

UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA

DESARROLLO DE SOFTWARE



EXTRACCIÓN DE CONOCIMIENTO EN BASES DE DATOS

**REPORTE DE SOLUCIÓN DE CASO DE ESTUDIO EN LA QUE PRESENTE
OBJETIVO, ALCANCE, JUSTIFICACIÓN DE LA METODOLOGÍA Y PLANEACIÓN
DE LAS ETAPAS PARA EL ANÁLISIS DE DATOS**

PRESENTA:

MILDRED VILLASEÑOR RUIZ

ÁNGEL RICARDO CHÁVEZ ZARAGOZA

DARON TARÍN GONZÁLEZ

RICARDO ALONSO RÍOS MONREAL

DOCENTE:

ING. LUIS ENRIQUE MASCOTE CANO

Chihuahua, Chih., 27 de septiembre de 2025

Contenido

| | |
|--------------------------------------|---|
| Objetivo del Proyecto | 3 |
| Alcance | 4 |
| Justificación de la metodología..... | 5 |
| Planeación de etapas | 7 |
| Conclusión | 9 |
| Referencias..... | 9 |

Ilustraciones

| | |
|--------------------|---|
| Ilustración 1..... | 3 |
| Ilustración 2..... | 8 |

Tablas

| | |
|--------------|---|
| Tabla 1..... | 6 |
|--------------|---|

Objetivo del Proyecto

El objetivo central de este proyecto es **diseñar un sistema de predicción de la demanda eléctrica con un horizonte de 24 horas**, aplicado a la ciudad ficticia de ElectroVille. La empresa distribuidora enfrenta problemas recurrentes durante períodos de calor extremo, donde la demanda energética se eleva de forma abrupta debido al uso masivo de aires acondicionados, ventiladores y otros dispositivos de refrigeración. Estos picos imprevistos generan riesgos de **sobrecarga en la red** y potenciales cortes de suministro, afectando tanto a la población como a la continuidad de los servicios esenciales.

El sistema propuesto busca anticipar estas fluctuaciones a través de un modelo predictivo que combine **series de tiempo del consumo eléctrico histórico por barrio** con variables externas como **temperatura, humedad relativa y calendario laboral** (incluyendo fines de semana, festivos y horarios de actividad económica). Con estas predicciones, la distribuidora podrá **ajustar la producción, activar reservas energéticas o redistribuir cargas** antes de que los problemas se manifiesten.

En suma, el proyecto persigue un doble beneficio: **optimizar la operación técnica del sistema eléctrico y garantizar la calidad y continuidad del servicio para los ciudadanos**.



Ilustración 1

ElectroVille, paisaje urbano ficticio con red de distribución eléctrica.

Fuente: Elaboración propia con IA (2025).

Alcance

El proyecto de pronóstico de demanda eléctrica en la ciudad de ElectroVille se centra en el diseño e implementación de un sistema predictivo con horizonte de 24 horas, enfocado en la red de distribución urbana. El alcance incluye las siguientes consideraciones:

Datos incluidos

- Series históricas de consumo eléctrico horario desagregadas por barrio.
- Variables meteorológicas: temperatura y humedad relativa.
- Información de calendario: fines de semana, días festivos y horarios de actividad económica (fábricas, centros de ocio, comercios).

Limitaciones

- El sistema se restringirá a un horizonte de predicción de 24 horas; no contempla proyecciones a mediano o largo plazo.
- Se trabajará únicamente con variables meteorológicas básicas (temperatura y humedad), sin integrar otras como velocidad del viento o radiación solar.
- La calidad del modelo dependerá de la completitud y confiabilidad de los datos históricos y meteorológicos disponibles.
- El despliegue se limitará al centro de control de la distribuidora, sin contemplar aplicaciones móviles o interfaces para usuarios finales.

