrico.....	23
Figura 6	Crecimiento de artículos por tema	23
Figura 7	Cronología del marco legal vigente.....	24
Figura 8	Transacciones en el sector eléctrico	25
Figura 9	Política energética nacional del Perú	25
Figura 10	Proyecto de reglamentación de la generación distribuida	26
Figura 11	Coeficiente de electrificación nacional y rural, 1995-2025 (Desde el 2018 son proyecciones)	27
Figura 12	Datos disponibles en un sistema eléctrico.....	30
Figura 13	Cadena de valor de los datos	38
Figura 14	Concepto de internet de las cosas	47
Figura 15	La investigación cuantitativa	52
Figura 16	La investigación cualitativa.....	53
Figura 17	Diagrama del ciclo de vida de Kimball.....	61
Figura 18	Pasos para crear una solución de BI.....	62
Figura 19	Fases del modelo de referencia CRISP-DM	63
Figura 20	Metodología de DM para aplicaciones de ingeniería como extensión de CRISP-DM.....	64
Figura 21	Caso1 - Archivo de datos en formato fuente	73
Figura 22	Caso 1 - Datos preparados en PowerBI	73
Figura 23	Caso 1 - Dashboard de la tensión	74
Figura 24	Caso 1 - Dashboard del Factor de Potencia.....	75
Figura 25	Caso 2- Matriz de datos.....	81
Figura 26	Caso 2 - Porcentaje de varianza explicada acumulada por cada componente principal.....	82

Figura 27	Caso 2 - Gráfico SPE - Muestra 1: Operación previa a calibración de álabes (26/02/2021 17:03)	85
Figura 28	Caso 2 - Gráfico T2 de Hotelling - Muestra 1: Operación previa a calibración de álabes (26/02/2021 17:03)	85
Figura 29	Caso 2 - Gráfico SPE - Muestra 2: Operación con vertedero funcionando (diferentes días).....	86
Figura 30	Caso 2 - .Gráfico T2 de Hotelling - Muestra 2: Operación con vertedero funcionando (diferentes días)	86
Figura 31	Caso 2 - Gráfico SPE - Muestra 3: Operación previa a limpieza de rejillas y calibración de álabes (03/02/2021 22:46)	87
Figura 32	Caso 2 - Gráfico T2 de Hotelling - Muestra 3: Operación previa a limpieza de rejillas y calibración de álabes (03/02/2021 22:46)	88
Figura 33	Caso 2 - Gráfico SPE - Muestra 4: Operación previa a parada por falla en Unidad de Alta Presión. Parada a las 20:48:45 del 10/03/2021	88
Figura 34	Caso 2 - Gráfico T2 de Hotelling - Muestra 4: Operación previa a parada por falla en Unidad de Alta Presión. Parada a las 20:48:45 del 10/03/2021	89
Figura 35	Caso 2 - Gráfico SPE - Muestra 5 - Simulación de alteración de variables BT2515, BT3403, BT3404, BT2516 y comportamiento en parada y arranque de unidad del 10/03/2021.....	89
Figura 36	Caso 2 - Gráfico T_2 de Hotelling - Muestra 5 - Simulación de alteración de variables BT2515, BT3403, BT3404, BT2516 y comportamiento en parada y arranque de unidad del 10/03/2021.....	90
Figura 37	Caso 2 - Captura de pantalla del despliegue del Gráfico SPE - Muestra 5 en Power BI	92
Figura 38	Caso 3 - Registro de producción mensual de la planta desde 1998	95
Figura 39	Caso 3 - Datos cargados después de limpieza, ordenamiento, transformaciones y cálculos	96
Figura 40	Caso 3: Modelado de los datos	97
Figura 41	Caso 3 – Predicción con el modelo Smooth Exponential.....	97
Figura 42	Caso 3 - Predicción con el modelo ARIMA	99
Figura 43	Caso 3 - Predicción con molo baso en Redes Neuronales	99
Figura 44	Caso 3 - Comparación de los modelos con la producción real	101

Introducción

En la actualidad la dirección y gestión empresarial basada en datos es un tema de enorme vigencia incrementado por la gran disponibilidad de datos provenientes de diferentes fuentes como la internet de las cosas, los sistemas de control y adquisición de datos, las redes sociales, los wearables, etc.

Por otro lado, el sector eléctrico es uno de los sectores importantes para el desarrollo de un país, su crecimiento va de la mano del crecimiento poblacional y del crecimiento del PBI de un país teniendo una importancia transversal en casi todos los sectores productivos.

El sector eléctrico en su cadena de valor está integrado por empresas generadoras, transmisoras, distribuidoras y comercializadoras, todas incluidas en este incremento en la disponibilidad de datos que lleva a muchos investigadores a afirmar que si una empresa de este sector no incluye el uso de los datos en su toma de decisiones está destinada a salir del mercado.

En el Perú el sector empresarial está formado mayoritariamente por pequeñas y medianas empresas, lo mismo sucede en el sector eléctrico; por lo tanto, investigar la manera de como las empresas de este sector puedan emplear herramientas que transformen los datos cada vez más disponibles en conocimiento que mejore su toma de decisiones lo consideramos útil.

Una de las herramientas que desde hace varios años transforma datos en conocimiento es la inteligencia empresarial, inteligencia de negocios o *business intelligence*. El presente trabajo tiene por objetivo proponer y probar una metodología que pueda ser utilizada por las empresas pequeñas y medianas del sector eléctrico para convertir sus datos en conocimiento con el fin específico de mejorar la gestión de su suministro eléctrico.

Para cumplir su objetivo, el trabajo está organizado en cinco capítulos. En el primer capítulo presentamos los conceptos generales y actual grado de desarrollo de las aplicaciones de inteligencia de negocios en el sector eléctrico, en el capítulo dos mostramos el marco teórico de la inteligencia de negocios incluyendo conceptos relacionados como ciencia de datos, big data e internet de las cosas, en el capítulo tres presentamos el marco metodológico utilizado, específicamente el de casos de estudio, en el capítulo cuatro mostramos la metodología elegida, la probamos en tres casos de estudio y mostramos los resultados

obtenidos, en el capítulo cinco realizamos una discusión y análisis de los resultados de la metodología en los casos de estudio y finalmente presentamos unas conclusiones producto del trabajo realizado.



Capítulo 1

Conceptos generales y actual grado de desarrollo de las aplicaciones de inteligencia de negocios al sector eléctrico

1.1 Definiciones básicas

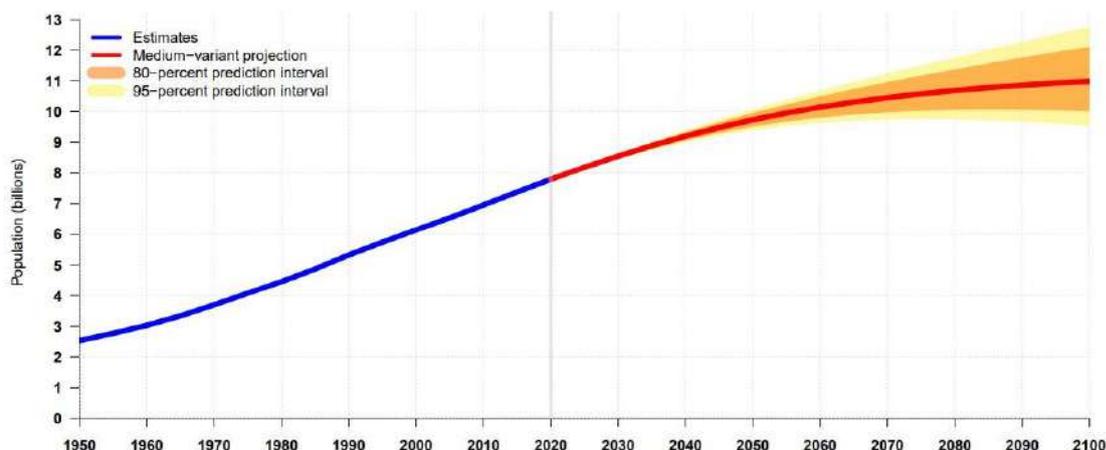
En la actualidad la gestión de la energía en general es un tema prioritario si queremos tener un suministro energético confiable, seguro y sostenible. Un adecuado sistema de gestión de la energía es la base para la implementación de medidas que incrementen la eficiencia en el uso de la energía de diferentes fuentes.

La prioridad de una adecuada gestión del suministro energético se basa en la intuición que la demanda energética mundial se verá incrementada teniendo en cuenta el crecimiento poblacional mostrado en la Figura 1 pues un aumento poblacional requiere más energía para la atención de necesidades básicas y de confort de la población.

Esta proyección de crecimiento de la demanda energética puede observarse y, por tanto, reafirmarse con la proyección de la demanda eléctrica mostrada en la Figura 2.

Figura 1

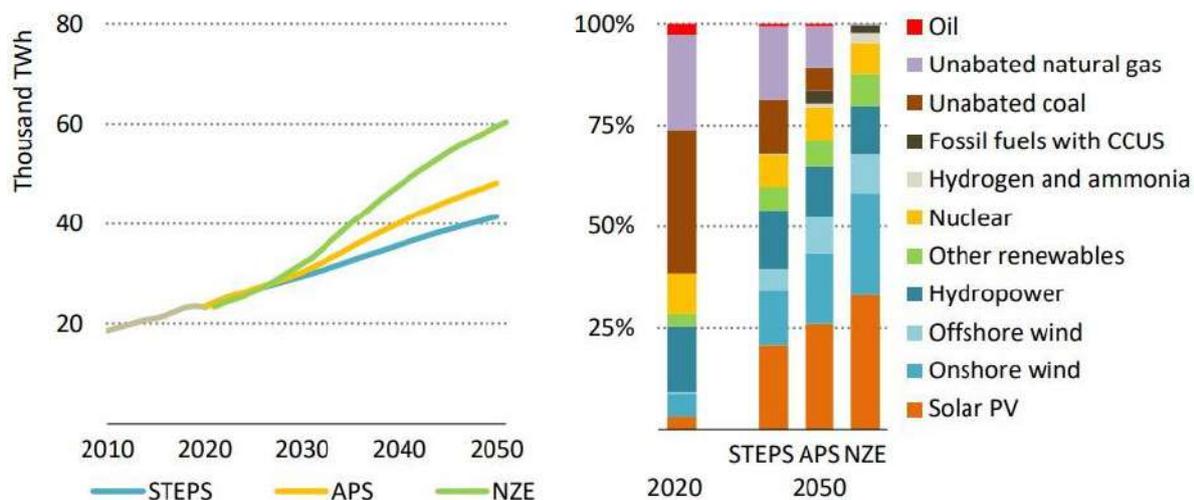
Población mundial – Proyección



Fuente: (United Nations, Department of Economic and Social Affairs, 2019)

Figura 2

Demanda de electricidad global y mix de generación por escenario de proyección



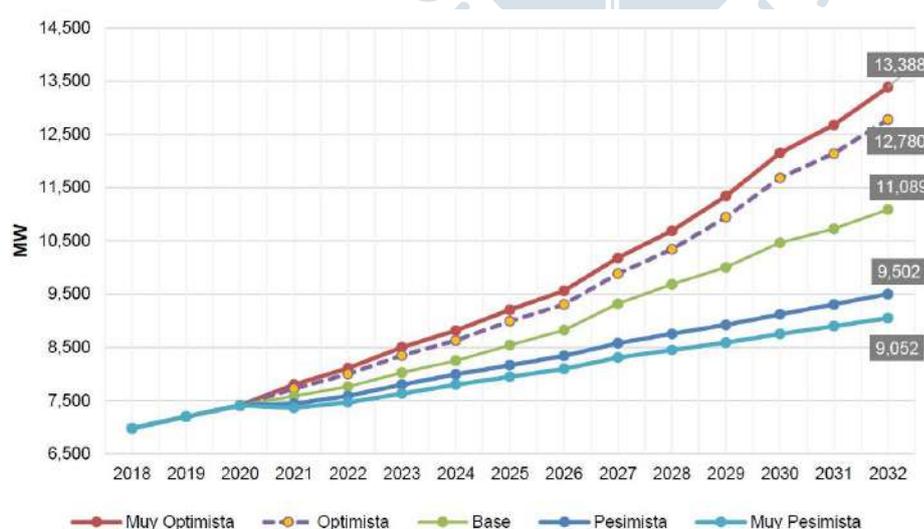
Fuente: (IEA, 2021)

La Figura 2 muestra una proyección de incremento de la demanda de electricidad en tres escenarios posibles STEPS (escenario de políticas establecidas), APS (escenario de promesas anunciadas) y NZE (escenario con emisiones netas cero de carbono). Notamos que en todos los escenarios superados los años de pandemia la proyección de la demanda de electricidad es a aumentar.

El Perú no tiene proyecciones diferentes y tal como se muestra en la Figura 3 aun en el más pesimista de los escenarios la demanda se espera que se incremente hacia el año 2032.

Figura 3

Proyección de la demanda global del Perú en escenarios (MW)



Fuente: (COES, 2021)

Las proyecciones mostradas en los párrafos anteriores muestran que a nivel mundial y a nivel nacional el sector eléctrico es un sector en crecimiento y por ende un sector muy importante. Esta importancia se muestra también volviendo a observar la Figura 4 y fijándose en el hecho que en el escenario de emisiones cero de carbono la demanda de electricidad aumentaría notoriamente pues muchas fuentes de energía para diferentes procesos serían cambiadas por la electricidad como fuente de energía.

Esta importancia hace que el desarrollo de investigaciones en este sector esté justificado más aun cuando, como veremos más adelante, no se aprovechan las nuevas tecnologías basadas en inteligencia artificial para una adecuada gestión del suministro eléctrico a diferente nivel y a nivel práctico no solo académico.

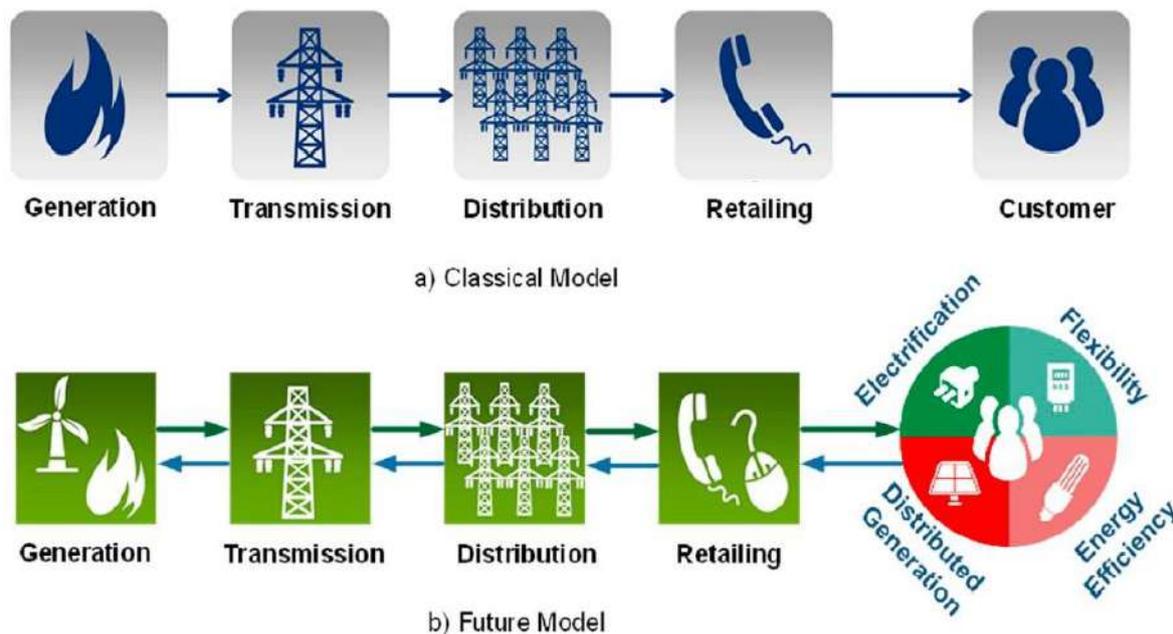
1.2 Aplicación de inteligencia de negocios y analítica de datos en empresas del sector eléctrico

Para la definición del nivel de aplicación de la inteligencia de negocios y de la analítica de datos en el sector eléctrico creemos conveniente explicar de manera resumida como está constituido el sector eléctrico en general y luego mencionaremos algunas de las aplicaciones frecuentes de analítica de datos en este sector. Teniendo en cuenta que en esta sección no definiremos a profundidad algunos conceptos de analítica de datos pues estos serán tratados a más profundidad en el capítulo 2.

Según (Scheidt et al., 2020) un sector eléctrico está constituido por la generación, la comercialización, la transmisión - distribución y el consumo. La generación se encarga de la producción de la energía en las diferentes plantas de generación, la comercialización controla la compra y venta de la energía eléctrica en los mercados mayoristas, la transmisión – distribución entrega la electricidad a través de redes eléctricas y el consumo es la demanda y el uso final de la energía eléctrica.

Para (Alavikia & Shabro, 2022), desde 1930 la estructura de una red de energía no ha cambiado mucho y consta de tres partes principales: la generación, la transmisión y la distribución en los que el flujo de energía eléctrica va desde los proveedores hacia los consumidores comerciales o domiciliarios.

(Dranka & Ferreira, 2020) presenta dos modelos de estructuras, uno clásico el que coincide con el de (Alavikia & Shabro, 2022) de tipo unidireccional y un modelo futuro que implica un sentido bidireccional del flujo de energía. En ambos modelos las partes principales son siempre la generación, transmisión, distribución y el consumo atendido por un comercializador al por menor.

Figura 4*Modelo clásico y futuro de una red eléctrica*

Fuente : (Dranka & Ferreira, 2020)

Consolidando las estructuras planteadas por los tres autores anteriores podemos definir que los integrantes principales de un sistema eléctrico son: los generadores, transmisores, distribuidores y consumidores. La comercialización no la consideramos pues en el Perú ésta la hace el generador o distribuidor con su cliente final y no existe como integrante individual.

Respecto a la aplicación de la inteligencia de negocios y analítica de datos en las empresas del sector o sistema eléctrico utilizaremos como referencia dos trabajos que se complementan y que resumen muy bien estas aplicaciones.

En el trabajo de (Zhang et al., 2018) se presenta a las ya muy conocidas redes inteligentes (o *smart grids* en inglés) que no es otra cosa que la red de energía eléctrica a la que se le ha integrado una capa de información que permite la comunicación bidireccional de información entre los controladores centrales (sean de la generación, transmisión y distribución) y los actuadores locales, así como las unidades lógicas para responder digitalmente a situaciones emergencia en los elementos físicos o cambios rápidos en la demanda eléctrica. En este trabajo se presenta la definición de la Unión Europea donde se indica que las redes inteligentes son “redes eléctricas que pueden integrar de manera inteligente el comportamiento y las acciones de todos los usuarios conectados a ella: Generadores, consumidores y aquellos que hacen ambos con el fin de proporcionar eficientemente suministros de electricidad sostenibles, económicos y seguros.”

Este intercambio de información suscitado en las redes inteligentes genera una gran cantidad de datos provenientes de diferentes fuentes y con diferentes estructuras que muy bien pueden ser definidos como big data (concepto que definiremos más adelante). Junto con el concepto de big data aparece el concepto de descubrir información valiosa a partir de datos recopilados masivamente en operaciones comerciales como ayuda al conocimiento para la toma de decisiones comerciales, propuesto en 1989 por Howard Dresner como “inteligencia de negocios” (en inglés *Business Intelligence*)(Zhang et al., 2018).

Las aplicaciones que (Zhang et al., 2018) se proponen para el análisis de datos en las redes eléctricas son los siguientes:

- Detección de fallas.
- Mantenimiento predictivo y basado en condición.
- Análisis de estabilidad transitoria.
- Estimación del estado de “salud” del dispositivo eléctrico.
- Monitoreo de la calidad de la energía eléctrica.
- Identificación de la topología de la red.
- Pronóstico de energía renovable.
- Pronóstico de demanda.
- Perfilado del diagrama de carga.
- Desagregación de carga o monitoreo de carga no intrusiva (NILM)
- Detección de pérdidas no técnicas.

Aun con la gran cantidad de trabajos de investigación en este tema y lo diverso de las posibles aplicaciones en toda la cadena de valor del sector eléctrico, el número de soluciones implementadas es bajo. Las causas principales según (Zhang et al., 2018), son: la falta de un formato estándar de datos tanto para el software como para las estructuras de las bases de datos, la interoperabilidad de los diferentes sistemas de información, la forma tradicional de almacenamientos aislados de los datos de los diferentes sistemas que impiden intercambios de información. Por otro lado, la confidencialidad de muchos de los datos en los diferentes sistemas hace difícil la investigación con datos reales.

Otro aspecto importante es que el análisis de datos en redes inteligentes es un campo completo y complicado, que no solo depende de los algoritmos o técnicas matemáticas, sino que también depende de la operación de los sistemas, los comportamientos de un gran número de usuarios autónomos, las tecnologías TIC (Tecnología de la Información y Comunicaciones), la experiencia del campo, etc. Por lo tanto, necesita la sinergia entre expertos de diferentes campos.

Y finalmente (Zhang et al., 2018) plantea que la falta de visión estratégica, diseños de aplicaciones, gran inversión y un poco reconocimiento del valor de los datos hace que las aplicaciones de analítica de datos en sistemas reales esté creciendo lentamente.

Aunque en (Zhang et al., 2018) se menciona que el desarrollo de las redes inteligentes es requerido ante la necesidad de descarbonizar, descentralizar y digitalizar (tres D) el sector eléctrico, es en el trabajo de (Scheidt et al., 2020) donde se plantea que estas tres D impulsan la importancia de la inteligencia artificial en general y del análisis de datos en particular. Scheidt plantea que existen unos factores de empuje como son la disminución de los costos de la tecnología de la información y comunicaciones, así como los avances en el poder de cómputo hacen que se dispongan de cada vez más datos y aparezcan nuevas oportunidades de análisis. También indica que existen unos factores de atracción como la creciente volatilidad de la generación de electricidad debido a una proporción cada vez mayor de fuentes de energía renovables y un número creciente de actores activos en el sistema eléctrico, que aumentan la complejidad y crean nuevas necesidades para el análisis de datos.

Respecto a las posibles aplicaciones, (Scheidt et al., 2020) plantea que con base en las aplicaciones típicas de la literatura de análisis de datos, se definen cuatro categorías:

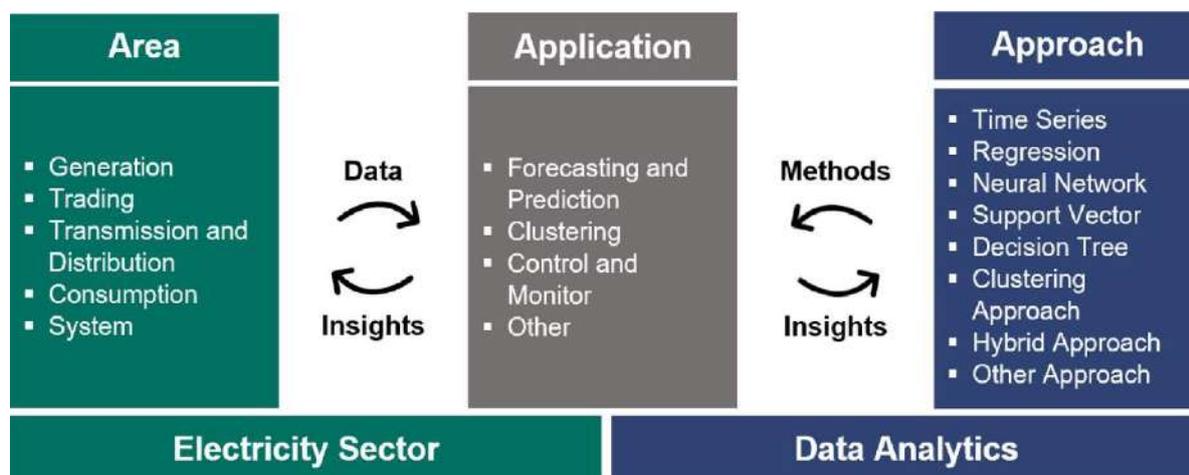
- Pronóstico y predicción mediante análisis de datos supervisados,
- Agrupamiento (*clustering* en inglés) mediante análisis de datos no supervisados,
- Monitoreo y control (mediante análisis tanto supervisados como no supervisados)
- Otros. Tanto el pronóstico como la predicción se ocupan de la estimación de resultados para datos no vistos en el futuro.

Se precisa que debido a que los términos 'predicción' y 'pronóstico' son utilizados como sinónimos por muchos autores, la primera categoría contempla ambas aplicaciones. El agrupamiento o *clustering*, por otro lado, es la incorporación de objetos en grupos homogéneos. En cuanto a Seguimiento y Control, ambos términos están relacionados e implican un proceso de observación y medición del desempeño para tomar acciones correctivas en caso de ser necesario.

Respecto a los métodos utilizados en cada categoría (Scheidt et al., 2020) define ocho grupos de enfoques: (i) serie de tiempo, (ii) regresión, (iii) redes neuronales, (iv) máquinas de vectores de soporte, (v) enfoques basados en árboles, (vi) enfoques de agrupamiento, (vii) enfoques híbridos y (viii) otros enfoques. La Figura 5 muestra un resumen de la relación del análisis de datos con el sector eléctrico a través de tres dimensiones: el área, la aplicación y el método o enfoque utilizado para el análisis.

Figura 5

Dimensiones del análisis de datos en el sector eléctrico

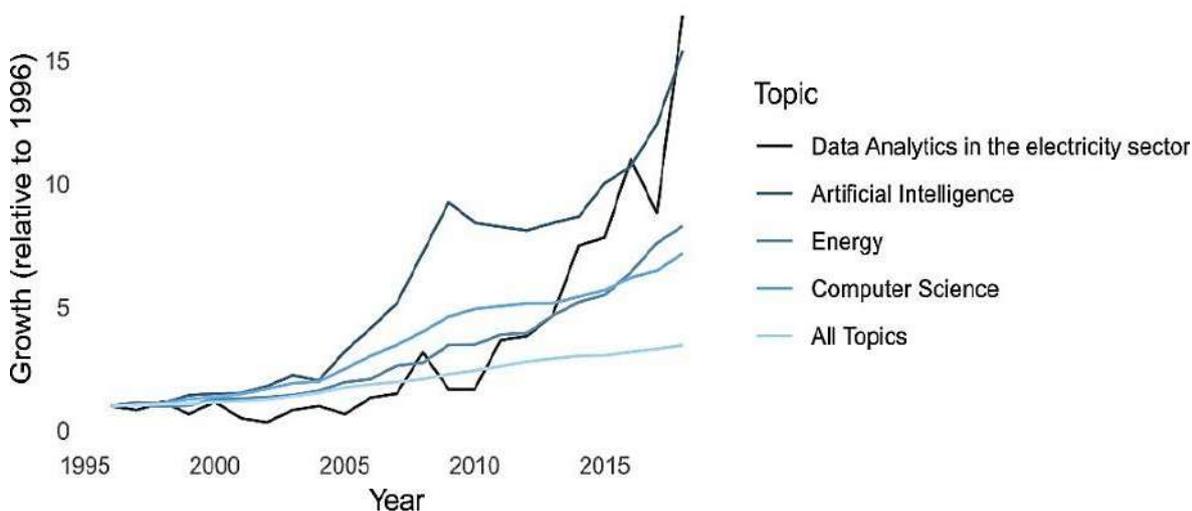


Fuente: (Scheidt et al., 2020)

Finalmente el trabajo de revisión sistemática realizado en (Scheidt et al., 2020) muestra que la cantidad de artículos relacionados con el análisis de datos en el sector eléctrico viene creciendo desde el año 2010 considerablemente, véase la Figura 6

Figura 6

Crecimiento de artículos por tema



Fuente: (Scheidt et al., 2020)

Vistas las posibles aplicaciones y enfoques con que el análisis de datos puede resolver diferentes problemas del sector eléctrico veamos ahora la situación de la analítica de datos en el Perú.

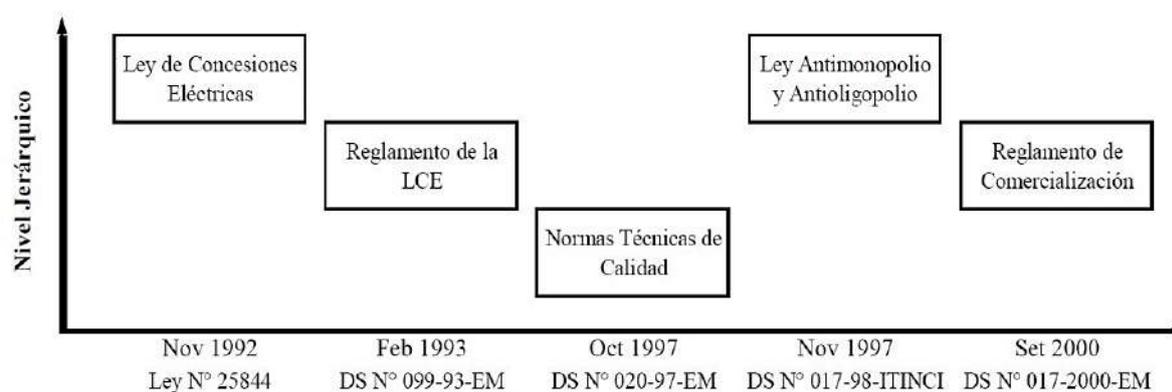
1.3 Situación de la analítica de datos en las empresas del sector eléctrico del Perú

Antes de exponer la situación de la analítica de datos en el sector eléctrico peruano conviene hacer un resumen de cómo está estructurado y normado el sector, además de mostrar la expectativa del desarrollo de las redes inteligentes en el sector.

En el informe (CEPA-NEGLI, 2016) se resume como es que opera el sector eléctrico hoy, se indica que este modo de operación nace de una gran reforma normativa realizada en 1992 con la publicación del Decreto Ley 25844, Ley de Concesiones Eléctricas (LCE), esto en el marco de una gran crisis que ocurría por aquellos años. La Ley y su Reglamento, promulgado en 1993 (RLCE), establece el marco legal para el desarrollo del negocio eléctrico, dividiéndolo en las actividades de generación, transmisión, distribución y comercialización de electricidad. Se definen también los roles de las entidades del estado: al Ministerio de Energía y Minas (MEM) se le asigna el rol de formulador de la política energética, al Organismo Supervisor de la Inversión en Energía y Minería (OSINERGMIN) se le encarga de inicio el rol regulador de las tarifas (antes lo hacía la Comisión de Tarifas Eléctricas (CTE)) y posteriormente (en 1996) se le suma el rol de fiscalización y supervisión en el sector, finalmente al Comité de Operación Económica del Sistema (COES), se le asigna el rol de operador del despacho económico del sistema eléctrico y administrar el Mercado de Corto Plazo (MCP). Además, participan hoy: el Organismo de Evaluación y Fiscalización Ambiental (OEFA) que supervisa los estándares de protección ambiental y el Instituto Nacional de Defensa de la Competencia y Protección de la Propiedad Intelectual (INDECOPI), que es la agencia de competencia transectorial y controla las fusiones y concentraciones en el mercado eléctrico. En (Dammert et al., 2005), se muestra una cronología del marco legal vigente a setiembre del 2000 que mostramos en la Figura 7.

Figura 7

Cronología del marco legal vigente

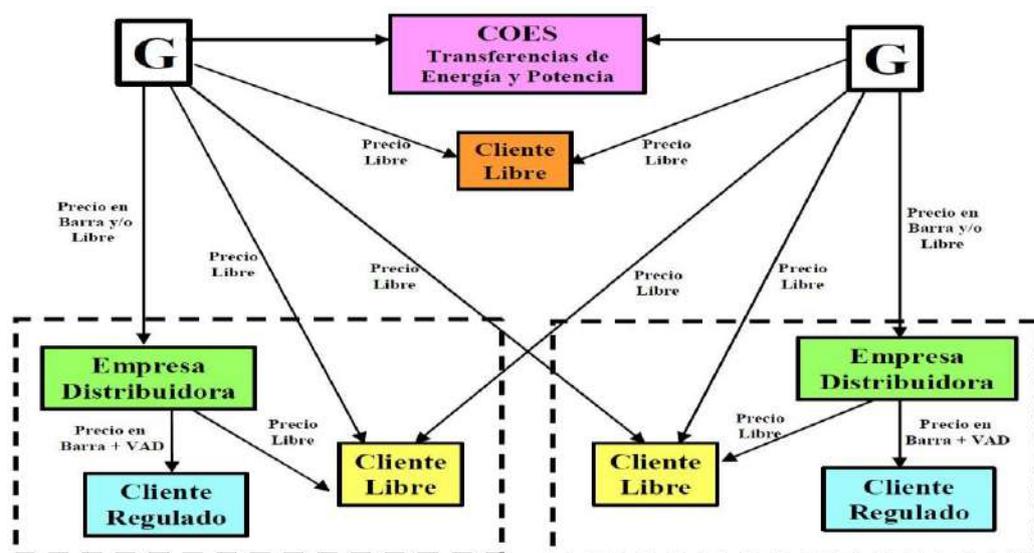


Fuente: (Dammert et al., 2005)

En el mismo trabajo (Dammert et al., 2005), los autores resumen cómo es que se realizan las transacciones en el mercado eléctrico peruano y los integrantes del sistema que intervienen mostrándolos en la Figura 8.

Figura 8

Transacciones en el sector eléctrico



Fuente: (Dammert et al., 2005)

Desde el inicio del cambio del marco normativo descrito, éste ha seguido actualizándose en un afán promotor de mejoras en el sector. En el año 2010 se plantea una política energética para un período hasta el 2040. En (Mitma, 2018), se resumen las políticas principales planteadas tal como se muestra en la Figura 9.

Figura 9

Política energética nacional del Perú

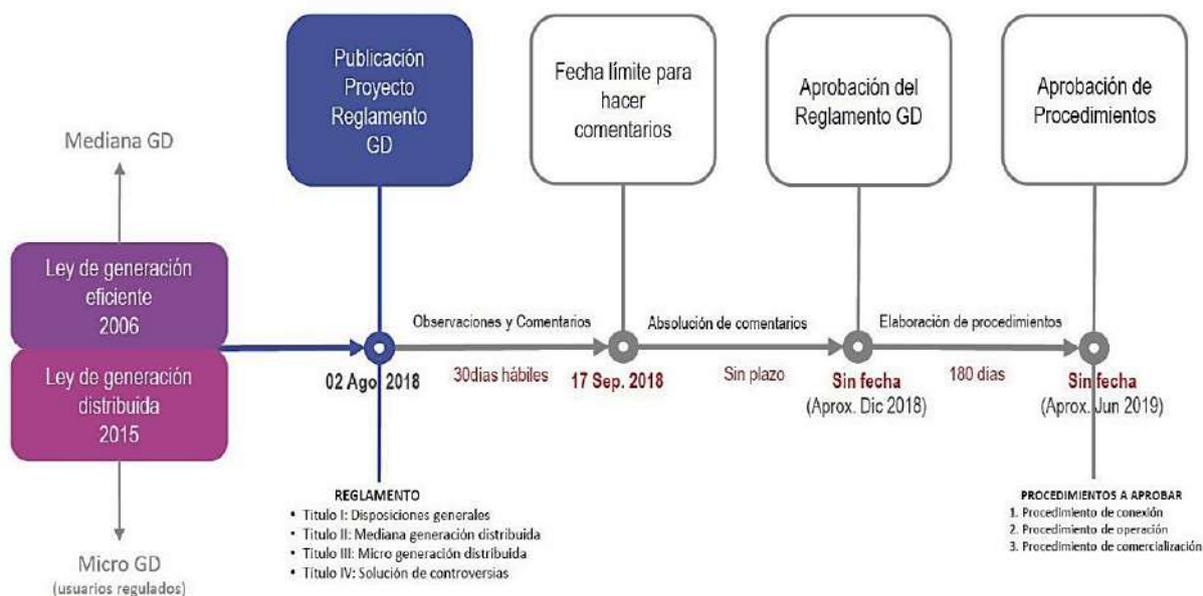


Fuente: (Mitma, 2018)

La Figura 9, muestra que la estrategia del sector apunta a la promoción de energías renovables, generación eficiente y generación distribuida. Sin embargo, tal como se muestra en la Figura 10, esta normativa está detenida.

