



UES

Universidad Estatal de Sonora
La Fuerza del Saber Estimulará mi Espíritu

MANUAL DE PRÁCTICAS DE LABORATORIO

Big Data y Minería de Datos

Laboratorio

Programa Académico
Plan de Estudios
Fecha de elaboración
Versión del Documento

Ing. en Software
2/07/2025
Ver. 1.0



Dra. Martha Patricia Patiño Fierro
Rectora

Mtra. Ana Lisette Valenzuela Molina
**Encargada del Despacho de la Secretaría
General Académica**

Mtro. José Antonio Romero Montaña
Secretario General Administrativo

Lic. Jorge Omar Herrera Gutiérrez
**Encargado de Despacho de Secretario
General de Planeación**

Tabla de contenido

INTRODUCCIÓN.....	4
IDENTIFICACIÓN	5
<i>Carga Horaria del alumno</i>	<i>5</i>
<i>Consignación del Documento</i>	<i>5</i>
MATRIZ DE CORRESPONDENCIA	6
NORMAS DE SEGURIDAD Y BUENAS PRÁCTICAS	7
<i>Reglamento general del laboratorio</i>	<i>7</i>
<i>Reglamento de uniforme.....</i>	<i>7</i>
<i>Uso adecuado del equipo y materiales.....</i>	<i>7</i>
<i>Manejo y disposición de residuos peligrosos.....</i>	<i>7</i>
<i>Procedimientos en caso de emergencia</i>	<i>7</i>
RELACIÓN DE PRÁCTICAS DE LABORATORIO POR ELEMENTO DE COMPETENCIA....	8
PRÁCTICAS.....	3
FUENTES DE INFORMACIÓN	7
ANEXOS	3

INTRODUCCIÓN

Como parte de las herramientas esenciales para la formación académica de los estudiantes de la Universidad Estatal de Sonora, se definen manuales de práctica de laboratorio como elemento en el cual se define la estructura normativa de cada práctica y/o laboratorio, además de representar una guía para la aplicación práctica del conocimiento y el desarrollo de las competencias clave en su área de estudio. Su diseño se encuentra alineado con el modelo educativo institucional, el cual privilegia el aprendizaje basado en competencias, el aprendizaje activo y la conexión con escenarios reales.

Con el propósito de fortalecer la autonomía de los estudiantes, su pensamiento crítico y sus habilidades para la resolución de problemas, las prácticas de laboratorio integran estrategias didácticas como el aprendizaje basado en proyectos, el trabajo colaborativo, la experimentación guiada y el uso de tecnologías educativas. De esta manera, se promueve un proceso de enseñanza-aprendizaje dinámico, en el que los estudiantes no solo adquieren conocimientos teóricos, sino que también desarrollan habilidades prácticas y reflexivas para su desempeño profesional.

Señalar en este apartado brevemente los siguientes elementos según corresponda:

- Propósito del manual
- Justificación de su uso en el programa académico
- Competencias a desarrollar
 - **Competencias blandas:** Habilidades transversales que se refuerzan en las prácticas, como la comunicación, el trabajo en equipo, el uso de tecnologías, etc.
 - **Competencias disciplinares:** Conocimientos específicos del área del laboratorio, incluyendo fundamentos teóricos y habilidades técnicas.
 - **Competencias profesionales:** Aplicación de los conocimientos adquiridos en escenarios reales o simulados, en concordancia con el perfil de egreso del programa.

IDENTIFICACIÓN

Nombre de la Asignatura		Big Data y Minería de Datos	
Clave	061CE007	Créditos	6
Asignaturas Antecedentes		Plan de Estudios	2021

Área de Competencia	Competencia del curso
Implementar soluciones e innovaciones tecnológicas, con el fin de contribuir a una planeación responsable de los recursos humanos, tecnológicos y financieros, con base en las necesidades, la problemática analizada y los estándares de calidad establecidos por la organización.	Evaluar técnicas avanzadas de minería de datos en proyectos de Big Data, con el objetivo de extraer información significativa y relevante para la toma de decisiones estratégicas en organizaciones, promoviendo así un enfoque de mejora continua y la adaptación a entornos empresariales dinámicos y complejos.

