

Cargo:	Docente Especialista		
Nombre:	M.Sc Elvis Pachacama		
Asignatura:	Machine Learning		
Carrera:	Desarrollo de Software	Nivel:	Cuarto Nivel
Estudiante:	Adriel Alejandro Bedoya Suárez		

ACTIVIDAD PRÁCTICO EXPERIMENTAL EN EL ENTORNO ACADÉMICO
ADMINISTRACIÓN DE SERVICIOS EN AMBIENTES VIRTUALES PARA
PYMES

Objetivos

Realizar una comparación entre Data Mining y Machine Learning

Planteamiento del problema

Se debe leer el artículo subido por el profesor en el EVA+ y una vez leído toca visualizar la diferencia entre Data Mining y Machine Learning

Pasos por realizar

- 1) Leer el artículo PDF
- 2) Hay que explicar que es Data Mining
- 3) Hay que explicar que es Machine Learning
- 4) Realizar una comparación entre Data Mining y Machine Learning



Desarrollo

DATA MINING

El punto es extraer patrones, conocimiento o información a partir de un conjunto de datos, su resultado principal es conocimiento estático (insignias, reglas, modelos descriptivos). Este modelo se usa principalmente para análisis puntual donde una vez extraído el conocimiento, el proceso suele terminar.

Además, se encarga de trabajar con datos históricos y no está diseñado para adaptarse automáticamente a cambios futuros. Su ciclo de vida es parecido al de Machine Learning:

- Preparar los datos
- Construir el modelo
- Evaluar resultados

Su calidad se mide por la precisión del modelo y la calidad de análisis final, además, está bien cubierto por CRISP-DM, que es un estándar de facto en la industria. ¿Cómo se puede diferenciar esto con Machine Learning? Esto es gracias a que nosotros tenemos de datos al modelo y del modelo al conocimiento generado.



MACHINE LEARNING

Su punto principal es que debe construir modelos que tomen decisiones y realicen predicciones automáticamente, será una aplicación activa que va a operar de forma continua en un entorno real. Este modelo se despliega en producción y se usa para inferir en resultados en tiempo real o de forma continua, este modelo forma parte de un sistema o producto

Este opera en entornos dinámicos y cambiantes, su rendimiento puede degradarse con el tiempo por lo que va a requerir un monitoreo y mantenimiento continuo. Su ciclo de vida es:

- Despliegue
- Monitoreo
- Reentrenamiento
- Actualización del Modelo

La calidad para este modelo se basa en el desempeño técnico, robustez, escalabilidad, explicabilidad, el impacto del negocio y se requiere una garantía de calidad en todas sus fases.

Usar CRISP-DM ya no es suficiente porque no considera monitoreo post-despliegue y no gestiona riesgos de degradación del modelo. Para este tipo de modelo se usa CRISP-ML(Q) como extensión específica para Machine Learning. Sus pasos para ejecutarse de la mejor manera son:

- 1) Datos
- 2) Entrenamiento
- 3) Modelo
- 4) Inferencia continua en producción



¿QUÉ RELACIÓN TIENEN?

Los conceptos son parecidos, pero son bastante diferentes, Data Mining busca entender los datos, pero Machine Learning busca actuar sobre el mundo real usando datos, de forma continua y controlada.

Aspecto	Data Mining	Machine Learning
Propósito principal	Descubrir patrones, relaciones y conocimiento oculto en los datos.	Construir modelos que aprendan de los datos y tomen decisiones automáticas.
Resultado final	Conocimiento, reglas, reportes o modelos descriptivos.	Un modelo desplegado que genera predicciones o acciones en tiempo real.
Uso del modelo	Análisis puntual o exploratorio; el modelo no siempre se reutiliza.	Uso continuo en sistemas o productos reales.
Relación con el tiempo	Enfocado en datos históricos.	Opera en entornos dinámicos con datos cambiantes.
Adaptación al cambio	No está diseñado para adaptarse automáticamente.	Requiere monitoreo, reentrenamiento y mantenimiento.
Ciclo de Vida	Termina tras el análisis y la evaluación.	Cubre todo el ciclo: desarrollo, despliegue, monitoreo y actualización.