Incluye

- Recolección, limpieza e integración de los datos.
- Construcción del pipeline analítico bajo la metodología CRISP-DM.
- Entrenamiento y validación de un modelo de predicción basado en series de tiempo enriquecidas con variables externas.
- Generación de predicciones diarias de demanda eléctrica con horizonte de 24 horas.
- Plan de despliegue del modelo en el centro de control, incluyendo mecanismos de retroalimentación y actualización periódica.

Excluye

- Expansión del sistema a predicciones de mediano y largo plazo.
- Incorporación de datos en tiempo real provenientes de sensores IoT de los hogares.
- Optimización directa de la operación de centrales eléctricas; el sistema se limita a entregar pronósticos como insumo para la toma de decisiones.
- Desarrollo de interfaces de usuario finales para clientes residenciales o comerciales.

De esta forma, el proyecto se mantiene acotado a una primera versión funcional y operativa del sistema de predicción, con foco en el uso de datos históricos, meteorológicos y de calendario para generar pronósticos útiles para la distribuidora.

Justificación de la metodología

Para el caso de pronóstico de demanda eléctrica en la red urbana de ElectroVille, proponemos usar la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining) como marco estructurante del proyecto analítico.

¿Por qué CRISP-DM?

1. Enfoque centrado en el negocio

CRISP-DM inicia con la fase de Comprensión del negocio (Business Understanding), lo cual obliga al equipo a vincular claramente los objetivos técnicos con las necesidades reales de la empresa (en nuestro caso, evitar sobrecargas, cortes y optimizar la generación). Esto minimiza el riesgo de emprender análisis técnicos que no agreguen valor al problema real.

2. Metodología genérica y neutral respecto a herramientas

CRISP-DM no prescribe herramientas o librerías específicas, lo que permite flexibilidad técnica para usar R, Python, plataformas en nube o software propietario según disponibilidad.

3. Iterativa y adaptable

Aunque CRISP-DM tiene seis fases “ordenadas”, permite retroceder de una fase a otra si surgen hallazgos nuevos (por ejemplo, si en modelado se identifica necesidad de nueva transformación de datos). En el pronóstico eléctrico, donde nuevas variables o anomalías emergen, esta flexibilidad es

clave.

4. Énfasis en la calidad y preparación de datos

Dado que el caso incluye múltiples fuentes (historias de consumo horario, meteorología, calendario de actividad económica), la fase de Preparación de datos es crítica para limpieza, integración y transformación. CRISP-DM dedica explícitamente esfuerzos a esa fase, lo cual es compatible con la complejidad del caso.

5. Integración del despliegue y operación

CRISP-DM contempla la fase de Despliegue (Deployment), asegurando que el modelo no quede solo como un prototipo, sino que se integre en el centro de control para generar predicciones diarias reales.

Esto encaja con la necesidad operativa del proyecto.

6. Reconocimiento y uso en aplicaciones energéticas similares

La literatura reciente sobre predicción de consumo energético ha adoptado CRISP-DM como marco de referencia para estructurar los proyectos. Por ejemplo, investigaciones de consumo energético en edificios han seguido las fases de CRISP-DM para modelado, evaluación y despliegue.

Tabla 1

Retos del caso abordados con CRISP-DM

| Reto del caso | Cómo CRISP-DM lo aborda |
|---|---|
| Integrar múltiples fuentes heterogéneas (consumo horario, clima, actividad económica) | En las fases de Comprensión de datos y Preparación, se pueden analizar correlaciones, detectar valores atípicos y realizar alineamientos temporales entre datasets. |
| Cambios dinámicos en los patrones debido al clima, días festivos, anomalías | La naturaleza iterativa de CRISP-DM permite reentrar a fases de modelado o preparación cuando nuevos patrones emergen. |
| Asegurar que el modelo sea operativo y confiable en producción | La fase de Despliegue y la retroalimentación del modelo permiten monitorear degradaciones y actualizar el modelo regularmente. |
| Balance entre exploración y entrega rápida de valor | Aunque CRISP-DM es sistemático, se pueden aplicar “slices verticales” (pequeños incrementos end-to-end) para entregar versiones iniciales de predicción, antes de perfeccionar el pipeline completo, en un estilo más ágil. |

Planeación de etapas

La planeación del proyecto seguirá la metodología CRISP-DM, adaptada al caso de ElectroVille. Cada fase se desglosa en actividades concretas y en los entregables que permitirán evidenciar el progreso.