Figura 10

Proyecto de reglamentación de la generación distribuida



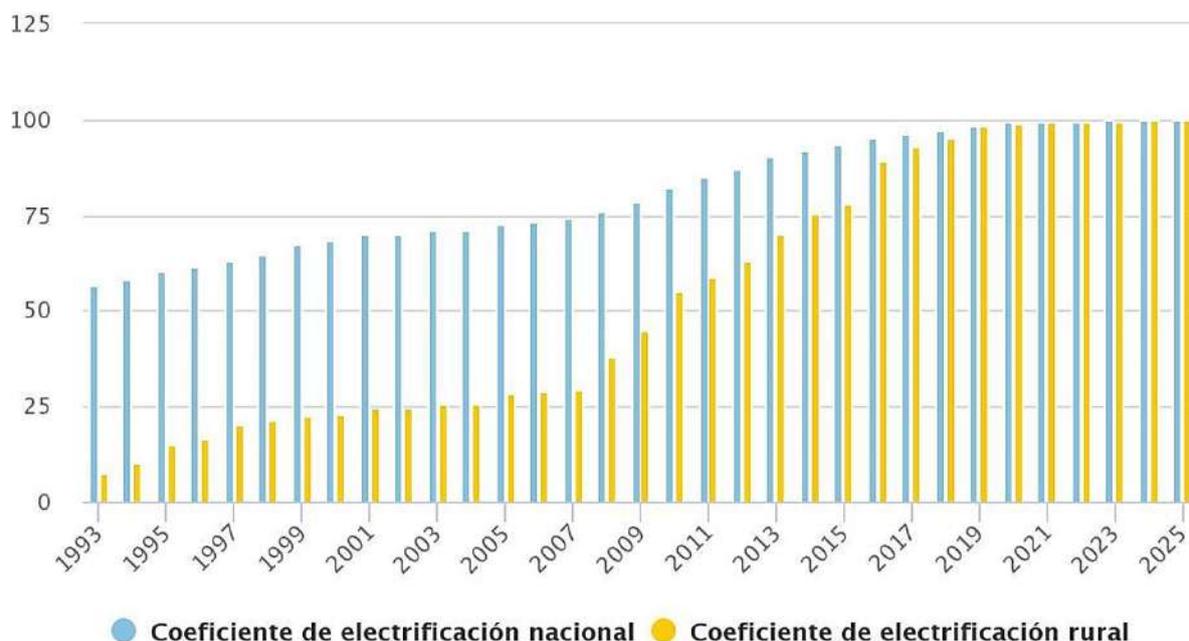
Fuente: (Mitma, 2018)

En este contexto, es claro que las políticas del sector apuntan a desarrollos que definitivamente deberían ser soportados por redes inteligentes. (Vasquez Cordano, 2017) indica, por ejemplo que en setiembre del 2010 se presentó el Proyecto de Ley 4335 en el que se proponía promover el desarrollo de las redes eléctricas inteligentes en el país, con los objetivos de lograr una mayor eficiencia del sistema eléctrico, mejorar el monitoreo y el control del consumo eficiente de electricidad, y contribuir a una mayor utilización de energías renovables y generación distribuida, proporcionando a los usuarios un beneficio mayor que los costos correspondientes. Este proyecto nunca salió a consulta.

Independientemente de las iniciativas por el desarrollo de las redes inteligentes, en la Figura 11, obtenida del Observatorio Minero Energético de OSINERGMIN, se muestra la evolución del coeficiente de electrificación desde el año 1993, se observa que desde el 2010 el coeficiente aumenta su tendencia de crecimiento hasta llegar al actual de alrededor del 98 %. Podríamos concluir que las políticas están logrando una mejora notable en el índice de electrificación y por ende el crecimiento de la demanda es de esperar.

Figura 11

Coeficiente de electrificación nacional y rural, 1995-2025 (Desde el 2018 son proyecciones)



Fuente: <https://observatorio.osinergmin.gob.pe/evolucion-coeficiente-electrificacion>

Sin un adecuado desarrollo de las redes inteligentes la aplicación de la analítica de datos aplicada al sector eléctrico en el Perú será insipiente.

Para nuestro caso, la pequeña y mediana empresa, nos interesa el desarrollo de medición inteligente y automatización en las redes de distribución, en este sentido la iniciativa de la distribuidora ENEL con 8500 medidores inteligentes instalados y 10000 más por instalar desde el 2021 es quizá la más ambiciosa. Por otro lado, al nivel de las distribuidoras la comunicación entre equipamiento principal de distribución (reconectores y subestaciones importantes) así como la inversión en iluminación led e inteligente es donde apuntan sus inversiones. Existen también proyectos piloto de instalación de medidores inteligentes en diferentes distribuidoras regionales.

Para el caso de las expectativas en empresas regionales se utilizó un Focus Group en una muestra intencional que nos permite acceder a personas con profundo conocimiento del tema. Las personas elegidas son la gerencia general, la gerencia comercial, la gerencia técnica y la jefatura de TI (Tecnologías de la Información) de una empresa regional de distribución. En el Anexo 1 se puede observar el diseño del Focus Group y sus resultados.

En resumen, los entrevistados consideran que el nivel de implementación de soluciones enmarcadas en redes inteligentes es medio a alto, consideran su principal avance la implementación de redes de comunicación e integración entre diversos sistemas, sin embargo, se consideran a distancia de las distribuidoras privadas como ENEL, por ejemplo. La causa de este nivel de implementación es principalmente las políticas sectoriales, el estar

sujetos a la ley de contrataciones del estado que hace que los proyectos se dilaten 1 o 2 años, el no reconocimiento de los costos en proyectos piloto y la no existencia del reconocimiento tarifario para la innovación lo que la dificulta en el corto plazo.

Para conocer la situación en los diferentes agentes, se preparó una encuesta abierta cuyos resultados se pueden observar en el Anexo 2. La encuesta indica que el 75% de los encuestados conocen poco o muy poco la inteligencia de negocios y analítica para la gestión del suministro eléctrico de sus empresas, el 83% no han usado herramientas de inteligencia de negocios y creen que la mayor utilidad sería en la gestión de operaciones y mantenimiento. Para la mayoría de los entrevistados los aspectos necesarios muy importantes para la implementación de una solución con BI&A (inteligencia de negocios y analítica), son: tener personal capacitado, tener un mayor conocimiento de las aplicaciones, contar con metodologías claras de implementación y contar con infraestructura tecnológica suficiente.

1.4 Limitaciones para el desarrollo de inteligencia de negocios en empresas del sector eléctrico en el país

De los resultados del Focus Group y de la encuesta realizada podemos resumir que las causas identificadas son:

- Políticas sectoriales inadecuadas.
- Corrupción y falta de valores.
- Políticas de reconocimiento de la inversión en innovación inadecuada.
- Para las empresas con régimen estatal, la ley de contrataciones dilata mucho las inversiones y no permite inversiones ágiles en planes piloto.
- Personal capacitado.
- Métodos claros.
- Conocimiento de las aplicaciones disponibles.
- Infraestructura tecnológica.

Surge entonces la necesidad de proponer una metodología de implementación de una solución basada en una plataforma de *Business Intelligence* asistida por *Business Analytics* aplicada a la gestión del suministro eléctrico de empresas del sector.

1.5 Planteamiento del problema

Nuestro tema de investigación está enmarcado en las ciencias de la computación específicamente en la inteligencia de negocios y analítica de datos.

El objeto es el desarrollo de una metodología aplicada a una pequeña o mediana empresa dedicada a la generación, distribución o consumo de energía eléctrica en el Perú regida por la normativa vigente a partir del 2020.

El problema es que en estas empresas que generan, distribuyen o consumen energía eléctrica, existe una gran cantidad de datos que pueden generar información relevante.

Dichos datos no se están utilizando actualmente para mejorar su gestión operativa y comercial.

Y la pregunta de investigación es: ¿Cómo con el desarrollo de una metodología de implementación de soluciones BI&A se puede contribuir a mejorar el desempeño organizacional de una pequeña o mediana empresa del sector eléctrico?

1.6 Objetivos de la investigación

1.6.1 Objetivo general

Proponer una metodología para la implementación de soluciones de analítica de datos para la toma de decisiones utilizando la inteligencia de negocios en pequeñas y medianas empresas del sector eléctrico peruano.

1.6.2 Objetivos específicos

- Realizar un análisis situacional de la inteligencia de negocios en las pequeñas y medianas empresas del sector.
- Proponer una metodología factible de aplicar a la realidad nacional.
- Identificar las limitaciones existentes para la aplicación de la metodología en el sector eléctrico objetivo.

1.7 Justificación de la investigación

El sector eléctrico es uno de los sectores importantes en el desarrollo de los países. Las empresas del sector eléctrico por tanto se enfrentan a constantes necesidades de mejora para la optimización de sus procesos.

Cuando nos referimos a empresas del sector eléctrico nos referimos a empresas componentes de toda la cadena de valor de un sistema eléctrico como son: generación, comercialización, transmisión, distribución y consumo, tal como se plantea en (Scheidt et al., 2020).

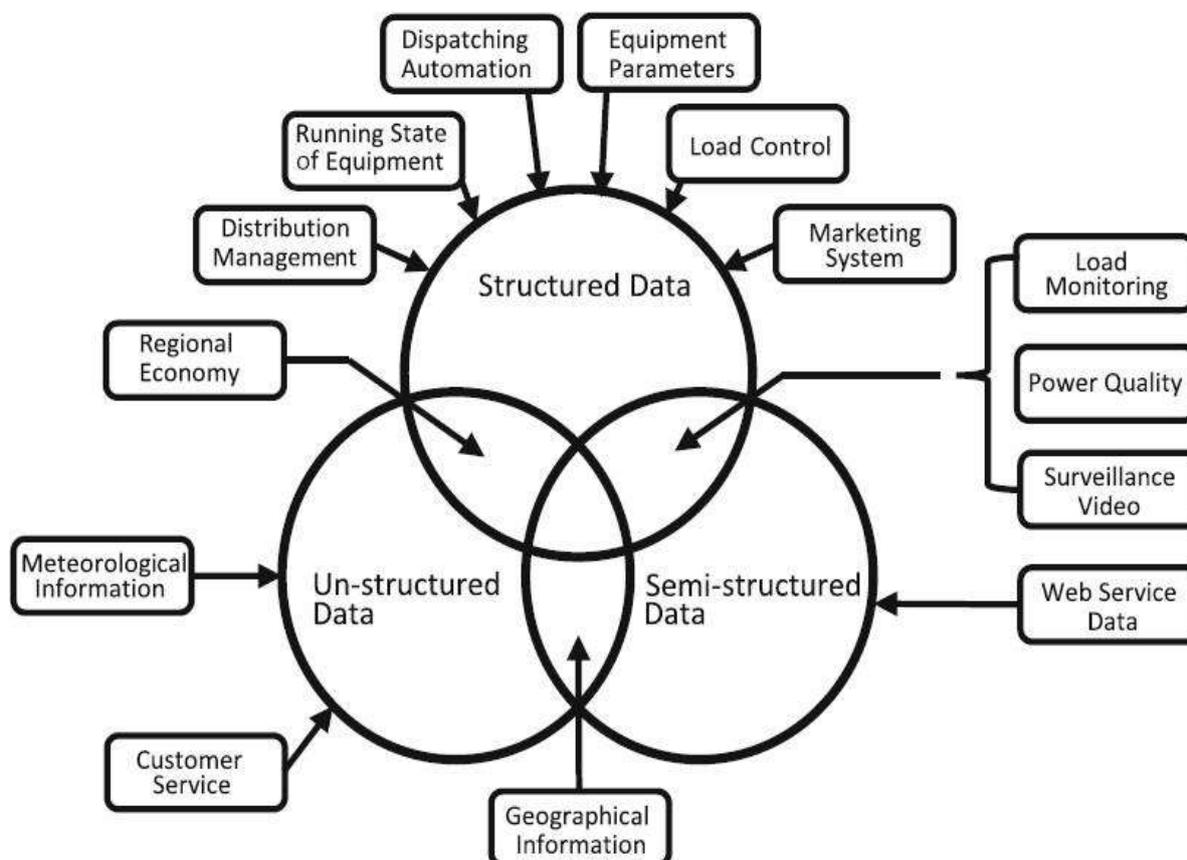
En la actualidad en las empresas de un sistema eléctrico existe una gran cantidad de datos producto de sus operaciones. Estos datos tienen un enorme potencial de generar información relevante para la oportuna toma de decisiones y normalmente no están siendo aprovechados para mejorar la gestión operativa y comercial de las mismas.

Afirmamos esto basándonos en la abundante literatura actual disponible sobre el incremento del enfoque de gestión manejada por datos de las empresas relacionadas con el sector eléctrico.

En (Zhang et al., 2018) por ejemplo se muestra los diferentes tipos de datos disponibles en un red de distribución eléctrica: datos estructurados, no estructurados, semiestructurados, tal como podemos ver en la Figura 12.

Figura 12

Datos disponibles en un sistema eléctrico



Fuente: (Zhang et al., 2018)

Por otro lado en (Arghandeh & Zhou, 2017) se afirma que para las empresas del sector eléctrico resulta inevitable convertirse en una empresa cuya gestión esté basada en datos en la era de Internet, la puesta en la nube, los teléfonos inteligentes y los recursos distribuidos. Esto teniendo en cuenta que el análisis de datos avanzado hace posible la innovación continua desbloqueando información nunca antes vista a través del aprendizaje de máquina (*Machine Learning*), el aprendizaje profundo (*Deep Learning*) y la inferencia estadística.

Para mostrar que el problema está vigente a la fecha mediante utilizamos el análisis cuantitativo de investigación en el tema realizado en (Scheidt et al., 2020) donde se puede concluir este crecimiento mostrado gráficamente en la Figura 6, donde se puede ver que la analítica de datos (*Data Analytics*) en el sector eléctrico está teniendo un crecimiento importante desde el año 2010.

Las aplicaciones de la *data analytics* en el sector eléctrico pasa por toda la cadena de valor y podemos mencionar las aplicaciones y subsistemas donde se aplican en la Tabla 1.

Tabla 1*Aplicaciones de data analytics por subsectores*

Subsistema	Aplicación	Artículos de referencia
Generación	El uso más generalizado es predicción	(Blaga et al., 2019),(Das et al., 2018)y (Mason et al., 2018)
Comercialización	La mayor aplicación es la previsión de precios	(Bento et al., 2018), (J. Wang et al., 2016) y (Liu & Shi, 2013)
Distribución	Las aplicaciones pasan por predicción y análisis de fallas y gestión de las pérdidas no técnicas.	(Kalu & Madueme, 2018), (Ahmad, 2017), (Nagi et al., 2010)
Consumidores	Análisis del consumo para los diferentes tipos de clientes (hogar, comercial e industrial)	(Zhou & Yang, 2016), (Kwac et al., 2014) y (F. Wang et al., 2018)

La Tabla 1 presenta solo una muestra pequeña de la gran cantidad de artículos escritos mostrándose una diversidad de campos con muy útiles aplicaciones.

Sin embargo, en nuestro país la aplicación de estas tecnologías en la pequeña y mediana empresa del sector eléctrico es todavía insipiente. Para validar la información se hizo una búsqueda en las bases de datos de Google Scholar y SCOPUS con los siguientes criterios:

- Criterio 1: "sector eléctrico" AND ("Perú" OR "peruano") AND ("business intelligence" OR "data analytics")
- Criterio 2: ("Perú" OR "peruano") AND ("business intelligence" OR "data analytics" OR "Machine learning" OR "data mining")
- Criterio 3: ("electric") AND ("Perú" OR "peruano" OR "peruvian") AND ("business intelligence" OR "data analytics" OR "Machine learning" OR "data mining")
- Criterio 4: ("peruvian") AND ("business intelligence" OR "data analytics" OR "Machine learning" OR "data mining")

Para la búsqueda se utilizó el software Harzing's Publish or Perish, tal como ya se ha hecho en (Banner et al., 2021) y (Lassio et al., 2021) . Los resultados de la búsqueda muestran que la cantidad de trabajos relacionados con nuestro tema es muy baja y más baja aun para trabajos relacionados con el sistema eléctrico, tal como se muestra en la Tabla 2.

Tabla 2*Número de artículos según criterios de búsqueda utilizados*

Criterios	Google Scholar		SCOPUS	
	En general	Relacionados con nuestro tema	En general	Relacionado con nuestro tema
Criterio 1	72	12	0	0
Criterio 2	43	1	16	2
Criterio 3	0	0	1	0
Criterio 4	11	0	6	0

Para relacionar la cantidad de trabajos con el número de empresas pequeñas y medianas existentes en el Perú daremos unas cifras del número de estas empresas. En el Perú según la Ley N°30056 están en la categoría pequeña y mediana empresa (en adelante le llamaremos PYME) a las empresas con ventas entre USA\$ 182,000 y 2'700,000 (Tipo de cambio 3.8 S/ por USA\$ y Unidad Impositiva Tributaria S/. 4600). El número de PYMES en el Perú según el Ministerio de la Producción peruano, al 2020 es de 73,100 PYMES, cifra que considera que por la coyuntura de pandemia el número de PYMES disminuyó, ya que en el 2019 operaban 98,900 aproximadamente, cifra que sin pandemia venía incrementándose.

Si bien era de esperar que en SCOPUS las búsquedas en español mostrarían esos resultados llama la atención en general lo poco escrito en este campo en el Perú o aplicando la investigación a la realidad peruana.

Por tanto, creemos que desarrollar investigaciones aplicadas que faciliten a las empresas la aplicación de estas herramientas en el sector redundará en su mejoramiento con el colateral efecto en la eficiencia energética, en la disminución de los daños al medio ambiente y en mejora en la calidad de los suministros eléctricos es justificado.

1.8 Delimitación de la investigación

El trabajo de investigación se desarrollará bajo las siguientes delimitaciones:

1.8.1 Delimitación geográfica

El trabajo se desarrollará en el Perú incidiendo en aplicaciones aplicables a empresas del norte del Perú como zona cercana al lugar de la investigación.

1.8.2 Delimitación temporal

El desarrollo de esta propuesta se lleva a cabo desde abril de 2019 hasta marzo de 2022.

1.8.3 Delimitación temática

El presente trabajo se enfoca en la propuesta metodológica para lograr obtener información relevante para una pequeña o mediana empresa del sector eléctrico de manera tal que esta información responda cuestiones relevantes para una adecuada gestión del suministro eléctrico mejorando así sus operaciones, mantenimiento y gestión comercial.

1.9 Limitaciones de la investigación

La investigación ha tenido la principal limitación ocasionada por el periodo de su ejecución. Este periodo ha caído durante la ocurrencia de la pandemia mundial por el COVID 19, que implicó aislamientos y cuarentena que han impedido el trabajo en instalaciones de las empresas y las entrevistas presenciales.

Así mismo el tratamiento de los datos obtenidos han sido de manera off-line pues la adquisición on-line ha sido imposible. El trabajo y análisis de los casos de estudio se han realizado en gabinete, por las mismas razones ya explicadas en el párrafo anterior.





Capítulo 2

Marco teórico de la inteligencia de negocios

2.1 Grado de desarrollo de la inteligencia de negocios

En (Srivastava et al., 2022) se presenta un buen resumen de la evolución de la inteligencia de negocios o *Business Intelligence* (BI), el trabajo narra que en 1865, Richard Miller Devens acuñó por primera vez el término *Business Intelligence*. Usó la expresión '*Business Intelligence*' (BI) en su libro titulado '*Business Anecdote Cyclopedia*', y la usó para caracterizar cómo un banquero utilizaba información para tomar decisiones de su negocio, informadas e ir por delante de sus competidores. En 1958, Hans Peter Luhn compuso un artículo para un diario de IBM titulado '*A Business Intelligence Program*'. El artículo presentó las bases del BI en términos fáciles de entender. En 1960, Computer-Data Adventure Companies, con el incremento del uso de las computadoras, comenzó a acumular una enorme cantidad de datos, pero entonces no había ningún sistema de software disponible, para extraer información comercial de esos datos. En 1970, los grandes "jugadores" entraron en el mercado, IBM y Siebel se unieron a la industria moderna de BI, reconociendo la creciente necesidad de un análisis de datos rápido y preciso.

En los años 1970-1988, la racionalización de la investigación de BI bajo marcos de almacenamiento de datos y otros avances habían impulsado los marcos de administración de bases de datos (DB). A lo largo de los años 1988-1990, los sistemas y tecnologías de almacenamiento de datos llevaron a que el enfoque se ponga en los sistemas de gestión de base de datos de tal forma de agilizar el procesamiento y simplificar el análisis de BI. Para entonces, nuevos proveedores habían entrado en el mercado empresarial. En 2001, donde ocurre lo que los autores llaman la segunda ola de BI, las aplicaciones se introdujeron como software basado en la web. Estas aplicaciones permiten a los usuarios recopilar de forma independiente información útil de plataformas basadas en la Web. El software basado en la web mejoró las eficiencias generales, la visualización de los datos, la flexibilidad y la alimentación de datos en tiempo real. En 2006, Cloud BI Sellers ofreció beneficios que dieron ventajas sobre la eficiencia de costos, la capacidad, la velocidad y un acceso más simple a la información empresarial. Se utilizaron acuerdos versátiles y empoderados, ya que permiten a los clientes utilizar sus dispositivos portátiles para hacer uso de las aplicaciones de BI. En 2007, el concepto de Big Data ganó protagonismo ante la aparición de una enorme cantidad de

información desorganizada y desestructurada en los negocios. En 2010, la tecnología In-Chip® se hizo popular, mientras que las PC de 64 bits se estaban volviendo avanzadas en soluciones de BI In-Chip.

(Srivastava et al., 2022) continúa narrando que, en 2016, aparecen ya sofisticadas herramientas de BI que con un único programa se satisface todas las necesidades de las principales unidades de negocio de la empresa con resultados favorables. En 2017, la adopción de la nube continuó superando la adopción en las instalaciones y en el aprendizaje de máquina o Machine Learning (ML). Aparece la Intelligent Digital Mesh (Red digital inteligente), es decir, la interconectividad de wearables (relojes, pulseras, zapatillas inteligentes), dispositivos móviles y sensores IOT que la gente usa para encontrar información o comunicarse en línea que incluye además el conjunto expandido de puntos finales de conexión que se utilizan para acceder a las aplicaciones, recopilar información o fomentar interacciones humanas, ampliándose la BI a un nuevo horizonte de BI digital y BI Colaborativo. En 2018, la analítica aumentada y la inteligencia artificial (AI) se dispararon, mientras que la visualización de datos en la nube entra a su apogeo. En 2019, el auge del lenguaje natural de la AI humaniza los datos, los análisis de investigación, las asociaciones de datos, los códigos de ética, la gestión de datos y la narración de datos. En 2020, las tendencias en BI fueron: “Agile, Cloud, Embedded, Mobile- BI, Collaborative, Real-time Analytics, Data Quality Control, Information Discovery, Prescriptive Analytical System [...] and Community-driven data”(Srivastava et al., 2022).

A continuación mostramos lo que (Srivastava et al., 2022) considera como algunas de las últimas tendencias en BI:

Automatización de datos: sin la automatización (análisis de datos), los aspectos de BI no estarían completos. Actualmente se generan, recopilan y preparan tantos datos que las empresas están buscando seriamente nuevas tecnologías para la automatización el manejo de grandes volúmenes de información recopilada.

Gobernanza de datos: la gobernanza efectiva de los datos es una preocupación urgente para muchas organizaciones. En el Perú por ejemplo el DL N°1412 permite intuir el concepto de gobernanza de los datos como “los instrumentos técnicos y normativos que establecen los requisitos mínimos que las entidades [...] deben implementar conforme a su contexto legal, tecnológico y estratégico para asegurar un nivel básico y aceptable para la recopilación, procesamiento, publicación, almacenamiento y apertura de los datos que administren.”(PCM, 2021)

AI explicable (XAI): el sistema de BI impulsado por AI mejora la claridad de los datos. Ayuda a los usuarios/gerentes a pensar claramente desde sus áreas de especialización. Por ello la utilidad de la XAI. “Inteligencia artificial explicable (XAI) es un conjunto de procesos y métodos que permite a los usuarios humanos comprender y confiar en los resultados y la información generados por algoritmos de machine learning. La IA

explicable se utiliza para describir un modelo de IA, su impacto esperado y posibles sesgos. Ayuda a caracterizar la precisión del modelo, la imparcialidad, la transparencia y los resultados en la toma de decisiones basada en IA. La IA explicable es crucial para una organización en la generación de confianza al poner en producción los modelos de IA. La explicación de la IA también ayuda a una organización a adoptar un enfoque responsable del desarrollo de la IA.”(IBM, 2021b)

Gestión de la calidad de los datos (DQM): En (IBM, 2021a) se define DQM como el proceso que comprueba los datos en busca de valores necesarios, tipos de datos válidos y códigos válidos y que implica también corregir los datos para proporcionar valores predeterminados, formatear números y fechas y agregar nuevos códigos. La gestión de la calidad de los datos, junto con la limpieza y estandarización de nombres y la limpieza y estandarización de direcciones, está diseñada para optimizar y mejorar la calidad de los datos. DQM es un proceso crucial para garantizar la integridad, la usabilidad y la calidad de los datos.

2.2 Conceptos generales

Presentaremos a continuación algunos de los conceptos relacionados con la inteligencia de negocios. Como se verá muchos conceptos están muy relacionados entre sí y esto hace que exista: una separación muy fina en sus definiciones, en algunos casos traslape de conceptos o que algunos de los conceptos estén incluidos en otros.

Empezaremos con lo que se conoce como ciencia de datos.

2.2.1 *Ciencia de datos*

(Xu et al., 2021) define la ciencia de datos primero con una explicación sociológica de la ciencia de datos, dice que es una ciencia sobre la era de los grandes datos; su objetivo es revelar los nuevos desafíos, oportunidades, pensamientos y modos en la era de los grandes datos. Para él, se trata de un conjunto de sistemas de conocimiento compuesto por nuevas teorías, métodos, modelos, tecnologías, plataformas, herramientas, aplicaciones y mejores prácticas en la era de los grandes datos. Esta explicación tiene un significado sociológico, es compleja y carece de una descripción sustantiva de lo que debería incluir la ciencia de datos. Esta disciplina debe definirse al menos desde tres dimensiones: el objeto de investigación, la metodología y la tarea u objetivo científicos.

Sobre esta base, la ciencia de datos es la teoría y metodología básica relacionada con la realización de la cadena de valor de los datos (Ver Figura 13), utiliza el modelado, el análisis, la computación y el aprendizaje para estudiar la conversión de datos a información, de información a conocimiento y de conocimiento a la toma de decisiones, realizando la cognición y la manipulación del mundo real. (Xu et al., 2021) explica que, en la ciencia de datos, "realizar la cadena de valor de los datos" es el objeto de investigación, lo que es más preciso y específico que decir: "los datos como objeto de investigación". Para Xu, el objetivo

de la ciencia de datos es realizar la cognición y el control del mundo real, y la tarea científica es completar la transformación de datos a información, de información a conocimiento y de conocimiento a toma de decisiones en el espacio de datos, resumiendo la tarea a “tres transformaciones y una realización”. La metodología de la asignatura de ciencia de datos, continua Xu, se resume como "un híbrido de modelado, análisis, computación y aprendizaje" lo que para él constituye una innovación metodológica. Aunque Xu afirma la relevancia multidisciplinaria de la ciencia de datos; evita reducir la definición de ciencia de datos como un término general de múltiples disciplinas.

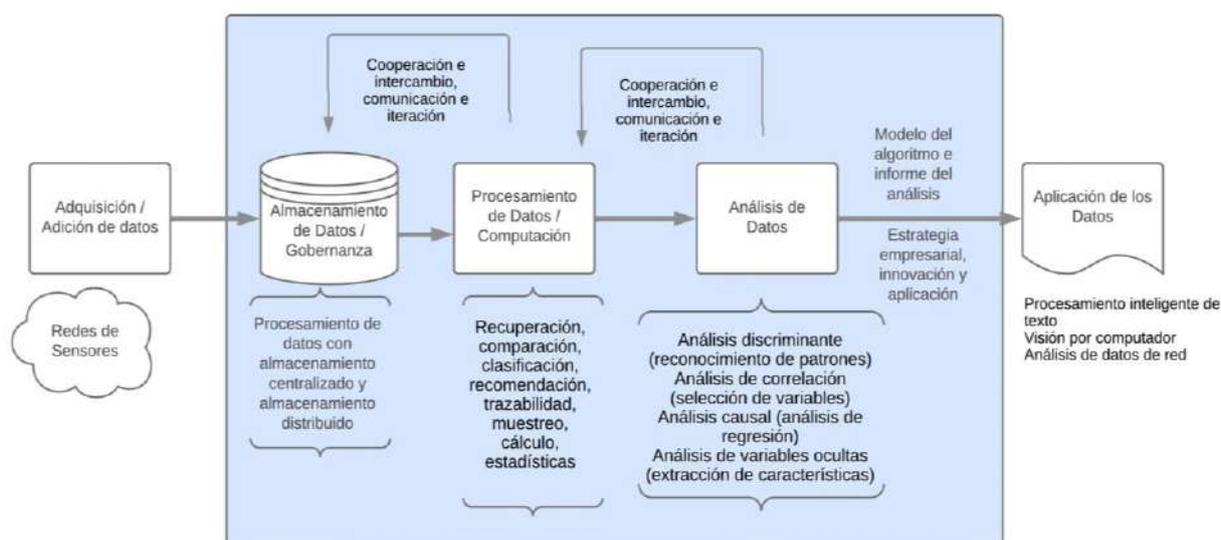
La ciencia de datos surgió en la década de 1960 y se desarrolló desde la década de 1990 hasta principios de este siglo. Sin embargo, como disciplina, se forma en la era actual de Big Data, según Xu. Para él, la etapa embrionaria fue entre 1960 y 1990 en que utiliza el término "datalogía" a esta ciencia. En 1966, Naur, científico informático danés y ganador del Premio Turing, propuso definir “la ciencia del procesamiento de datos” con el término “datalogía”, cuyo objeto de investigación son los datos digitales. La aparición del concepto de “datalogía” otorga al procesamiento de datos un estatus científico y es el origen de su concepto.

Entre 1990 y 2010, Xu considera que ocurre, la etapa de formación multiperspectiva. Naur presenta por primera vez el concepto de ciencia de datos que no se había discutido durante mucho tiempo y nota que diferentes materias utilizan diferentes términos uno tras otro y proporcionan la interpretación bajo la perspectiva de sus respectivas disciplinas.

Del 2010 en adelante Xu, considera la etapa de establecimiento de la nueva disciplina. Nota que, desde principios de este siglo, el Big Data ha promovido el rápido desarrollo de la ciencia de datos, convirtiéndola en una disciplina emergente ampliamente aceptada.

Figura 13

Cadena de valor de los datos



Fuente: Adaptado de(Xu et al., 2021)

2.2.2 Inteligencia de negocios

El término inteligencia de negocios aparece desde 1958 como veremos. En (Phillips-Wren et al., 2021) se muestran una serie de definiciones a través del tiempo, para el término inteligencia de negocios o Business Intelligence (BI). Podemos ver estos conceptos en la Tabla 3.

Podemos notar en la Tabla 3 que desde sus inicios el término se asocia a asimilar, comprender los hechos de un negocio para guiar la acción, en el 2004 entran en juego los datos como inputs para, luego de almacenarlos y gestionarlos, generar informes que ayuden a la toma de decisiones. En todos los conceptos lo común es la ayuda a la toma de decisiones y a través del tiempo se van incluyendo nuevas herramientas, métodos, procesos, tipos de análisis que apuntan hacia ese objetivo.

Tabla 3

Definiciones de Business Intelligence en el tiempo

Año	Autor	Definición
1958	Luhn	La capacidad de aprehender las interrelaciones de los hechos presentados de tal manera que guíe la acción hacia una meta deseada.
2004	Negash	Los sistemas de BI combinan la recopilación de datos, el almacenamiento de datos y la gestión del conocimiento con herramientas analíticas para presentar información interna y competitiva compleja a los planificadores y tomadores de decisiones.
2006	Davenport	El término que la gente de TI usa para los procesos y el software de análisis y generación de informes. BI abarca una amplia gama de procesos y software que se utilizan para recopilar, analizar y difundir datos, todo ello en aras de una mejor toma de decisiones.
2008	Hayen	Las aplicaciones de BI incluyen las actividades de los Sistemas de Soporte de Decisiones (DSS): consultas e informes, procesamiento analítico en línea (OLAP), análisis estadístico, pronóstico y minería de datos, siendo entonces, una actividad principal dentro de la amplia gama de implementaciones de DSS
2014	Davenport	Herramientas para apoyar decisiones basadas en datos, con énfasis en informes
2016	Larson & Chang	Gartner (2013) y Halpern (2015) han ampliado BI para que sea un término general que incluye aplicaciones, herramientas, infraestructura y prácticas para permitir el acceso y el análisis de la información para optimizar el rendimiento y la toma de decisiones.
2018	Power & Heavin	Término general que describe un conjunto de conceptos y métodos para mejorar la toma de decisiones comerciales mediante el uso de sistemas de soporte de decisiones basados en hechos; también se refiere a un conjunto de herramientas de software que se pueden utilizar para extraer y analizar datos de bases de datos corporativas
2019	Arnott, Gao, Lizama, Meredith & Song	Los sistemas de BI son sistemas a gran escala que combinan tecnologías de la información, informes de datos y procesos analíticos para respaldar la toma de decisiones en una organización.

Fuente: Adaptado de (Phillips-Wren et al., 2021)

Creemos también que una definición muy fácil de fijar y entender es la presentada (Ramesh & Ramakrishna, 2018) que la resume en: BI es un sistema que permite a los usuarios la posibilidad de convertir datos en información y completa con la definición del Data-Warehousing Institute, que resume: la inteligencia de negocios o BI incluye “herramientas, tecnologías y procesos necesarios para convertir los datos en información y la información en conocimiento y planes que optimizan las acciones comerciales”.

Finalmente (Tumpa et al., 2020) agrega: la extracción de información valiosa y el descubrimiento de patrones de datos subyacentes permite que BI tome decisiones rápidas y efectivas para diferentes organizaciones.

2.2.3 *Analítica de datos aplicada a los negocios (Business Analytics)*

Para definir primero lo que sería la analítica de datos nos referimos a (Amalina et al., 2020) que en su artículo indica que según IBM la analítica de datos sería el uso de técnicas como aprendizaje automático, análisis predictivo, minería de datos, estadísticas, análisis de texto y análisis de aprendizaje profundo en la gran cantidad de datos disponible (Big Data) para tomar buenas decisiones.

Asociando ahora la analítica de datos a los negocios, nos podemos referir al trabajo de (Radovilsky et al., 2018) que resume varias definiciones de "análisis de datos" (DA), específicamente en conjunto con "análisis de negocios" (BA), y propone que en las definiciones de DA y BA destacan cuatro aspectos clave: datos, tecnología, análisis estadístico y cuantitativo, y apoyo a la toma de decisiones. También se preocupa en diferenciarse de las definiciones de ciencia de datos (DS) que incluyen otros cuatro aspectos clave: datos, bases de datos, sistemas informáticos (tomar un conjunto de entradas, procesarlas para producir salidas) y análisis avanzados (por ejemplo, aprendizaje automático, procesamiento del lenguaje natural, inteligencia artificial), incluidas las estadísticas. Resalta entonces que lo que diferencia ambos conceptos es que DS enfatiza los sistemas informáticos, los algoritmos y las habilidades de programación de computadoras, mientras que DA/BA tiene un enfoque sustancial en el análisis estadístico y cuantitativo de datos y el apoyo a la toma de decisiones.

Para completar estos conceptos podemos observar la Tabla 4 donde se muestran algunas definiciones según el trabajo de (Phillips-Wren et al., 2021).

Observando las definiciones que se dan a BA en la Tabla 4, la diferencia con BI es imperceptible, podemos sin embargo intuir, que la BA está definida con el objetivo de utilizar la DA a la mejora del desempeño de los negocios en todos los departamentos de éstos incluida, por lo tanto, la gestión del suministro eléctrico.

Tabla 4*Definiciones de Business Analytics en el tiempo*

Año	Autor	Definición
2007	Davenport & Harris	BA se preocupa por el uso extensivo de datos, análisis estadístico y cuantitativo, modelo explicativo y predictivo, y gestión basada en hechos para impulsar decisiones y acciones.
2012	Boyd	La analítica es el proceso científico de transformar datos en información para tomar mejores decisiones.
2014	Davenport	Centrarse en el análisis estadístico y matemático para la toma de decisiones.
2014	Holsapple, Lee-Post & Pakath	Reconocimiento y resolución de problemas basados en evidencia que ocurren dentro del contexto de situaciones comerciales
2017	Seddon, Constantinidis, Tamm, & Dod	Business Analytics (BA) como el uso de datos para tomar decisiones comerciales más sólidas y basadas en evidencia.
2018	Denle & Ram	BA es el arte y la ciencia de descubrir información, mediante el uso de métodos matemáticos, estadísticos, de aprendizaje automático y de ciencia de redes sofisticados junto con una variedad de datos y conocimientos expertos, para respaldar una toma de decisiones mejor, más rápida y oportuna.

Fuente: Adaptado de (Phillips-Wren et al., 2021)

2.2.4 Inteligencia artificial

Este concepto es quizá el más amplio y el que está vigente y en pleno crecimiento actualmente. En el trabajo de (Peter Norvig, 2021) se establece el contexto cultural en el que se ha desarrollado la Inteligencia Artificial (AI), se plantea que para definirla hay que tener claro bajo que aspecto se quiere considerar la inteligencia artificial: por el pensamiento o por el comportamiento, para Norvig, de acuerdo con lo que él ha llamado el modelo estándar, la IA se ocupa principalmente de la acción racional, según la cual, un agente inteligente ideal toma la mejor acción posible en una situación, es decir, estudia el problema de construir agentes que sean inteligentes en este sentido.

(Peter Norvig, 2021) , refina dos ideas: primero, la capacidad de cualquier agente, humano o no, para elegir acciones racionales está limitada por la capacidad computacional de hacerlo; y segundo, el concepto de una máquina que persigue un objetivo definido debe ser reemplazado por el de una máquina que persigue objetivos para beneficiar al ser humano, pero sin tener claro cuáles son esos objetivos.