Carga Horaria de la asignatura

Horas Supervisadas			Horas Independientes	Total de Horas
Aula	Laboratorio	Plataforma		
3	1	1	2	7

Consignación del Documento

Unidad Académica	Unidad Académica Hermosillo
Fecha de elaboración	02/07/2025
Responsables del diseño	Sergio Ramón Rossetti López, Gabriel Garcia Corral
Validación	
Recepción	Coordinación de Procesos Educativos

EMATRIZ DE CORRESPONDENCIA

Señalar la relación de cada práctica con las competencias del perfil de egreso

PRÁCTICA	PERFIL DE EGRESO
EC2 Exploración y Limpieza de Datos para Minería de Datos	<p>Aplicar soluciones e innovaciones tecnológicas con la finalidad de automatizar los procesos, atendiendo los principios de la organización y gestión efectiva de la información en los departamentos que así lo requieran, poniendo en práctica sus habilidades de trabajo en equipo y planeación.</p>
EC2 Aplicación de la Minería de Datos para Análisis Predictivo	
EC3 Técnicas de Minería de Datos	
EC3 Práctica: Aplicación de Técnica de Minería con Python (Clúster)	

NORMAS DE SEGURIDAD Y BUENAS PRÁCTICAS

Reglamento general del laboratorio

El presente reglamento establece los lineamientos básicos de comportamiento, seguridad, uso adecuado del equipo y respuesta ante emergencias en las sesiones de laboratorio realizadas en centros de cómputo de la universidad.

Reglamento de uniforme

En esta institución pública no existe una política obligatoria de uniforme. Sin embargo, se solicita a los estudiantes presentarse con ropa adecuada y respetuosa del entorno académico. Está prohibido asistir con prendas que representen riesgos para la seguridad personal o de otros, tales como sandalias en laboratorios técnicos, camisetas con mensajes ofensivos o ropa que impida la movilidad. Se debe mantener una imagen que refleje el compromiso con la actividad académica.

Uso adecuado del equipo y materiales

- El equipo de cómputo debe utilizarse exclusivamente con fines académicos o institucionales.
- Queda prohibida la instalación de software no autorizado, el cambio de configuraciones del sistema, el uso de dispositivos externos sin permiso o la manipulación del hardware.
- Los usuarios deben respetar la estación de trabajo asignada, no desconectar cables, ni intercambiar periféricos.
- Cualquier falla, daño o comportamiento inusual en el equipo debe reportarse de inmediato al docente o responsable del laboratorio.
- Está estrictamente prohibido comer, beber o fumar dentro del laboratorio.
- Al finalizar la sesión, el estudiante debe dejar su estación de trabajo limpia, ordenada y cerrando sesión adecuadamente.

Manejo y disposición de residuos peligrosos

En los centros de cómputo no se generan residuos peligrosos de tipo biológico o químico. No obstante, los residuos electrónicos menores (pilas, memorias dañadas, cables, etc.) que se utilicen en prácticas especiales deberán entregarse al docente para su correcta disposición conforme a los lineamientos institucionales de reciclaje electrónico.

No deben desecharse estos materiales en botes comunes, ni ser retirados del laboratorio por los alumnos.

Procedimientos en caso de emergencia

En caso de una emergencia como sismo, incendio o amenaza, los estudiantes deberán suspender sus actividades de inmediato, evacuar el laboratorio de forma ordenada siguiendo las indicaciones del docente y dirigirse al punto de reunión establecido. No se debe regresar al aula ni manipular equipos hasta que las autoridades lo indiquen. Se debe mantener la calma en todo momento y colaborar con el grupo.