Fase 1. Comprensión del negocio

- *Actividades:* Revisión del contexto energético en ElectroVille, análisis de riesgos asociados a picos de demanda, definición de objetivos y métricas de éxito.
- *Entregables:* Documento de requerimientos del negocio, métricas clave de desempeño (ej. error medio absoluto en pronóstico).

Fase 2. Comprensión de los datos

- *Actividades:* Exploración de series históricas de consumo por barrio, análisis de correlación con variables meteorológicas y de calendario, identificación de valores atípicos y datos faltantes.
- *Entregables:* Reporte exploratorio de datos (EDA), gráficos de tendencias y correlaciones iniciales.

Fase 3. Preparación de los datos

- *Actividades:* Limpieza de datos inconsistentes, imputación de valores faltantes, alineación temporal de series, integración de datasets meteorológicos y de calendario.
- *Entregables:* Dataset integrado y limpio, diccionario de variables transformadas.

Fase 4. Modelado

- *Actividades:* Selección de técnicas de predicción de series de tiempo (ej. SARIMAX, Prophet, LSTM), entrenamiento de modelos con validación cruzada, ajuste de hiperparámetros.
- *Entregables:* Modelos entrenados y evaluados, comparación de métricas de desempeño (ej. RMSE, MAE).

Fase 5. Evaluación

- *Actividades:* Validación final de resultados frente a datos no vistos, análisis de errores en escenarios críticos (olas de calor, fines de semana festivos), verificación con objetivos de negocio.
- *Entregables:* Informe de evaluación del modelo, recomendación del modelo final a implementar.

Fase 6. Despliegue

- *Actividades:* Implementación del pipeline en el centro de control de ElectroVille, automatización de la generación diaria de predicciones, diseño de un tablero simple de monitoreo.

- *Entregables:* Sistema de pronóstico operativo, manual de usuario técnico, protocolo de mantenimiento y actualización del modelo.

Cronograma tentativo

- Fases 1 y 2: Semanas 1–2
- Fase 3: Semanas 3–4
- Fase 4: Semanas 5–6
- Fase 5: Semana 7
- Fase 6: Semanas 8–9

Este desglose asegura que el proyecto avance de manera ordenada y que en cada etapa se generen productos intermedios que validen la calidad y el alineamiento del sistema con los objetivos de la empresa distribuidora.

