



DE LA INFORMACIÓN AL CONOCIMIENTO:

MINERÍA DE DATOS COMO
HERRAMIENTA PARA
TRANSFORMAR REALIDADES

Dra. Doricela Gutiérrez Cruz



SOY SPEAKER EN

DEV DAY 4
Women



DRA. DORICELA GUTIÉRREZ CRUZ

**Docente—Investigadora de Tiempo
Completo · Ciencia y Minería de Datos**

 **NOVARTIS**



SOY SPEAKER EN

DEV DAY 4
Women



DRA. DORICELA GUTIÉRREZ CRUZ

¿Qué cambia cuando los datos
dejan de describir
y empiezan a orientar decisiones?

 NOVARTIS



01 · FUNDAMENTOS

Qué es la minería de datos y cómo transforma datos crudos en conocimiento accionable.

03 · METODOLOGÍA

El proceso de un proyecto de datos y los retos éticos que implica.

02 · APLICACIONES

Investigación aplicada en salud, educación y seguridad ciudadana.

04 · MUJERES STEAM

Empoderamiento y una invitación a sumar más mujeres a la ciencia de datos.



¿Que es la minería
de datos?



NOVARTIS



devday4w.com



NOVARTIS

Minería de datos

Proceso de descubrir patrones y conocimiento útil oculto en grandes volúmenes de datos, uniendo estadística, aprendizaje automático y bases de datos.

La minería de datos conecta evidencia, contexto y acción.



De los datos al conocimiento

La escalera del valor: cada nivel agrega significado y capacidad de decisión.



NIVEL 1

NIVEL 2

NIVEL 3

NIVEL 4

DATO

Hechos crudos sin contexto: un número, una fecha, un registro.

INFORMACIÓN

Datos organizados y con contexto: responden qué, quién y cuándo.

CONOCIMIENTO

Patrones y relaciones que explican el porqué y permiten actuar.

SABIDURÍA

Decisiones y políticas que transforman la realidad de las personas.

¿Por qué ahora?

Cada día generamos cerca de 2.5 trillones de bytes de datos y buena parte del mundo digital es reciente. El reto ya no es tener datos, sino convertirlos en decisiones.



| El proceso: CRISP-DM

Un proyecto de minería de datos es iterativo, no un camino lineal.



FASE 1



COMPRENDER

Entender el problema y el contexto: negocio y datos disponibles.

FASE 2



PREPARAR

Limpiar, integrar y transformar los datos. Suele ser el 70% del trabajo.

FASE 3



MODELAR

Aplicar algoritmos de minería y entrenar modelos sobre los datos.

FASE 4



DESPLEGAR

Validar, interpretar y poner el conocimiento en acción.

TODO INICIA CON UNA PREGUNTA DE IMPACTO

- ¿Qué riesgo queremos anticipar?
- ¿Qué patrón necesitamos explicar?
- ¿Qué decisión cambiará con el resultado?
- ¿Cómo sabremos si realmente mejoró la realidad?



CUATRO TAREAS ESENCIALES

CLASIFICACIÓN

Asigna categorías a nuevos casos a partir de ejemplos. Ej.: ¿es un paciente de alto riesgo?

ASOCIACIÓN

Halla reglas «si ocurre A, suele ocurrir B». Ej.: factores que coinciden en un fenómeno.

AGRUPAMIENTO

Descubre grupos naturales sin etiquetas previas. Ej.: segmentar zonas o estudiantes por perfil.

PREDICCIÓN

Estima valores o hechos futuros con datos históricos. Ej.: demanda, deserción o incidencia.

Herramientas y tecnologías



◆ El ecosistema de código abierto que convierte datos en conocimiento ◆



Python / R

Análisis y modelado



SQL

Gestión de datos



ML

Aprendizaje automático



BI

Visualización



Herramientas gratuitas, abiertas y con enormes comunidades:
al alcance de cualquier investigadora.

¿Qué método elegir y cómo evaluar?

Selección del modelo y criterios para confiar en sus resultados.

1 MÉTODOS



Árboles y reglas

Útiles cuando la explicación es prioritaria: rutas de decisión, umbrales y perfiles.



Regresión

Adecuada para estimar probabilidades o valores y comunicar efectos de variables.