RELACIÓN DE PRÁCTICAS DE LABORATORIO POR ELEMENTO DE COMPETENCIA

Elemento de Competencia al que pertenece la práctica	II y III
---	-----------------

PRÁCTICA	NOMBRE	COMPETENCIA
Práctica No. 1	EC2 Exploración y Limpieza de Datos para Minería de Datos	Analizar tendencias y relaciones significativas en grandes volúmenes de datos mediante técnicas avanzadas de minería, bajo criterios de precisión y relevancia, para respaldar la toma de decisiones estratégicas en entornos empresariales complejos y desarrollar pensamiento crítico y habilidades de resolución de problemas.
Práctica No. 2	EC2 Aplicación de la Minería de Datos para Análisis Predictivo	Analizar tendencias y relaciones significativas en grandes volúmenes de datos mediante técnicas avanzadas de minería, bajo criterios de precisión y relevancia, para respaldar la toma de decisiones estratégicas en entornos empresariales complejos y desarrollar pensamiento crítico y habilidades de resolución de problemas.
Práctica No. 3	EC3 Técnicas de Minería de Datos	Evaluar técnicas avanzadas de minería para grandes conjuntos de datos, con el fin de identificar patrones y relaciones cruciales que respalden decisiones estratégicas e informadas en entornos empresariales altamente complejos, desarrollando pensamiento crítico en el análisis de datos y toma de decisiones estratégicas.
Práctica No. 4	EC3 Práctica: Aplicación de Técnica de Minería con Python (Clúster)	Evaluar técnicas avanzadas de minería para grandes conjuntos de datos, con el fin de identificar patrones y relaciones cruciales que respalden decisiones estratégicas e informadas en entornos empresariales altamente complejos, desarrollando pensamiento crítico en el análisis de datos y toma de decisiones estratégicas.



PRÁCTICAS

NOMBRE DE LA PRÁCTICA	Exploración y Limpieza de Datos para Minería de Datos
COMPETENCIA DE LA PRÁCTICA	Analizar tendencias y relaciones significativas en grandes volúmenes de datos mediante técnicas avanzadas de minería, bajo criterios de precisión y relevancia, para respaldar la toma de decisiones estratégicas en entornos empresariales complejos y desarrollar pensamiento crítico y habilidades de resolución de problemas.

FUNDAMENTO TEÓRICO
<p>La calidad de los datos es un aspecto crítico en cualquier proceso de minería de datos. Antes de aplicar algoritmos, es esencial realizar una adecuada exploración y limpieza de los datos para detectar errores, valores faltantes, inconsistencias, valores atípicos y sesgos que puedan afectar los modelos.</p> <p>Las técnicas de exploración incluyen el análisis descriptivo, visualización y resumen estadístico de variables. La limpieza de datos puede implicar imputación, eliminación de registros, corrección de errores o transformación de datos.</p> <p>En Python, bibliotecas como pandas, numpy, matplotlib y seaborn son fundamentales para estas tareas.</p>

MATERIALES, EQUIPAMIENTO Y/O REACTIVOS
<ul style="list-style-type: none"> • Computadora con entorno Python (recomendado: Jupyter Notebook o Google Colab) • Paquetes: pandas, numpy, matplotlib, seaborn • Dataset con inconsistencias (se puede usar Titanic dataset o dataset simulado) https://www.kaggle.com/c/titanic/data • Conexión a internet

PROCEDIMIENTO O METODOLOGÍA
<ol style="list-style-type: none"> 1. Formación de equipos Trabajo colaborativo en parejas o tríos. 2. Carga del conjunto de datos Descargar el dataset propuesto y cargarlo en Python usando pandas. 3. Exploración de datos (EDA) Identificar tipo de variables, estructura general, valores únicos, valores nulos. Visualizar la distribución de datos por variable usando histogramas, diagramas de caja y gráficos de barras. 4. Detección de errores y valores atípicos

Usar `.describe()`, `.info()`, `.isnull()`, `.duplicated()`, entre otros comandos de pandas.