Resulta muy interesante mostrar el resumen que desde diferentes disciplinas del saber y precisando el aporte histórico de estas, se realiza en (Peter Norvig, 2021). Así por ejemplo indica que los filósofos (que se remontan al año 400 aC) hicieron que la IA fuera concebible al sugerir que la mente es de alguna manera como una máquina, que opera sobre el conocimiento codificado en algún lenguaje interno y que el pensamiento se puede usar para elegir qué acciones tomar.

Indica, además, que los matemáticos proporcionaron las herramientas para manipular declaraciones de certeza lógica, así como declaraciones probabilísticas inciertas y por ende sentaron las bases para comprender la computación y el razonamiento sobre los algoritmos.

Plantea también que los economistas formalizaron el problema de tomar decisiones que maximicen la utilidad esperada para el tomador de decisiones, los neurocientíficos descubrieron algunos hechos sobre cómo funciona el cerebro y las formas en que es similar y diferente de las computadoras, los psicólogos adoptaron la idea de que los humanos y los animales pueden considerarse máquinas de procesamiento de información, los lingüistas demostraron que el uso del lenguaje encaja en este modelo, los ingenieros informáticos proporcionaron las máquinas cada vez más potentes que hacen posibles las aplicaciones de IA, y los ingenieros de software las hicieron más utilizables. (Peter Norvig, 2021)

En la misma ruta, la teoría del control se ocupa de diseñar dispositivos que actúen de manera óptima sobre la base de la retroalimentación del entorno. Inicialmente, las herramientas matemáticas de la teoría del control eran bastante diferentes de las utilizadas en la IA, pero los campos se están acercando. (Peter Norvig, 2021)

Para Norvig, la IA ha madurado considerablemente en comparación con sus primeras décadas, tanto teórica como metodológicamente. A medida que los problemas con los que se ocupa la IA se volvieron más complejos, el campo pasó de la lógica booleana al razonamiento probabilístico, y del conocimiento artesanal al aprendizaje automático a partir de datos. Esto ha llevado a mejoras en las capacidades de los sistemas reales y una mayor integración con otras disciplinas.

Todos estos avances hacen, según Norvig, que a medida que los sistemas de IA encuentran aplicación en el mundo real, se ha vuelto necesario considerar una amplia gama de riesgos y consecuencias éticas. Concluyendo que a largo plazo, nos enfrentamos al difícil problema de controlar los sistemas de IA superinteligentes que pueden evolucionar de manera impredecible. Resolver este problema parece requerir un cambio en nuestra concepción de la IA, plantea (Peter Norvig, 2021).

En (Campeato, 2020) aborda el asunto de definir la AI definiendo primero lo que es inteligencia, citando a R. Sternberg, que en un texto sobre la conciencia humana, proporciona la siguiente definición útil de inteligencia: "La inteligencia es la capacidad cognitiva de un individuo para aprender de la experiencia, razonar bien, recordar información importante y hacer frente a las demandas de la vida diaria" y complementa la definición citando a Raphael quien dice: la inteligencia artificial es la ciencia de hacer que las máquinas hagan cosas que requerirían inteligencia si lo hiciera el hombre.

(Campeato, 2020) plantea que actualmente hay dos campos principales con respecto a la IA. El enfoque de IA débil que asocia con el Instituto de Tecnología de Massachusetts cuyo enfoque es cualquier sistema que exhiba un comportamiento inteligente como un ejemplo de

IA. Este campo se centra en que un programa funcione correctamente, independientemente de si el artefacto realiza su tarea de la misma manera que lo hacen los humanos. Es aquel enfoque en que los resultados de los proyectos de IA en ingeniería eléctrica, robótica y campos relacionados se refieren principalmente a un rendimiento satisfactorio.

Para Campesato, el otro enfoque de la IA, que llama: el fuerte o de integridad biológica, está asociado con la Universidad Carnegie-Mellon. De acuerdo con este enfoque, cuando un artefacto exhibe un comportamiento inteligente, su rendimiento debe basarse en las mismas metodologías utilizadas por los humanos. Pone como ejemplo, considerar un sistema capaz de oír: para los defensores de una IA fuerte, el sistema debe aspirar a lograr el éxito simulando el sistema auditivo humano (esta simulación incluiría los equivalentes a la cóclea, el canal auditivo, el tímpano y otras partes del oído, cada una realizando sus tareas requeridas en el sistema.), mientras que los para los defensores de la IA débil la preocupación debe centrarse simplemente por el rendimiento del sistema y que la razón de ser de la investigación de la IA es resolver problemas difíciles independientemente de cómo se resuelvan realmente

Un aporte importante del trabajo (Campesato, 2020) es el resumen de las partes importantes de la IA, agrupamiento muy aceptado en los trabajos de investigación actualmente. Para él la IA está constituida por:

- ML (Aprendizaje automático)
- DL (Aprendizaje Profundo)
- PNL (Procesamiento del Lenguaje Natural)
- RL (Aprendizaje por Refuerzo)
- DRL (Deep Reinforcement Learning)

A continuación, mostramos las definiciones mostradas en el trabajo de Camposato.

2.2.4.1 Aprendizaje automático o Machine Learning (ML). A mediados del siglo XX, el aprendizaje automático (como subconjunto de la IA) se basaba principalmente en los datos para optimizar y "aprender" cómo realizar tareas, a menudo acompañado de algoritmos nuevos o mejorados, como regresión lineal, k-NN, árboles de decisión, bosques aleatorios y SVM; con la excepción de la regresión lineal, todos los demás algoritmos son clasificadores.

Puede intuirse entonces, que el aprendizaje automático es un campo diverso que incluye otros subcampos.

En el ML los datos (en lugar de las reglas) son muy y función de estos, el aprendizaje automático generalmente es de uno de los siguientes tipos:

- Aprendizaje supervisado (Se cuenta muchos datos etiquetados)
- Aprendizaje semisupervisado (Se cuenta muchos datos parcialmente etiquetados)

- Aprendizaje no supervisado: Se cuenta con muchos datos no etiquetado por ende realiza agrupación en clústeres
- Aprendizaje por refuerzo: Se basa en el método: prueba, retroalimentación y mejora

Según la manera de categorizar los datos, los algoritmos de aprendizaje automático se pueden clasificar en los siguientes tipos principales:

- Clasificadores (para imágenes, spam, fraude, etc.)
- Regresión (precio de las acciones, precio de la vivienda, etc.)
- Agrupación en clústeres (clasificadores no supervisados)

2.2.4.2 Aprendizaje profundo. Es un subcampo importante del aprendizaje automático, que también tiene sus raíces a mediados del siglo XX. Las arquitecturas de aprendizaje profundo se basan en el perceptrón como base de las redes neuronales, a menudo involucrando conjuntos de datos grandes o masivos, sus arquitecturas involucran heurística y resultados empíricos. Hoy en día, el aprendizaje profundo puede superar a los humanos para alguna clasificación de imágenes.

Camposato, precisa además que mientras el aprendizaje automático involucra MLP (perceptrones multicapa), el aprendizaje profundo introduce redes neuronales profundas, con nuevos algoritmos y nuevas arquitecturas (por ejemplo, redes neuronales convolucionales, RNN y LSTM).

2.2.4.3 Aprendizaje por refuerzo. El aprendizaje por refuerzo (también un subconjunto del aprendizaje automático) implica prueba y error para maximizar una recompensa para un llamado agente. El aprendizaje por refuerzo profundo combina las fortalezas del aprendizaje profundo con el aprendizaje por refuerzo. En particular, el agente en el aprendizaje por refuerzo se reemplaza con una red neuronal.

El aprendizaje por refuerzo profundo tiene aplicaciones en muchos campos diversos, y tres de los más populares son: Juegos (Go, Ajedrez, etc.), Robótica, PNL

2.2.4.4 Procesamiento del lenguaje natural (PNL). El PNL es un área de la informática y la IA que implica la interacción entre las computadoras y los lenguajes humanos. En los primeros días, la PNL involucraba técnicas basadas en reglas o técnicas estadísticas. El PNL y el aprendizaje automático pueden procesar/analizar volúmenes de datos de lenguaje natural usando programas informáticos realizando este procesamiento.

Hay muchas tareas de PNL que se resuelven con técnicas de aprendizaje automático. Algunas áreas de interés que involucran la PNL incluyen: traducción entre idiomas, encontrar información significativa a partir del texto, resumen de documentos, detección del discurso de odio, etc.

Sobre IA, otro trabajo interesante es el de (Zohuri, 2020), en su trabajo se presenta también una IA que denomina débil y que está diseñada para realizar tareas específicas (por ejemplo, solo reconocimiento facial o solo búsquedas en Internet o solo conducir un automóvil). Sin embargo, se reconoce que el objetivo a largo plazo de muchos investigadores es crear IA más general o como se denomina en el trabajo, más fuerte. Esta IA más fuerte superaría a los humanos en casi todas las tareas cognitivas a diferencia de la IA débil que solo puede superar la inteligencia humana en tareas específicas.

En (Zohuri, 2020), se sigue explicando que una de las grandes ventajas de la Inteligencia Artificial es aprender de la experiencia, es decir, que las máquinas tengan la capacidad de adaptarse a nuevas entradas y realicen tareas similares a las humanas. Plantea también que, la mayoría de los ejemplos de IA de los que se oye hablar hoy en día, desde computadoras que juegan al ajedrez hasta automóviles autónomos, dependen en gran medida del aprendizaje profundo y la PNL (procesamiento del lenguaje natural) con estas, las computadoras pueden ser entrenadas para realizar tareas específicas procesando grandes cantidades de datos y reconociendo patrones en los datos.

Además (Zohuri, 2020), plantea que la IA desempeñará un rol importante en la vida moderna mencionando las siguientes características y particularidades importantes:

La IA automatiza el aprendizaje repetitivo y el descubrimiento a través de los datos. En este caso la IA realiza tareas frecuentes y de alto volumen, así como tareas automatizadas de manera confiable y sin agotamientos. Y se resalta que, para este tipo de automatización, la investigación humana sigue siendo esencial para configurar el sistema y definir las preguntas correctas y objetivos correctos.

La IA agrega inteligencia a los productos existentes. Es decir, la IA no se desarrollará como productos independientes, sino que agregará prestaciones a productos y servicios existentes como el caso de SIRI en Apple o Alexa de Amazon a esto se agregan las plataformas conversacionales, los “bots” y máquinas inteligentes que utilizan los datos disponibles para generar soluciones para el hogar y la industria en campos como la seguridad y análisis de inversiones, por ejemplo.

La IA se adapta a través de algoritmos de aprendizaje progresivo para permitir que los datos hagan la programación. En la IA su algoritmo encuentra patrones en los datos de tal forma que finalmente se convierte en un clasificador o un predictor. Este modelo se adapta cuando se le ingresan o adquiere más datos, así el algoritmo puede enseñarse a sí mismo. En este proceso la retropropagación es una técnica de IA que permite que el modelo se ajuste, a través del entrenamiento y los datos agregados, cuando la primera respuesta no es la esperada.

La IA analiza cada vez más datos y de manera más profunda utilizando redes neuronales que tienen muchas capas ocultas. Esto se ha logrado con el aumento de la potencia

de cómputo y la Big Data. Esto aumenta el requerimiento de datos pues se necesitan muchos datos para entrenar los modelos de aprendizaje profundo debido a que aprenden directamente de los datos. Cuantos más datos se puedan ingresar al algoritmo, en tiempo real o al menos casi en tiempo real, más precisos se volverán estos algoritmos.

La IA logra una precisión increíble a través de redes neuronales profundas, lo que antes era imposible. Por ejemplo, las interacciones con Alexa, las búsquedas de Google y Google Fotos se basan en el aprendizaje profundo, son más precisas cuanto más las usamos. Otro ejemplo es en el campo médico, donde la clasificación de imágenes y reconocimiento de objetos ahora se pueden usar para detectar cáncer en resonancias magnéticas con la misma precisión que los radiólogos altamente capacitados.

La IA saca el máximo provecho de los datos. Cuando los algoritmos son de autoaprendizaje, los datos en sí mismos pueden convertirse en propiedad intelectual. Cómo las respuestas están en los datos; con solo aplicar IA obtendríamos las respuestas buscadas. Esto indica que el rol de los datos es ahora más importante que nunca y por tanto puede crear una ventaja competitiva: quien tiene los mejores datos tendrán enormes ventajas en una industria competitiva, incluso si todos están aplicando técnicas similares.

Como se puede observar en las declaraciones anteriores, la IA ya está integrada a nuestra vida moderna. Ayuda al ser humano en la industria, banca, economía, seguridad nacional, gobierno, etc., donde están capitalizando los avances en Inteligencia Artificial (IA), así como sus otros componentes, como el Aprendizaje Automático (ML) que es impulsado por el Aprendizaje Profundo (DL).

Tal como se ha expuesto la AI y ML están trabajando mano a mano junto con DL para tratar datos históricos, así como de nuevos datos entrantes a nivel de diferentes fuentes generando los que se denomina Big Data (BD) que explicaremos a continuación.

2.2.5 Big data e internet de las cosas

Hemos agrupado estos dos conceptos teniendo en cuenta que la Internet de las Cosas (IoT) es un gran generador de datos que en muchos casos implican o forman parte de lo que se conoce como Big Data.

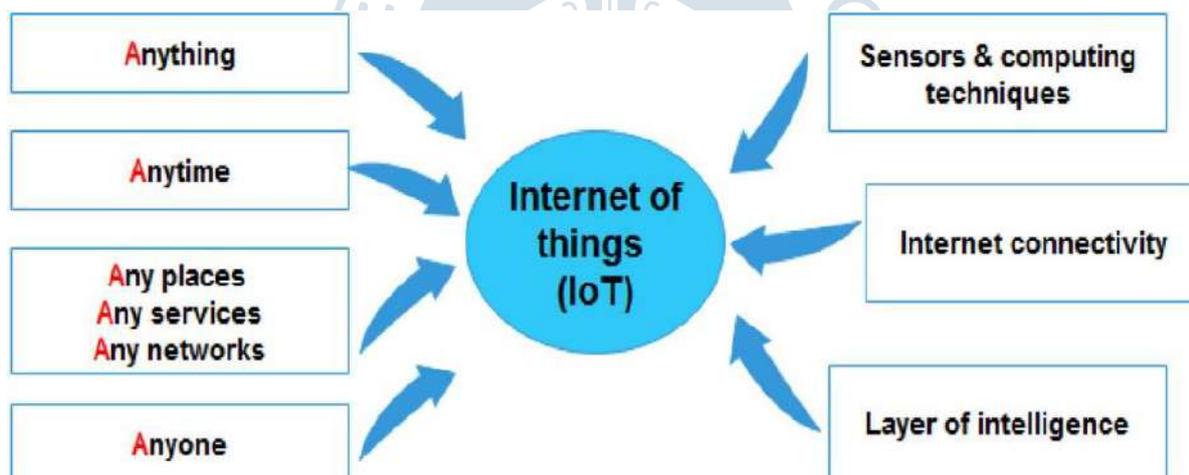
En (Hajjaji et al., 2021), el autor plantea que los conceptos fundamentales de IoT fueron inventados por Kevin Ashton en 1999, cuando introdujo la comunicación de dispositivo a dispositivo en una escala más amplia de lo que era posible anteriormente. Según él y citando a Atzori y otros, manifiesta que "IoT es el resultado de la convergencia de tres visiones: orientada a las cosas, orientada a Internet y orientada a la semántica". Específicamente en términos de semántica, IoT es una "red mundial de objetos interconectados". Añade además que IoT se puede definir como "una infraestructura de red global dinámica y como tal, puede identificar, controlar y monitorear cada objeto en la tierra a través de Internet de acuerdo con

un protocolo de acuerdo específico, y a través de la interconexión de cosas físicas y virtuales basadas en la interoperabilidad de las tecnologías de la información y la comunicación”.

(Hajjaji et al., 2021) indica también que el objetivo principal de IoT es ayudar a compartir información en tiempo real a través de actores autónomos en red. Los autores explican mediante la Figura 14, como es que un sensor con capacidades computacionales inteligentes se coloca en un lugar que acceso a una conexión a Internet. Este sensor podrá comunicarse con cualquier otro dispositivo con similares capacidades, en cualquier momento y desde cualquier lugar dentro de la red. Los sistemas de recopilación de datos ubican y transfieren datos a través de una gran cantidad de dispositivos de comunicación dentro de la infraestructura de IoT, lo que facilita el proceso de recopilación de datos. Varias soluciones de comunicación, como WIFI, ZigBee, Bluetooth y GSM, permiten la interconexión de dispositivos que utilizan varias redes de acceso, incluida la identificación por radiofrecuencia (RFID), dispositivos con sensores inalámbricos y cualquier objeto inteligente conectado a Internet sobre una red física.

Figura 14

Concepto de internet de las cosas



Fuente : (Hajjaji et al., 2021)

La realidad indica que hoy en día, IoT genera enormes volúmenes de datos, Estos datos, a menudo denominados "Big Data", se refieren a datos a gran escala que exigen nuevas arquitecturas y tecnologías para la gestión de datos (captura y procesamiento) para permitir la extracción de valor para una mejor comprensión y toma de decisiones.(Hajjaji et al., 2021)

Por otro lado, la Big Data se reconoce por sus características y por unas propiedades que los trabajos de investigación reconocen, como son: alto volumen, alta velocidad, alta variedad y alta veracidad.

Se espera este gran volumen de datos por la proyección de cantidad de dispositivos conectados a la IoT. Respecto a esta proyección, aunque se esperaba que para el 2020 el

número de equipos conectados bordee los 50 000 millones o más de dispositivos, la realidad dice que al 2021 según (Sinha, 2021) hay 12 300 millones de dispositivos conectados y la expectativa al 2025 es que se llegue a 27 000 millones. Aun cuando la diferencia con la proyección inicial es considerable de todas maneras el volumen de datos esperado sigue siendo alto.

Según el mismo trabajo de (Hajjaji et al., 2021) recientemente, IoT se ha aplicado en entornos inteligentes, lo que permite a los usuarios comprender y controlar mejor su entorno a través de una gama de dispositivos interconectados. En las aplicaciones de entornos inteligentes, IoT se emplea para construir una red de monitoreo ecológico integral, de varios niveles y completamente cubierta, que se puede lograr mediante la integración de sensores en todos los niveles aprovechando IoT con información espacial y temporal, y construyendo una plataforma masiva con un centro de datos y soporte de servicio unificado. La tecnología IoT y su integración con big data se han aplicado ampliamente en diversos campos, como ciudades inteligentes, atención médica inteligente, sistemas de alerta inteligentes y gestión de desastres. Por lo tanto, la construcción y aplicación de IoT y Big Data en campos ambientales se ha convertido en una medida crucial, particularmente para el desarrollo, promoción y gestión de un nuevo entorno estratégico en la industria.

Aunque se ha logrado un progreso significativo en IoT para aplicaciones ambientales inteligentes, aún sigue siendo un reto capturar, almacenar, procesar y analizar los grandes volúmenes de datos con técnicas de inteligencia artificial. Para lograr los resultados deseados, las aplicaciones de entornos inteligentes requieren un procesamiento en tiempo real o casi en tiempo real sobre flujos de datos de alta velocidad, lo que a su vez requiere un procesamiento rápido, eficiente y a gran escala.

Casi en la misma línea (Tang et al., 2022) indica que con el auge de las técnicas de internet y la informática, ha surgido y se ha acumulado una variedad de datos en formatos no estructurados o semi-estructurados, formando la Big Data que permite describir al mundo desde diferentes perspectivas. Para este año el autor aun considera que no existe una definición uniforme para la Big Data todavía, pero al igual que la mayoría de los investigadores, caracteriza la Big Data por las 5V, es decir, volumen, variedad, velocidad, valor y veracidad. Precisa que, en las últimas décadas, la Big Data se ha introducido en diversos campos de investigación, trayendo innovaciones alentadoras a las teorías y tecnologías asociadas. Por un lado, los macrodatos informativos han proporcionado nueva información y conocimientos, respaldando una nueva o mejor comprensión de los problemas específicos, desafiando e incluso remodelando las teorías básicas que se basaban en los datos tradicionales. Por otro lado, en un formato diferente (no estructurado o semi-estructurado), esta Big Data ha requerido una renovación sustancial de las técnicas de procesamiento y análisis e incluso han dado lugar a nuevas metodologías, particularmente en el campo de la informática.

2.2.6 Resumen

Podemos observar que los conceptos definidos en este capítulo aun cuando antiguos tienen definiciones planteadas según cada investigador o trabajo. Hemos tratado de utilizar aquellas que más pueden ser aplicadas a nuestro objetivo, la pequeña y mediana empresa.

Nos quedamos por tanto con los siguientes conceptos:

Para ciencia de datos nos quedamos con su metodología y tarea. Es decir, es la ciencia en la que el modelado de los datos, el análisis de los mismos, la computación y el aprendizaje realiza la tarea científica de transformar estos datos en información, la información en conocimiento y del conocimiento se pase a la toma de decisiones.

Para el BI nos quedamos con la generalización planteada en (Ramesh and Ramakrishna, 2018) que resume: BI incluye “herramientas, tecnologías y procesos necesarios para convertir los datos en información y la información en conocimiento y planes que optimizan las acciones comerciales”.

Para la analítica de datos hemos orientado nuestras definiciones hacia la analítica de datos aplicada a los negocios o business analítics (BA), desde este punto de vista nos quedamos con la definición de: arte y ciencia de descubrir información, mediante el uso de métodos matemáticos, estadísticos, de aprendizaje automático y de ciencia de redes sofisticados junto con una variedad de datos y conocimientos expertos, para respaldar una toma de decisiones mejor, más rápida y oportuna.

Inteligencia Artificial es quizá la definición más difícil de resumir por todo lo que implica. Para nuestros fines la definimos como la disciplina integrada por el machine learning (ML), aprendizaje profundo (DL), el procesamiento del lenguaje natural, el aprendizaje por refuerzo, el aprendizaje por refuerzo profundo que intenta hacer que las máquinas hagan cosas que requerirían inteligencia si las hiciera el hombre. Teniendo en cuenta que inteligencia es la capacidad cognitiva de aprender de la experiencia, razonar bien, recordar información importante y hacer frente a nuevas situaciones.

Los conceptos de Big Data e Internet de las Cosas tienen una relación por el hecho que la IoT como red mundial de objetos interconectados con el fin de compartir información en tiempo real a través de actores autónomos en red, genera un gran volumen de datos con características 5V gran volumen, variedad grande, velocidad elevada, con información valiosa y veraz que es la característica principal del Big Data.



Capítulo 3

Marco metodológico

3.1 Investigación cuantitativa e investigación cualitativa

Para explicar las diferencias entre investigación cuantitativa y cualitativa creemos importante definir primero lo que es investigación. Para ellos nos apoyamos en la siguiente definición: “conjunto de procesos sistemáticos, críticos y empíricos que se aplican al estudio de un fenómeno o problema con el resultado (o el objetivo) de ampliar su conocimiento. Esta concepción se aplica por igual a los enfoques cuantitativo, cualitativo y mixto”(Sampieri, 2018).

El mismo autor en (Sampieri, 2018), plantea que, refiriéndose a los enfoques cuantitativos, cualitativos o mixtos, “ninguna ruta es mejor que otra, sino solo más apropiada para llegar al lugar que quieres (el que pretendes indagar, el problema de investigación) y todas requieren de diferentes herramientas (que son los métodos de investigación) y un mapa (el diseño de investigación). Además, la ruta a seleccionar depende de tus conocimientos y el entrenamiento que hayas recibido.”

Asimismo, (Sampieri, 2018) indica que la distinción entre la investigación cuantitativa y cualitativa es relativa y que en realidad las rutas se entrelazan y comparten ciertas cuestiones, es decir, no son caminos completamente independientes.

Los tres enfoques utilizan procesos sistemáticos, reflexivos y empíricos en su esfuerzo de generar conocimiento, valiéndose de las siguientes estrategias:

1. Observación y evaluación de fenómenos.
2. Establecimiento de suposiciones como consecuencia de la observación y evaluación.
3. Demostración del grado en que las suposiciones tienen o no fundamento o son ciertas en determinado contexto, mediante análisis y pruebas.
4. Proponer nuevas observaciones y evaluaciones para consolidar, esclarecer o modificar las suposiciones; o incluso para generar otras.

3.1.1 La investigación cuantitativa

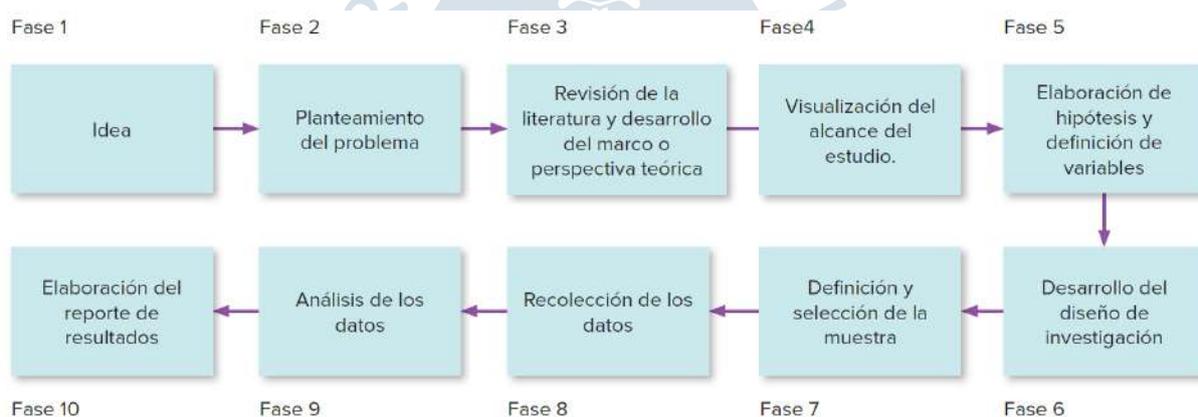
(Sampieri, 2018) explica muy bien lo que implica lo que él llama la ruta cuantitativa:

“El significado original del término cuantitativo (del latín “quantitas”) se vincula a conteos numéricos y métodos matemáticos (Niglas, 2010).³ Actualmente, representa un conjunto de procesos organizado de manera secuencial para comprobar ciertas suposiciones. Cada fase precede a la siguiente y no podemos eludir pasos, el orden es riguroso, aunque desde luego, podemos redefinir alguna etapa. Parte de una idea que se delimita y, una vez acotada, se generan objetivos y preguntas de investigación, se revisa la literatura y se construye un marco o perspectiva teórica. De las preguntas se derivan hipótesis y determinan y definen variables; se traza un plan para probar las primeras (diseño, que es como “el mapa de la ruta”); se seleccionan casos o unidades para medir en estas las variables en un contexto específico (lugar y tiempo); se analizan y vinculan las mediciones obtenidas (utilizando métodos estadísticos), y se extrae una serie de conclusiones respecto de la o las hipótesis.”

Este proceso se puede observar en la Figura 15

Figura 15

La investigación cuantitativa



Fuente: (Sampieri, 2018)

3.1.2 *La investigación cualitativa*

Igualmente (Sampieri, 2018) explica muy bien lo que constituye una investigación cuantitativa.

(Sampieri, 2018) recuerda lo ya definido por Niglas, 2010 acerca que el término cualitativo tiene su origen en el latín “qualitas”, el cual hace referencia a la naturaleza, carácter y propiedades de los fenómenos.

“Bajo el enfoque cualitativo se estudian fenómenos de manera sistemática. Sin embargo, en lugar de comenzar con una teoría y luego “voltear” al mundo empírico para confirmar si esta es apoyada por los datos y resultados, el investigador comienza el proceso examinando los hechos en sí y revisado los estudios previos, ambas acciones de manera simultánea, a fin de generar una teoría que sea consistente con lo que está

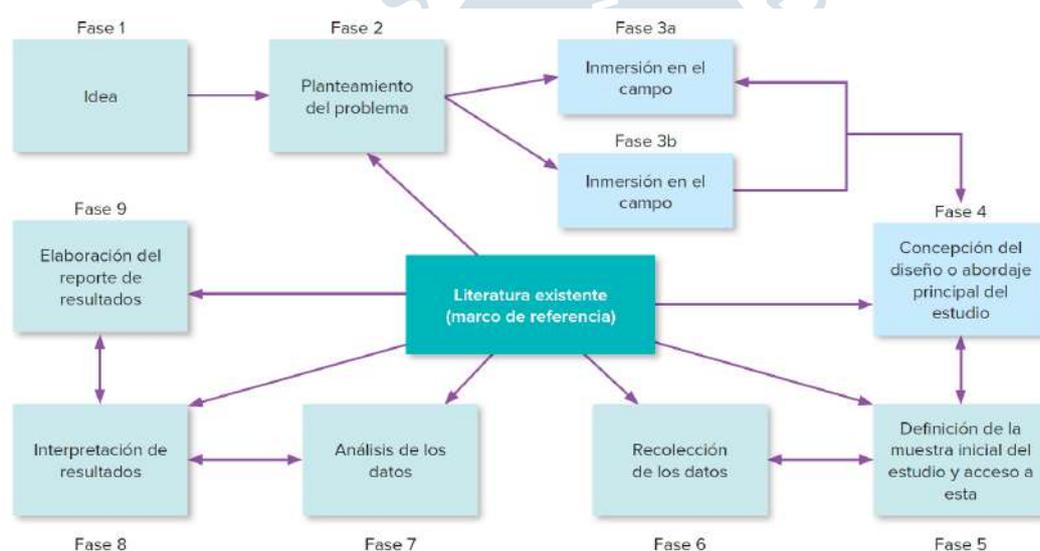
observando que ocurre. [...]De igual forma, se plantea un problema de investigación, pero normalmente no es tan específico como en la indagación cuantitativa. Va enfocándose paulatinamente. La ruta se va descubriendo o construyendo de acuerdo al contexto y los eventos que ocurren conforme se desarrolla el estudio.”(Sampieri, 2018)

(Sampieri, 2018) completa su definición indicando que “las investigaciones cualitativas suelen producir preguntas antes, durante o después de la recolección y análisis de los datos. La acción indagatoria se mueve de manera dinámica entre los hechos y su interpretación, y resulta un proceso más bien “circular” en el que la secuencia no siempre es la misma, puede variar en cada estudio.”

La Figura 16 muestra el proceso de manera referencial pues en la realidad este proceso es flexible.

Figura 16

La investigación cualitativa



Fuente: (Sampieri, 2018)

La Figura 16 muestra que en la investigación cualitativa la revisión de la literatura es permanente y que algunas etapas son iterativas, es decir, según sus resultados puede ser necesario volver a etapas previas; además las fases de muestra, recolección y análisis se realizan en forma simultánea y van influyéndose entre sí. Para(Sampieri, 2018) esta investigación implica “sensibilizarse con el ambiente o entorno en el cual se llevará a cabo el estudio, identificar informantes que aporten datos y guíen al investigador por el lugar, adentrarse y compenetrarse con la situación de investigación, además de verificar la factibilidad del estudio.”

Dada la característica de nuestro estudio mostraría que este tipo de investigación sería la que más se ajusta a los objetivos que pretendemos alcanzar. Uno de los tipos de

investigación cualitativa para ciertos autores, como en (Hancock, 2006), es la investigación por estudio de casos.

En (Yin, 2018) se presenta una perspectiva opuesta, sugiere que “la investigación de estudios de casos puede estar separada de la investigación cualitativa. Los estudios de caso pueden necesitar seguir sus propios procedimientos de investigación personalizados, como para identificar y definir el caso a estudiar, junto con muchos otros procedimientos” y plantea también esta cuestión como un tema de investigación pendiente a la luz de diferentes disciplinas.

Nosotros enfocaremos nuestra investigación bajo el enfoque cualitativo respaldado por la rigurosidad que le reconoce Yin, al método por estudios de casos.

3.2 El método de estudio de casos como estrategia de investigación

Para (Hancock, 2006), los estudios de casos se diferencian de otros tipos en que son análisis intensivos y descripciones de una sola unidad o sistema delimitado por el espacio y el tiempo. Los temas que a menudo se examinan en los estudios de casos incluyen individuos, eventos o grupos. A través de estudios de casos, los investigadores esperan obtener una comprensión profunda de las situaciones y el significado para los involucrados. Sugiere que las ideas obtenidas de los estudios de casos pueden influir directamente en las políticas, los procedimientos y la investigación futura. Ver Tabla 5

Aunque los estudios de casos se analizan extensamente en la literatura y se emplean con frecuencia en la práctica, se ha escrito poco sobre los pasos específicos que se pueden seguir para planificar, realizar y compartir con éxito los resultados de un proyecto de estudio de casos.

Tabla 5

Diferencias entre los métodos Cualitativos - Cuantitativos y estudios de caso

Estudios Cuantitativos	Estudios Cualitativos	Casos de estudio
El investigador identifica el tema o la(s) pregunta(s) de interés y selecciona a los participantes y organiza procedimientos que proporcionan respuestas que se aceptan con un grado de confianza predeterminado; Las preguntas de investigación a menudo se expresan en hipótesis que se aceptan o rechazan mediante pruebas y análisis estadísticos.	El investigador identifica el tema o la(s) pregunta(s) de interés; recopila información de diversas fuentes, a menudo como observador participante; y acepta la tarea analítica como una de descubrir respuestas que surgen de la información disponible como resultado del estudio.	La investigación identifica el tema o la(s) pregunta(s) de interés, determina la unidad apropiada para representarlo y define lo que se sabe sobre la base de un análisis cuidadoso de múltiples fuentes de información sobre el "caso".
El proceso de investigación puede variar mucho desde el contexto que se está investigando.	El proceso de investigación está diseñado para reflejar, en la medida de lo posible, el contexto natural y continuo que se está investigando; la información a menudo es recopilada por observadores participantes.	El proceso de investigación se define por una serie sistemática de pasos diseñados para proporcionar un análisis cuidadoso del caso.

Estudios Cuantitativos	Estudios Cualitativos	Casos de estudio
La recopilación de información puede durar unas pocas horas o unos días, pero generalmente es de corta duración utilizando medidas cuidadosamente construidas diseñadas específicamente para generar información válida y confiable en las condiciones del estudio.	La recopilación de información puede durar unos meses o el tiempo que sea necesario para que surja una respuesta adecuada; el marco de tiempo para el estudio a menudo no está definido en el momento en que se lleva a cabo la investigación.	La recopilación de información puede durar unas pocas horas, unos días, unos meses o el tiempo que sea necesario para "definir" adecuadamente el caso.
El informe de los resultados del proceso es generalmente expositivo, que consiste en una serie de respuestas estadísticas a las preguntas bajo investigación.	El informe de los resultados del proceso es generalmente narrativo, que consiste en una serie de "páginas de la historia" o "capítulos del libro".	El informe de los resultados del proceso es generalmente de naturaleza narrativa, que consiste en una serie de descripciones ilustrativas de aspectos clave del caso.

Fuente: (Hancock, 2006)

En (Hancock, 2006) y basado en la información disponible en esa fecha, la investigación de estudios de caso representa una visión mucho más amplia. Significa realizar una investigación empírica de un fenómeno contemporáneo dentro de su contexto natural utilizando múltiples fuentes de evidencia, concepto obtenido del trabajo de Yin, 2003. Los temas de investigación de estudios de caso varían ampliamente.

Hancock presente tres aspectos a tener en cuenta en un estudio de caso:

Primero, aunque la investigación de estudios de casos a veces se centra en un representante individual de un grupo, más a menudo aborda un fenómeno (por ejemplo, un evento, situación, programa o actividad en particular). Por ejemplo, un especialista en tecnología puede buscar una mayor comprensión de los procesos de toma de decisiones que influyen en la adopción de programas de software en su organización (actividad). Estos fenómenos representan el foco de la mayoría de los estudios de casos, pero no son mutuamente excluyentes.

En segundo lugar, el fenómeno que se está investigando se estudia en su contexto natural, limitado por el espacio y el tiempo. El especialista en tecnología restringirá su investigación de los procedimientos de adopción de software a las prácticas de su propia organización desde la compra de un mainframe informático. Claramente, el contexto es importante en la investigación de estudios de casos, y sus beneficios son la fortaleza de hacer investigaciones intensivas de individuos o grupos, así como eventos, situaciones, programas, actividades y otros fenómenos de interés.

En tercer lugar, la investigación de estudios de casos es ricamente descriptiva, porque se basa en fuentes de información profundas y variadas. El especialista en tecnología cita ejemplos de prácticas existentes que influyen en la toma de decisiones organizacionales con

respecto a la adopción de paquetes de software. Por lo tanto, la información se explora y extrae en el entorno del estudio de caso para un examen más exhaustivo del fenómeno dado.

Las similitudes y diferencias adicionales que a veces se encuentran en otras formas de investigación también caracterizan la investigación de estudios de casos. Por ejemplo, en contraste con la investigación experimental, la investigación de estudios de casos es generalmente más exploratoria que confirmatoria; es decir, el investigador del estudio de caso normalmente busca identificar temas o categorías de comportamiento y eventos en lugar de probar relaciones o probar hipótesis. Debido a que implica recopilar y analizar información de múltiples fuentes, como entrevistas, observaciones y documentos existentes, la investigación de estudios de casos a veces requiere que el investigador pase más tiempo en el entorno que se investiga que en el caso de otros tipos de investigación. Finalmente, como con la mayoría de las investigaciones, hacer estudios de casos crea oportunidades para que el investigador explore preguntas adicionales por el acto de investigar un tema en detalle.