Clustering

Permite descubrir segmentos, trayectorias o zonas con comportamientos semejantes.



Ensamblés

Random Forest y otros combinan modelos para mejorar estabilidad y desempeño.

2 EVALUACIÓN



Desempeño

Accuracy, precisión, recall, F1, ROC-AUC y error según el tipo de problema.



Generalización

Separar entrenamiento y prueba; usar validación cruzada y validación externa.



Explicabilidad

Entender qué variables influyen, con qué dirección y bajo qué condiciones.



Utilidad

Un modelo técnicamente bueno debe producir una decisión o intervención viable.



Elegir el método adecuado y evaluarlo con rigor asegura modelos confiables, explicables y útiles para decisiones que generan impacto.



**Una trayectoria con
propósito:
Investigar para
transformar**



 NOVARTIS



devday4w.com

TRES ÁMBITOS, UNA MISMA INTENCIÓN

Usar evidencia para comprender problemas complejos y orientar decisiones.



SALUD: APLICACIONES DE MINERÍA DE DATOS

Del expediente clínico al apoyo para decisiones de atención

- 1 **320 expedientes** de pacientes con cáncer en contexto de cuidados paliativos (2018–2023).
- 2 **Integración y revisión** de variables clínicas, sociales y de atención.
- 3 Minería de datos para identificar **perfiles, asociaciones y rutas de priorización**.
- 4 El objetivo no es reemplazar el juicio clínico, sino **fortalecer el cuidado** y la **toma de decisiones**.



PREPARACIÓN

Integración, codificación y control de calidad de datos.



ASOCIACIONES

Reglas Apriori para explorar combinaciones frecuentes.



DECISIÓN

Árboles REPTree para representar perfiles comprensibles.



CAUTELA

Los hallazgos requieren interpretación clínica y validación.



Encontrar patrones para cuidar mejor.

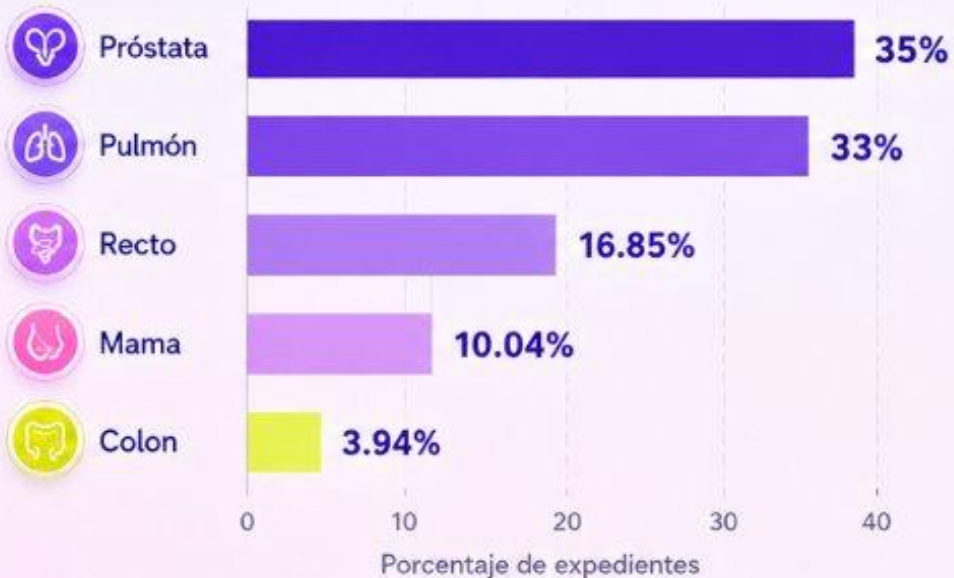
SALUD: DE LA EVIDENCIA A LA PRIORIZACIÓN

Cómo convertimos expedientes clínicos en hallazgos útiles para orientar el cuidado.



BASE CLÍNICA 2018–2023

320 expedientes de pacientes con cáncer en contexto de cuidados paliativos.



ANÁLISIS APLICADO



Preparación

Integración, codificación y calidad de variables.



Asociaciones

Reglas Apriori para explorar combinaciones frecuentes.



Decisión

Árboles REPTree para representar perfiles comprensibles.



