

Sigla Asignatura	FMY0100	Nombre de la Asignatura	Fundamentos de Machine Learning	Tiempo	40 minutos
Experiencia de Aprendizaje N° 1	Fundamentos de estadística				
Actividad N° 1.3	Análisis exploratorio de datos				
Nombre del Recurso Didáctico	1.3.6 Actividad Aplicando Metodologías				

1. Aprendizajes e indicadores de logro

Aprendizajes (Procedimentales, Actitudinales y conceptuales)	Indicadores de logro
Reconocer las características, fases y tareas de las metodologías (CRISP-DM, KDD, SEMMA) usadas para la extracción de datos.	Identifica las principales características, diferencias entre las fases de las metodologías para su aplicación dependiendo de las necesidades del proyecto. Distingue las principales tareas de las metodologías disponibles para su aplicación en el análisis de datos.
Reconocer conceptos de estadística descriptiva orientada al manejo de datos, de acuerdo al tipo de herramienta utilizada.	Reconoce conceptos de estadística básica orientada al manejo de datos, de acuerdo al tipo de herramienta utilizada. Reconoce los diferentes modelos matemáticos aplicables al manejo de datos, según su uso en la industria.
Demuestra rigurosidad en el seguimiento de las etapas de la metodología de análisis de los datos.	Demuestra capacidad de análisis en las etapas de aplicación de la metodología. Es riguroso en la adhesión a las etapas de la metodología, dándole continuidad al proceso de análisis de datos.

2. Descripción general actividad

1. Esta actividad tiene **carácter formativo**, es decir: permitirá reforzar lo aprendido en base a un análisis de un dataset dado.
2. Esta actividad potenciará el desarrollo de soluciones basadas en la ciencia de datos, para satisfacer las necesidades de la industria.
3. Esta actividad se debe desarrollar en forma individual.

Docente Diseñador	Gmo. Pinto F.	Revisor metodológico	Manuela Jiménez
-------------------	---------------	----------------------	-----------------

PASO 1

Carga en una planilla Excel el archivo CSV “adults-with-diabetes-per-100-lghc-indicator-23”.

Este archivo es un conjunto de estadísticas históricas del estado de California, USA, sobre ciertas características de la población afectada por diabetes.

PASO 2

Analiza los datos y genera una descripción de ellos en forma escrita. Luego, responde lo siguiente también en forma escrita:

1. ¿Qué muestran en cada columna?
2. ¿Cuántos años de información muestran?
3. Por cada año ¿qué datos se muestran?
4. ¿Podrías indicar cuál sería un posible objetivo de minería de datos en este caso?
5. Indica para cada metodología (KDD, SEMMA y CRISP-DM) que fase estás ejecutando en este paso.

PASO 3

6. Si tuvieras que eliminar algún dato a nivel año, ¿cuál elegirías? ¿por qué?
7. ¿Se puede decir que existen datos irrelevantes?
8. Haz un resumen de los datos por año e indica si existen “outliers”.
9. Busca si hay “missing values”. ¿A qué año corresponden?
10. Indica para cada metodología (KDD, SEMMA y CRISP-DM) que fase estás ejecutando en este paso.
11. ¿Qué fase vendría a continuación?

PASO 4

Prepara un breve informe con tus respuestas para exponer en la clase.

- ✓ **Cuentas con 40 minutos para realizar este trabajo.**
- ✓ **Importante es que consideres que el docente elegirá algún trabajo para exponer los resultados.**
- ✓ **Para finalizar, el docente realiza una retroalimentación de toda la Experiencia de Aprendizaje.**

Docente Diseñador	Gmo. Pinto F.	Revisor metodológico	Manuela Jiménez
-------------------	---------------	----------------------	-----------------