Hacer una investigación de estudio de caso significa identificar un tema que se presta a un análisis en profundidad en un contexto natural utilizando múltiples fuentes de información. Una vez que se ha establecido el escenario, debemos determinar lo que se sabe y lo que no se sabe sobre el tema para crear una pregunta de investigación importante. El propósito al volver a ver la literatura es establecer la base conceptual para el estudio, definir y establecer la importancia de la pregunta de investigación, identificar las fortalezas y debilidades de los modelos y diseños que otros han utilizado para estudiarla, e identificar el estilo y la forma utilizados por los expertos para ampliar la base de conocimientos que rodea la pregunta.

Debemos indicar que en (Hancock, 2006) se presenta el método con un enfoque más orientado a estudios sociales. Sin embargo en (Yin, 2018), sexta edición de un muy citado libro sobre la investigación por casos de estudio presenta una interesante trilogía que hace factible el uso de esta metodología como un método serio de investigación y la separa de los casos de estudios populares o casos de estudio para enseñanza. Esta trilogía separa:

- Investigación de estudio de caso (el modo de investigación),
- Estudios de casos (el método de investigación, o método de investigación utilizado para hacer investigación de estudios de casos), y
- Caso(s) (la unidad habitual de investigación en un estudio de caso).

Pone analogías que explican esta trilogía. Por ejemplo, investigación experimental (modo), experimentos (método) y temas (unidades); o investigación de encuestas (modo), encuestas (método) y encuestados (unidades); y posiblemente más especulativamente, investigación estadística (modo), modelado estadístico (método) y variables (unidades).

De esta manera, el presente trabajo plantea usar este marco de trabajo pensado y utilizado principalmente en ciencias sociales a una aplicación tecnológica basándonos en el

hecho que una investigación de estudio de caso, en comparación con otros métodos puede aplicarse cuando (1) sus principales preguntas de investigación son preguntas de "cómo" o "por qué", (2) tiene poco o ningún control sobre los eventos de comportamiento, y (3) su enfoque de estudio es un fenómeno contemporáneo (en lugar de completamente histórico), un "caso".

Vemos también que para el caso de ingeniería viene siendo muy utilizado también. Por ejemplo, para los años 2020, 2021 y 2022 en la base de artículos y libros de *Science Direct* existen más de 172,000 trabajos con la búsqueda "Case study" filtrada por la subárea "Engineering", para el caso de la base de artículos IEEE Xplore para los mismos años muestra más de 27,000 trabajos entre libros, artículos para revistas y conferencias. Estos números muestran que el uso de casos de estudio está siendo una estrategia muy útil en trabajos de ingeniería también.

3.3 Selección de los casos de estudio

Tal como se hizo (Ingemarsdotter et al., 2021) realizaremos el estudio de tres casos alineados con los objetivos de la investigación. En cada uno de los casos estudiados, estaremos interesados en los desafíos y soluciones relacionados con el desarrollo y la implementación de marcos de trabajo de BI aplicados a la gestión óptima del suministro eléctrico en una pequeña y mediana empresa. A continuación, describimos cómo se seleccionaron los casos y cómo se recopilaban y analizaban los datos.

Para ser considerado un caso relevante para este estudio, se debían cumplir los siguientes criterios.

- Los datos para analizar provienen de una empresa del sector eléctrico: generador, distribuidor o consumidor, perteneciente a la pequeña o mediana empresa peruana.
- La información por descubrir debe permitir la mejora de un aspecto importante de sus operaciones, mantenimiento y/o facturación comercial.
- Deben disponer de un mínimo de información accesible en algún formato posible de tratar con herramientas informáticas.

3.4 Recopilación de datos

La recopilación de los datos ha sido ejecutada mediante el contacto directo con los propietarios de la información, la misma ha sido remitida en formatos fuente ya sea de los equipos que la registraron o desde los reportes disponibles por los operadores o encargados de la gestión de la información.

Se realizaron entrevistas semiestructuradas e informales con representantes de las empresas y con los usuarios de los datos y se identificó las posibles aplicaciones a desarrollar.

3.5 Análisis de los datos

Para la analítica de los datos se utilizó de preferencia software libre o de fácil licenciamiento, de bajo costo y de uso común en la mayoría de las pequeñas y medianas empresas.

- Básicamente los archivos fuentes han sido archivos Excel, .CSV o .TXT.
- Para el preprocesamiento se ha utilizado las aplicaciones: Excel, Power BI y R .

Los despliegues se han mostrado en computador local, aunque con posibilidades de escalar a la internet con licenciamientos relativamente económicos en el caso de Power BI.

3.6 Desarrollo del marco

Para analizar los casos, resolver la pregunta de “cómo implementar una solución de BI&A” en cada caso se investigó las alternativas disponibles para los tres casos y se llegó a la conclusión que el modelo de referencia CRISP-DM es un modelo que puede ser aplicado en las empresas investigadas y puede ser generalizado a otras empresas con relativa facilidad.

De la misma manera a lo realizado en (Bordeleau et al., 2020) nos encargamos de seleccionar casos representativos para poder implementar de manera empírica, soluciones que cumplan un objetivo útil en la empresa objeto de cada estudio. Aunque el número de estudios es pequeño y por ende no permitiría una generalización amplia; sin embargo, este estudio exploratorio tiene igualmente, como objetivo una generalización lógica y contextual más que estadística.

3.7 Descripción de los casos

Los casos de estudio han sido elegidos para aplicar tres campos del *data analytics* en auge actualmente: la clasificación, el reconocimiento de patrones y la predicción o pronóstico.

3.7.1 Caso 1: Calidad de energía y clasificación

Paper: “Use of a business intelligence framework in the management of the quality of electricity supply in small and medium-sized companies. “

Resumen:

El objetivo de este estudio es presentar una metodología basada en inteligencia de negocios en pequeñas y medianas empresas. Se seleccionaron y evaluaron tres métodos, y se utilizó como referencia el proceso estándar intersectorial para la minería de datos (CRISP-DM). La metodología se aplicó a un caso de estudio real: una empresa agroindustrial, utilizando una aplicación comercial de inteligencia de negocios, para encontrar medidas correctivas para mejorar la gestión del suministro eléctrico de la empresa. Con base en los resultados de este estudio, se puede concluir que con herramientas computacionales de bajo costo es posible visualizar y controlar parámetros del suministro eléctrico, agrupar el consumo de energía reactiva para una adecuada selección de pasos en un banco de capacitores. Además, con unos

conocimientos mínimos de modelización, análisis de datos y del sector eléctrico, estas herramientas no solo se aplican en materia comercial sino también en la gestión del suministro eléctrico.

3.7.2 Caso 2: Análisis de patrones de operación

Paper: “Aplicación de técnicas de minería de datos para determinar un modelo de operación normal de un turbogenerador hidráulico”

Resumen:

Este artículo presenta la aplicación de un modelo de referencia para la implementación de minería de datos como parte de un entorno de inteligencia de negocios con el objetivo de analizar la operación de un turbogenerador hidráulico. El análisis parte de los datos provenientes del sistema de control y adquisición de datos de una central hidroeléctrica, con ellos se generó un modelo utilizando análisis de componentes principales, luego se evaluó el modelo con nuevas observaciones utilizando los gráficos de control estadístico multivariante cuadrado del error de predicción y T2 de Hotelling con los que se implementó el análisis conectando las tablas generadas en una herramienta convencional de inteligencia de negocios. El modelo se probó en 5 muestras correspondientes a periodos previos a una parada del turbogenerador. Concluimos que el modelo de referencia utilizado puede ser aplicado al análisis de la operación de estos equipos, que los gráficos de control son capaces de detectar algunos comportamientos de operación y que es necesario el conocimiento del proceso para una adecuada interpretación de los resultados.

3.7.3 Caso 3: Predicción de producción

Paper: “Predicción de energía hidroeléctrica en una empresa peruana utilizando una plataforma de inteligencia de negocios”

Resume:

El presente trabajo presenta un método para aplicar una plataforma basada en *Business Intelligence* en la predicción de la producción de la energía mensual de una central hidroeléctrica en una empresa peruana. Para ello utilizamos como referencia la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining) a la que le hemos agregado un paso adicional. Aplicamos la metodología al central objetivo con información desde el año 1998 hasta el año 2017 y proyectamos la producción de energía mensual del año 2018. Mediante la implementación de modelos embebidos en *Power BI*: *Smooth Exponential*, ARIMA y Neural Network obtenemos unas proyecciones que comparamos con la producción real evaluando el Mean Square Error. Finalmente comparamos todos los modelos incluida la actual forma de calcular la proyección obteniendo que el modelo con *Smooth Exponential* tiene el mejor performance.

Los casos de estudio tal como han sido presentados pueden ser observados en los Apéndices de este trabajo.



Capítulo 4

Propuesta metodológica aplicada a una PYME del sector eléctrico peruano

4.1 Introducción

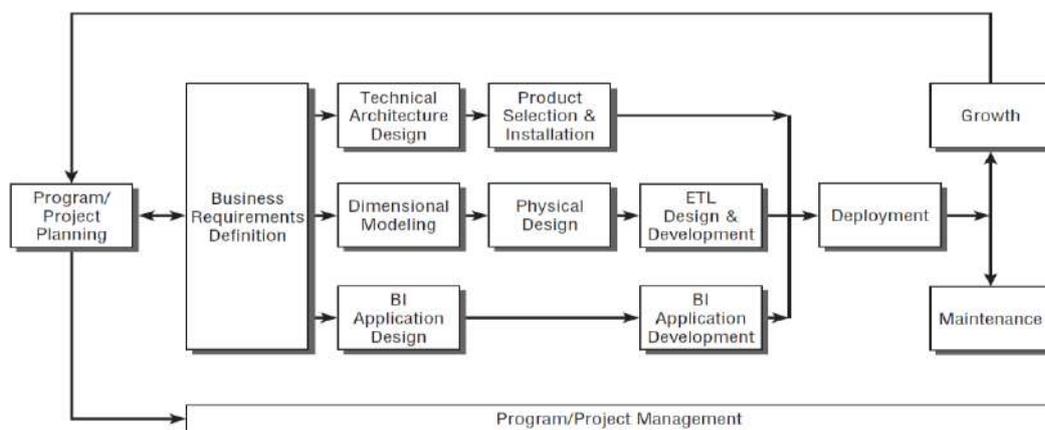
Ante presentar nuestra propuesta, en esta sección presentamos cuatro antecedentes metodológicos que hemos utilizado como referencia para la elaboración de una alternativa que se ajuste a nuestros requerimientos.

La primera metodología, generalmente aceptada en la industria como una de las mejores prácticas, es el ciclo de vida de Kimball planteado en (Ralph Kimball, 2013), que sigue una doctrina general para centrarse en el valor de negocio, colaborar con la empresa y desarrollarse de forma incremental.

La hoja de ruta de esta metodología se muestra en la Figura 17. Para este método, el primer paso es la definición de los requisitos de negocio, seguido de desarrollos paralelos. Por un lado, el diseño, desarrollo y despliegue de la aplicación de BI debe estar soportado en un adecuado modelado dimensional de los datos. Por otro lado, se requiere un diseño físico, así como el diseño y desarrollo de una solución ETL (extracción, transformación y carga de datos). El enfoque de Kimball es muy robusto, pero en nuestra opinión se puede simplificar dependiendo de las características de la solución que busquemos, más aún cuando se aplica a una pequeña o mediana empresa.

Figura 17

Diagrama del ciclo de vida de Kimball

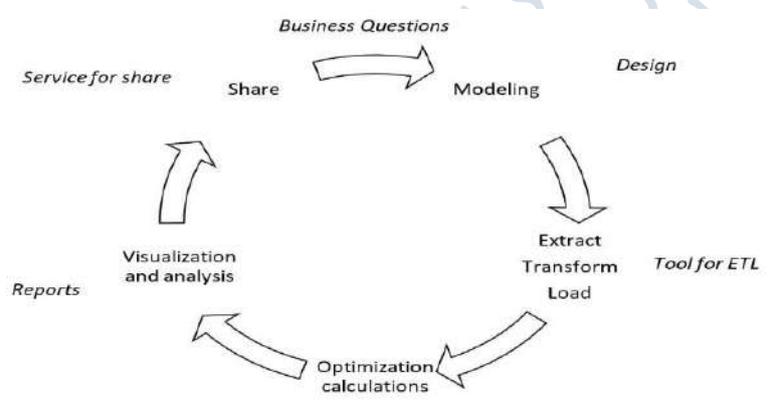


Fuente : (Ralph Kimball, 2013)

Otra alternativa metodológica es la propuesta por Ramos en (Ramos, 2019). Se puede resumir esta metodología en los pasos descritos en la Figura 18. Se puede observar que estos pasos son bastante lógicos y se ajustan mucho a nuestros objetivos y pueden ser aplicados en una PYME. Lo planteado por Ramos coincide con la primera metodología en cuanto a la definición de las cuestiones de negocio, sin embargo, el enfoque de Ramos va hacia el uso de una herramienta informática definida (PowerBI) y al ser pasos exclusivos para estos despliegues, no se fija en procesos paralelos como: diseño de arquitecturas o diseños físicos. Otro aspecto por resaltar es que para Ramos el modelado se refiere solo al modelado de los datos, no a la generación de modelos de IA o DM para encontrar información o conocimiento detrás de los datos.

Figura 18

Pasos para crear una solución de BI



Fuente: Adaptado de(Ramos, 2019)

Otra metodología encontrada es la que utiliza los conceptos adicionales estrechamente relacionados con BI: análisis de negocios (BA) y minería de datos (DM).

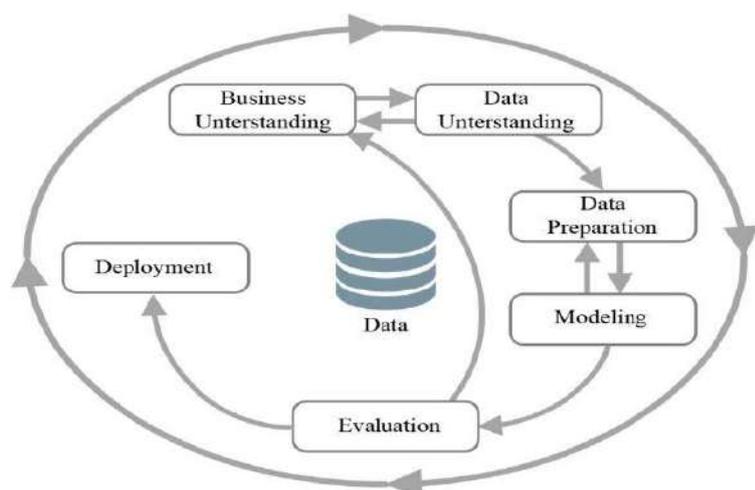
La sutil diferencia entre BI y BA ya la hemos explicado en capítulos precedentes. Respecto a DM, (Mishra et al., 2017) la considera como el proceso de “descubrimiento de conocimiento”, que existe en la base de datos y puede ser utilizado en la toma de decisiones. Por lo tanto, para él, la minería de datos proporciona información sobre escenarios pasados y presentes y ayuda a comprender posibles resultados futuros que pueden proporcionar resultados efectivos y ayudar a descubrir hechos desconocidos presentes en la base de datos.

En 1996, para el desarrollo de una solución de DM, cuatro líderes del mercado emergente de DM concibieron un modelo de referencia para el desarrollo de soluciones con DM, al que denominaron: “CRISP-DM (Cross-Industry Standard Process for Data Mining)” (Huber et al., 2019). Se trataba de un modelo genérico, documentado y de libre acceso que permitió a las organizaciones lograr mejores resultados de minería de datos y promover las mejores prácticas en la industria (Shearer, 2000)

La Figura 19 muestra las fases de un proceso de DM, según el modelo de referencia CRISP-DM, en el que las seis fases del proceso son: “comprensión del negocio, comprensión de los datos, preparación de datos, modelado, evaluación e implementación” (Huber et al., 2019).

Figura 19

Fases del modelo de referencia CRISP-DM



Fuente: (Huber et al., 2019)

Los autores de (Ali et al., 2018) han proporcionado una explicación de la Figura 19, en la que se puede observar la secuencia de pasos y la interrelación entre ellos. Además, la naturaleza cíclica del proceso está representada por la gran elipse. Esto indica que después de que se han identificado algunas soluciones basadas en el conocimiento encontrado, se pueden generar nuevas soluciones a partir de los descubrimientos anteriores.

En (Wiemer et al., 2019), se presenta una extensión del modelo de referencia CRISP-DM considerando la aplicación de DM a soluciones de ingeniería. La Figura 20 muestra esta propuesta.

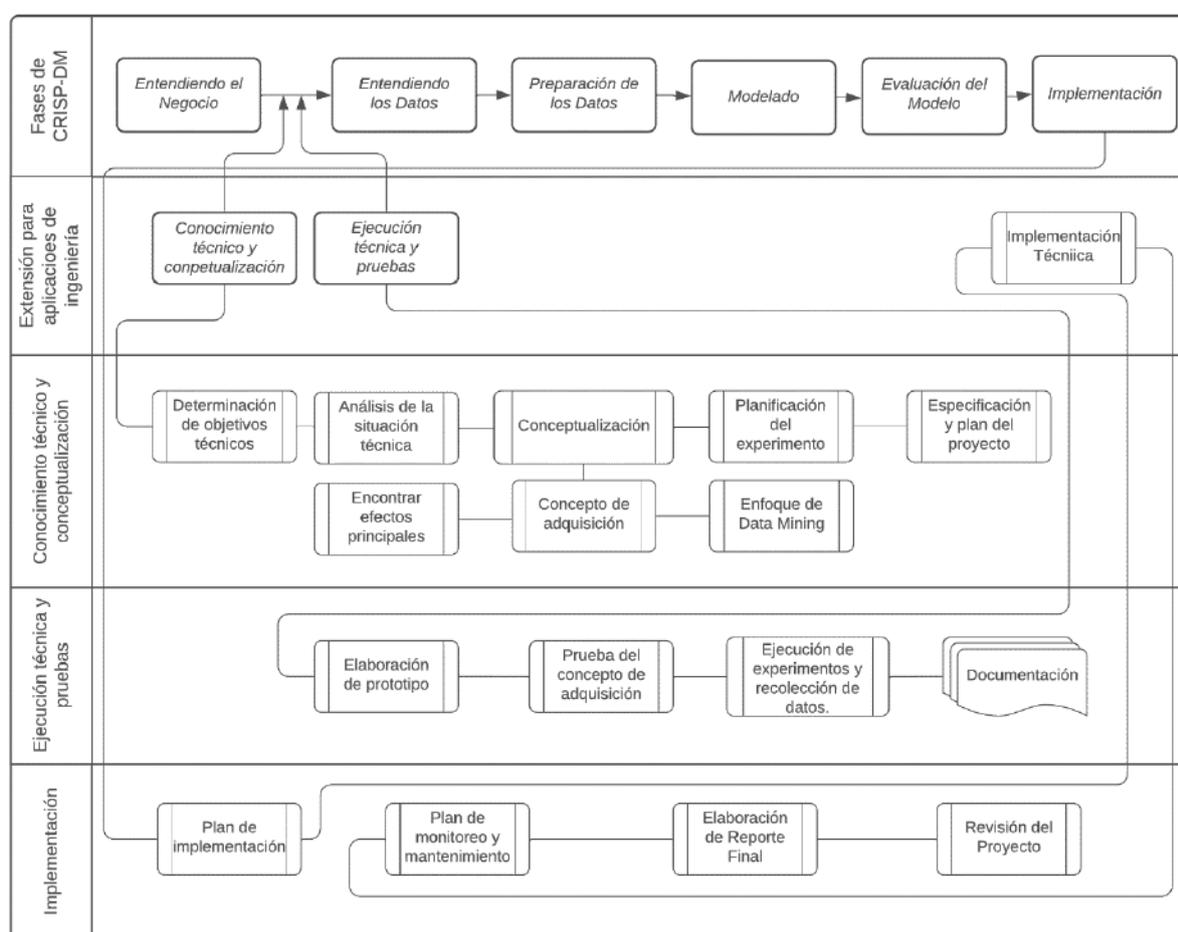
En resumen, Wiemer parte del desarrollo de su propuesta de la afirmación que los modelos conocidos de procedimientos de DM asumen que los datos ya están disponibles y que, bajo este supuesto, estos enfoques se centran en el análisis de datos solamente. El propone, en lo que llama un enfoque holístico, incluir entre la fase de entender el negocio y entender los datos, unos pasos de conceptualizar técnicamente el proceso y objetivo a analizar y desarrollar un plan de pruebas de concepto mediante prototipos cuyos resultados deben ser documentados. Además, en la fase de implementación incluye una fase de implementación técnica que incluye el proceso de transferencia del modelo en prototipo al proceso en producción considerando los problemas que la producción genere (datos reales y en línea, limitaciones de comunicación, etc.) así como la capacitación a los usuarios finales entre otros detalles.

Esta propuesta holística es interesante por dos razones: 1) muestra que a pesar de lo relativamente antiguo del modelo CRISP-DM (data del año 2000), éste sigue siendo investigado y usado (Wiemer no es el único que lo usa para mejorarlo o extenderlo) y 2) con ciertas reducciones consideramos que puede ser aplicada a procesos y objetivos de la PYME, tal como podemos observar en los casos de estudio que hemos desarrollado en los que la secuencia lógica de pasos aplicados siguen el modelo de referencia CRISP-DM y en cierta medida consideran el enfoque holístico de Wiemer, pues es nuestro objetivo pyme del sector eléctrico y en éstas el conocimiento técnico y conceptualización es importante.

La ejecución técnica y pruebas en nuestro caso se ha realizado en cada caso según el objetivo perseguido y para cada caso la implementación se ha hecho siempre fuera de línea.

Figura 20

Metodología de DM para aplicaciones de ingeniería como extensión de CRISP-DM



Fuente: Adaptado de (Wiemer et al., 2019)

Teniendo en cuenta el objetivo general de nuestro estudio: encontrar un método para implementar un marco de inteligencia empresarial que se pueda aplicar a las empresas eléctricas o usuarios de electricidad, especialmente para las pequeñas o medianas empresas, realizamos unos comentarios a cada una de las metodologías analizadas, como se muestra en la Tabla 6.

Tabla 6*Resumen de métodos analizados*

Metodología	Comentario
Kimball:	La hoja de ruta de Kimball es una hoja de ruta integral para un proyecto de almacenamiento de datos e inteligencia empresarial. Su metodología debe incluir necesariamente el tratamiento de datos de forma separada y paralela al desarrollo de la aplicación de BI. Para aplicaciones en pequeñas y medianas empresas, esta separación puede complicar la implementación de la solución.
Ramos	El método Ramos está diseñado específicamente para aplicaciones de BI desarrolladas en una herramienta específica. En esta metodología, el término modelado lo usa al modelado de datos, y no al modelado de la solución a implementar. Al estar diseñado considerando una solución comercial específica (Power BI), pierde generalidad.
CRISP-DM	La secuencia de pasos integra los métodos anteriores en pasos secuenciales y paralelos que facilitan la implementación. El modelado si se refiere a la generación del modelo que representa el proceso a implementar. La característica de esta metodología permite su implementación para temas simples o complejos y se puede adaptar a varios sectores.
DMME extensión del CRISP-DM	Es una extensión que considera una extensión del modelo CRISP-DM Pensado en soluciones para procesos de ingeniería, aunque generalizable a la pequeña y mediana empresa.

4.2 Justificación

Los negocios de la pequeña y media empresa de la cadena de valor del sector eléctrico tienen unas características notorias:

- Conocen poco o nada la utilidad de la inteligencia de negocios en los diferentes departamentos de la gestión de sus negocios. La mediana empresa ya cuenta con áreas definidas como operación o producción, mantenimiento, logística, recursos humanos y en todas estas áreas existen aplicaciones de inteligencia de negocios que pueden mejorar el desempeño de estas áreas.
- No existe cultura de gestión de los datos. Los datos normalmente están en formatos distintos según la aplicación que los genera, el departamento que los procesa o el reporte que la alta dirección de cada negocio necesita.
- Ya a nivel de control de operación las pyme disponen de datos de operación de sus procesos y en especial del suministro eléctrico. Estos datos provienen de sus sistemas SCADA, de medidores inteligentes, de analizadores de redes y en un futuro no muy lejano de dispositivos IoT. Estos datos no se aprovechan utilizando las herramientas BI soportadas con aplicaciones de inteligencia artificial de algunos de los tipos mencionados en el marco conceptual.

Los casos de estudio planteados han demostrado que, con herramientas disponibles actualmente y probadas ampliamente, se puede desarrollar soluciones BI aplicables a estas empresas con costos relativamente bajos, que justificaría muchas veces su implementación y desarrollo.

4.3 Propuesta

La metodología que planteamos tiene los siguientes requerimientos:

Debe poder ser aplicada a una pequeña o mediana empresa.

Una PYME normalmente no cuenta con un área de TI por lo tanto no debe implicar complejidades informáticas.

El poder de cómputo necesario para el pre-procesamiento, procesamiento y modelado debe ser bajo. Los algoritmos y modelos deben poder ejecutarse en PCs de escritorio convencionales.

Nuestra propuesta considera que una forma sistemática de desarrollar una solución que implique inteligencia de negocios es seguir el modelo de referencia CRISP-DM. Para aproximar los pasos del modelo que proponemos utilizamos las fases de este modelo mostradas en (Schröer et al., 2021) cuyo resumen se muestra en la Tabla 7.

Tabla 7

Etapas del Modelo CRISP-DM

Fase	Breve descripción
Entendiendo el negocio	Evaluación y/o identificación de la situación del negocio o de parte del negocio para obtener una visión general de los recursos disponibles y necesarios. Determinación del objetivo de la aplicación BI, definiendo el tipo herramienta a aplicarse y los criterios de éxito de esta. Se crea un plan de proyecto.
Entendiendo los datos	Recopilación de datos desde sus fuentes, exploración, descripción y verificación de la calidad de los datos. Es posible utilizar análisis estadístico para la exploración inicial de los datos.
Preparación de los datos	Selección de los datos mediante la definición de criterios de inclusión y exclusión. La mala calidad de los datos se puede manejar limpiando los datos. Dependiendo del modelo utilizado (definido en la primera fase) se deben construir los atributos derivados. Esta preparación es posible realizarla con diferentes métodos dependiendo del modelo.
Modelado	Consiste en seleccionar la técnica de modelado, construir el caso de prueba y el modelo. Se pueden utilizar todas las técnicas aplicables a BI. En general, la elección depende del problema a resolver o investigar y de los datos. Lo importante es cómo explicar la elección. Para construir el modelo, se deben establecer parámetros específicos. Para evaluar el modelo, es apropiado comparar los resultados de este contra los criterios de evaluación y seleccionar los mejores.
Evaluación	Se realiza la comparación de los resultados con los objetivos definidos. Por lo tanto, los resultados deben interpretarse y según la interpretación deben definirse acciones adicionales. Esto podría requerir revisar el proceso en general, es un proceso iterativo.
Implementación	Puede ser un informe final, un componente de software, un tablero de control, etc. Esta fase consiste en planificar la implementación, la supervisión y el mantenimiento.

Fuente: Adaptado de (Schröer et al., 2021)

Pasamos a detallar los pasos definidos acotando el negocio a empresas del sector eléctrico y utilizando las herramientas de BI a la gestión del suministro eléctrico.

4.3.1 Paso uno: Comprender el negocio

En este paso, es necesario comprender los objetivos del proyecto desde la perspectiva de la empresa.

Gestionar el suministro eléctrico en una empresa del sector eléctrico independiente del lugar que ocupe en la cadena de valor requiere de los siguientes controles:

- Asegurarse que la calidad del servicio eléctrico que brindamos o recibimos se encuentre dentro de lo normado, esto incluye el suministro (controlar las interrupciones) y el producto (la tensión, pflickers y armónicos).
- Evitar el consumo, generación o distribución de energía reactiva inductiva en la red que limita la capacidad de las redes de distribución y provoca en los clientes facturación por una energía que no se aprovecha.
- Controlar el diagrama de carga tanto de producción como de demanda. Para los consumidores este control les permite: 1) proyectar sus consumos, 2) elegir una adecuada opción tarifarias y 3) proyectar sus consumos y por ende sus pagos. Para los generadores y distribuidores en cambio: 1) les permite optimizar sus recursos en función de la demanda, 2) proyectar inversiones focalizadas según la demanda, 3) detectar consumos anómalos, 4) proyectar su producción.
- Monitorear y mejorar la eficiencia de sus procesos. Para un generador la gestión del suministro eléctrico pasa por una adecuada gestión de su mantenimiento, reducción de paradas intempestivas mediante el mantenimiento predictivo o bajo condición. En la distribución corresponde reducir las pérdidas técnicas y no técnicas, así como mejorar sus indicadores de calidad de suministro. Para un consumidor corresponde también el monitoreo de los equipos eléctricos importantes (monitoreo de sus condiciones de operación), identificar patrones de consumo y consumos atípico así predecir posibles fallas intempestivas de estos.

Definidos estos requerimientos, entonces es necesario traducir estos objetivos en objetivos de BI para mediante minería de datos, por ejemplo, desarrollar un plan o estrategia preliminar.

Los resultados de esta fase se pueden utilizar para aclarar los objetivos, definir el problema, el plan preliminar y la estrategia a seguir.

En este paso, entendemos el negocio a través de las preguntas que pretendemos resolver e identificamos si tenemos los datos para posteriormente describirlos, explorarlos y verificarlos. Aquí hay que definir el tipo de análisis a aplicar para solucionar las cuestiones planteadas.

Al definir explícitamente el tipo de análisis, preparamos adecuadamente los datos en la siguiente etapa. Se pueden realizar los siguientes tipos de análisis:

- **Descriptivo:** generalmente se realiza en la fase inicial del análisis para lograr una comprensión completa de la forma y los patrones de los datos. Se enfoca en descubrir "lo que ha sucedido". Los procesos típicos incluyen la segmentación, el agrupamiento y la clasificación. El análisis descriptivo es ampliamente utilizado en el análisis del comportamiento de parámetros o variables, brindando un medio eficiente para comprender el comportamiento de varias variables y crear acciones para cada segmento de variables en lugar de para todo el sistema. En cuanto a la precisión y el rendimiento, los algoritmos de análisis descriptivo son los más naturales.
- **Diagnóstico:** el análisis de diagnóstico, que a veces se combina con el análisis descriptivo, evalúa las correlaciones. Las pruebas de diagnóstico suelen ser complejas, lo que da como resultado información más valiosa. Sin embargo, las correspondencias no logran demostrar de manera convincente la causalidad mediante algoritmos de aprendizaje automático porque la correspondencia solo se basa en el conjunto de datos de aprendizaje limitado utilizado. La información resultante de los análisis descriptivos y de diagnóstico se considera comúnmente como retrospectiva, ya que se refiere principalmente a hechos pasados. A diferencia del análisis descriptivo, el análisis de diagnóstico intenta responder: "¿Por qué sucedió?"
- **Predictivo:** El análisis predictivo utiliza información pasada para comprender la incidencia probable del futuro y se centra en responder a la pregunta: "¿Qué pasará?". Los modelos complejos y el aprendizaje automático son vitales para realizar el razonamiento inductivo.
- **Prescriptivo:** el análisis prescriptivo proporciona información sobre los pasos que se pueden tomar para intensificar la probabilidad de que ocurra un resultado deseado, es decir, "¿qué debo hacer?". Sin embargo, este es el análisis más difícil. Los análisis predictivos y descriptivos ahora son más frecuentes que antes. Los avances tecnológicos en software y hardware permiten un análisis más elaborado. También hay muchas fuentes de datos altamente digitalizados en todo el mundo y con la llegada de IoT. Dichos factores, junto con la expansión de las mejoras en el software y el hardware, implican que hay muchos más datos disponibles para el análisis que en el pasado.

4.3.2 Paso dos: comprensión de los datos

En el paso dos, comenzamos con la adquisición y la familiarización con los datos disponibles. Debemos detectar datos alineados con nuestros objetivos iniciales e identificar datos de mala calidad o irrelevantes. "La fase de comprensión de los datos consta de cuatro pasos, que incluyen la recopilación de datos iniciales, la descripción de los datos, la exploración de los datos y la verificación de la calidad de los datos" (Shearer, 2000).

Para esta fase en las empresas del sector eléctrico los datos corresponden a:

- Diagramas de carga: normalmente potencia activa, reactiva y aparente cada 15min, en algunos casos energía activa, reactiva y aparente por intervalos o acumulada. Control de los bloques horarios de Horas de Punta y Horas Fuera de Punta.
- Parámetros eléctricos: Tensión, Corriente, Frecuencia. Cada 15min o 10 min.
- Registros de Pflckers y Armónicos (40 armónicos) obtenidos de analizadores de redes o de medidores analizadores.
- Reportes de facturación de potencia generada o comprada: potencias y energías, según marco regulatorio para clientes regulados o contratos bilaterales para clientes libres.
- Contratos bilaterales y pliegos tarifarios según sea el caso.
- Reportes de sistemas SCADA, reportes de eventos y alarmas y tendencias de todos los parámetros bajo control para los equipos importantes.

4.3.3 Paso tres: preparar los datos

Una vez que se ha definido el tipo o tipos de análisis a realizar, el paso cuatro involucra todas las acciones requeridas para construir el conjunto de datos final o el conjunto de datos (que ingresaría a la herramienta de modelado) a partir de los datos sin procesar iniciales. Este paso implica acciones como la selección de tablas, registros y atributos, así como la conversión y limpieza de datos para la herramienta de modelado. Normalmente, hay cinco pasos en la preparación de datos: selección de datos, limpieza, construcción, integración y formateo.

En las PYME del sector este paso es muy importante y laborioso debido a que normalmente no existe una cultura de gestión de los datos en general y en especial de los datos que permitan aplicar BI a la gestión del suministro. Los datos se encuentran en reportes independientes, no están en un solo repositorio, no existe política de almacenado de los mismos, no existe la gobernanza de estos, no existe la cultura del manejo de registros, no se registra datos no directamente relacionados con la gestión eléctrica (por ejemplo, no se aprovecha la gestión de las redes sociales para el control de las interrupciones en las distribuidoras).

Por ello en esta etapa (ya se ha avanzado algo en la Fase 1) se recopila y ordena toda la data necesaria y disponible acorde al objetivo de la BI. Aquí se hacen todas las transformaciones necesarias a los datos para luego puedan ser cargados al modelo propuesto. En esta fase se puede incluir la reducción dimensional por PCA (por ejemplo) para identificar patrones de operación.

4.3.4 Paso cuatro: Modelado

Aquí, se seleccionan y aplican varias técnicas de modelado y sus parámetros correspondientes se calibran a valores óptimos. Normalmente, existen varios parámetros para

el mismo tipo de problema BI. En esta etapa se incluyen todos los posibles desarrollos de AI y minería de datos. Ciertas técnicas presentan demandas específicas en forma de datos. Por lo tanto, puede ser necesario volver a la etapa de preparación de datos. Las etapas de modelado implican la selección de la técnica de modelado, la generación del diseño de prueba, la creación de modelos y la evaluación de modelos.

4.3.5 Paso cinco: Evaluación

Cuando se define el modelo final, antes de su implementación o despliegue, se debe realizar una evaluación adicional del modelo y una revisión de la construcción del modelo para garantizar que los objetivos establecidos se logran adecuadamente. En este punto, es vital establecer si los problemas colaterales significativos se han abordado a fondo. El responsable del desarrollo puede decidir exactamente cómo se utilizarán los resultados del análisis al final de este paso. Los pasos principales incluyen la evaluación de los resultados, la revisión del proceso y la determinación de los próximos pasos.

4.3.6 Paso seis: Implementación

En este punto, el conocimiento adquirido debe organizarse y presentarse de tal manera que sea útil para el usuario final. Esto implica frecuentemente la aplicación de modelos “online” en el proceso de toma de decisiones dentro de una organización.

Según las preguntas a responder, la fase de implementación podría ser tan simple como generar un informe o tan compleja como implementar un proceso de análisis de datos repetible en toda la empresa. Es crucial que el usuario final tenga una comprensión previa de las acciones que deben tomarse para usar los modelos creados. Los pasos clave dentro de esta fase son la implementación del plan, el seguimiento y mantenimiento del plan, la emisión del informe final y la revisión del proyecto.

En la Tabla 8 presentamos un resumen del requerimiento que una PYME debe asumir si quiere implementar una solución BI. Para esta elaboración nos hemos basado en:

- La experiencia del desarrollo de las tres aplicaciones mostradas en los casos de estudio.
- El método de diseño en ingeniería planteado en (Grech, 2013)
- El concepto de científico de datos que no necesariamente implica una persona: Según (Irizarry, 2020) decir científico de datos implica: un ingeniero de datos, un analista de datos e ingeniero de aprendizaje automático. Nuestra propuesta considera que para una PYME del sector eléctrico una persona con conocimiento básicos de ciencia de datos es suficiente para generar soluciones útiles como han sido los casos de estudio.