Cautela

Interpretación clínica y validación.

ruta de valor



Identificar

Definir una necesidad de atención o riesgo.



Analizar

Explorar perfiles, asociaciones y variables.



Priorizar

Traducir hallazgos en alertas, rutas o criterios.



Evaluar

Revisar utilidad clínica, equidad, seguridad y resultados.



La minería de datos no reemplaza el juicio clínico: lo fortalece con evidencia.



EDUCACIÓN: APLICACIONES DE MINERÍA DE DATOS

De las trayectorias escolares al acompañamiento oportuno



1

98 estudiantes y 18 variables académicas y de trayectoria.



2

Programas analizados:
ICO y LIA.



3

Objetivo:
clasificar niveles de riesgo para
orientar acompañamiento oportuno.



4

El riesgo académico no es una etiqueta:
es una oportunidad de intervención.



SEÑALES CLAVE



Situación
académica



Semestre
activo



Promedio



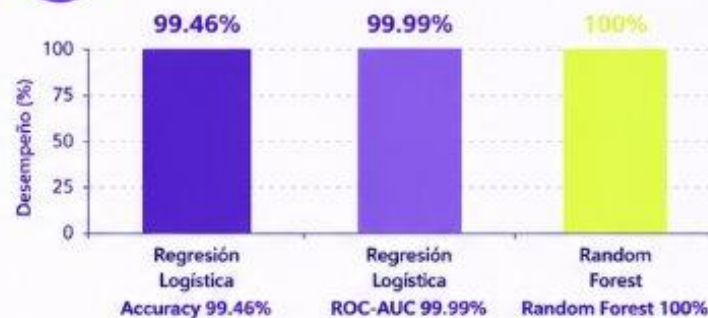
Avance
académico



La situación académica y el semestre activo muestran la **relación más alta**. El promedio y el avance aportan señales adicionales.



RESULTADOS EXPERIMENTALES



Desempeño **muy alto** de Regresión Logística y Random Forest. Resultados consistentes incluso con muestra pequeña.



DEL MODELO AL ACOMPAÑAMIENTO



1

DETECTAR

Identificar señales tempranas y casos que requieren revisión.



2

CONTEXUALIZAR

Contrastar el resultado con historia académica y situación personal.



3

INTERVENIR

Tutoría, asesoría, regularización, orientación y seguimiento.



4

APRENDER

Medir resultados y actualizar el modelo con nuevos semestres.



El ciclo se retroalimenta para mejorar decisiones y resultados.



EDUCACIÓN: PENSAMIENTO COMPUTACIONAL

Del artículo académico a estrategias para fortalecer el aprendizaje y la resolución de problemas.



¿QUÉ APORTA EL PENSAMIENTO COMPUTACIONAL?



1

Descomposición de problemas complejos.



2

Reconocimiento de patrones y relaciones.



3

Abstracción para identificar lo esencial.



4

Diseño de algoritmos y pasos de solución.



HALLAZGOS DEL ARTÍCULO



Favorece el razonamiento lógico y crítico.



Mejora la resolución de problemas en distintos contextos.



Fortalece creatividad, análisis y toma de decisiones.



Puede integrarse en actividades de programación y sin programación.



COMPONENTES CLAVE



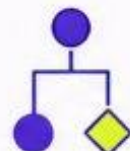
Descomponer



Patrones



Abstraer



Algoritmos



APLICACIÓN EDUCATIVA



1

EXPLORAR

plantear retos o problemas.



2

DISEÑAR

construir estrategias de solución.



3

PROBAR

aplicar, comparar y ajustar.



4

TRANSFERIR

usar lo aprendido en nuevos contextos.



El pensamiento computacional no solo enseña tecnología: enseña a pensar mejor.



SEGURIDAD CIUDADANA: MINERÍA DE DATOS PARA LEER EL TERRITORIO

Aplicaciones en Nezahualcóyotl para comprender patrones, orientar prevención y apoyar decisiones.

-  1 La incidencia tiene dimensión espacial, temporal y social.
-  2 Los patrones ayudan a reconocer concentraciones, recurrencias y cambios.
-  3 El análisis debe incorporar contexto territorial y perspectiva de género.
-  4 El objetivo es priorizar prevención, servicios y atención comunitaria.