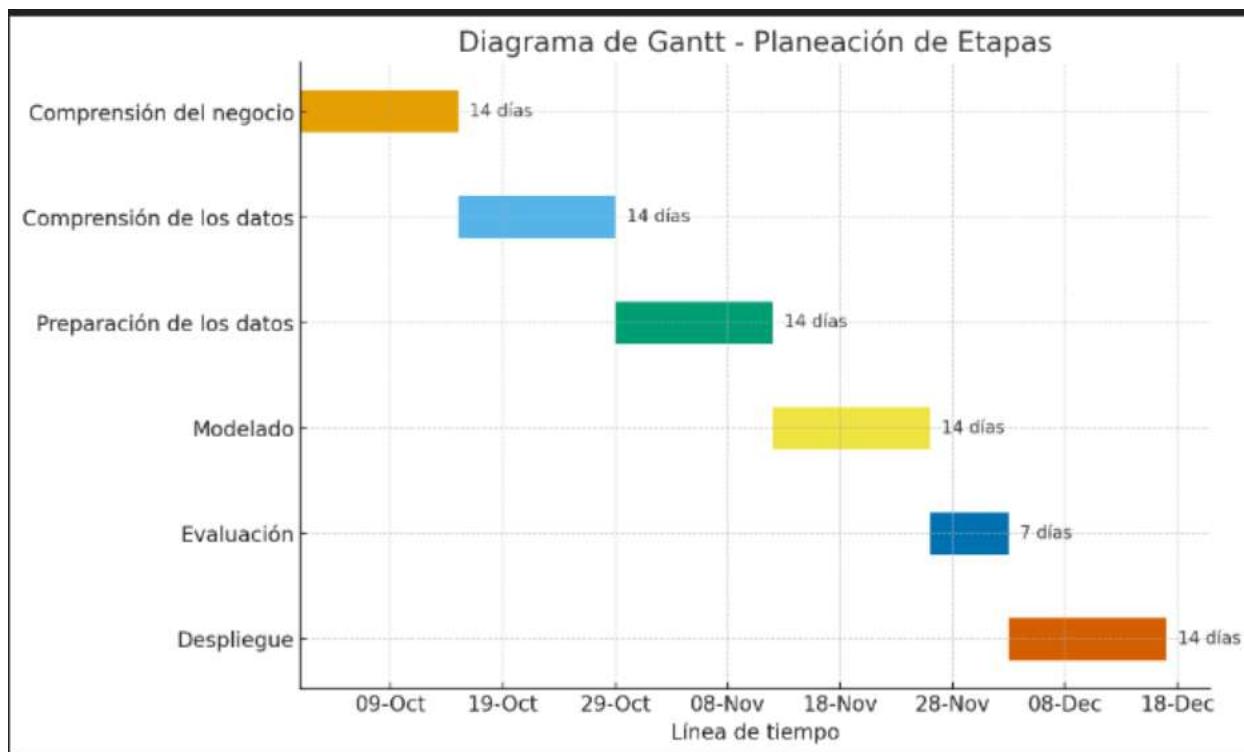


Ilustración 2

Diagrama de Gantt

Conclusión

El proyecto de predicción de demanda eléctrica en ElectroVille ofrece una solución **práctica y estratégica** para anticipar picos de consumo en la red. Al combinar modelos de series de tiempo con variables meteorológicas y de calendario, se busca lograr una predicción precisa que permita **garantizar la continuidad del servicio eléctrico y optimizar el uso de recursos energéticos**.

La metodología **CRISP-DM** asegura un desarrollo estructurado, iterativo y alineado con los objetivos de negocio, lo que aumenta la viabilidad del proyecto en entornos reales.

No obstante, se identifican algunos **riesgos** relevantes:

- Dependencia de la calidad de los datos meteorológicos externos.
- Posible degradación del modelo con el tiempo si no se actualiza con nuevos patrones de consumo.
- Necesidad de soporte técnico constante para el mantenimiento del pipeline de predicción.

Por lo que la propuesta representa una **estrategia viable y escalable** para enfrentar los retos de la gestión de demanda energética en contextos urbanos.

Referencias

Chapman, P., Clinton, J., Kerber, R., Khabeza, T., Reinartz, T., & Wirth, R. (1997). The CRISP-DM Process Model. In <https://keithmccormick.com/wp-content/uploads/CRISP-DM%20No%20Brand.pdf>. <https://keithmccormick.com/wp-content/uploads/CRISP-DM%20No%20Brand.pdf>

IBM SPSS Modeler. (2025, January 6). <https://www.ibm.com/docs/en/spss-modeler/18.6.0?topic=dm-crisp-help-overview>

IBM SPSS Modeler Subscription. (2021, August 17). <https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>

Roberto, C. (2022, May 31). *Crisp-DM: las 6 etapas de la metodología del futuro*. Blog MBA Esalq USP. <https://blog.mbauspapel.com/es/crisp-dm-las-6-etapas-de-la-metodologia-del-futuro/>

Zúñiga, J. J. E. (2020). Aplicación de metodología CRISP-DM para segmentación geográfica de una base de datos pública. *Ingeniería Investigación Y Tecnología*, 21(1), 1–13.

<https://doi.org/10.22201/fi.25940732e.2020.21n1.008>