Tabla 8

Fases de la implementación

Fase	Descripción	Entradas o Fuentes de Datos	Conocimientos	Capacidades , habilidades	Herramientas Informáticas Necesarias	Resultados o entregables	Tiempo	Personas
1	Comprendiendo el negocio	Focus group Entrevistas semiestructuradas Encuestas Visita de campo	Técnicas para manejo de grupos. Técnicas de comunicación. Conocimiento del proceso. Definición de Key Drivers. Conocimiento de metodologías ágiles.	Liderazgo, generar confianza. Capacidad analítica. Espíritu de observación. Capacidad de síntesis.	Ofimática básica Aplicaciones para videoconferencias Aplicaciones para trabajo colaborativo	Objetivo(s) claro(s) Key drivers Tipo de análisis a aplicar Plan de trabajo	1 semana	Especialista del Proceso Científico de Datos Especialista en sistemas eléctricos
2	Comprendiendo los datos	Medidores electrónicos Medidores inteligentes Analizadores de redes Sistemas SCADA	Técnicas para realizar e interpretar encuestas. Estadística. Probabilidad . Cálculo numérico.	Conocimiento de análisis exploratorio Conocimiento de sistemas eléctricos Capacidad de analítica. Espíritu de observación.	Hoja de cálculo (Excel o similar) Power BI (Visualización) R	Datos necesarios Faltan datos Calidad de datos OK.	1 semana	Científico y/o Analista de Datos Especialista en sistemas eléctricos
3	Preparando los datos	Archivos en formato origen Reportes de los sistemas de gestión Registros de tendencia y eventos	Estadística. Probabilidad. Cálculo numérico. Conocimiento del proceso de extracción, transformación y carga de datos (ETL). Programación básica (R, DAX, M, Python). Conexión de datos a herramientas BI.	Capacidad analítica. Espíritu de observación. Creatividad.	Hoja de cálculo (Excel o similar) Power BI- Transformación de datos (DAX, M) R, librerías para transformación de datos, eliminación de atípicos, nulos, etc.	Datos listos para el modelo Definición de datos de entrenamiento Definición de datos de prueba	2 semanas	Científico y/o Analista de Datos
4	Modelado	Datos listos Modelos predefinidos en librerías o en embebidos	Métodos de investigación: método científico. Conocimientos de AI: Modelos de predicción, estadística multivariable, modelos de clasificación, etc. Codificación de modelos en R, Python y conexión a herramientas BI	Capacidad analítica. Capacidad investigativa. Creatividad.	R con librerías especializadas Python con librerías especializadas Power BI	Modelo listo para evaluación	2 semanas	Científico de Datos
5	Evaluación	Modelo definido Datos de prueba	Método de investigación: método científico. Conocimiento de criterios de evaluación de modelos de AI: MSE, Gráficos estadísticos, Matrices de Confusión, etc. Técnicas de pensamiento convergente.	Imparcialidad. Respeto a los criterios. Actitud crítica. Lógica.	R con librerías especializadas Python con librerías especializadas Power BI Hojas de Cálculo (Excel o similar)	Modelo listo para implementar: capaz de predecir, clasificar o detectar patrones	1 semana	Científico de Datos Especialista en SE
6	Implementación	Modelo validado y listo para producción. Nuevos datos u observaciones. Requerimientos finales del usuario.	Técnicas de manejo de tiempos. Finanzas y costos. Normas para elaboración de manuales. Conocimientos de la normatividad del sector. Conocimiento de técnicas efectivas de comunicación oral y escrita. Conocimiento de programas de presentación efectiva. Conocimiento en desarrollo de dashboards. Conocimiento de conexión inter aplicaciones. Conocimientos de conexiones a la Nube.	Capacidad analítica y sintética. Capacidad de convencimiento. Capacidad de expresión clara de ideas complejas. Lógica y sentido común. Perfeccionismo.	Power BI (Reportes) Excel colaborativo	Reporte Dashboard Agrupamientos Proyecciones	2 semana	Científico de Datos y/o Analista de datos con conocimiento de desarrollo de dashboards Especialista SE Especialista de Proceso

4.4 Aplicación de la metodología a los casos de estudio

4.4.1 Caso 1: Calidad de energía eléctrica y clasificación

Tal como fue expuesto en el capítulo 3 este caso tiene por objetivo analizar la calidad del suministro eléctrico de una empresa Agroindustrial. En este caso aplicamos los 6 pasos del método propuesto como aplicable a una pyme.

4.4.1.1 Paso uno: Entendiendo el negocio. La empresa es una empresa agroindustrial, procesadora de semillas que opera las 24 horas del día. Cuentan con suministro eléctrico desde una distribuidora local, pero para controlar su facturación, utilizan generación propia para las horas de punta.

Los objetivos del estudio son: 1) evaluar el comportamiento de los principales parámetros eléctricos para verificar la calidad de suministro, el efecto sobre el equipamiento de producción y sobre la facturación de la energía eléctrica y 2) establecer medidas de control en la operación y mantenimiento que puedan reducir los costes del suministro eléctrico.

Dos aspectos debemos resaltar en este paso: 1) hemos hecho una reducción del “negocio” a una parte del negocio en general. La empresa agroindustrial tiene muchas áreas, sin embargo, nuestro análisis se basa en lo relacionado a la parte del negocio relacionada solo al suministro eléctrico y 2) los objetivos son específicos y claros. Ambos aspectos encajan muy bien en los objetivos que el CRISP-DM tiene para la DM como lo plantea (Martinez-Plumed et al., 2021).

De acuerdo con las características de los objetivos buscados, dos tipos de análisis fueron aplicados: un análisis descriptivo y un análisis prescriptivo.

El análisis descriptivo muestra el comportamiento de los parámetros eléctricos principales envueltos en el monitoreo de la calidad del suministro eléctrico y el análisis prescriptivo provee unas recomendaciones para mejorar la calidad de este suministro y disminuir la facturación.

En un inicio y tal como lo propusimos en el artículo aprobado sobre este caso el tipo de análisis a realizar lo hicimos en un paso tres, pero creemos que este paso encaja muy bien en el paso uno y por ende en esta propuesta final lo estaríamos incluyendo aquí.

4.4.1.2 Paso dos: Entendiendo los datos. Los datos obtenidos para este estudio provinieron de un analizador de redes que registra los parámetros eléctricos principales como son: tensión, corriente, potencia activa y reactiva, flickers y armónicos. Estos parámetros fueron registrados cada 15 min por 1 semana. Los datos se obtienen en formato fuente extraído desde el analizador y entregado por el analista de calidad en formato xls. Una muestra puede observarse en la Figura 21.

El método exige que en este paso se evalúe la utilidad del modelo usado, para nuestro caso solo implicaría comprobar que los dashboards son útiles y muestran información útil cara a los objetivos perseguidos.

Para la parte prescriptiva, producto del análisis descriptivo se detecta muchos intervalos con tensiones por encima del valor límite superior y factores de potencia muy bajos, es decir, la empresa paga consumos de energías reactivas realmente inútiles.

Esta información prescribe una revisión de los TAPs del transformador o un reclamo al concesionario de tensión, respecto al alto consumo de energía reactiva el análisis prescriptivo indicaría la instalación de un banco de condensadores. La aplicación de un modelo de clusterización o agrupamiento de los consumos de energía reactiva, permitirá una adecuada selección de los pasos de compensación del banco.

4.4.1.5 Paso cinco: Evaluación. En este caso la aplicación del método y la información obtenida de los datos analizados generan conocimiento útil para la toma de decisiones que mejoren la calidad del suministro y disminuyan costos por el mismo.

Los *dashboards* muestran muchos intervalos con tensiones por encima del límite de +5% de la tensión nominal siendo en una semana casi un 40% de los intervalos fuera de rango (Ver Figura 23), también muestran que existen muchos intervalos con un factor de potencia menor al límite que evita el pago de compensaciones por consumo de energía reactiva (Ver Figura 24). Podemos concluir que existe una mala calidad de suministro y un inadecuado consumo de energía reactiva.

Figura 23

Caso 1 - Dashboard de la tensión

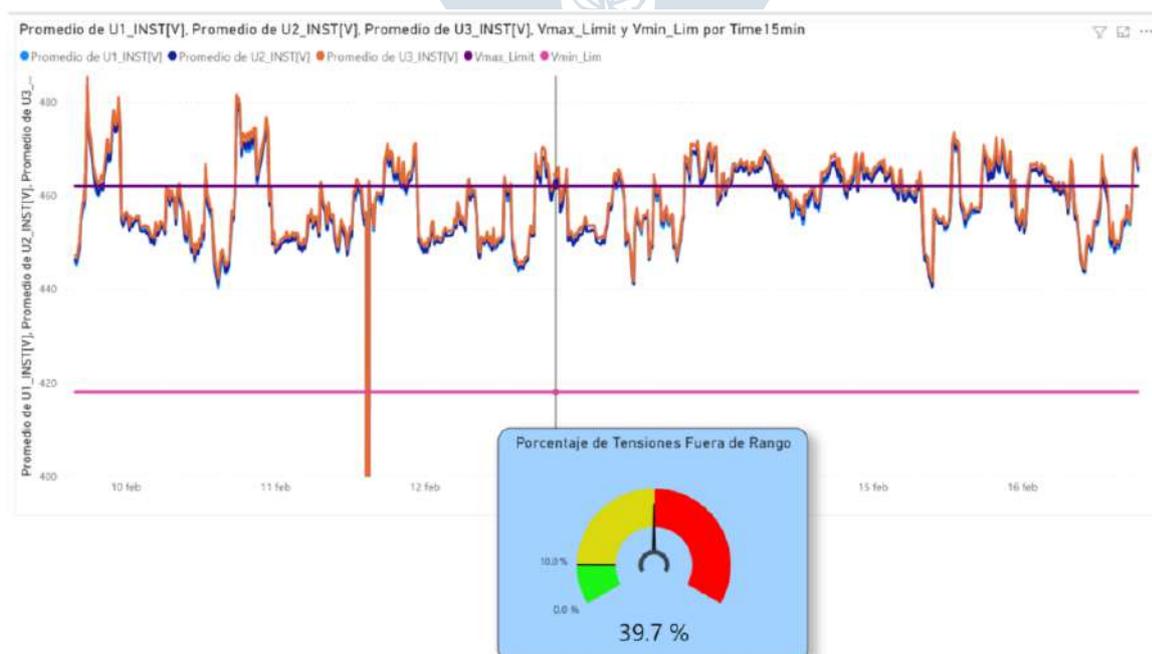
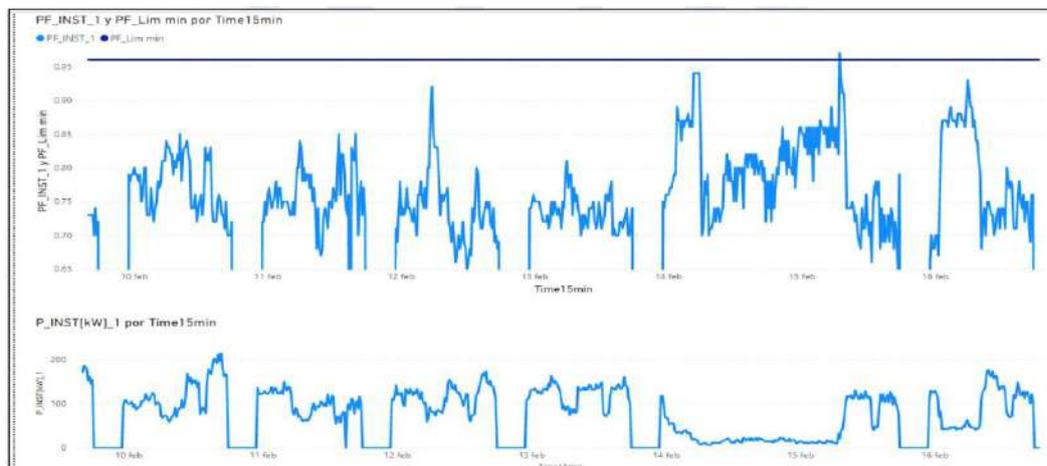


Figura 24

Caso 1 - Dashboard del Factor de Potencia



Para el agrupamiento se usó un modelo de agrupamiento (clustering) sobre los datos de consumo de energía reactiva. Para ello se utilizó la herramienta analítica de agrupación en clústeres incorporada a Power BI, esta agrupación en clústeres permitió encontrar rápidamente grupos de consumo de energía reactiva fuera de rango. Se seleccionaron 6 grupos y por ende 6 pasos de compensación posibles. La Tabla 9 muestra el agrupamiento y la Tabla 10 los pasos posibles del banco.

Tabla 9

Caso 1 - Resultado del agrupamiento en clusters de los factores de potencia

Clusters	FP	P [W]	Q [VAr]	ϕ	Qc[kVAr]
Cluster 1	0.75	102911	90069	0.72	60.75
Cluster 2	0.15	3694	1981	1.42	23.43
Cluster 3	0.83	51958	35568	0.59	20.40
Cluster 4	0.78	76467	62327	0.68	40.55
Cluster 5	0.71	125204	123804	0.78	88.42
Cluster 6	0.73	115552	108139	0.75	75.39

Tabla 10

Caso 1 - Pasos del banco de condensadores

Steps	SE Model
1	20.403
2	23.427
3	40.550
4	60.751
5	75.390
6	88.425

4.4.1.6 Paso seis: Implementación. La implementación en este caso implica la presentación de estos resultados en un informe que muestre los hallazgos y la implementación de las prescripciones realizadas: movimiento de los TAPs del transformador para ajustar la tensión, así como la instalación de un banco de capacitores en seis pasos.

Estos pasos deben ajustarse a los pasos comerciales. Mediante el cálculo de potencias promedios y máximas donde se prescribe un banco de capacitores de 120 kVAr, se infiere que el capacitor está sobredimensionado y puede causar sobretensión cuando se opera en un solo paso.

4.4.1.7 Conclusiones del caso. Concluimos que el procesamiento y análisis de los datos disponibles genera información útil y relevante para la toma de decisiones. Esto demuestra cómo es que siguiendo una metodología de BI&A contribuimos a una mejor gestión de la calidad del suministro eléctrico y, por ende, a un mejor desempeño organizacional.

Esto se logra mediante:

- Visualización de los parámetros de funcionamiento. Esto nos permite controlarlos y ajustarlos a los parámetros correctos que no causen daño a los equipos y operaciones.
- Uso de agrupamiento en *clusters* para determinar el banco óptimo de capacitores multipaso que evite sobrecostos y daños al equipo debido al alto consumo de energía reactiva.

Las herramientas computacionales utilizadas pueden desarrollarse con hardware de escritorio de costo medio y software libre. Por lo tanto, pueden ser implementados tanto por medianas como por pequeñas empresas.

Asimismo, los contadores multitarifa convencionales pueden proporcionar esta información en línea, lo que permite aplicar estas herramientas en tiempo real.

Se requiere un conocimiento mínimo de modelización, análisis de datos y del sector para identificar y transmitir los hallazgos y prescripciones recomendadas a los departamentos de dirección de las empresas.

El caso muestra como contribución mostrar conceptos actuales de BI&A basados en los conceptos de DSS y mostrar también que, aunque en la mayoría de los casos las herramientas de BI se aplican a identificación de conocimiento que mejore la gestión comercial de un negocio, el caso demuestra que también se pueden aplicar a soluciones de mejora de la gestión del suministro eléctrico.

4.4.1.8 Comentarios a los resultados obtenidos. En este caso se ha demostrado la conexión de datos desde diferentes fuentes y desde archivos provenientes de analizadores de redes para mostrar en un *dashboard* los parámetros eléctricos del suministro de una empresa agroindustrial. Se comprueba que con la conexión de los datos desde el formato

original a una herramienta de BI convencional y comercial como PowerBI, resulta fácil visualizar parámetros fuera de rango y por ende controlar la calidad del suministro eléctrico.

Así mismo, se hizo un análisis de la facturación del consumo de energía reactiva (un análisis muy común en la PYME del sector), se mostró que complementar las herramientas de visualización con un modelo de AI, clustering (agrupamiento) de los diferentes consumos de energía reactiva, se puede especificar los pasos de un banco de condensadores, para la compensación de la energía reactiva inductiva que ocasiona muchas veces una facturación innecesaria en las PYME. Además, con el resultado conseguido se muestra que es posible incluir estos modelos de AI en el análisis del suministro eléctrico y que está al alcance de la pyme.

Para este caso y en general para el control de calidad de la energía, la aplicación permite a cualquier PYME, mediante un equipo permanentemente instalado o durante un período de control, poder visualizar el comportamiento de los parámetros dentro de los rangos de calidad adecuados. Esto logra dos objetivos: 1) preservar sus equipos, al no someterlos a tensiones eléctricas fuera de rango o llena de armónicos eléctricos y 2) identificar la mala calidad del suministro para tomar acción sobre los equipos o procesos que ocasionan esta mala calidad o exigir al suministrador eléctrico las compensaciones del caso y ejecute las medidas correctivas del caso.

Para este caso, con un modelo de agrupamiento en *clusters (clustering)* básico y embebido en la plataforma de *Power BI* se pudo identificar los pasos adecuados en un banco de condensadores. La práctica habitual y simple es dimensionar el banco de condensadores de compensación con una potencia reactiva nominal calculada en base a la potencia reactiva determinada por la facturación mensual. Los datos provenientes del analizador de redes registran los consumos de reactivo cada 1, 5 o 15 min según sea configurado y permite agrupar consumos de energía reactiva en una semana típica de operación lo que permite agrupar estos consumos y con ello definir unos pasos más precisos.

4.4.2 Caso 2: Determinación de patrones de operación normal

En este caso se aplica la metodología para determinar un patrón de operación normal en un turbogenerador hidráulico a partir de parámetros registrados desde el sistema SCADA del centro de control de la central hidroeléctrica. Para la evaluación de los parámetros se utilizan dos gráficos de control: el cuadrado del error de predicción (SPE) y el T cuadrado de Holelling, gráficos muy usados en el control estadístico de procesos multivariantes. Los gráficos de control son aplicados luego de una reducción dimensional de registros mediante análisis de componentes principales.

A continuación, describimos cómo es que aplicamos la metodología a este caso.

4.4.2.1 Paso 1: Entendiendo el negocio. Tal como se hizo en el caso 1, para aplicar la metodología a la gestión del suministro eléctrico, reducimos el “negocio” a lo directamente

relacionado con nuestro objetivo. En este caso el objetivo de nuestro negocio es lograr una mayor productividad del turbogenerador detectando patrones anormales de operación que eviten paradas intempestivas no programadas.

El proceso por analizar, entonces, es la operación de una unidad de generación hidráulica (turbogenerador) de una central hidroeléctrica de 12.5 MW en la ciudad de Piura, Perú. La central cuenta con dos unidades de generación cada una compuesta por una turbina Kaplan de 6.25 MW y un generador síncrono de 7.5 MVA acoplados en eje vertical. La central opera las 24 horas de día los 7 días de la semana en función de la disponibilidad hídrica y cuenta con un sistema SCADA que recolecta información de diferentes sensores que permiten el control y monitoreo de la operación de las unidades.

El objetivo de la aplicación de la minería de datos es encontrar un patrón de operación normal de cada unidad de generación en función a los datos proporcionados por los sensores de la unidad a fin de alertar cualquier comportamiento anormal que pueda implicar una parada intempestiva o no programada. Para el análisis y determinación del patrón de operación se planeó realizar primero una reducción dimensional utilizando uno de los métodos tradicionales de control estadístico de procesos multivariantes como es el análisis de componentes principales, que es un método basado en datos, cuantitativo y estadístico según (Mirnaghi & Haghghat, 2020)

El patrón de operación se proyecta modelar utilizando las gráficas de control multivariante para el control estadístico de calidad: T^2 de Hotelling y el Error de Predicción al Cuadrado (SPE, en inglés). Para evaluar el proceso seguimos el procedimiento planteado en (Kourti, 2020) para ellos se utiliza un set de datos consistentes de proceso en operación normal como conjunto de datos de entrenamiento con los que se genera un modelo. Luego con datos proveniente de nuevas observaciones y utilizando información del modelo se calculan parámetros y nuevos puntos en los gráficos T^2 de Hotelling y SPE, si estos puntos no superan los valores límite, el proceso está en operación normal. En esta fase debemos tomar en cuenta el conocimiento técnico de la operación de las unidades en esta central a fin de elegir los datos necesarios para un adecuado modelado. Es importante indicar que un adecuado modelado debe considerar periodos de operación continua, los mismos que se dan en épocas cuando el caudal es constante y suficiente para una operación de las unidades a potencia nominal y sin interrupciones. Esta época es en los meses entre enero y mayo de cada año normalmente.

4.4.2.2 Paso 2: Entendiendo los datos. Los datos son obtenidos de las siguientes fuentes: del sistema SCADA en archivos con variables del proceso proveniente de los sensores de la unidad de generación (archivos generados automáticamente cada día con registros de datos cada 10 segundos de todo un año de operación, de junio de 2020 a junio de 2021. Ejemplo: "U1trend (20200609 0000).DBF". De los reportes de control de paradas de unidad de todo un año de operación. Mismo periodo de la fuente anterior. Se trata de archivos

Excel generados por los operadores de la central. Ejemplo: “SEM 27 - Reportes de paradas - 2020.xls”. De la documentación de la planta correspondiente a la información tal como quedó construido el sistema de control (esta información es la llamada información “Como Construido” o “CC”) se obtiene el archivo de definición de señales del sistema de control. (“748_14_PLC_IO_B+translation (23.11.1600).xls”

4.4.2.3 Paso 3: Preparando los datos. La preparación consistió en elegir periodos de operación normal en los meses de alta disponibilidad del recurso hídrico, que para el 2021 correspondió a los meses entre febrero y abril. En la Tabla 11 se muestran estos períodos, obtenidos en base a la información fuente obtenida. De la documentación de la planta correspondiente a la información Como Construido se identificó la descripción de las medidas de cada sensor según su código, etiqueta o “tag”. La lista de las 40 variables utilizadas en la generación del modelo se puede observar en la Tabla 12.

Tabla 11

Caso 2: Períodos de operación del Turbogenerador

Hora Inicio	Hora Fin	Tipo de operación
01/02/2021 00:00	02/02/2021 00:00	Operación normal
02/02/2021 00:00	02/02/2021 22:46	Operación (previa a limpieza de rejillas)
02/02/2021 22:46	03/02/2021 09:27	Limpieza de rejillas / calibración de álabes.
03/02/2021 10:00	11/02/2021 09:00	Operación normal
11/02/2021 10:06	11/02/2021 11:03	Parada para búsqueda de cadáver en rejillas
11/02/2021 12:00	25/02/2021 00:00	Operación normal.
25/02/2021 00:00	26/02/2021 17:03	Operación previa a calibración de álabes.
26/02/2021 17:03	26/02/2021 21:09	Parada manual para calibración de álabes.
27/02/2021 00:00	03/03/2021 00:00	Operación normal
03/03/2021 00:30	03/03/2021 10:40	Parada por falla de LT
03/03/2021 12:00	10/03/2021 00:00	Operación normal
10/03/2021 00:00	10/03/2021 20:48	Operación normal previa a falla en UAP
10/03/2021 20:48	10/03/2021 23:34	Parada por MI ante falla UAP
11/03/2021 00:00	20/04/2021 00:00	Operación normal
20/04/2021 00:00	22/04/2021 16:53	Operación previa a calibración de álabes.
22/04/2021 16:53	22/04/2021 20:05	Parada manual para calibración de álabes.
23/04/2021 00:00	30/04/2021 23:59	Operación normal

Tabla 12*Caso 2 - Variables registradas en el sistema SCADA*

N°	Tag o Etiqueta	Descripción
1	BC3101	Posición del distribuidor
2	BC3102	Posición alabes rodete
3	BF1	Caudal De Turbina
4	BH1	Frecuencia del Generador
5	BH2	Frecuencia de Red
6	BP1	Potencia activa del generador
7	BP3201	Presión en Espiral
8	BQ4101	Nivel de agua en cámara de carga
9	BQ4102	Nivel de agua en cámara de descarga
10	BT2502	Temperatura de Hierro Del Estator 2
11	BT2503	Temperatura de Hierro Del Estator 3
12	BT2507	Temperatura de Bobina Del Estator U
13	BT2508	Temperatura de Bobina Del Estator V
14	BT2509	Temperatura de Bobina Del Estator W
15	BT2514	Temperatura de Aire Frio De Generador 1
16	BT2515	Temperatura de Aire Frio De Generador 2
17	BT2516	Temperatura de Aire Frio De Generador 2
18	BT2517	Temperatura de Aire Tibio De Generador 1
19	BT2518	Temperatura de Aire Tibio De Generador 1
20	BT2519	Temperatura de Aire Tibio De Generador 2
21	BT3201	Temperatura de Cojinete Guía 1 de Turbina
22	BT3202	Temperatura de Cojinete Guía 2 de Turbina
23	BT3203	Temperatura de Aceite de Cojinete Guía Turbina
24	BT3301	Temperatura de Cojinete Guía Gen. Superior metal 1
25	BT3302	Temperatura de Cojinete Guía Gen. Superior metal 2
26	BT3303	Temperatura de Cojinete Guía Gen Superior metal 3
27	BT3305	Temperatura de Gen. Cojinete Carga metal 1
28	BT3306	Temperatura de Gen. Cojinete Carga metal 2
29	BT3307	Temperatura de Gen. Cojinete Carga metal 3
30	BT3309	Temperatura de Aceite Cojinete Comb. Gen.
31	BT3310	Temperatura de Cojinete Guía Inferior Gen. Metal 1
32	BT3311	Temperatura de Cojinete Guía Inferior Gen. Metal 2
33	BT3312	Temperatura de Cojinete Guía Inferior Gen. Metal 3
34	BT3314	Temperatura de Aceite Cojinete Guía Inferior Gen.
35	BT3401	Temperatura de Cojinete Guía Gen. Descarga Agua Ref.
36	BT3402	Temperatura de Enfriador De Aire Gen. Entrada Agua Ref.
37	BT3403	Temperatura de Enfriador De Aire Gen. Descarga Agua Ref.
38	BT3404	Temperatura de Cojinete Combinado Gen. Descarga Agua Ref.
39	RT_SPD	Velocidad de la unidad
40	SP3201	Caudal De Turbina

Para la preparación de los datos se utilizó el software R y sus librerías “foreign” y “openxlsx” para cargar los datos desde los archivos fuente en extensión “*.dbf” y “*.xlsx”. Los datos fueron limpiados manualmente retirando aquellas variables cuyos sensores presentan valores nulos, negativos o en unidades no configuradas. Con esto se quedan finalmente 40 variables. Además, se considera operación normal aquella operación que no vierte agua por el vertedero de demasías, operación que se registra con la variable BQ4101 menor o igual a 63.90 msnm.

4.4.2.4 Paso 4: Modelado. En esta fase incluimos el análisis de componentes principales a la matriz de datos, tal como lo planteó (Lee & Jemain, 2021) y como se muestra en la Figura 25. El modelo es el planteado en la ecuación (1).

Figura 25

Caso 2- Matriz de datos

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1Y} \\ x_{21} & x_{22} & \ddots & x_{2Y} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NY} \end{bmatrix}$$

$$X = TP' + E$$

(1)

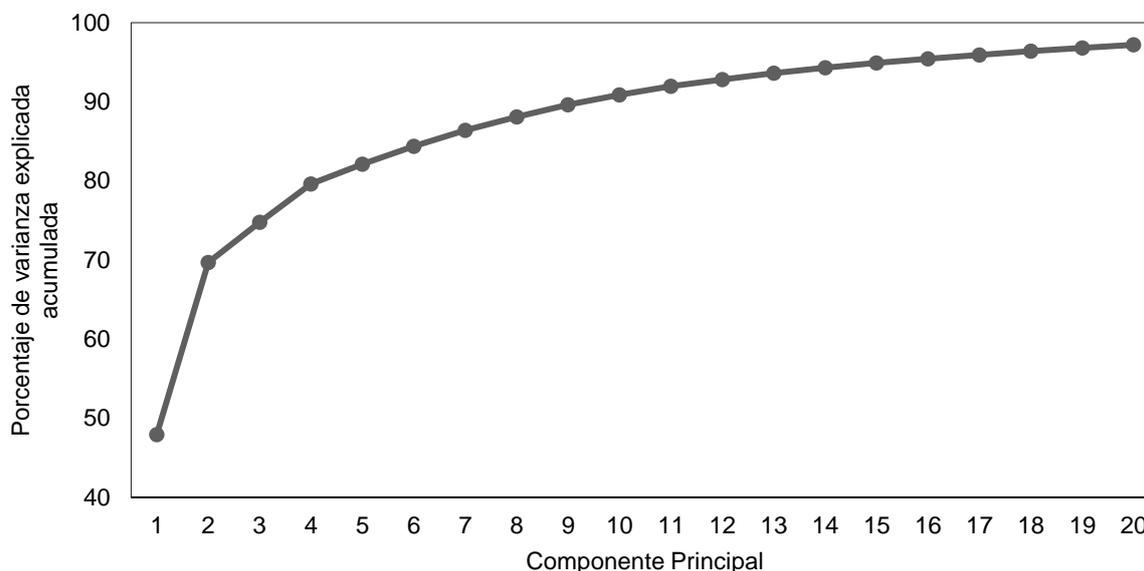
En la ecuación (1), X es una matriz de (N x Y) con las variables estandarizadas (media cero y desviación estándar 1), T es una matriz (N x A) de N “scores” o proyecciones sobre A componentes principales, P es una matriz (P x A) de P “loadings” o pesos de las A componentes principales y E es una matriz (N x P) residuos. El cálculo de las matrices se hace en R. El estandarizado se realiza con la librería “scales”, las matrices T y P se calculan con la función “prcomp”, esta función calcula la matriz “rotation” que contienen los pesos o “loadings” y la matriz “P” que contiene las proyecciones sobre las componentes principales o “principal component scores”. Para mostrar gráficamente la varianza explicada de cada componente principal utilizamos la librería “factoextra”.

Para el modelo se han seleccionado 3 semanas de cada mes entre febrero y abril con operación continua, de esta muestra se han considerado para el modelo, aquellas muestras con nivel de agua al ingreso de la central con cota menor o igual a 63.90msnm, esto generó una muestra con N = 168798 observaciones. Se consideró esta cota pues la cota desde la cual el vertedero de demasías empieza a operar, es decir, a descargar agua que no será aprovechada para la generación de energía. El número de A componentes principales que

utilizamos para el modelo fue aquel que tenía una varianza explicada acumulada mayor al 90% que en este caso se logra con 10 componentes principales tal como se muestra en la Figura 26. Podemos calcular que la suma acumula de varianza de los 10 componentes es de 90.9%.

Figura 26

Caso 2 - Porcentaje de varianza explicada acumulada por cada componente principal



El modelo resultante fue el indicado en la ecuación (2) y las nuevas muestras se modelaron según la ecuación (3).

$$\hat{X} = TP' \quad (2)$$

$$\hat{X}_{new} = T_{new}P' \quad (3)$$

4.4.2.5 Paso 5: Evaluación. En este paso realizamos tanto la evaluación del modelo como la evaluación de la herramienta cara a los objetivos perseguidos.

Para la evaluación del modelo y las nuevas muestras, utilizamos dos gráficos de control estadístico utilizados en procesos multivariantes, se trata de el T^2 de Hotelling y el cuadrado del error de predicción (SPE) planteados en (Kourti, 2020), según las ecuaciones (4) y (5). Nótese que en la ecuación (5) el T^2 de Hotelling corresponde a A componentes principales.

$$SPE = (X - \hat{X})^2 \quad (4)$$

$$T_A^2 = \sum_{i=1}^A \frac{t_i^2}{\lambda_i} = \sum_{i=1}^A \frac{t_i^2}{s_{t_i}^2} \quad (5)$$

Para el cálculo de los límites de control superior (nombramos a estos límites con el subíndice UCL) utilizamos expresiones matemáticas planteadas en (Kourti, 2020), que son: la simplificación de Nomikos y MacGregor. La ecuación (6) muestra la expresión para cálculo del límite de control superior del SPE y la ecuación (7) expresión para el cálculo del límite de control superior para el T^2 de Hotelling.

$$SPE_{UCL} = \frac{v}{2m} X_{\alpha}^2 \left(\frac{2m^2}{v} \right) \quad (6)$$

Donde v es la varianza de los datos, m es la media de los SPEi y X_{α}^2 es la distribución Chi cuadrado calculada con niveles de significancia α .

$$T_{A,UCL}^2 = \frac{(N^2 - 1)A}{N(N - A)} F_{\alpha(A,N-A)} \quad (7)$$

Los cálculos y gráficos de control han sido realizados en Excel, ingresando los resultados obtenidos en R y utilizando las funciones mostradas en la Tabla 13. En la ecuación (7) el límite superior es calculado en base a los A primeros componentes principales y derivadas de N observaciones obtenidas usando una distribución F con niveles de significancia α .

Tabla 13

Caso 2 -Funciones Excel utilizadas para el cálculo de la distribución Chi cuadrado y distribución F

Función por calcular	Función en Excel
$X_{0.05}^2$	= INV. CHICUAD. CD(<i>probabilidad, grados de libertad</i>)
$F_{0.05(A,N-A)}$	= INV. F. CD(<i>probabilidad; grado de libertad 1; grado de libertad 2</i>)

Tabla 14

Caso 2 - Valores calculados utilizando Excel y sus funciones, para los límites de control superior del SPE

m	v	$\frac{v}{2m}$	$\frac{2m^2}{v}$	$X_{0.05}^2 \left(\frac{2m^2}{v} \right)$	$X_{0.01}^2 \left(\frac{2m^2}{v} \right)$	$SPE_{UCL-95\%}$	$SPE_{UCL-99\%}$
3.66	35.90	5.04	1	3.84	6.63	19.38	33.47

Tabla 15

Caso 2 - Valores calculados utilizando Excel y sus funciones, para los límites de control superior del T2 de Hotelling

A	N	$F_{0.05(A,N-A)}$	$F_{0.01(A,N-A)}$	$T_{UCL_{95\%}}^2$	$T_{UCL_{99\%}}^2$
10	168798	1.83	2.32	18.31	23.21

Con el modelo generado en base a las 168798 observaciones y considerando las 10 primeras componentes principales se analizaron 5 conjuntos de observaciones (muestras) descritas en la Tabla 16. Los resultados de las evaluaciones de cada una de las muestras para los dos gráficos estadísticos de control se pueden observar en las figuras desde la 27 hasta la 36. En estas figura hemos nombrado SPE_X al cuadrado del error de predicción de la observación a evaluar y T2_X al valor T2 de Hotelling para la observación a evaluar, los límites de control en cada grafico han sido nombrados como SPE_(UCL_95%) y SPE_(UCL_99%) para el caso del cuadrado del error de predicción y T2_(UCL_95%) y T2_(UCL_99%) para el caso del T2 de Hotelling, ambos límites se han definido con un nivel de confiabilidad de 95% (alpha=0.05) y 99% (alpha=0.01).

Tabla 16

Caso 2 - Conjuntos de muestras analizadas

Muestra	Descripción del conjunto de datos	Inicio	Fin
1	Operación previa a la calibración de álabes directrices de turbina	26/02/2021 00:00:05	26/02/2021 16:30:04
2	Operación con vertedero funcionando (Observaciones obtenidas filtrando muestras con nivel de agua al ingreso de la turbina mayor o igual a 63.90msnm).	NA	NA
3	Operación previa a limpieza de rejillas y calibración.	02/02/2021 21:00:07	02/02/2021 22:59:57
4	Operación previa a una falla en la Unidad de Alta Presión del Gobernador de Turbina. Parada de unidad a las 20:48:45.	03/10/2021 18:00:05	03/10/2021 20:59:35
5	Simulaciones creadas con alteración de variables BT2515, BT3403, BT3404, BT2516 en un periodo con parada de unidad.	03/10/2021 18:00:05	03/10/2021 23:40:05

Los resultados de las evaluaciones de cada una de las muestras para los dos gráficos estadísticos de control se pueden observar a continuación. La Figura 27 muestra que la operación está dentro del modelo generado, mientras que la Figura 28 muestra que las observaciones previas a la parada de unidad si se encuentra en muchos puntos por encima de los límites T2 de Hotelling, que indican una operación anormal varias horas anteriores a la parada para este ajuste.