¿QUÉ BUSCA LA MINERÍA EN SEGURIDAD?



DÓNDE

Zonas o corredores con concentración relativa de eventos.



CUÁNDO

Horarios, días, temporadas y cambios en el tiempo.



CÓMO

Tipologías, recurrencias y combinaciones de variables.



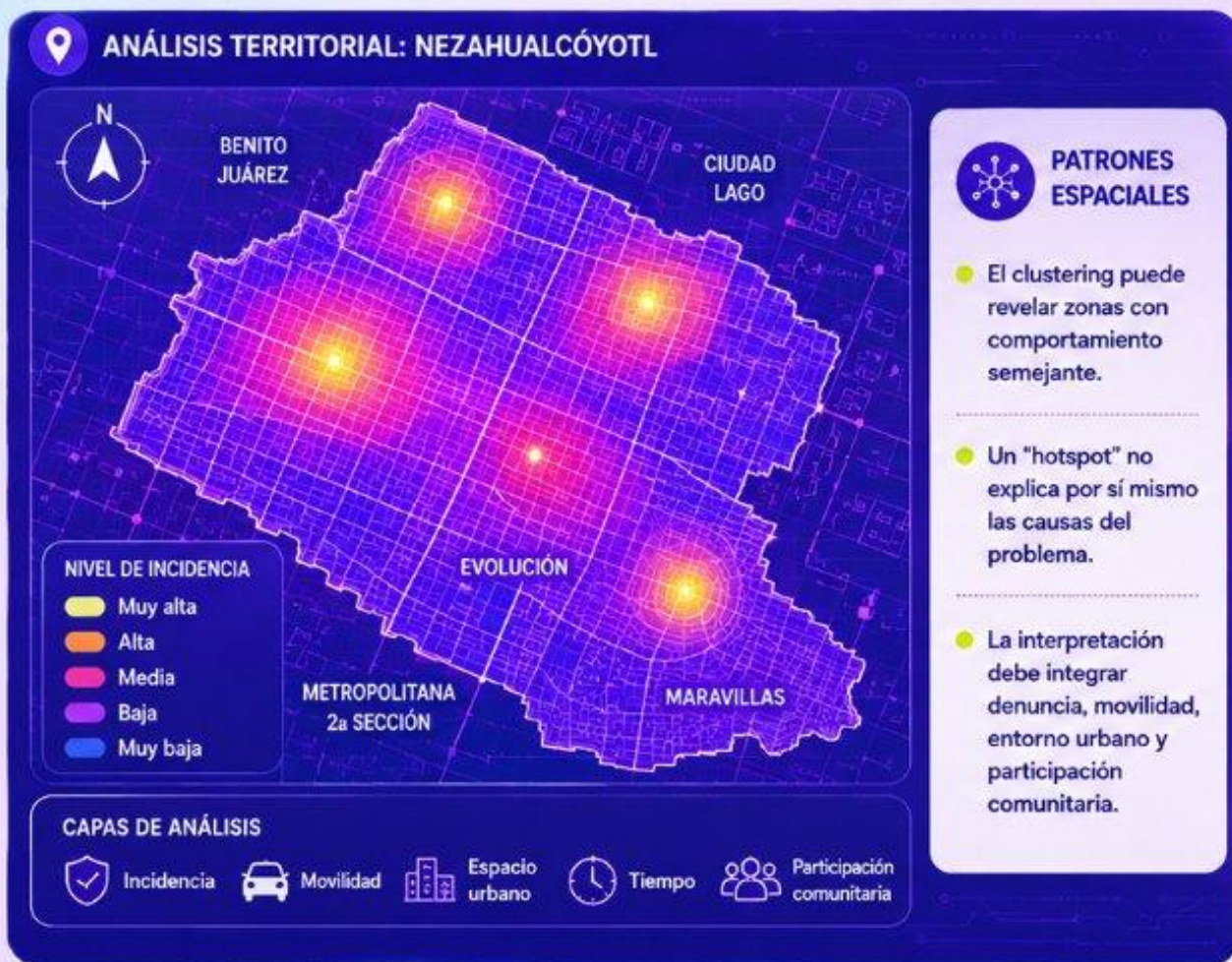
PARA QUÉ

Priorizar prevención, servicios, iluminación, movilidad y atención comunitaria.