Despliegue en producción	Generalmente no obligatorio.	Esencial para cumplir su objetivo.
Riesgo principal	Interpretaciones incorrectas de los datos.	Degradación del modelo, sesgos, fallos en producción.
Aseguramiento de la calidad (QA)	Limitado; se enfoca en la calidad del resultado final.	Integrado en todas las fases del proceso.
Impacto en el negocio	Apoya la toma de decisiones humanas.	Automatiza decisiones que impactan directamente al negocio.
Modelo de proceso común	CRISP-DM.	CRISP-ML(Q) (extensión de CRISP-DM).
Necesidad de monitoreo	No necesaria una vez terminado el análisis.	Crítica para evitar pérdida de desempeño.
Ejemplo	Análisis de clientes para segmentación.	Sistema de recomendación en producción.

Para poder entender esto, necesitamos saber que es el modelo CRISP-DM y el modelo CRISP-ML(Q) (extensión de CRISP-DM).

1. CROSS-INDUSTRY STANDARD PROCESS FOR DATA MINING (CRISP-DM)

Es un modelo estándar para los proyectos que usan minería de datos, creado para guiar de manera estructurada el desarrollo de proyectos cuyo objetivo es la extracción de conocimiento a partir de datos. Su propósito es ordenar el trabajo de



los proyectos, reducir errores al seguir con un proceso definido y asegurar que los resultados estén alineados con los objetivos del negocio; es un modelo iterativo (ósea que puede volver a fases anteriores) y está enfocado en análisis y descubrimiento de conocimiento, no en sistemas activos en producción

Sus limitaciones es que no está diseñado para aplicaciones de Machine Learning en producción ya que no considera el monitoreo del modelo, no tiene en cuenta la degradación del desempeño con el tiempo y tampoco incluye una metodología en QA

2. CROSS-INDUSTRY STANDARD PROCESS FOR MACHINE LEARNING WITH QUALITY ASSURANCE (CRISP-ML(Q))

Es un modelo propuesto específicamente para aplicaciones de Machine Learning, que extiende y mejora el CRISP-DM, su objetivo es guiar el desarrollo de aplicaciones de Machine Learning en producción, incorporar QA en todas sus fases y reducir el alto índice de fracaso de proyectos con este modelo.

¿En qué se diferencian?

Se diferencia de CRISP-DM porque este modelo toma decisiones automáticamente en los entornos reales, pero el otro solo descubre conocimiento, además en la parte del ciclo de vida de CRISP-DM es que el proyecto termina tras realizar el análisis, por otro lado, CRISP-ML(Q) el que el proyecto continúa mientras el modelo esté en uso



CONCLUSIONES

- 1) Entre Data Mining y Machine Learning comparten técnicas y bases estadísticas similares, pero siguen objetivos distintos: el primero se orienta al descubrimiento de conocimiento, mientras que el segundo busca la automatización de decisiones en entornos reales y dinámicos.
- 2) CRISP-DM es un modelo sólido y muy aceptado en proyectos de minería de datos, pero presenta limitaciones bastante importantes cuando se aplica a proyectos de Machine Learning en producción.
- 3) El ciclo de vida del Machine Learning no termina con la construcción del modelo, ya que el desempeño puede degradarse con el tiempo debido a cambios en los datos y en el contexto del negocio.
- 4) CRISP-ML(Q) surge como una evolución necesaria de CRISP-DM, incorporando monitoreo continuo, mantenimiento del modelo y aseguramiento de la calidad en todas las fases del proceso.
- 5) La falta de un enfoque sistemático de calidad y gestión de riesgos es una de las principales causas del fracaso de proyectos de Machine Learning, lo cual justifica la adopción de metodologías como CRISP-ML(Q).



RECOMENDACIONES

- 1) Seleccionar el modelo de proceso adecuado según el objetivo del proyecto, utilizando CRISP-DM para análisis de datos y CRISP-ML(Q) para aplicaciones de Machine Learning que se desplieguen en producción.
- 2) Integrar el aseguramiento de la calidad desde las etapas iniciales del proyecto, definiendo métricas, criterios de éxito y riesgos técnicos antes de construir los modelos.
- 3) Planificar el monitoreo y mantenimiento del modelo como parte esencial del proyecto, considerando la detección de degradación y la necesidad de reentrenamiento.
- 4) Fortalecer la alineación entre negocio, datos y tecnología, asegurando que los objetivos del modelo de Machine Learning generen un impacto real y medible en la organización.
- 5) Capacitar a los equipos en metodologías específicas para Machine Learning, promoviendo el uso de estándares como CRISP-ML(Q) para reducir errores, costos y fallos en producción.