Figura 27

Caso 2 - Gráfico SPE - Muestra 1: Operación previa a calibración de álabes (26/02/2021 17:03)

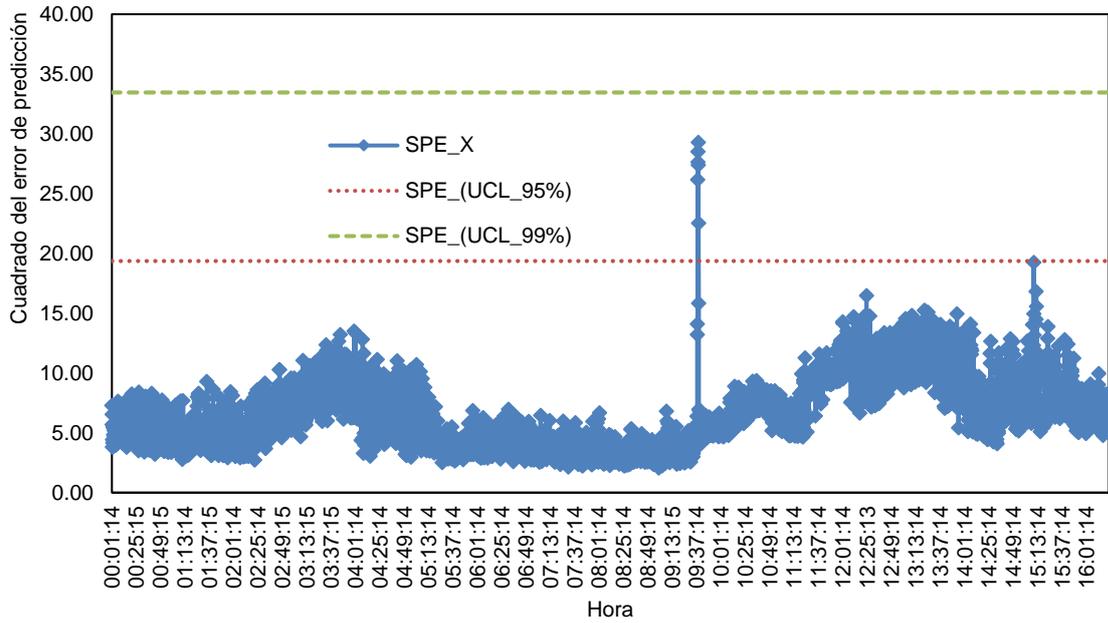


Figura 28

Caso 2 - Gráfico T2 de Hotelling - Muestra 1: Operación previa a calibración de álabes (26/02/2021 17:03)

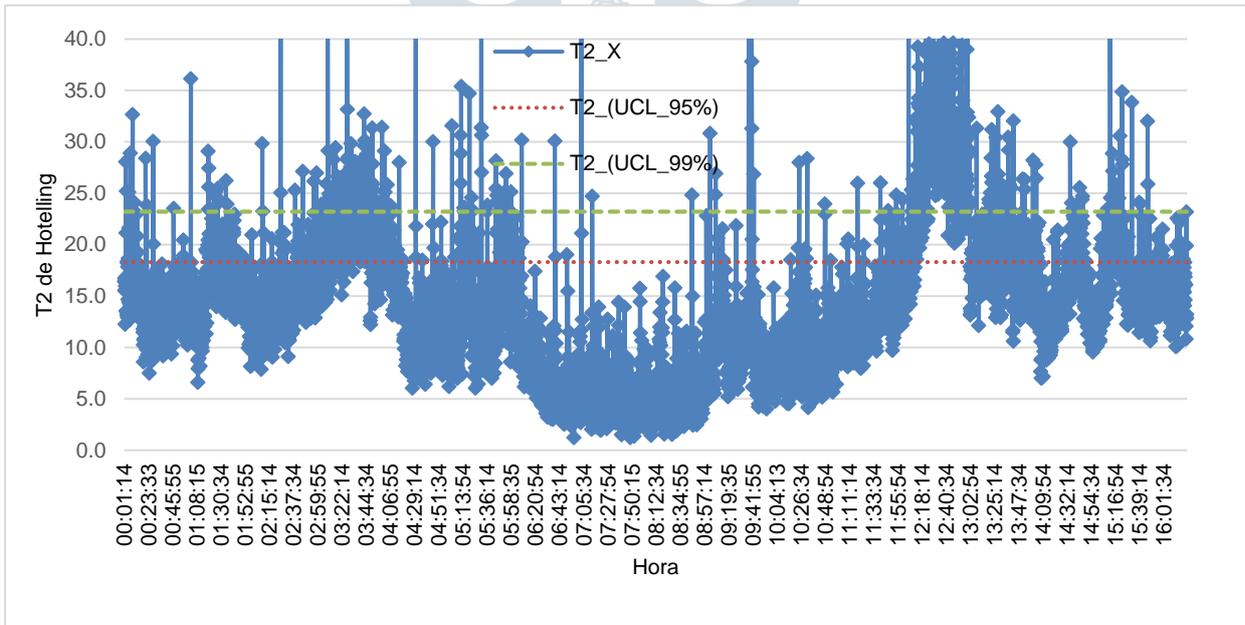


Figura 29

Caso 2 - Gráfico SPE - Muestra 2: Operación con vertedero funcionando (diferentes días)

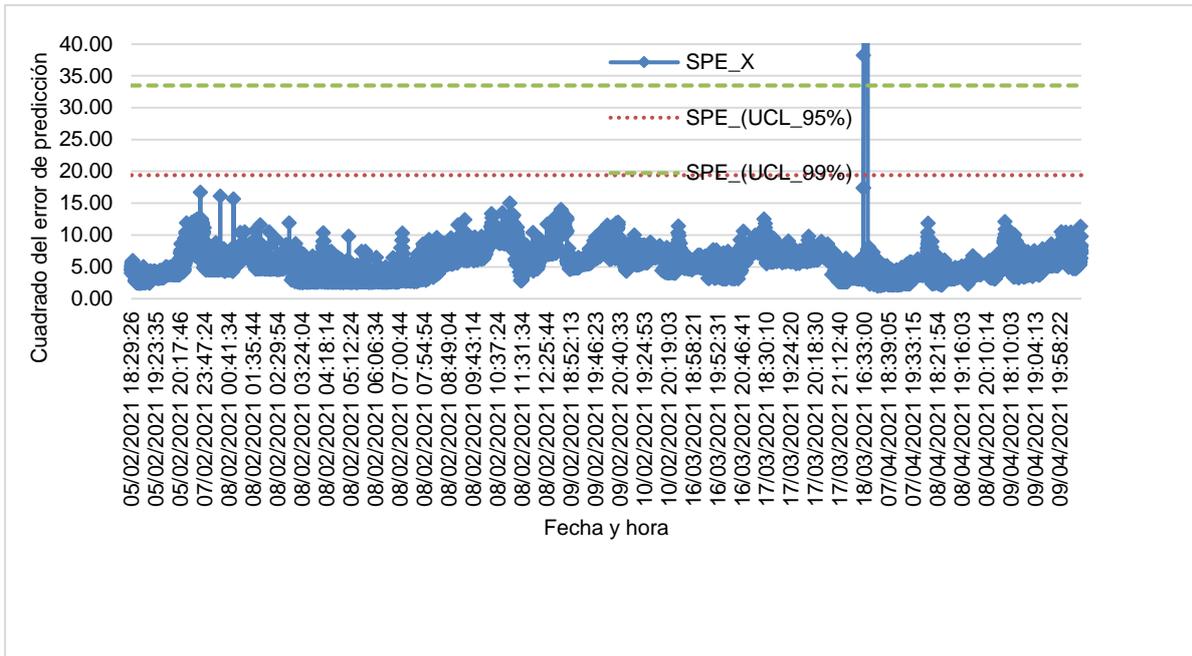
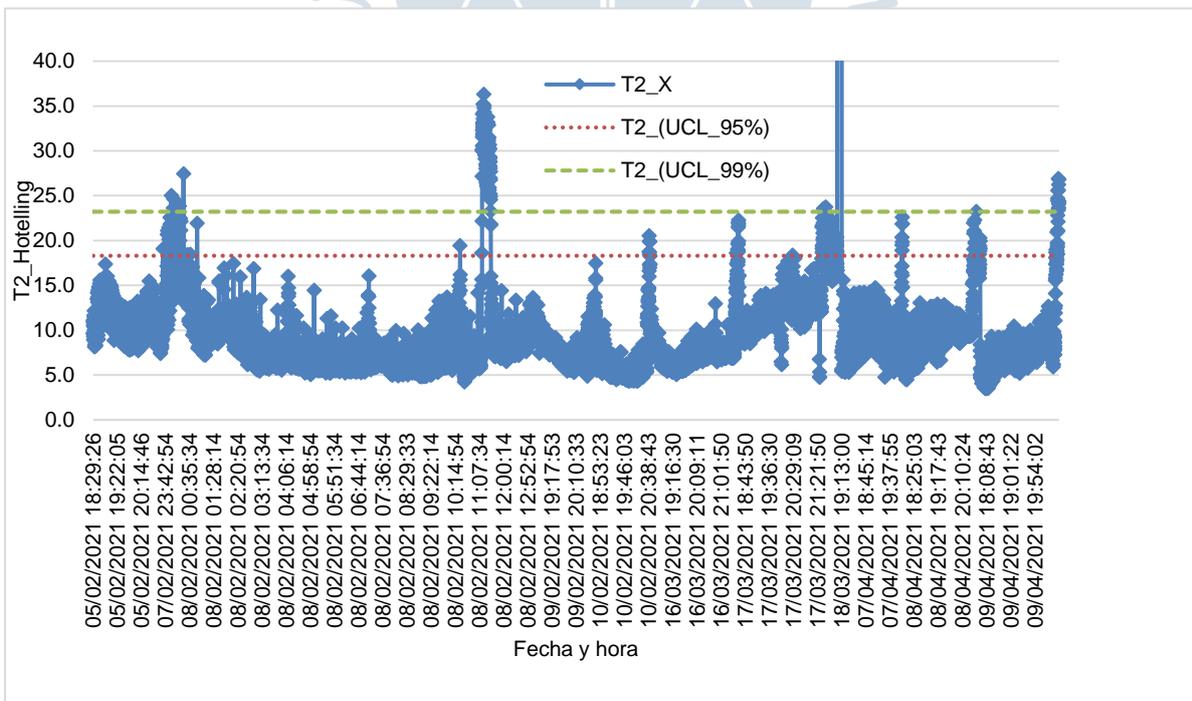


Figura 30

Caso 2 - .Gráfico T2 de Hotelling - Muestra 2: Operación con vertedero funcionando (diferentes días)



La

Figura 29 y la

Figura 30 corresponden a una operación de la unidad descargando agua por el vertedero de demasías al existir más recurso hídrico que el que la turbina puede utilizar. En la figura 13 el modelo muestra un pico entre el 17/03 y 07/04 y corresponde a una parada de unidad ocurrida el 18/03 entre las 16:30 y 16:38. La Fig. 14 en cambio muestra que los puntos cuando se superan los límites superiores de control ocurren cuando el caudal descargado es alto (nivel de agua en una cota mayor a 64.00 msnm).

La Figura 31 y la Figura 32 muestra que estas observaciones previas a la parada manual para limpieza y calibración se no superaban los límites de operación normal, el incremento mostrado a las 22:49:17 corresponde al proceso de parada de la unidad. Las

Figura 33 y la Figura 34 muestran que la parada intempestiva por fallas en la unidad de alta presión del gobernador de la turbina no puede ser previsto con los datos disponibles. Las variables registradas no están incluyendo presiones del gobernador, por ejemplo. Correspondería sugerir la adquisición de estos datos como se tiene en instalaciones más modernas. En las Figura 35 y Figura 36 lo que se muestra es el comportamiento del modelo ante variaciones datos manuales (simuladas). Estos incrementos pueden apreciarse con más claridad en el gráfico SPE, mientras que en el gráfico T2 de Hotelling estas variaciones se muestran más levemente.

Figura 31

Caso 2 - Gráfico SPE - Muestra 3: Operación previa a limpieza de rejillas y calibración de álabes (03/02/2021 22:46)

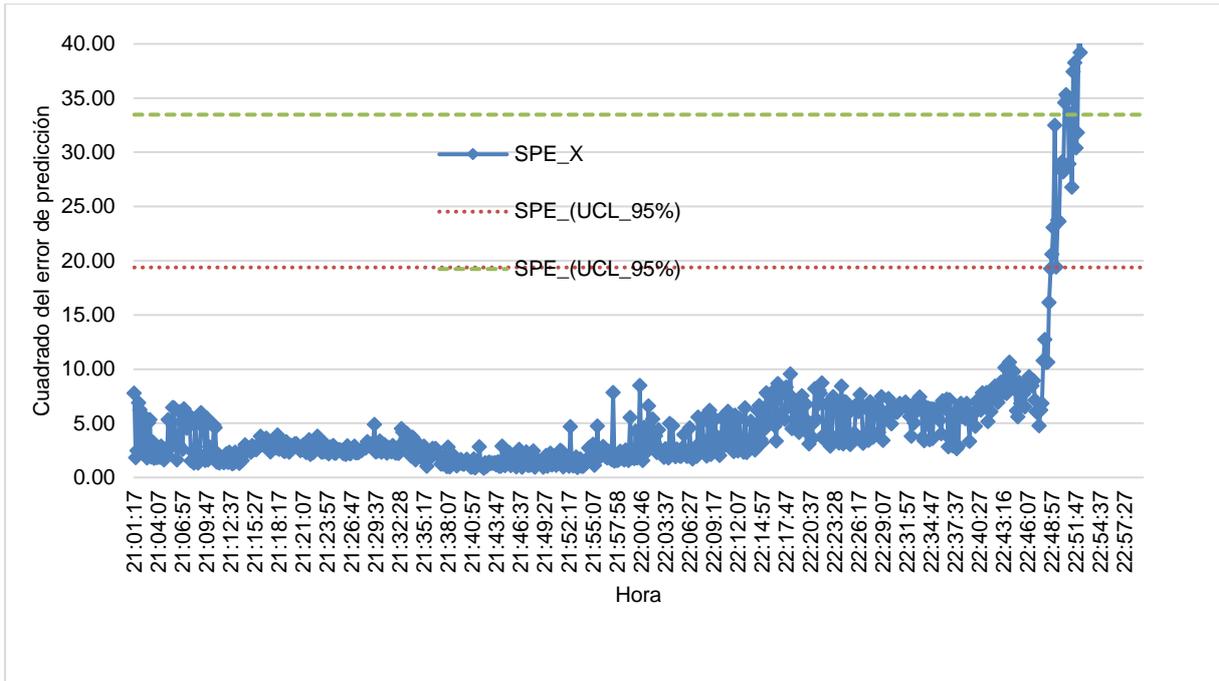


Figura 32

Caso 2 - Gráfico T2 de Hotelling - Muestra 3: Operación previa a limpieza de rejillas y calibración de álabes (03/02/2021 22:46)

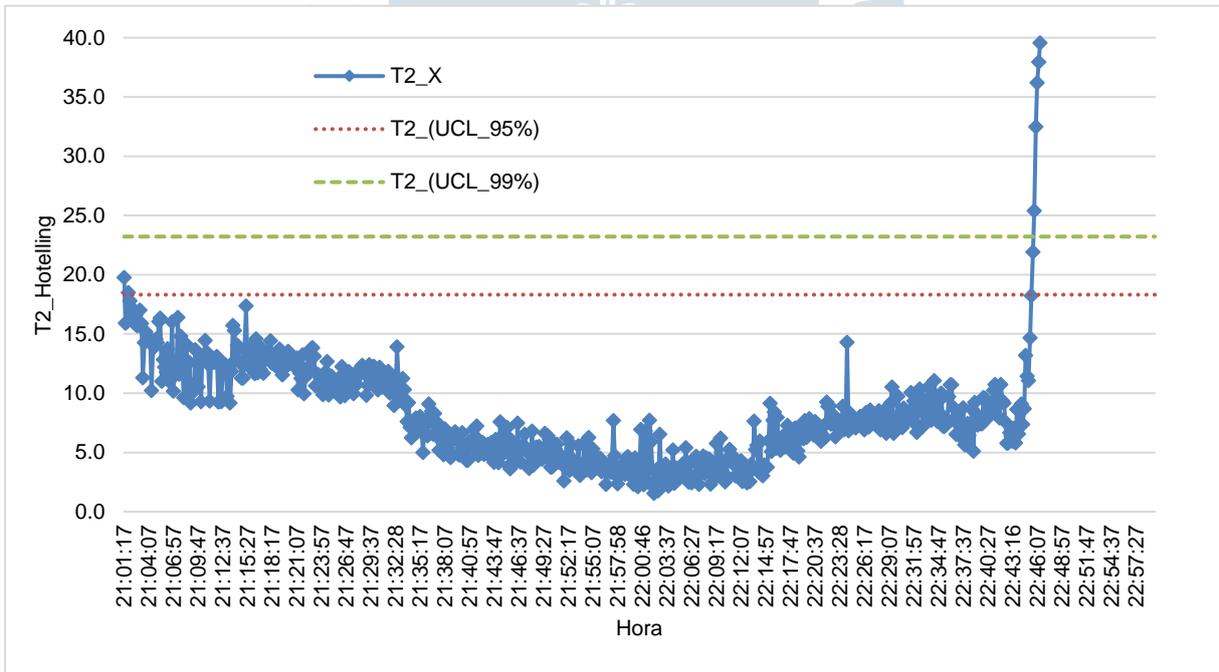


Figura 33

Caso 2 - Gráfico SPE - Muestra 4: Operación previa a parada por falla en Unidad de Alta Presión. Parada a las 20:48:45 del 10/03/2021

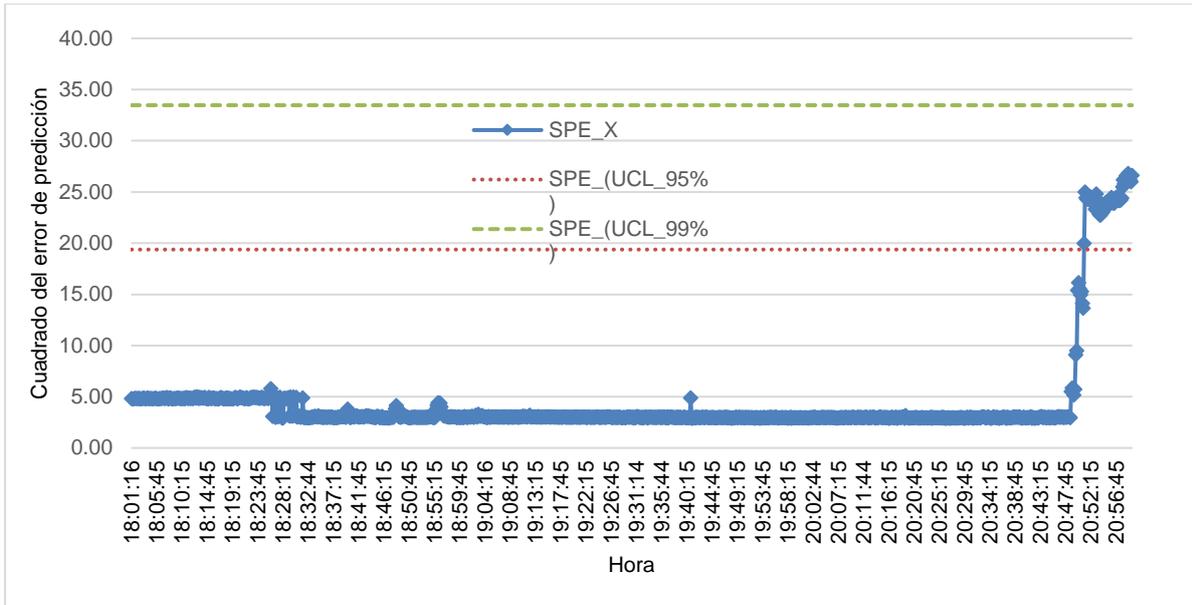


Figura 34

Caso 2 - Gráfico T2 de Hotelling - Muestra 4: Operación previa a parada por falla en Unidad de Alta Presión. Parada a las 20:48:45 del 10/03/2021

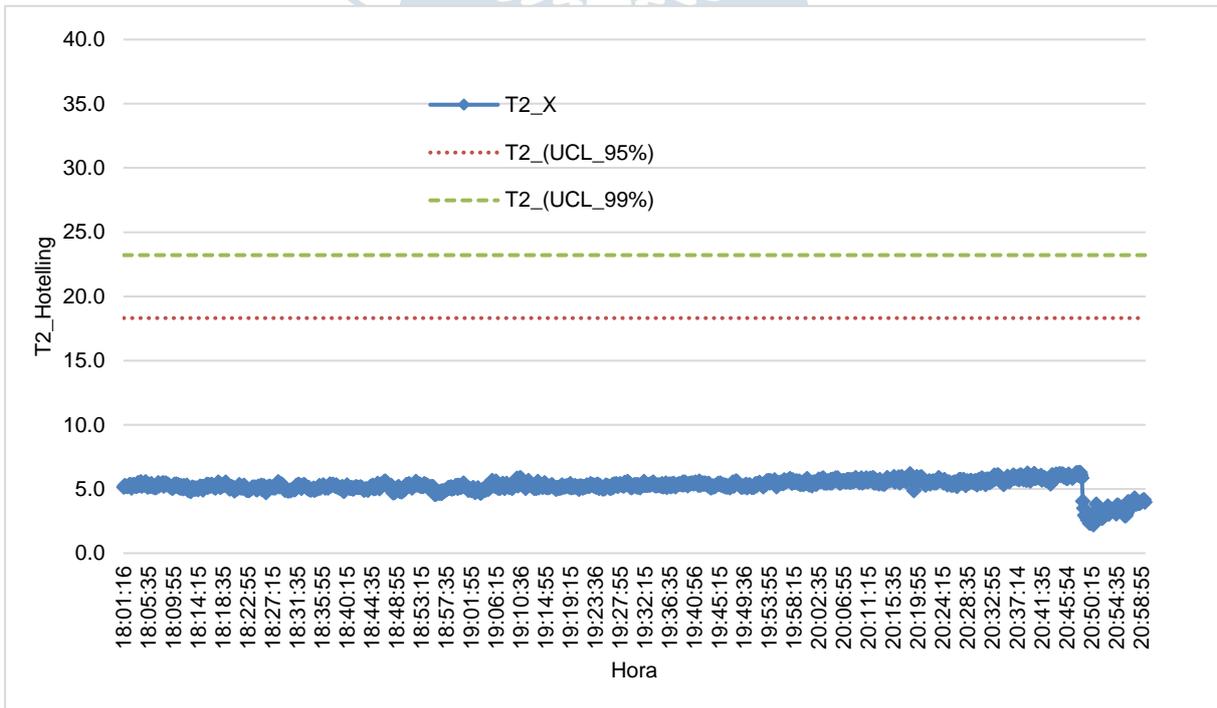


Figura 35

Caso 2 - Gráfico SPE - Muestra 5 - Simulación de alteración de variables BT2515, BT3403, BT3404, BT2516 y comportamiento en parada y arranque de unidad del 10/03/2021

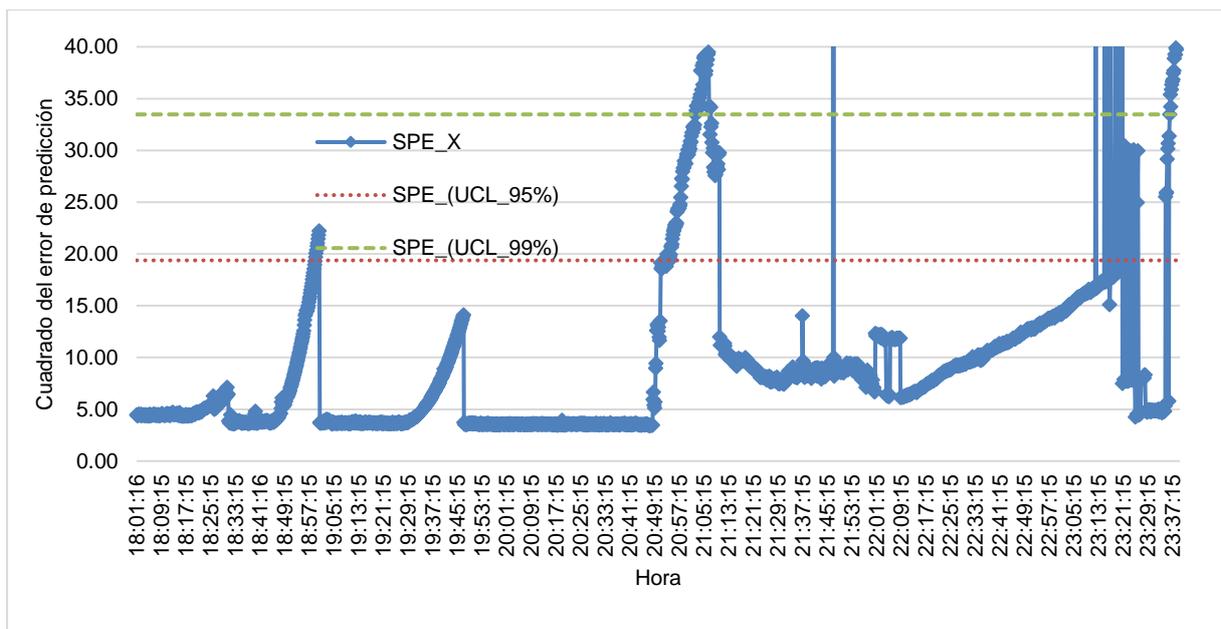
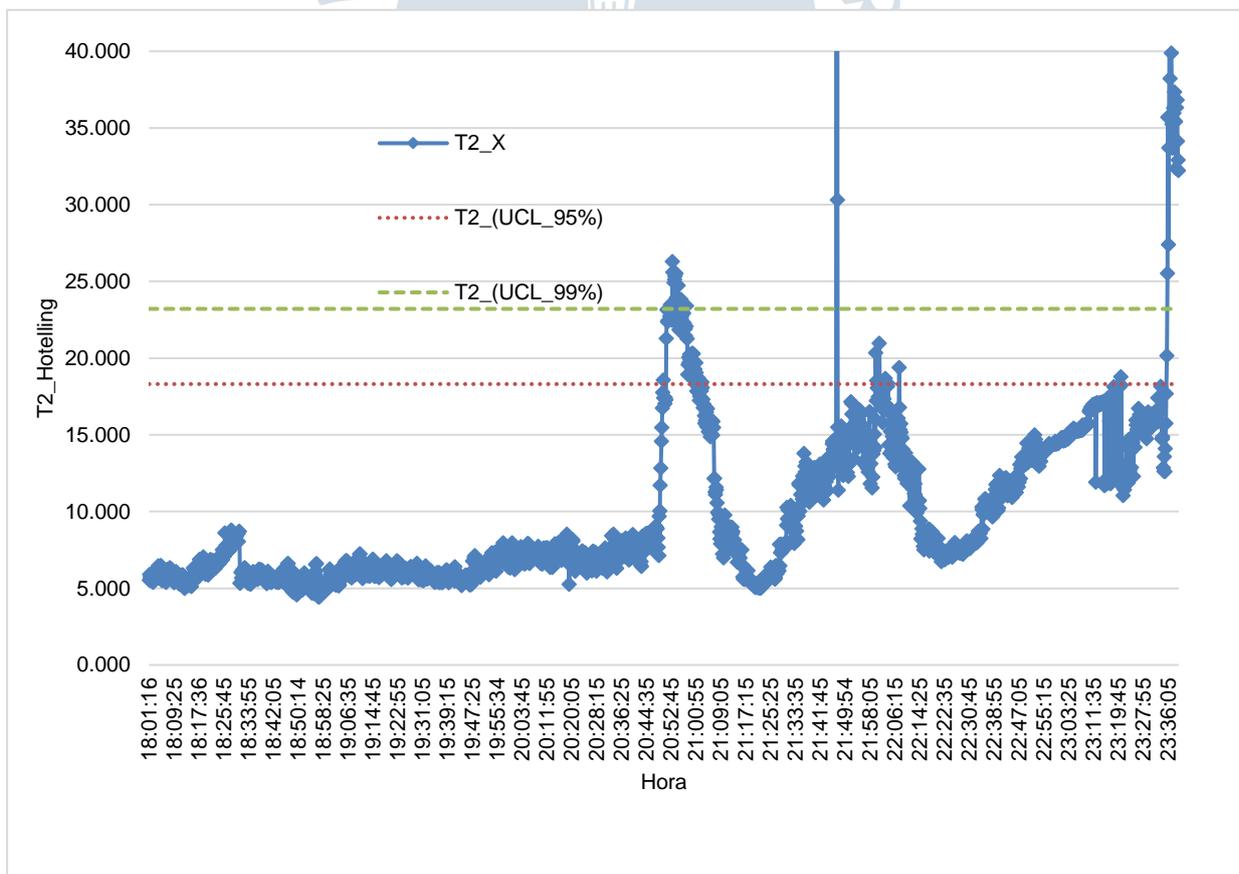


Figura 36

Caso 2 - Gráfico T₂ de Hotelling - Muestra 5 - Simulación de alteración de variables BT2515, BT3403, BT3404, BT2516 y comportamiento en parada y arranque de unidad del 10/03/2021

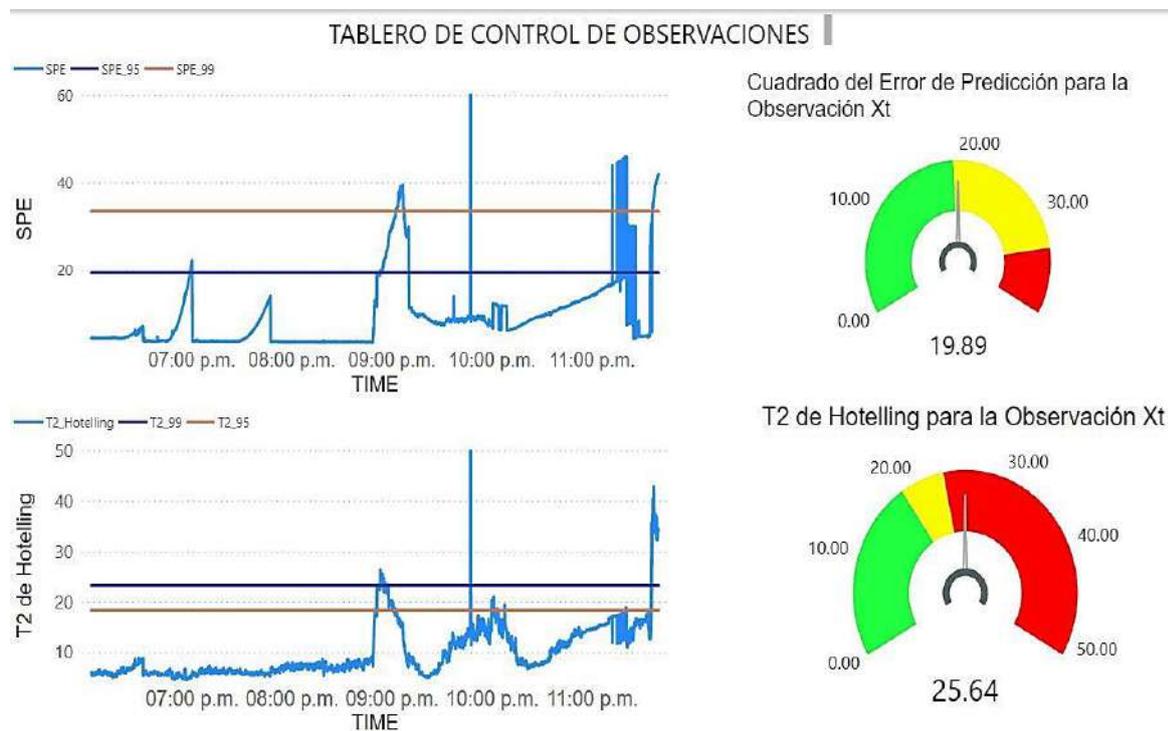


4.4.2.6 Paso 6: Implementación. La implementación, fase que tiene la finalidad de visualización del comportamiento de nuevas observaciones con fines de supervisión y mantenimiento, se ha elaborado con las tablas generadas en R (archivos “*.xlsx”) conectadas a una herramienta de inteligencia de negocios, en este caso, Microsoft Power BI, tal como ya se ha hecho en (Ayad et al., 2020). En Power BI se elaboraron dos visualizaciones correspondientes a las dos graficas de control más dos visores con los límites de control integrados en un tablero de control de observaciones. En la Figura 37 se muestra una captura de pantalla del tablero de control de observaciones con los datos de la muestra 5. La visualización con la herramienta de BI resulta ser de más fácil despliegue y de rápido seguimiento y mantenimiento.



Figura 37

Caso 2 - Captura de pantalla del despliegue del Gráfico SPE - Muestra 5 en Power BI



4.4.2.7 Conclusiones del caso. De acuerdo con los resultados del caso se pueden extraer las siguientes tres conclusiones principales: 1) El modelo de referencia CRISP-DM empleado como modelo de desarrollo de una solución de minería de datos, puede ser utilizada en sistemas de control de unidades de generación hidráulica con un SCADA implementado, de tal forma que se puede aprovechar los datos generados para generar la información y conocimiento necesario para tomar decisiones oportunas. 2) No todas las situaciones y comportamientos anómalos pueden ser detectados, esto depende de que las variables del sistema causante de la anomalía estén contempladas en el modelo y 3) Aunque los modelos y despliegues pueden ser desarrollados con relativa facilidad con herramientas informáticas como las utilizadas en este trabajo, de todas maneras, el conocimiento del proceso es muy importante para un adecuado análisis e interpretación de resultados.

4.4.2.8 Comentarios a los resultados obtenidos. En este caso se ha aplicado la metodología en una empresa que cuenta con un sistema SCADA. Se trata de una central de mediana potencia donde se analizó uno de sus turbo generadores. Los datos del sistema SCADA se utilizaron para determinar un patrón de operación normal. La anomalía de la operación fue evaluada mediante gráficos estadísticos. Se mostró que la conexión de los datos hacia la herramienta BI utilizada fue relativamente sencilla junto con la creación de los dashboards (tableros de control) correspondientes.

El caso muestra que el monitoreo de parámetros de operación tiene también una enorme utilidad. Hoy esto ya no es inalcanzable a una PYME. Los sistemas SCADA eléctricos y

de diferentes procesos están cada vez más desplegados. Sin embargo, el uso que se da a estos datos es limitado. Sirven para el análisis de fallas luego de ocurridas permitiéndoles identificar la causa de la falla y con ello tomar las medidas correctivas, arrancar el proceso nuevamente y planificar acciones que eviten que vuelva a ocurrir. Con el método aplicado y los datos disponibles de un turbogenerador (pero escalable a cualquier otro proceso) con el control de gráficos estadísticos muy utilizados y de libre acceso se pudo implementar un modelo de verificación de patrón normal de operación. Los cálculos y modelos se generaron con aplicaciones de acceso abierto o de amplio uso como R, Excel, y Power BI, se usó también algoritmos de reducción dimensional ampliamente desarrollados y aplicados como es el análisis de componentes principales (PCA) que ya tiene librerías desarrolladas en R y en otros lenguajes de programación.

El caso también mostró que aun cuando la herramienta funciona, no deja de tener importancia el registro de datos de sensores relacionados con la parte del proceso que puede ocasionar la causa. Los resultados pueden motivar a las empresas a medir aquellas partes del proceso que han ocasionado paradas intempestivas en el pasado a fin de que ingresen al modelo y pueda ser pronosticado un funcionamiento anormal o fuera de patrón futuros.

4.4.3 Caso 3: Predicción de producción hidráulica

En este caso se plantea aplicar la metodología propuesta en una empresa de generación eléctrica de mediana potencia peruana. Estas empresas requieren anualmente proyectar su producción del año siguiente tanto para reportar esta proyección a los reguladores y operadores del sistema y para proyectar sus ingresos.

La práctica más usual es sacar una proyección mensual de los años anteriores y utilizar estos promedios como proyecciones.

Basándonos en los registros de producción de años pasados planteamos aplicar la metodología propuesta para afinar estas proyecciones y mejorar la precisión de estas.

4.4.3.1 Paso 1: Entendiendo el negocio. La empresa es una empresa concesionaria de generación eléctrica que debe proyectar anualmente la producción de sus plantas de generación. Elegimos arbitrariamente una de sus plantas con 12,5 MW de capacidad instalada. La planta ha estado operando desde 1998. El objetivo final es estimar la producción mensual del año siguiente.

Los tipos de análisis a realizar serían uno descriptivo para visualizar el comportamiento pasado y otro predictivo para poder estimar la producción al año siguiente.