SEGURIDAD CIUDADANA: DE LA INCIDENCIA A LA PREVENCIÓN

Cómo convertimos patrones territoriales en acciones preventivas con enfoque comunitario.



RUTA DE ANÁLISIS



EL PATRÓN COMÚN

UNA PREGUNTA HUMANA
Cuidar mejor, acompañar a tiempo o prevenir con mayor precisión.

DATOS IMPERFECTOS
Registros incompletos, sesgos, escalas distintas y necesidad de limpieza.

MODELOS EXPLICABLES
Resultados comprensibles para autoridades, profesionales y comunidades.

IMPACTO MEDIBLE
Decisiones responsables, seguimiento y aprendizaje continuo.



La minería de datos orienta la prevención cuando convierte el territorio en **conocimiento accionable**.





Mujeres STEAM

DEV DAY 4
Women

 NOVARTIS



devday4w.com

EL TALENTO CRECE CUANDO ENCUENTRA RED



La brecha: cerca de 3 de cada 10 profesionales en ciencia y tecnología son mujeres, y aún menos en datos e IA. **El talento existe; faltan oportunidades y referentes.**

- La ciencia de datos y los espacios STEM/STEAM necesitan perspectivas diversas.
- Mentoría, colaboración y visibilidad multiplican oportunidades.
- No se trata de ocupar un lugar aislado, sino de abrir caminos para otras.



Más mujeres en datos significa más ciencia, más innovación y más impacto social.



BARRERAS QUE DEBEMOS CONVERTIR EN PALANCAS

Empoderar no es solo inspirar: es crear condiciones para participar, crecer y liderar.



AISLAMIENTO
→ **RED**

Crear comunidades de práctica, mentoría y colaboración interinstitucional.



DUDA
→ **CONFIANZA**

Aprender haciendo, presentar resultados y reconocer logros propios y colectivos.



INVISIBILIDAD
→ **VOZ**

Publicar, participar, liderar proyectos y compartir referentes cercanos.



COMPETENCIA
→ **ALIANZA**

Sumar saberes, instituciones y generaciones para ampliar el impacto.



FORMACIÓN

Acceso temprano a datos, programación y matemáticas.



REDES

Comunidades y mentorías que acompañan.



REFERENTES

Mostrar mujeres científicas y tecnólogas.



OPORTUNIDADES

Espacios reales de investigación, liderazgo y decisión.

Cuando una mujer avanza sola, **progres**a; cuando avanza en red, **transforma**.

UNA INVITACIÓN A SUMAR ESFUERZOS

Cada persona puede activar una parte del cambio para que más niñas y mujeres participen en ciencia de datos y STEM.

E

ESTUDIANTES

Atrévete a formular preguntas, construir proyectos y pedir mentoría.



D

DOCENTES

Abre datos, retos, laboratorios y espacios de publicación.



I

INVESTIGADORAS

Conecta redes, comparte metodologías y acompaña nuevos liderazgos.



A

INSTITUCIONES

Financia, reconoce y sostiene trayectorias diversas en STEM.



APRENDE

una herramienta de datos



COMPARTE

tu historia inspira



MENTORA

acompaña a quien inicia



CONECTA

a una red de mujeres STEAM

La transformación se acelera cuando el conocimiento circula y la oportunidad se comparte.



TRANSFORMAR DATOS ES TRANSFORMAR POSIBILIDADES

Información → conocimiento → decisiones → realidades,
con propósito, evidencia y colaboración.

1

1. El problema guía al algoritmo.

2

2. Limpiar datos también es investigar.

3

3. Correlación no significa causalidad.

4

4. Validar es tan importante como modelar.

5

5. Explicar genera confianza.

6

6. La ética se diseña desde el inicio.

7

7. El contexto completa al dato.

8

8. El impacto debe medirse.

9

9. Las redes multiplican capacidades.

10

10. Más mujeres STEM significa mejor ciencia.

“

*La pregunta no es cuántos datos tenemos,
sino qué decisión mejoramos.*



Del dato a la acción,
y de la acción a la **transformación**.



**Los datos transforman
realidades**



 NOVARTIS



devday4w.com

PREGUNTAS



¡Gracias!
Sumemos más mujeres a la
ciencia de datos