4.4.3.2 Paso 2: Entendiendo los datos. Los datos provienen de un archivo generado por la administración de operaciones y actualizado mes a mes, como puede verse en la

4.4.3.3 Figura 38.



Figura 38

Caso 3 - Registro de producción mensual de la planta desde 1998

PRODUCCIÓN HISTÓRICA MENSUAL DE SINERSA A : DICIEMBRE 2020														TOTAL SINERSA	TOTAL SINERSA
														MWh	GWh
														3,268,696	3,269
CENTRAL HIDROELECTRICA CURUMUY														ACUMULADO TOTAL MWh	ACUMULADO TOTAL GWh
		ENERO	FEBRERO	MARZO	ABRIL	MAYO	JUNIO	JULIO	AGOSTO	SEPTIEMBRE	OCTUBRE	NOVIEMBRE	DICIEMBRE		
AÑO 1998	MWh	790	1785	4867	8242	4335	4189	1700	5958	5063	4988	4484	5180	49562	50
AÑO 1999	MWh	2290	5418	7851	8050	8541	8266	7053	4575	3853	2369	1023	1516	60806	61
AÑO 2000	MWh	5288	5127	7315	8104	8412	6125	4301	4401	3848	4110	3598	3401	64033	64
AÑO 2001	MWh	5867	7165	8035	8122	8515	8436	8645	7734	5453	2888	3287	3757	77903	78
AÑO 2002	MWh	4048	4715	8305	6899	5875	5555	6843	4949	3926	4490	4441	3720	63763	64
AÑO 2003	MWh	4730	4828	3799	5951	7043	4759	1872	3804	3456	3555	2734	2776	48306	48
AÑO 2004	MWh	5794	4043	5804	5547	4758	2628	1174	1677	1568	2060	1021	2221	38234	38
AÑO 2005	MWh	5372	4772	8402	7770	4552	2485	563	2171	4551	3460	2599	1965	48643	49
AÑO 2006	MWh	2441	7437	8145	7828	5740	4144	1144	888	1201	1435	165	4414	44982	45
AÑO 2007	MWh	5732	7460	4088	6046	4932	4053	1549	2742	3341	3424	2073	3403	48843	49
AÑO 2008	MWh	4843	7539	7819	7752	8427	3013	5025	1812	3329	3367	4466	3131	60524	61
AÑO 2009	MWh	7033	7574	8352	7958	6325	7400	6585	5317	4601	3951	3559	2868	71523	72
AÑO 2010	MWh	4235	7071	7542	7368	7431	4283	786	3040	4314	3733	2581	1963	54347	54
AÑO 2011	MWh	2628	3797	6807	6040	6201	4710	5950	3608	2856	2734	2189	5195	52096	52
AÑO 2012	MWh	8541	7618	8062	7875	8404	5526	5108	2607	3386	2827	3266	2650	65871	66
AÑO 2013	MWh	6718	7237	5989	5285	5702	153	267	1958	5821	3420	2013	1993	47535	43
AÑO 2014	MWh	1409	1955	4783	6127	6685	3473	2176	3185	3520	3400	2760	2541	41794	42
AÑO 2015	MWh	2184	5750	8201	7792	7312	6711	2481	2475	4036	3507	2437	2878	55761	56
AÑO 2016	MWh	1266	6245	5680	7067	7574	4239	2220	3030	3823	3537	1480	1296	47457	47
AÑO 2017	MWh	1422	7099	7137	5090	7810	8030	7034	3600	2109	2362	2354	1671	55620	56
AÑO 2018	MWh	2771	5906	6511	5865	6557	2581	2377	1203	2384	2135	767	279	39136	39
AÑO 2019	MWh	2703	6018	5838	5106	6031	5196	1842	2588	3442	3257	2856	1938	46815	47
AÑO 2020	MWh	2114	6223	7553	6043	6551	3110	1807	2842	2987	3205	1935	2483	46853	47
														1,225,409	1,225

4.4.3.4 Paso tres: Preparación de los datos. Los datos requirieron una transformación de formato para adaptarlo a una serie temporal que se puede ver en la Figura 39. Además, se creó una tabla de planta de energía adicional para generar un modelo de estrella de vértice único y una tabla de medición. El modelo relacionaba dos tablas por código de planta y se creó una tabla de medición. Ver

4.4.3.5 Figura 40.

Figura 39

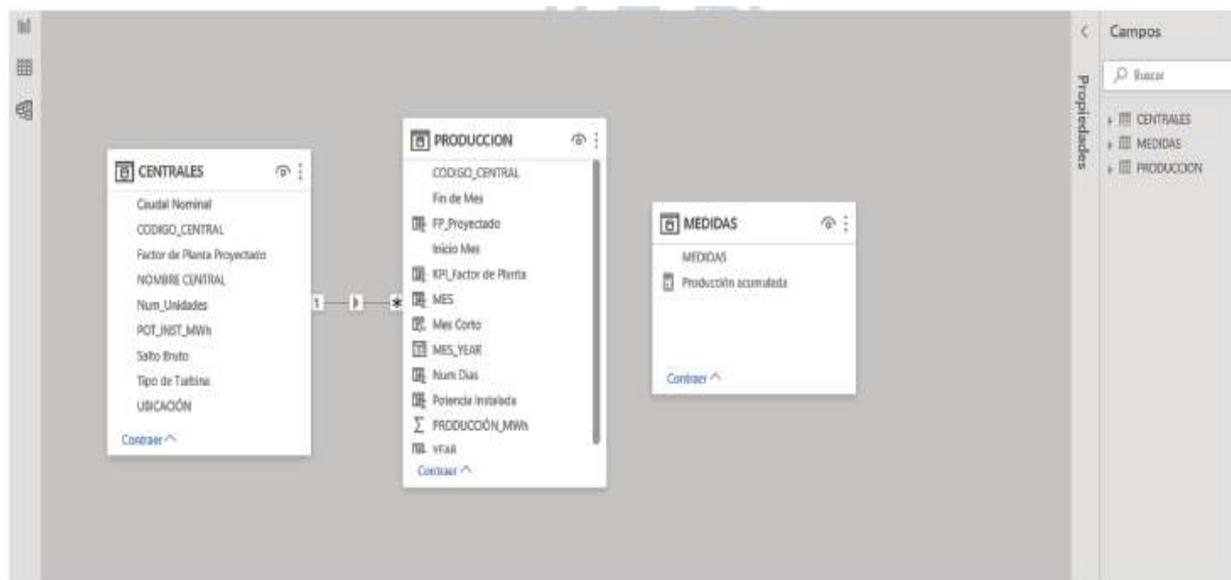
Caso 3 - Datos cargados después de limpieza, ordenamiento, transformaciones y cálculos

CODIGO_CENTRAL	MES_YEAR	PRODUCCION_MWH	Potencia Instalada	MES	PLANT	Inicio Mes	Fin de Mes	Num Dias	KPI_Factor de Planta	Mes Corto	FP_Proyectado
CHC	enero de 1998	289.685	12.5	3	1998	01/01/1998	31/01/1998	31		0.08	0.0
CHC	febrero de 1998	2785.804	12.5	3	1998	01/02/1998	28/02/1998	28		0.23	0.0
CHC	marzo de 1998	4847.438	12.5	3	1998	01/03/1998	31/03/1998	31		0.53	0.0
CHC	abril de 1998	6242.844	12.5	4	1998	01/04/1998	30/04/1998	30		0.69	0.0
CHC	mayo de 1998	4535.000	12.5	3	1998	01/05/1998	31/05/1998	31		0.47	0.0
CHC	junio de 1998	4105.000	12.5	6	1998	01/06/1998	30/06/1998	30		0.47	0.0
CHC	julio de 1998	2700.134	12.5	7	1998	01/07/1998	31/07/1998	31		0.18	0.0
CHC	agosto de 1998	5837.740	12.5	8	1998	01/08/1998	31/08/1998	31		0.64	0.0
CHC	septiembre de 1998	5053.078	12.5	9	1998	01/09/1998	30/09/1998	30		0.56	0.0
CHC	octubre de 1998	4847.556	12.5	30	1998	01/10/1998	31/10/1998	31		0.54	0.0
CHC	noviembre de 1998	4844.133	12.5	11	1998	01/11/1998	30/11/1998	30		0.50	0.0
CHC	diciembre de 1998	5108.405	12.5	11	1998	01/12/1998	31/12/1998	31		0.55	0.0
CHC	enero de 1999	2309.979	12.5	3	1999	01/01/1999	31/01/1999	31		0.25	0.0
CHC	febrero de 1999	3417.760	12.5	2	1999	01/02/1999	28/02/1999	28		0.64	0.0
CHC	marzo de 1999	7851.254	12.5	3	1999	01/03/1999	31/03/1999	31		0.84	0.0
CHC	abril de 1999	8040.834	12.5	4	1999	01/04/1999	30/04/1999	30		0.89	0.0
CHC	mayo de 1999	8540.833	12.5	5	1999	01/05/1999	31/05/1999	31		0.92	0.0
CHC	junio de 1999	8258.131	12.5	6	1999	01/06/1999	30/06/1999	30		0.92	0.0
CHC	julio de 1999	7053.081	12.5	7	1999	01/07/1999	31/07/1999	31		0.76	0.0
CHC	agosto de 1999	4575.365	12.5	8	1999	01/08/1999	31/08/1999	31		0.40	0.0
CHC	septiembre de 1999	3833.360	12.5	9	1999	01/09/1999	30/09/1999	30		0.43	0.0
CHC	octubre de 1999	2889.234	12.5	30	1999	01/10/1999	31/10/1999	31		0.25	0.0
CHC	noviembre de 1999	2024.590	12.5	11	1999	01/11/1999	30/11/1999	30		0.11	0.0
CHC	diciembre de 1999	2315.788	12.5	11	1999	01/12/1999	31/12/1999	31		0.16	0.0
CHC	enero de 2000	5257.888	12.5	3	2000	01/01/2000	31/01/2000	31		0.97	0.0
CHC	febrero de 2000	3122.494	12.5	3	2000	01/02/2000	29/02/2000	29		0.59	0.0
CHC	marzo de 2000	7316.805	12.5	3	2000	01/03/2000	31/03/2000	31		0.79	0.0



Figura 40

Caso 3: Modelado de los datos



4.4.3.6 Paso cuatro: modelado. El modelo más relevante es el modelo predictivo. Para ello, se utilizaron modelos integrados en la plataforma PowerBI. Se utilizó el exponencial liso, ARIMA y uno basado en un modelo de Red Neuronal. Los resultados gráficos se pueden ver en la Figura 41, Figura 42 y

4.4.3.7 Figura 43.

Figura 41

Caso 3 – Predicción con el modelo Smooth Exponential

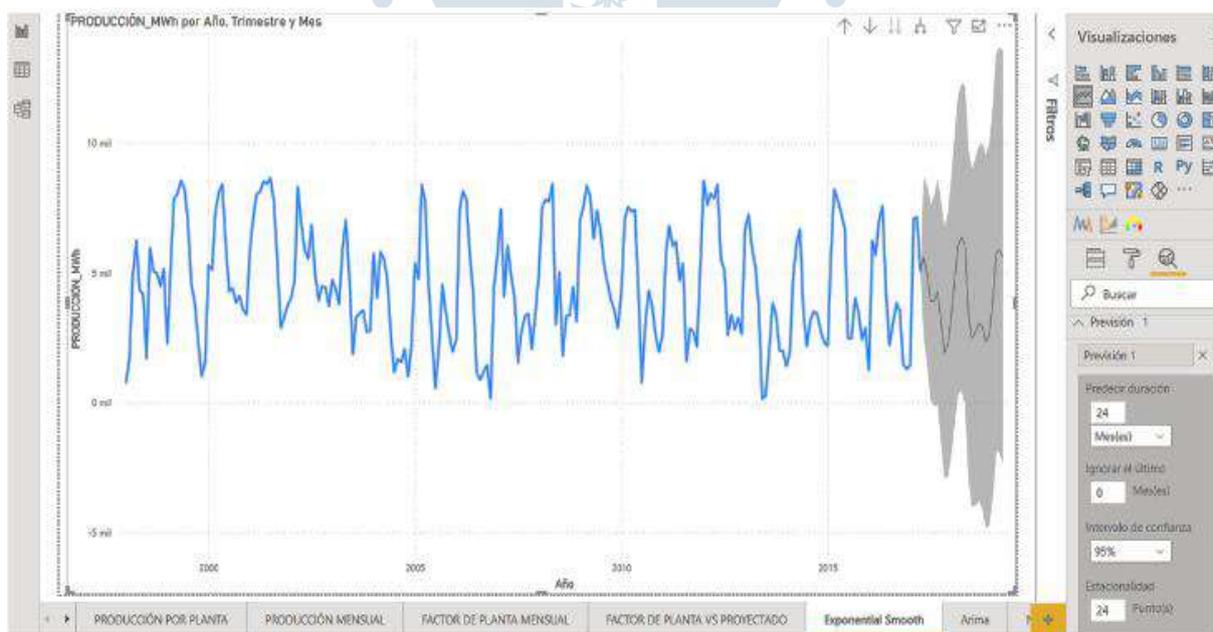
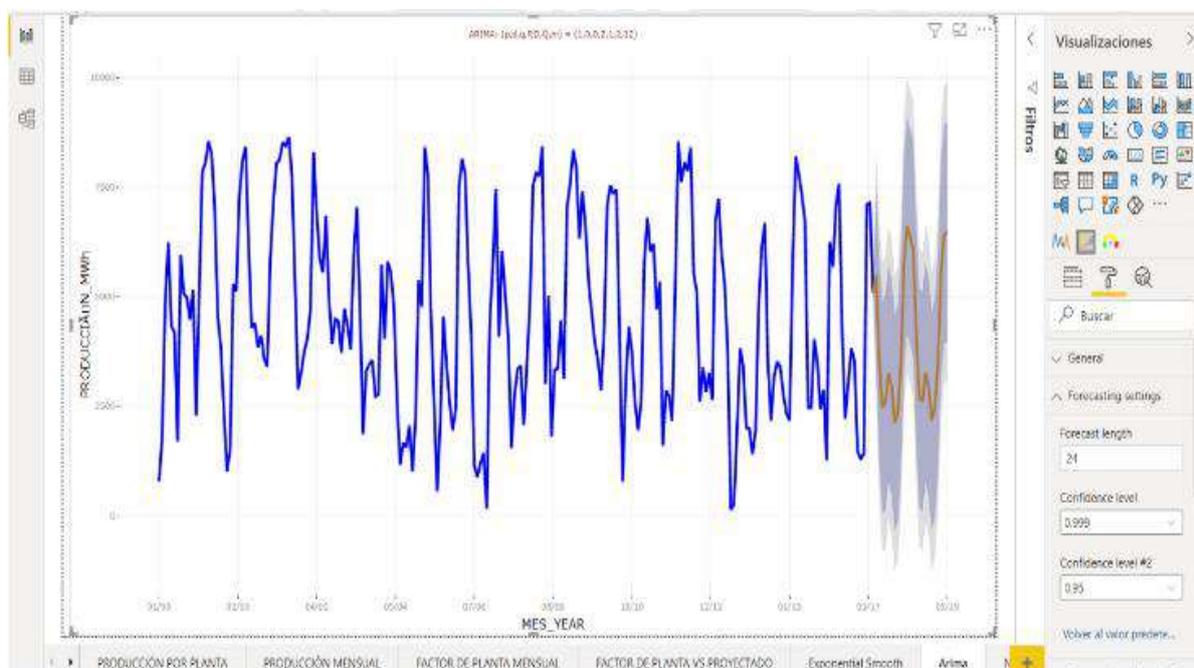
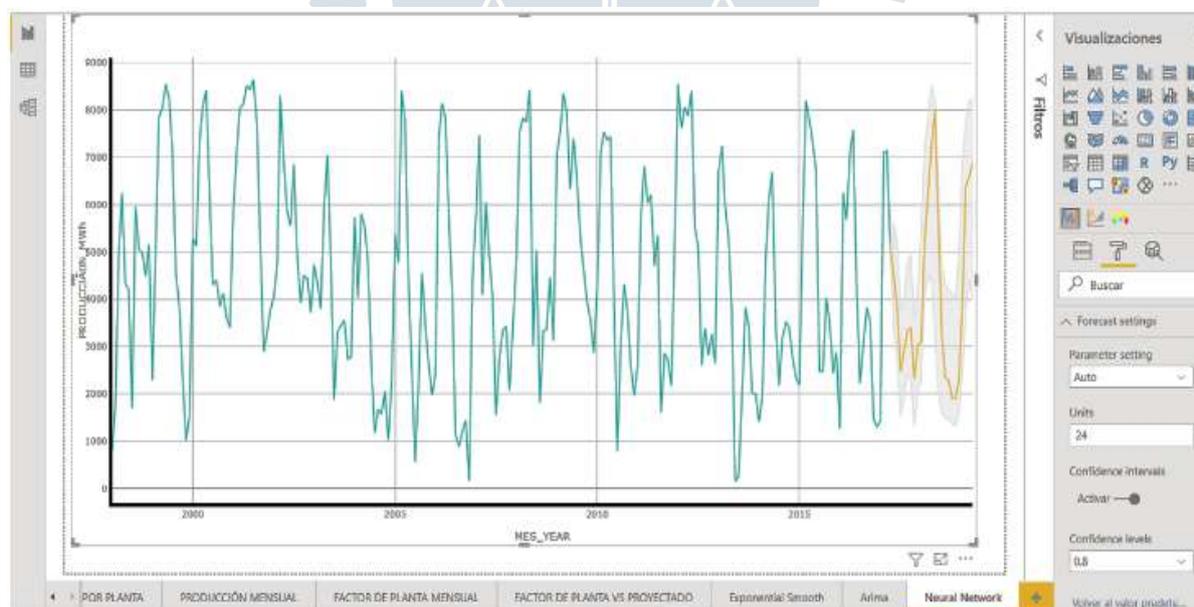




Figura 42*Caso 3 - Predicción con el modelo ARIMA***Figura 43***Caso 3 - Predicción con molo baso en Redes Neuronales*

Teniendo en cuenta que el objetivo de este trabajo es probar la metodología de implementación de un marco de trabajo de inteligencia de negocios basado en minería de datos en un caso concreto, no consideramos necesaria la explicación y profundización matemática de los modelos.

4.4.3.8 Paso cinco: Evaluación. Este paso presenta la posibilidad de confusión. Realmente, la evaluación que requiere el método es la evaluación frente a los objetivos del negocio o la minería de datos, no específicamente de los modelos. Sin embargo, en este caso concreto y en muchos otros, esta evaluación puede abarcar ambos aspectos. En este caso el mejor modelo es a la vez el que mejor responde a los objetivos de minería de datos, es decir, la predicción de la producción del próximo año, por lo tanto, la incluimos en esta fase.

La evaluación se realiza comparando los resultados de cada modelo con la producción real del año siguiente. Hemos generado el modelo con datos conocidos de 1998 a 2017 y comparado el modelo con la producción real del año 2018.

Esta comparación se puede ver en la Figura 44 y los resultados numéricos pueden observarse en la



Tabla 17 y en la Tabla 18.

La Tabla 18 muestra que el modelo con mejor performance evaluado bajo el criterio de MSE es el Smooth Exponential Model (SE Model).

Figura 44

Caso 3 - Comparación de los modelos con la producción real

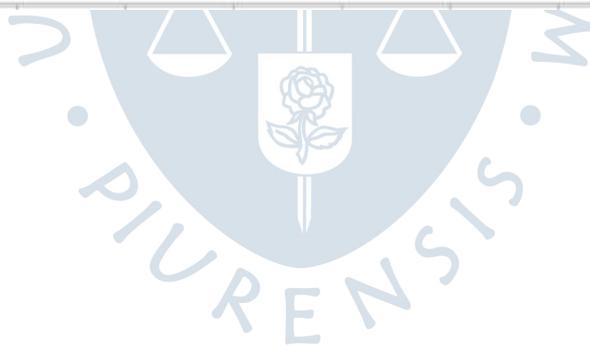
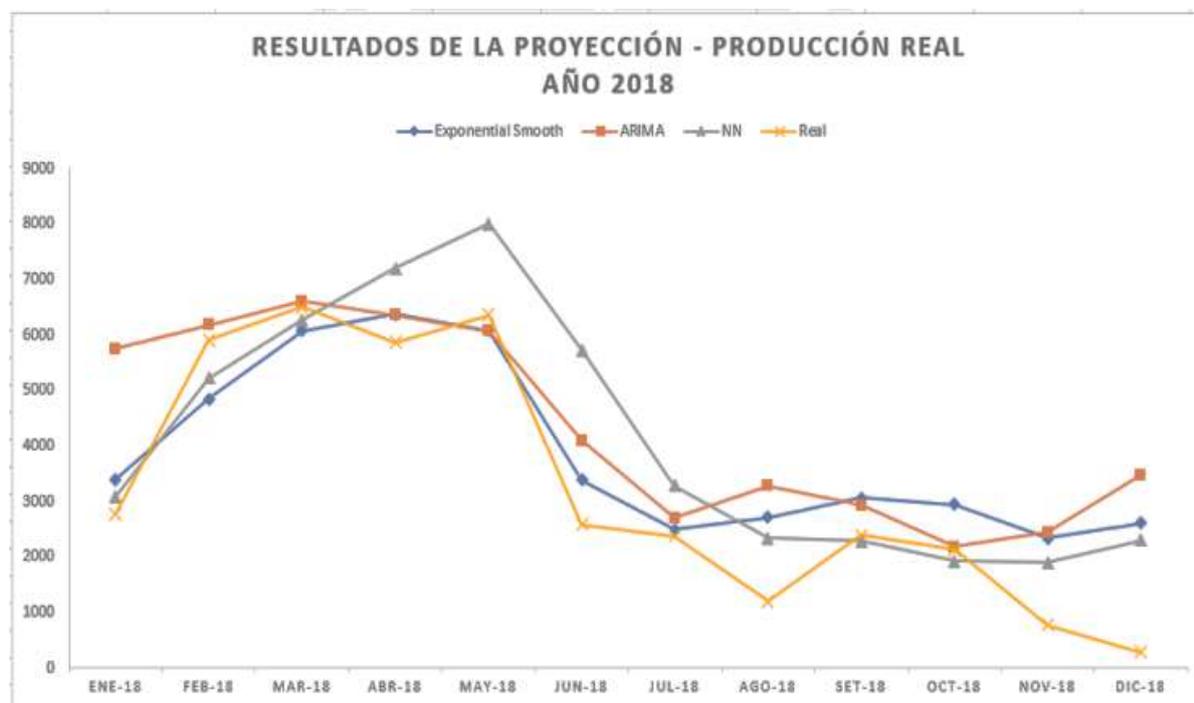


Tabla 17*Caso 3 - Resultados numéricos de las predicciones*

Predictions Results for 2018 in MWh				
Month	SE Model	ARIMA Model	NN Model	Real Production
01/2018	3393	5752	3090	2771
02/2018	4845	6183	5227	5906
03/2018	6084	6614	6271	6511
04/2018	6373	6365	7205	5865
05/2018	6073	6077	8005	6357
06/2018	3392	4097	5731	2581
07/2018	2494	2700	3288	2377
08/2018	2718	3280	2341	1203
09/2018	3061	2938	2280	2384
10/2018	2949	2184	1925	2131
11/2018	2334	2446	1894	767
12/2018	2603	3480	2303	279

Tabla 18*Caso 3 - Error cuadrático medio (MSE) de las predicciones*

Mean Square Error for the models				
Error	SE Model	ARIMA Model	NN Model	Mean Projection
MSE	1 165 351	2 449 482	1 883 277	2 266 310

4.4.3.9 Paso seis: Implementación. La implementación corresponde a presentar anualmente o mes a mes estas predicciones. Esto se haría según los criterios de la empresa o de los requerimientos de la alta dirección.

Lo relevante en este caso es que, al estar implementada en una plataforma de BI, como en este caso Power BI, esto permite una conexión al archivo con los datos actualizados lo que a su vez permitiría la posibilidad de una actualización manual o automática con la que se recalculará los parámetros del modelo obteniéndose una nueva proyección.

4.4.3.10 Conclusiones del caso. El presente caso muestra el uso de la metodología de minería de datos en un entorno de inteligencia de negocios para otro objetivo concreto en una empresa del sector eléctrico, obteniéndose resultados útiles y operables.

Los resultados no requieren software o hardware sofisticado, por lo que están al alcance de las pymes.

Aunque el objetivo del caso es mostrar la aplicación de la metodología y no profundizar en predicción en sí, fue posible definir un modelo adecuado de predicciones que mejora el rendimiento de la práctica de utilizar los promedios de años anteriores.

Finalmente, se muestra que cuando los objetivos de la minería de datos son claros y definidos, el modelo de referencia utilizado (CRISP-DM) sigue siendo válido.

4.4.3.11 Comentarios a los resultados obtenidos. En este caso se ha aplicado predicción de demanda utilizando series de tiempo y modelos de predicción muy utilizados (ARIMA, *Exponential Smoothing* y *Red Neuronal*), estos modelos pueden ser aplicados en la pequeña y mediana empresa que requiere soluciones probadas, de fácil aplicación y de mínimo costo o de acceso abierto.

La aplicación es útil en toda la cadena de valor del sistema eléctrico, por ejemplo: el generador puede prever sus posibles ingresos y planificar sus inversiones y flujos de caja. Le permite reportar adecuadamente sus proyecciones de producción a las entidades correspondiente con el efecto que esto tiene en la planificación de la operación de todo el sistema eléctrico nacional.

- Por otro lado, el pronóstico de la demanda permite a un distribuidor, planificar también sus inversiones en infraestructura, priorizando las áreas donde el pronóstico de demanda muestra mayores incrementos y evitando sobre inversión en zonas donde el incremento pronosticado no lo justifica.
- Finalmente, en el caso de un consumidor el pronóstico le permite identificar áreas de alto consumo y posibilidades de optimización de los procesos productivos, variar el diagrama de carga evitando los consumos en horas de punta o de facturación más alta. Esto les permite evaluar y por ende determinar una posible mejor opción tarifaria.

4.4.4 Requerimientos y comentarios sobre la metodología

La metodología ha sido ejecutada con equipamiento informático convencional y al alcance de un pyme.

Se utilizó una laptop con las siguientes características:

- Procesador : Intel(R) Core(TM) i7-9750HF CPU @ 2.60GHz 2.59 GHz
- RAM instalada : 16.0 GB
- Sistema Operativo : Windows 11 Home Single Language

Para el pre-procesamiento y procesamiento de los datos, el modelado y visualización de los resultados de la metodología empleada se ha utilizado:

- Se utilizó una cuenta de Office 365 – Familia con licenciamiento anual.
- Una cuenta Power BI – Tipo de licencia: Gratuita

- RStudio – 2021.09.0 Build 351
- R versión 4.1.1 (2021-08-10)
- La adquisición de datos se hizo fuera de línea.

Los datos tuvieron las siguientes características:

- Reportes en formato Excel. Para la predicción.
- Archivos en formatos dbf, provenientes del sistema SCADA archivos diarios de registro de parámetros de operación cada 10 seg.
- Para la calidad de energía se usó el archivo en formato CSV, obtenido del analizador de redes HIOKI 3169, luego de una medición semanal.

El trabajo de ingeniería de datos, análisis de datos y desarrollos de AI fueron desarrollados por el investigador, haciendo uso de las herramientas aprendidas en el programa de doctorado.

Para el análisis estadístico si se requirió el apoyo de un especialista en estadística para adquirir el conocimiento de las herramientas y gráficos utilizados.

Si bien es cierto la metodología utilizada se basa en un modelo de referencia de más de veinte años de creación, los casos han mostrado que sigue siendo muy útil su uso como modelo de referencia.

Los casos han mostrado también que la fase de evaluación del modelo y de evaluación del proceso de minería de datos se pueden traslapar para aplicaciones específicas como es el caso de la gestión de suministro eléctrico, donde el objetivo es muy concreto y definido.

La literatura actual como en (Martinez-Plumed et al., 2021), sigue planteando que para cuando los objetivos de la minería de datos son muy concretos y específicos, el modelo de referencia CRISP-DM sigue siendo útil y tiene vigencia.

En el mismo trabajo (Martinez-Plumed et al., 2021), plantea que un alto porcentaje de proyectos de *data science* entre los que están, obviamente los de inteligencia de negocios, no siguen una metodología. También se indica que alrededor del 43% utiliza CRISP-DM. De 19 metodologías analizadas el punto débil de CRISP-DM es la de integración del trabajo en equipo y sus puntos fuertes es su centralización en el proyecto, en los datos y la información. Teniendo en cuenta esto, su elección para ser usado como modelo de referencia para proyectos de inteligencia de negocios basado en minería de datos en proyectos específicos de mejora de la gestión del suministro eléctrico nos parece válida, más aún cuando en una pyme quizá la limitación del trabajo en equipo no afecte los resultados pues el número de integrantes y áreas del mismo es muy bajo y por tanto, con posibilidades a ser desarrollada por una o dos personas.

Conclusiones

De acuerdo con los resultados de este trabajo y de su discusión, de la comparación con trabajos de otros autores y del análisis presentado se pueden extraer las siguientes conclusiones principales:

Aunque hay mucho escrito sobre la importancia de los datos como generador de conocimiento mediante el uso de inteligencia artificial, minería de datos, inteligencia de negocios, etc.; en nuestro país y en especial en la pequeña y mediana empresa, no se aprovecha este potencial.

La brecha entre el estado de las investigaciones sobre el uso de la inteligencia de negocios y todos sus derivados y su uso en la industria nacional, especialmente el sector PYME es grande.

Una metodología simple, probada, de fácil entendimiento y aplicación resulta útil para poder implementar estas herramientas de inteligencia de negocios en la PYME. Esta metodología ha sido implementada de manera empírica en dos tipos de empresa, pero a aspectos de la gestión del suministro eléctrico que puede ser generalizado a otras empresas del sector.

Los casos de estudio desarrollados han mostrado que, con hardware y software accesible a la PYME, se puede seguir la metodología y obtener resultados útiles para la toma de decisiones.

Desde que se mencionó y nombró por primera vez la inteligencia de datos o desde que el modelo usado (CRISP-DM) fue lanzado, hasta la actualidad nuevas herramientas han surgido, las investigaciones y usos crecen día a día y actualmente lo que la inteligencia de negocio y minería de datos realiza están incluidos en lo que se conoce como ciencia de datos aplicada.

La ciencia de datos requiere habilidades integradoras que actualmente no están muy disponibles en el mercado laboral y mucho menos en la PYME. El trabajo ha demostrado que se requieren conocimiento de inteligencia artificial, estadística aplicada, manejo de datos para poder lograr resultados útiles. Esto es quizá la principal limitación para que la PYME implemente estas herramientas.

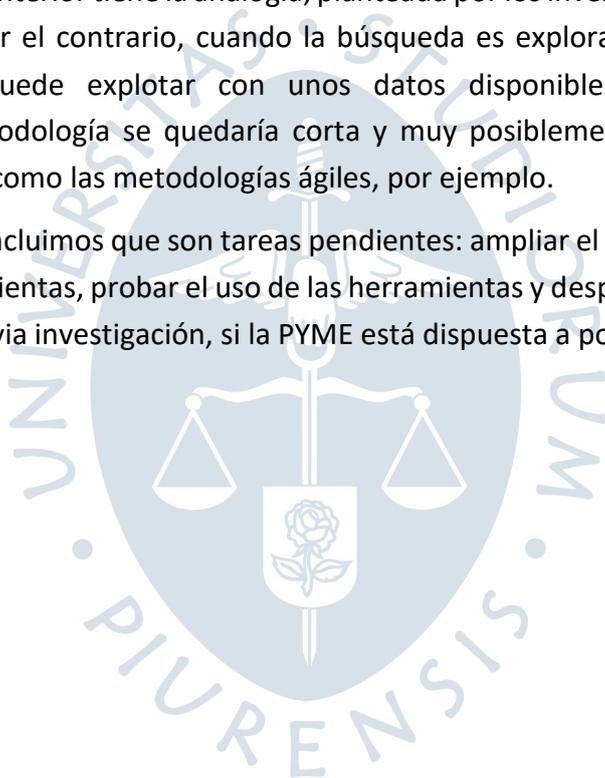
Otra limitación importante para la aplicación de la metodología en el sector objetivo es la poca o nula cultura de gestión de los datos y la falta de gobernanza de estos. Uno de los aportes del trabajo ha sido mostrar a las empresas la importancia de tener datos en el formato adecuado, de almacenarlos correctamente y de incluir nuevas variables de control.

La metodología propuesta muestra una ruta ampliamente documentada y probada que ha sido seguida en los casos de estudio y por tanto se valida su uso.

Se valida también lo que los investigadores están concluyendo actualmente al indicar que cuando la búsqueda de conocimientos está basada en datos disponibles y con objetivos claros y definidos el modelo de referencia propuesto y la metodología que se desprende de él, sigue estando vigente. Esto es aplicable a las empresas PYME del sector eléctrico.

La conclusión anterior tiene la analogía, planteada por los investigadores, con una mina ya lista a explotar. Por el contrario, cuando la búsqueda es exploratoria, es decir, se debe encontrar qué se puede explotar con unos datos disponibles: estructurados y no estructurados, la metodología se quedaría corta y muy posiblemente se requieran incluir nuevas herramientas como las metodologías ágiles, por ejemplo.

Finalmente concluimos que son tareas pendientes: ampliar el uso a más aplicaciones y con diferentes herramientas, probar el uso de las herramientas y despliegues con información en la nube (con la previa investigación, si la PYME está dispuesta a poner sus datos ahí).



Referencias bibliográficas

- Ahmad, T. (2017). Non-technical loss analysis and prevention using smart meters. *Renewable and Sustainable Energy Reviews*, 72, 573–589. <https://doi.org/10.1016/j.rser.2017.01.100>
- Alavikia, Z., & Shabro, M. (2022). A comprehensive layered approach for implementing internet of things-enabled smart grid: A survey. *Digital Communications and Networks*. <https://doi.org/10.1016/j.dcan.2022.01.002>
- Ali, M., Ali, R., Khan, W. A., Han, S. C., Bang, J., Hur, T., Kim, D., Lee, S., & Kang, B. H. (2018). A Data-Driven Knowledge Acquisition System: An End-to-End Knowledge Engineering Process for Generating Production Rules. *IEEE Access*, 6, 15587–15607. <https://doi.org/10.1109/ACCESS.2018.2817022>
- Amalina, F., Targio Hashem, I. A., Azizul, Z. H., Fong, A. T., Firdaus, A., Imran, M., & Anuar, N. B. (2020). Blending Big Data Analytics: Review on Challenges and a Recent Study. *IEEE Access*, 8, 3629–3645. <https://doi.org/10.1109/ACCESS.2019.2923270>
- Arghandeh, R., & Zhou, Y. (2017). *Big Data Application in Power Systems*.
- Ayad, H. A., Al-Kilani, L. A., Arshad, R., Al-Obadi, M. A., Hussein, H. T., & Kucukvar, M. (2020). Developing an Interactive Data Visualization Platform to Present the Adaption of Electrical Vehicles in Washington, California and New York. *2020 IEEE 7th International Conference on Industrial Engineering and Applications (ICIEA)*, 731–737. <https://doi.org/10.1109/ICIEA49774.2020.9101928>
- Banner, M., Alosert, H., Spencer, C., Cheeks, M., Farid, S. S., Thomas, M., & Goldrick, S. (2021). A decade in review: use of data analytics within the biopharmaceutical sector. *Current Opinion in Chemical Engineering*, 34, 100758. <https://doi.org/10.1016/j.coche.2021.100758>
- Bento, P. M. R., Pombo, J. A. N., Calado, M. R. A., & Mariano, S. J. P. S. (2018). A bat optimized neural network and wavelet transform approach for short-term price forecasting. *Applied Energy*, 210, 88–97. <https://doi.org/10.1016/j.apenergy.2017.10.058>

- Blaga, R., Sabadus, A., Stefu, N., Dughir, C., Paulescu, M., & Badescu, V. (2019). A current perspective on the accuracy of incoming solar energy forecasting. *Progress in Energy and Combustion Science*, 70, 119–144. <https://doi.org/10.1016/j.pecs.2018.10.003>
- Bordeleau, F.-E., Mosconi, E., & de Santa-Eulalia, L. A. (2020). Business intelligence and analytics value creation in Industry 4.0: a multiple case study in manufacturing medium enterprises. *Production Planning & Control*, 31(2–3), 173–185. <https://doi.org/10.1080/09537287.2019.1631458>
- Campeato, O. (2020). *Artificial Intelligence, Machine Learning, and Deep Learning*. MERCURY LEARNING & INFORMATION. https://www.ebook.de/de/product/38398221/oswald_campeato_artificial_intelligence_machine_learning_and_deep_learning.html
- CEPA-NEGLI. (2016). *Revisión del marco regulatorio del sector eléctrico peruano*. [http://minem.gob.pe/minem/archivos/Estudio Marco Regulatorio_CEPA-NEGLI\(1\).pdf](http://minem.gob.pe/minem/archivos/Estudio_Marco_Regulatorio_CEPA-NEGLI(1).pdf)
- COES. (2021). *Anexo C.1 Futuros de Demanda - INFORME COES/DP-01-2021*. <https://www.coes.org.pe/Portal/Planificacion/PlanTransmision/ActualizacionPTH>
- Dammert, A., Gallardo, J., & García, R. (2005). *Reformas Estructurales en el Sector Eléctrico Peruano* (Issue Documento de Trabajo N°5).
- Das, U. K., Tey, K. S., Seyedmahmoudian, M., Mekhilef, S., Idris, M. Y. I., Van Deventer, W., Horan, B., & Stojcevski, A. (2018). Forecasting of photovoltaic power generation and model optimization: A review. *Renewable and Sustainable Energy Reviews*, 81, 912–928. <https://doi.org/10.1016/j.rser.2017.08.017>
- Dranka, G. G., & Ferreira, P. (2020). Towards a smart grid power system in Brazil: Challenges and opportunities. *Energy Policy*, 136, 111033. <https://doi.org/10.1016/j.enpol.2019.111033>
- Grech, P. (2013). *Introducción a la ingeniería: un enfoque a través del diseño*. Pearson Educación.
- Hajjaji, Y., Boulila, W., Farah, I. R., Romdhani, I., & Hussain, A. (2021). Big data and IoT-based applications in smart environments: A systematic review. *Computer Science Review*, 39, 100318. <https://doi.org/10.1016/j.cosrev.2020.100318>
- Hancock, D. (2006). *Doing case study research: a practical guide for beginning researchers*. Teachers College Press.
- Huber, S., Wiemer, H., Schneider, D., & Ihlenfeldt, S. (2019). DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP-DM model. *Procedia CIRP*, 79, 403–408. <https://doi.org/10.1016/j.procir.2019.02.106>

- IBM. (2021a). *Data Quality Management (DQM)*. <https://www.ibm.com/docs/en/iii/9.0.0?topic=recognize-data-quality-management-dqm>
- IBM. (2021b). *IA Explicable*. <https://www.ibm.com/pe-es/watson/explainable-ai>
- IEA, I. E. A. (2021). *World Energy Outlook 2021*. <https://www.iea.org/reports/world-energy-outlook-2021>
- Ingemarsdotter, E., Kambanou, M. L., Jamsin, E., Sakao, T., & Balkenende, R. (2021). Challenges and solutions in condition-based maintenance implementation - A multiple case study. *Journal of Cleaner Production*, 296, 126420. <https://doi.org/10.1016/j.jclepro.2021.126420>
- Irizarry, R. A. (2020). The Role of Academia in Data Science Education. *Harvard Data Science Review*. <https://doi.org/10.1162/99608f92.dd363929>
- Kalu, O. O., & Madueme, T. C. (2018). Application of artificial neural network (ANN) to enhance power systems protection: a case of the Nigerian 330 kV transmission line. *Electrical Engineering*, 100(3), 1467–1479. <https://doi.org/10.1007/s00202-017-0599-y>
- Kourti, T. (2020). Multivariate Statistical Process Control and Process Control, Using Latent Variables. In *Comprehensive Chemometrics* (pp. 275–303). Elsevier. <https://doi.org/10.1016/B978-0-12-409547-2.14887-5>
- Kwac, J., Flora, J., & Rajagopal, R. (2014). Household Energy Consumption Segmentation Using Hourly Data. *IEEE Transactions on Smart Grid*, 5(1), 420–430. <https://doi.org/10.1109/TSG.2013.2278477>
- Lassio, J. G., Magrini, A., & Castelo Branco, D. (2021). Life cycle-based sustainability indicators for electricity generation: A systematic review and a proposal for assessments in Brazil. *Journal of Cleaner Production*, 311, 127568. <https://doi.org/10.1016/j.jclepro.2021.127568>
- Lee, L. C., & Jemain, A. A. (2021). On overview of PCA application strategy in processing high dimensionality forensic data. *Microchemical Journal*, 169, 106608. <https://doi.org/10.1016/j.microc.2021.106608>
- Liu, H., & Shi, J. (2013). Applying ARMA–GARCH approaches to forecasting short-term electricity prices. *Energy Economics*, 37, 152–166. <https://doi.org/10.1016/j.eneco.2013.02.006>
- Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N., Ramirez-Quintana, M. J., & Flach, P. (2021). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 33(8), 3048–3061. <https://doi.org/10.1109/TKDE.2019.2962680>

- Mason, K., Duggan, J., & Howley, E. (2018). Forecasting energy demand, wind generation and carbon dioxide emissions in Ireland using evolutionary neural networks. *Energy*, *155*, 705–720. <https://doi.org/10.1016/j.energy.2018.04.192>
- Mirnaghi, M. S., & Haghghat, F. (2020). Fault detection and diagnosis of large-scale HVAC systems in buildings using data-driven methods: A comprehensive review. *Energy and Buildings*, *229*, 110492. <https://doi.org/10.1016/j.enbuild.2020.110492>
- Mishra, B. K., Hazra, D., Tarannum, K., & Kumar, M. (2017). Business Intelligence using Data Mining techniques and Business Analytics. *Proceedings of the 5th International Conference on System Modeling and Advancement in Research Trends, SMART 2016*, 84–89. <https://doi.org/10.1109/SYSMART.2016.7894496>
- Mitma, R. (2018). *Visión política y regulación de la generación distribuida y/o autoproducción*.
- Nagi, J., Yap, K. S., Tiong, S. K., Ahmed, S. K., & Mohamad, M. (2010). Nontechnical Loss Detection for Metered Customers in Power Utility Using Support Vector Machines. *IEEE Transactions on Power Delivery*, *25*(2), 1162–1171. <https://doi.org/10.1109/TPWRD.2009.2030890>
- PCM, S. de G. D. (2021). *¿Qué es Gobernanza de Datos?* <https://www.datosabiertos.gob.pe/¿qué-es-gobernanza-de-datos>
- Peter Norvig, S. R. (2021). *Artificial Intelligence: A Modern Approach, Global Edition*. Pearson. https://www.ebook.de/de/product/40508318/peter_norvig_stuart_russell_artificial_intelligence_a_modern_approach_global_edition.html
- Phillips-Wren, G., Daly, M., & Burstein, F. (2021). Reconciling business intelligence, analytics and decision support systems: More data, deeper insight. *Decision Support Systems*, *146*, 113560. <https://doi.org/10.1016/j.dss.2021.113560>
- Radovilsky, Z., Hegde, V., Acharya, A., & Uma, U. (2018). Skills requirements of business data analytics and data science jobs: A comparative analysis. *Journal of Supply Chain and Operations Management*, *16*(1), 82–101.
- Ralph Kimball, M. R. (2013). *The Data Warehouse Toolkit*. Wiley John + Sons. https://www.ebook.de/de/product/20197316/ralph_kimball_margy_ross_the_data_warehouse_toolkit.html
- Ramesh, B., & Ramakrishna, A. (2018). Unified Business Intelligence Ecosystem: A Project Management Approach to Address Business Intelligence Challenges. *2018 Portland International Conference on Management of Engineering and Technology (PICMET)*, 1–10. <https://doi.org/10.23919/PICMET.2018.8481744>
- Ramos, S. (2019). *Power BI - 6 pasos para crear mejores soluciones*. <https://www.elfuturodelosdatos.com/power-bi-6-pasos-para-crear-mejores-soluciones/>

- Sampieri, R. (2018). *Metodología de la investigación : las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill Education.
- Scheidt, F. vom, Medinová, H., Ludwig, N., Richter, B., Staudt, P., & Weinhardt, C. (2020). Data analytics in the electricity sector – A quantitative and qualitative literature review. *Energy and AI*, 1, 100009. <https://doi.org/10.1016/j.egyai.2020.100009>
- Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, 181, 526–534. <https://doi.org/10.1016/j.procs.2021.01.199>
- Shearer, C. (2000). The CRISP-DM Model: The New Blueprint for Data Mining. In *Journal of Data Warehousing*. <https://www.bibsonomy.org/bibtex/28efa10127187a460bc800ed4a055d028/kdubiq>
- Sinha, S. (2021). *State of IoT 2021: Number of connected IoT devices growing 9% to 12.3 billion globally, cellular IoT now surpassing 2 billion*. IOT Analytics. <https://iot-analytics.com/number-connected-iot-devices/>
- Srivastava, G., S, M., Venkataraman, R., V, K., & N, P. (2022). A review of the state of the art in business intelligence software. *Enterprise Information Systems*, 16(1), 1–28. <https://doi.org/10.1080/17517575.2021.1872107>
- Tang, L., Li, J., Du, H., Li, L., Wu, J., & Wang, S. (2022). Big Data in Forecasting Research: A Literature Review. *Big Data Research*, 27, 100289. <https://doi.org/10.1016/j.bdr.2021.100289>
- Tumpa, Z. N., Saifuzzaman, M., Rabby, S. F., Crearie, L., & Stansfield, M. (2020). Understanding Business Intelligence in The Context of Mental Healthcare Sector of Bangladesh for Improving Health Services. *2020 IEEE 8th R10 Humanitarian Technology Conference (R10-HTC)*, 1–4. <https://doi.org/10.1109/R10-HTC49770.2020.9357023>
- United Nations, Department of Economic and Social Affairs, P. D. (2019). *No Title*. https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/files/documents/2020/Feb/un_2019_wpp_databooklet.pdf
- Vasquez Cordano, A. (2017). *Aspectos económicos de la implementación de redes inteligentes (smart grids) en el sector eléctrico peruano*.
- Wang, F., Li, K., Duić, N., Mi, Z., Hodge, B.-M., Shafie-khah, M., & Catalão, J. P. S. (2018). Association rule mining based quantitative analysis approach of household characteristics impacts on residential electricity consumption patterns. *Energy Conversion and Management*, 171, 839–854. <https://doi.org/10.1016/j.enconman.2018.06.017>

- Wang, J., Liu, F., Song, Y., & Zhao, J. (2016). A novel model: Dynamic choice artificial neural network (DCANN) for an electricity price forecasting system. *Applied Soft Computing*, 48, 281–297. <https://doi.org/10.1016/j.asoc.2016.07.011>
- Wiemer, H., Drowatzky, L., & Ihlenfeldt, S. (2019). Data Mining Methodology for Engineering Applications (DMME)—A Holistic Extension to the CRISP-DM Model. *Applied Sciences*, 9(12), 2407. <https://doi.org/10.3390/app9122407>
- Xu, Z., Tang, N., Xu, C., & Cheng, X. (2021). Data science: connotation, methods, technologies, and development. *Data Science and Management*, 1(1), 32–37. <https://doi.org/10.1016/j.dsm.2021.02.002>
- Yin, R. K. (2018). *Case Study Research and Applications*. Sage Publications Ltd. https://www.ebook.de/de/product/29024918/robert_k_yin_case_study_research_and_applications.html
- Zhang, Y., Huang, T., & Bompard, E. F. (2018). Big data analytics in smart grids: a review. *Energy Informatics*, 1(1), 8. <https://doi.org/10.1186/s42162-018-0007-5>
- Zhou, K., & Yang, S. (2016). Understanding household energy consumption behavior: The contribution of energy big data analytics. *Renewable and Sustainable Energy Reviews*, 56, 810–819. <https://doi.org/10.1016/j.rser.2015.12.001>
- Zohuri, B. (2020). *Artificial intelligence driven by machine learning and deep learning*. Nova Science Publishers.

Apéndices





Apéndice A. Exploración de la situación actual de la empresa – Focus Group

Exploración de la situación actual de la empresa – Focus Group

Para la realización se utilizó una muestra intencional para acceder a personas con profundo conocimiento del tema como lo realizado en (Khrais, 2020).

Para la recopilación de la información preliminar se optó por utilizar como instrumento el Focus Group. Este instrumento ya ha sido utilizado en (Soland et al., 2018) y (Sareen & Rommetveit, 2019).

1.1. Elección de la muestra.

Como el estudio está orientado a una empresa de unas características específicas se revisó los organigramas de empresas similares se determinó las personas objetivo acorde con la investigación.



Figura 1: Organigrama empresa ENSA

Fuente: Memoria anual 2019

En la figura una se muestra las personas elegidas para el focus Group a quienes se les agregó la jefatura de TI

1.2. Diseño del instrumento

El diseño del instrumento se elaboró siguiendo la propuesta de (Wilson, 2014).

Los Focus Groups, instrumento elegido, son útiles durante la definición de problemas, la recopilación de requisitos, el diseño conceptual y la implementación. (Wilson, 2014)

El instrumento elegido tiene las siguientes ventajas útiles para la investigación:

Ventajas	Desventajas
Es relativamente barato (en comparación con muchas entrevistas individuales) y se puede organizar rápidamente.	Las conclusiones que se pueden extraer de un grupo de enfoque son limitadas porque la muestra puede no ser representativa de la población más grande.
Utilizados al principio de un proyecto pueden producir información y respuestas a partir de la interacción entre diferentes usuarios [...]de productos o servicios. Este tipo de sinergia podría no ocurrir con entrevistas, encuestas u otros métodos individuales.	Pueden surgir conflictos y luchas de poder entre participantes con personalidades fuertes.
Se pueden modificar las preguntas del grupo de enfoque sobre la marcha, pedir a las personas que den opiniones en privado e investigar los nuevos temas que surjan durante la sesión.	Los datos de los focus group pueden ser difíciles de transcribir si varias personas intentan hablar al mismo tiempo.

Analizando las ventajas y desventajas determinamos que las ventajas nos permiten utilizar el instrumento y las desventajas no son significativas ni tienen relevancia en esta investigación.

1.2.1. Guía del Focus Group.

En la tabla siguiente se muestra la guía que se empleó en el Focus Group, acorde con los objetivos de la investigación exploratoria.



Guía del Moderador				
24/09/2020	Moderador:	Redy Risco		
	Focus Group	Número 1		
Tópico	Descripción	Materiales /preguntas	Duración	
Preparación	Coordinaciones previas.	Carta formal. Correos de confirmación. Preparación de sala de videoconferencia		
Introducción	Presentación. Informe del Objetivo de la entrevista. Exposición del método de la entrevista. Consentimiento informado.		10	Min
Tópicos estructurados	Tópico 1: Background personal	¿Cuál es la gerencia/ área / unidad / departamento de trabajo?	40	Min
		¿Cuál es cargo o puesto de trabajo?		
		¿Participas en recomendación o decisión e proyectos de innovación tecnológica en tu área o empresa?		
	Tópico 2: Indicadores de Gestión	¿Qué indicadores de gestión maneja tu área?		
		¿Qué datos (fuente) se utilizan para los cálculos de estos indicadores?		
		¿Con qué frecuencia se evalúan los indicadores?		
		La data fuente ¿se adquiere manualmente (reportes de otras áreas) o automáticamente de una base integrada a algún otro sistema?		
	Tópico 3: Smart Grid	¿Sabes si en la empresa o en tu área se ha implementado alguna solución Smart Grid?		
		¿Crees que se deberían implementar más soluciones Smart Grid en la empresa? ¿Por qué?		
		¿Cuál consideras es el nivel de implementación de soluciones Smart Grid en la empresa: Bajo, Medio, Alto?		
¿Por qué crees se tiene este nivel de implementación?				
Preguntas generales y diálogo abierto con el entrevistado			10	Min
Comentarios de cierre y finalización de cualquier reporte (recibos, cuestionario final, etc.)			10	Min
			70	Min

1.2.2. Resultados del Focus Group 1.

	PARTICIPANTE	E1	E2	E3
Tópico 1: Background personal	¿Cuál es la gerencia/ área / unidad / departamento de trabajo?	Gerente Regional	Gerente Técnico	Gerente Comercial
	¿Cuál es cargo o puesto de trabajo?	Gerente Regional	Gerente Técnico	Gerente Comercial
	¿Participas en recomendación o decisión e proyectos de innovación tecnológica en tu área o empresa?	SI	SI	SI
Tópico 2: Indicadores de Gestión	¿Qué indicadores de gestión maneja tu área?	Principales ROE / ROA SAIDI / SAIFI Pérdidas de distribución de energía. ISCAL - CIER	SAIFI / SAIDI Inductores: Plan de Mantenimiento, por UNN y áreas Producción de Generación (para Gen Distribuida). Nivel de Tensión Calidad de Alumbrado Público.	Incremento de Clientes Incremento de la venta de energía (en MT/BT/Mercado Libre) Indicadores de Facturación Indicadores de Cobranza Cobrabilidad Morosidad. ISCAL : Calidad de suministro, Calidad Imagen Atención al cliente
	¿Qué datos (fuente) se utilizan para los cálculos de estos indicadores?	ROE y ROA, resultados de la unidad de contabilidad La empresa tiene SAP. Se usa Estado de GG y PP , utilidad neta y EBITDA. Formula administrativa. Patrimonio , se maneja con el Balance.	SAIDI y SAIFI , diario . En CCO se registra cada una de las interrupciones desde acometida hasta nodo de transmisión. Cada turno se evalúa. Ver geografía y donde hay mayor problemas se van atendiendo.	Datos en un Data MART.
	¿Con qué frecuencia se evalúan los indicadores?	Mensual / Trimestral	Diario / Semanal	Semanal/Quincenal/ Mensual

	La data fuente ¿se adquiere manualmente (reportes de otras áreas) o automáticamente de una base integrada a algún otro sistema?	Los datos del ROE y ROA, salen del SAP	Es automática de una base integrada , va un sistema técnico comercial MSGC. Ingreso es manual y la gestión de la data es automática y se tiene unos aplicativos que genera reportería. Les ayuda a tomar decisiones en el momento	Cubos informáticos , Data Mart. Registros de clientes, ventas de energía, miles de soles, sale para pérdidas , consumos ceros. Emiten reportes y se puede gestionar.
Tópico 3: Smart Grid	¿Sabes si en la empresa o en tu área se ha implementado alguna solución Smart Grid?	No es completamente integral. Hay muy buenos avances. Les falta la integración.	Esta concentrado en los grandes nodos. En circuitos alimentadores. Hasta medición en cabecera, y todo motorizado. Para trabajar de manera remota. Falta la medición de las SE de distribución a los CCO y la información hacia los clientes. Falta la ultima milla de integración. y desarrollar sistemas de comunicación. 60%	Comunicación Zig Bee, En comercial se tiene a clientes mayores , medidores en Zig Be, se tienen 60 puntos. A nivel corporativa medición inteligente por GPRS. El piloto para enlazar clientes en BT lo están evaluando a nivel de la corporación , para salir a concurso. Gestionando un tema a nivel de acometida, están haciendo pilotos para hacer un corte efectivo a nivel de suministro (en lonchera), que indique , corriente, si esta en corte, su ubicación geográfica.
	¿Crees que se deberían implementar más soluciones Smart Grid en la empresa? ¿ Por qué?	Se obvió esta pregunta en base a respuestas anteriores se entendió la importancia que loes entrevistados le daban a la SG		
	¿Cuál consideras es el nivel de implementación de soluciones Smart Grid en la empresa: Bajo, Medio, Alto? Y en el País.	Tienen Zig Bee. Les da ventaja. A nivel país es un nivel bajo.	Un poco más del medio , ya casi en el nivel alto. Lo reflejan las reuniones las empresa de FONAFE a nivel país. A distancia de ENEL	Medio , acercándonos al Alto. La integración es corto plazo. Han pasado el camino duro. La integración esta alineado con una estandarización del grupo Distriluz.

	<p>¿Por qué crees se tiene este nivel de implementación?</p>	<p>Políticas sectoriales. Esas paredes y barreras, responden principalmente a corrupción y falta de valores. Como tema de fondo saber elegir.</p>	<p>Políticas de reconocimiento. La tele gestión de AP, en la fijación del VAD recién han reconocido algo. Los proyectos de inversión son freno, ley de contrataciones del estado, el <i>invierte.pe</i>, procesos que duran de 1 a 2 años. Los contratistas conocen leguleyadas para objetar procesos.</p>	<p>Empresa del estado, se debe fundamentar una inversión, mucho más para que se tenga un reconocimiento tarifario, se lucha técnica y comercialmente. El ente regulador, busca precio optimo para cliente, a pesar que este mayor precio pueda beneficiar a clientes finales. Dentro de la ley de contrataciones no existen pilotos, se busca a proveedores. El estudio de mercado se dilata mucho tiempo. Problemas en concursos. Complican el avance y nuevas tecnologías. Hacer pruebas, masificarlas y buscar especialistas para evitar problemas legales. Es bastante complicada la innovación en el corto plazo.</p>
--	--	---	--	--



1.2.3. Análisis preliminar de los resultados.

Consideramos que las primeras conclusiones que podemos obtener del Focus Group son las siguientes:

- a) Se ha comprobado que los indicadores de calidad: SAIDI y SAIFI son importantes para la dirección de la empresa.
- b) Otro indicador relevante es el control de pérdidas de energía.
- c) El tercer indicador relevante es el ISCAL calculado por el CIER.
- d) En el ámbito comercial se incluyen los indicadores anteriores a los que se le agregan control de facturación y gestión de clientes.
- e) Es también interesante que contrario a lo que se pensaba los entrevistados consideran un nivel de implementación de las tecnologías Smart Grid, como medio cerca del alta.
- f) Se cuenta con Data Marts que nos permitiría la obtención de datos para prueba de los algoritmos que la metodología pueda implementar.
- g) La menor frecuencia de cálculo de los indicadores de gestión es daría lo que daría un buen margen de operación de los algoritmos.
- h) Preocupa que factores éticos y procedimientos administrativos gubernamentales no alienten la innovación.

Apéndice B. Resultados de la encuesta

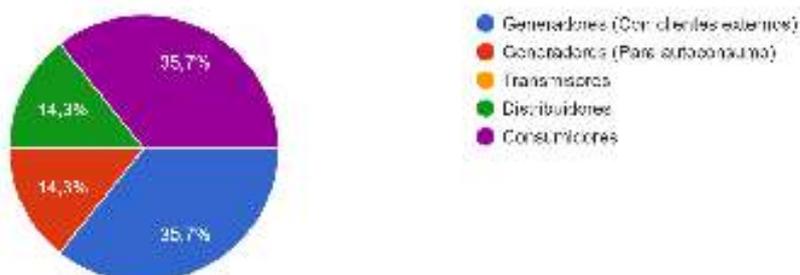
RESULTADOS DE ENCUESTA:

“Uso de BI&A en el sistema eléctrico peruano.”

TIPO DE INTEGRANTE

El sector eléctrico peruano está integrado por generadores, transmisores, distribuidores y consumidores. Según esto, tu empresa estaría dentro del grupo de:

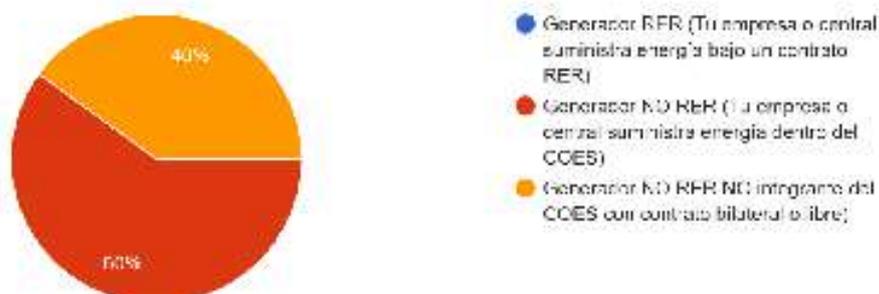
14 respuestas



TIPO DE GENERADOR

¿Qué tipo de generador es tu empresa?

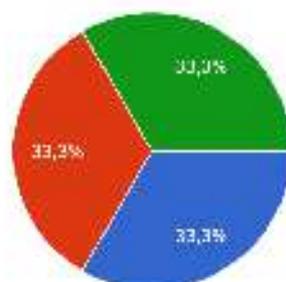
5 respuestas



TIPO DE CONSUMIDOR

Sector de Consumo

3 respuestas

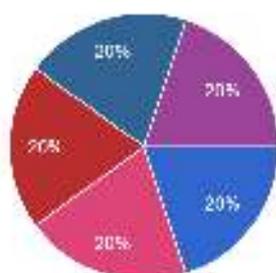


- Agroindustria
- Pesca
- Minería
- Petróleo y gas
- Alimentos, Bebidas y Tabaco
- Servicios financieros
- Educación
- Retail y Multisector

▲ 1/2 ▼

Sector del Consumidor

4 respuestas



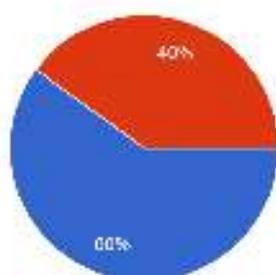
- Agroindustria
- Pesca
- Minería
- Petróleo y gas
- Alimentos, Bebidas y Tabaco
- Servicios financieros
- Educación
- Retail y Multisector

▲ 1/2 ▼

TIPO DE USUARIO

¿Qué tipo de usuario es tu empresa?

4 respuestas

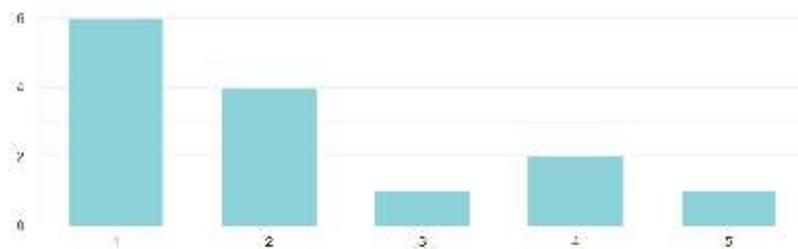


- Usuario regulado (Pagas un precio regulado por OS/NERGVIH a un distribuidor)
- Usuario Libre (Pagas el precio establecido en tu contrato con tu suministrador)

NIVEL DE CONOCIMIENTO DE BI&A**1: Muy Bajo 5: Muy alto**

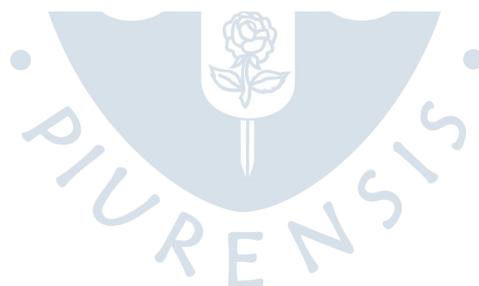
¿Cuál crees que es el nivel de conocimiento en tu empresa del BI&A para la gestión del suministro eléctrico?

14 respuestas

**NIVEL DE UTILIZACIÓN****1: No mucho - 5 Mucho**

¿Has utilizado alguna herramienta o plataforma de BI&A para la gestión del suministro eléctrico?

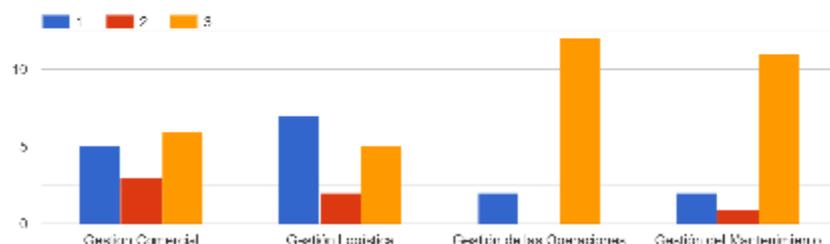
14 respuestas



AREAS DE APLICACIÓN

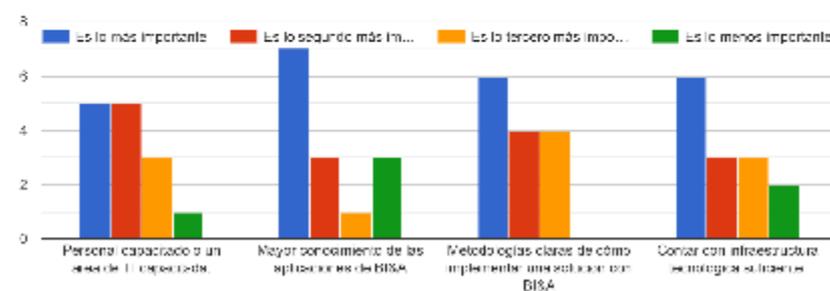
1: Ninguna aplicación 2: Alguna aplicación 3: Mucha aplicación

¿En qué área de tu empresa crees que tendría mayor aplicación las herramientas de BI&A?



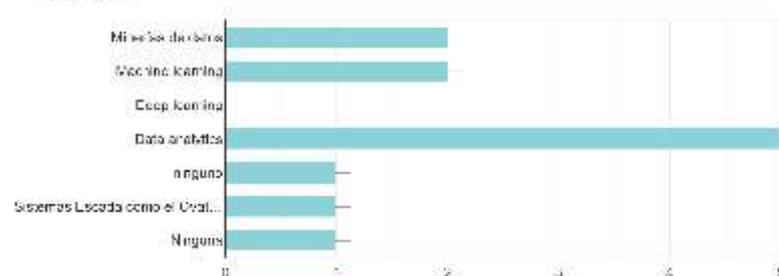
ASPECTOS IMPORTANTES PARA LA IMPLEMENTACIÓN

A continuación mostramos 4 aspectos necesarios para la implementación de herramientas BI&A. Para cada aspecto indica el nivel de importancia que para ti tiene cada aspecto.



HERRAMIENTAS UTILIZADAS

Si en tu empresa han utilizado alguna de las siguientes tecnologías marca cuál. Puedes marcar más de una, ninguna o indicar alguna otra tecnología...ión que conozcas y sepas que se ha implementado.



Apéndice C1. Paper caso 1 Uso de un nuevo marco de inteligencia empresarial en la gestión de la calidad del suministro eléctrico en pequeñas y medianas empresas



Use of a business intelligence framework in the management of the quality of electricity supply in small and medium-sized companies

Redy Risco-Ramos, Daniel Pérez-Aguilar, Luis Casaverde-Pacherez & Edilberto Vásquez-Díaz

Laboratorio de Sistemas Automáticos de Control, Universidad de Piura, Piura, Perú. redy.risco.r@alum.udep.edu.pe, daniel.perez.a@alum.udep.edu.pe, luis.casaverde@alum.udep.edu.pe, edilberto.vasquez@udep.edu.pe

Received: October 18th, 2021. Received in revised form: February 24th, 2022. Accepted: March 14th, 2022.

Abstract

The objective of this study is to present a methodology based on business intelligence in small and medium-sized companies. Three methods were selected and evaluated, and the cross-industry standard process for data mining (CRISP-DM) was used as a reference. The methodology was applied to a real case study: an agro-industrial company, using a commercial business intelligence application, to find corrective measures to improve the management of the company's electricity supply. Based on the results of this study, it can be concluded that with low-cost computational tools, it is possible to visualize and control parameters of the electrical supply, group the consumption of reactive energy for an adequate selection of steps in a capacitor bank. Additionally, with a minimum knowledge of modeling, data analysis, and the electricity sector, these tools are not only applied in commercial matters but also in the management of electricity supply.

Keywords: analytics; business intelligence; clustering; CRISP-DM; data mining.

Uso de un marco de inteligencia empresarial en la gestión de la calidad del suministro eléctrico en pequeñas y medianas empresas

Resumen

El objetivo del presente artículo es presentar una metodología basada en Business Intelligence en una pequeña y mediana empresa. Se eligieron y evaluaron tres métodos y de ellos se tomó como referencia el CRISP-DM (Cross-Industry Standard Process for Data Mining). Se aplicó la metodología a un caso de estudio real: una empresa agroindustrial, utilizando una aplicación de Business Intelligence comercial encontrándose medidas correctivas que mejorarán la gestión del suministro eléctrico de la empresa. Basado en los resultados del estudio se puede concluir que con herramientas computacionales de bajo costo se puede visualizar y controlar parámetros del suministro eléctrico, agrupar los consumos de energía reactiva para una adecuada selección de pasos en un banco de condensadores y que, con unos mínimos conocimientos de modelado, análisis de datos y del sector eléctrico, estas herramientas no solo tienen aplicación en materia comercial sino también en la gestión del suministro eléctrico.

Palabras clave: analítica; inteligencia de negocios; agrupamiento; CRISP-DM; minería de datos.

1. Introduction

For decades, small and medium-sized industries in Latin America in general and in Peru in particular, have not taken advantage of the large amount of data generated by their operations. A large amount of data is simply stored without any further use. As of 2010, it is known that there are processes, architectures, and technologies capable of transforming this data into relevant data that can enable them to make profitable business operational and commercial decisions.

This set of processes, architectures, and technologies is known as Business Intelligence (BI) [1,2].

With BI, the closely related concept of data mining (DM) arises. It is a subdomain of the concept of artificial intelligence, which refers to the process aimed at generating knowledge from data and providing users with comprehensive findings [3,4].

Generation of knowledge in the context of DM can be regarded as the finding of patterns, relationships and trends relevant to the user. As a process, DM mainly involves the collection, selection, pre-processing, and analysis of data

How to cite: Risco-Ramos, R., Pérez-Aguilar, D., Casaverde-Pacherez, L. and Vásquez-Díaz, E. Use of a business intelligence framework in the management of the quality of electricity supply in small and medium-sized companies. DYNA, 89(221), pp. 31-40, April - June, 2022.

© The author; licensee Universidad Nacional de Colombia 
 Revista DYNA, 89(221), pp. 31-40, April - June, 2022, ISSN 0012-7353
 DOI: <https://doi.org/10.15446/dyna.v89n221.99085>

Apéndice C2. Paper caso 2: Aplicación de técnica de minería de datos para determinar un modelo de operación normal de n turbogenerador hidráulico

Aplicación de técnicas de minería de datos para determinar un modelo de operación normal de un turbogenerador hidráulico

Redy H. Risco¹, Daniel A. Perez¹, Luis A. Casaverde¹ y Gerson La-Rosa¹

¹ Universidad de Piura, Doctorado en Ingeniería, Piura - Perú. (correo-e: redy.risco.r@alum.udep.edu.pe; daniel.perez.a@alum.udep.edu.pe; luis.casaverde@alum.udep.edu.pe; gerson.larosa@udep.edu.pe)

* Autor a quien debe ser dirigida la correspondencia

RESUMEN

Este artículo presenta la aplicación de un modelo de referencia para la implementación de minería de datos como parte de un entorno de inteligencia de negocios con el objetivo de analizar la operación de un turbogenerador hidráulico. El análisis parte de los datos provenientes del sistema de control y adquisición de datos de una central hidroeléctrica, con ellos se generó un modelo utilizando análisis de componentes principales, luego se evaluó el modelo con nuevas observaciones utilizando los gráficos de control estadístico multivariante cuadrado del error de predicción y T^2 de Hotelling con los que se implementó el análisis conectando las tablas generadas en una herramienta convencional de inteligencia de negocios. El modelo se probó en 5 muestras correspondientes a periodos previos a una parada del turbogenerador. Concluimos que el modelo de referencia utilizado puede ser aplicado al análisis de la operación de estos equipos, que los gráficos de control son capaces de detectar algunos comportamientos de operación y que es necesario el conocimiento del proceso para una adecuada interpretación de los resultados.

Palabras clave: minería de datos; CRISP-DM; control estadístico multivariante; análisis de componentes principales; SCADA.

Application of data mining techniques to determine a normal operation model of a hydraulic turbogenerator

ABSTRACT

This paper presents the application of a reference model for the implementation of data mining as part of a business intelligence environment to analyze the operation of a hydraulic turbogenerator. The analysis starts from the data from the control and data acquisition system of a hydroelectric power plant, with them a model was generated using principal component analysis, then the model was evaluated with new observations using multivariate statistical control graphs square prediction error and Hotelling's T^2 with which the analysis was implemented by connecting the tables generated in a conventional business intelligence tool. The model was tested on 5 samples corresponding to periods prior to a turbogenerator shutdown. We conclude that the reference model used can be applied to the analysis of the operation of this equipment, that the control charts can detect some operating behaviors and that knowledge of the process is necessary for an adequate interpretation of the results.

Keywords: data mining; CRISP-DM; multivariate statistical control; principal component analysis; SCADA.

Apéndice C3. Paper Caso 3: Prediction of hidropower energy in a Peruvian company using a business intelligence framework

Prediction of hydropower energy in a Peruvian company using a business intelligence framework

Redy Risco
Faculty of Engineering
Universidad de Piura
Piura, Peru
redy.risco.r@ahum.udep.edu.pe

Daniel Pérez
Faculty of Engineering
Universidad de Piura
Piura, Peru
daniel.perez.a@ahum.udep.edu.pe

Luis Casaverde
Faculty of Engineering
Universidad de Piura
Piura, Peru
luis.casaverde@ahum.udep.edu.pe

Edilberto Vásquez
Faculty of Engineering
Universidad de Piura
Piura, Peru
edilberto.vasquez@udep.edu.pe

Abstract—The objective of this work is to show how to use a business intelligence framework for the prediction of energy production in a medium hydropower plant. For this, the CRISP-DM (Cross-Industry Standard Process for Data Mining) reference model is used as a feasible methodology to use. In the methodology, smooth exponential, ARIMA, and neural networks prediction models are evaluated under the Mean Square Error criterion. The evaluation showed that the model with Smooth Exponential has the best performance, the application of the methodology in all its stages, and the prediction results obtained were also shown. Based on the results, it is concluded that the method used obtains useful results, is implementable with hardware accessible to SMEs and that when the mining objectives are clear, the reference model used is still valid.

Keywords—Business Intelligence, CRISP-DM, Mean Square Error, Predictive Analysis.

I. INTRODUCTION

At present, small and medium-sized enterprises in Peru still do not take advantage of the great value that the enormous amount of data available as a result of their operations can provide. As indicated in [1] since 1865, we try to take advantage of the data available in business through processes, architectures and technologies that can transform this data into relevant information that allows them to make important decisions for the efficient management of their businesses.

This set of processes, architectures and technologies is known as business intelligence (BI) [2][3].

Another concept closely related to BI is Data Mining (DM) which is a subdomain of Artificial Intelligence defined as a process that aims to generate knowledge from data presenting the findings in an integral way to the user [4].

The generation of knowledge in the context of DM can translate into the discovery of patterns, relationships, and trends relevant to the user. DM as a process involves, in essence, the collection and selection of data, the preprocessing of data, the analysis of data itself, including the visualization of results, the interpretation of findings, and the application of the knowledge obtained. To pre-process and analyze the data, machine learning (ML) and statistical methods are implemented [4].

Currently, data analytics, BI and DM platform providers are increasing including capabilities within their platforms that go far beyond their traditional markets. Analytics and BI platforms increasingly include functionality to perform augmented DM tasks. Predictive models run without having direct contact with the user, but delivering information and insights appear naturally within the BI analysis and process flow [5].

Our work aims to answer the research question by developing a DM methodology in a medium-generation

company, thus demonstrating its feasibility and application, which would also constitute its contribution.

Next, we describe the development of this work: in section 2 we present the approach of the problem supporting this approach, in section 3 we present the method we propose to use, in section 4 we explain how we have applied the methodology in this work, and finally, in section 5 we present some conclusions of the work

II. PROBLEM STATEMENT

Peru, since approximately 2010, electric generators with renewable resources, including hydroelectric plants with installed capacities of up to 20 MW, have been entering commercial operation in the Peruvian electricity system. [6]

These plants have the requirement to report the estimated production of the following year to the regulatory, operator, and regulatory entities.

In addition, this estimated production is very useful for the projection of economic income for the following year. Generating companies usually make an estimate based on the averages of previous years. We can summarize the problem in Table I.

TABLE I. PROBLEM STATEMENT

Topic, Object, Problem, and (Question Research)	
Topic	Business Intelligence
Object	Development of a methodology applied to an electric power generation company in Peru governed by the regulations in force in 2020
Problem	In companies that generate electricity, the next year's production can be estimated using a BI platform.
Research Question	How to develop a BI solution in a generation company to estimate the energy production for the following year?

III. PROPOSED METHODOLOGY

For the development of a Data Mining solution in 1996, four leaders of the nascent DM market conceived the CRISP-DM (Cross-Industry Standard Process for Data Mining) model, as a non-proprietary, documented, and freely available model that allowed organizations to achieve better data mining results and promote best practices in the industry. This model is described in [7][8].