Visualizar valores extremos con `boxplot()` de seaborn.

5. Limpieza de datos

Eliminar duplicados.

Imputar o eliminar valores nulos justificadamente.

Corregir formatos erróneos (fechas, cadenas, etc.).

Estandarizar variables categóricas.

6. Documentación del proceso

Explicar cada decisión de limpieza tomada y justificar su impacto en la calidad del análisis posterior.

RESULTADOS ESPERADOS

Al concluir la práctica, se espera que los estudiantes obtengan un conjunto de datos completamente limpio y depurado, preparado para su posterior análisis mediante técnicas de minería de datos. El dataset deberá haber pasado por un proceso riguroso de inspección, detección y corrección de errores, incluyendo el manejo adecuado de valores nulos, duplicados, datos atípicos y variables mal codificadas o tipificadas. Asimismo, el estudiante deberá entregar un reporte técnico detallado que documente todo el proceso de limpieza realizado, mostrando claramente el estado inicial de los datos, los procedimientos aplicados y el resultado final, acompañado de las justificaciones correspondientes a cada decisión tomada. Dicho reporte deberá incluir visualizaciones gráficas comparativas como histogramas, diagramas de caja, matrices de calor o gráficos de dispersión que evidencien las mejoras logradas en la estructura y calidad del conjunto de datos. En conjunto, estos productos permitirán demostrar la comprensión del impacto que tiene una adecuada limpieza de datos sobre la fiabilidad de los modelos analíticos y la toma de decisiones basada en datos.

ANÁLISIS DE RESULTADOS

Los alumnos analizarán cómo el proceso de limpieza de datos influye directamente en la calidad de los análisis posteriores, en la interpretación de patrones y en la toma de decisiones basada en datos. Se espera que, a partir de la comparación entre el estado original del conjunto de datos y su versión depurada, los estudiantes identifiquen los beneficios de una adecuada preparación de los datos.

Se deberán responder las siguientes preguntas:

- ¿Qué tipo de errores o inconsistencias encontramos en el dataset original?

Los estudiantes deberán identificar la presencia de valores nulos, duplicados, formatos incorrectos (por ejemplo, fechas mal codificadas), variables categóricas mal estandarizadas y valores atípicos que puedan sesgar el análisis.

- ¿Qué estrategias de limpieza fueron aplicadas y por qué se eligieron esas soluciones?
Aquí se debe reflexionar sobre las técnicas utilizadas (como eliminación de registros, imputación de datos, codificación categórica, normalización, entre otras), justificando por qué fueron adecuadas para el tipo de datos tratados y el objetivo del análisis.

- ¿Qué cambios observamos en la estructura y consistencia del dataset tras la limpieza?
Se debe analizar cómo cambió la distribución de los datos, la reducción de sesgo, la mejora en la homogeneidad de los registros y la eliminación de información redundante. Este punto debe estar respaldado por visualizaciones comparativas.

- ¿Cómo influye la limpieza en la etapa posterior de minería de datos?
Finalmente, los estudiantes deberán reflexionar sobre cómo la preparación adecuada de los datos impacta en la precisión, confiabilidad y relevancia de los modelos de minería que se implementarán más adelante. También se evaluará si la limpieza ayudó a reducir la complejidad del modelo o a mejorar su capacidad de generalización.

Este análisis permitirá a los alumnos comprender que la calidad del proceso analítico depende en gran medida de la calidad de los datos iniciales, y que la limpieza no es una tarea secundaria, sino una fase crítica en todo proyecto de ciencia de datos.

CONCLUSIONES Y REFLEXIONES

El profesor dirigirá una retroalimentación grupal para discutir los diferentes enfoques presentados por los equipos, identificando ventajas y áreas de mejora. Reflexionar sobre la importancia de la limpieza de datos antes de la minería. Valorar el uso de herramientas como pandas y seaborn en procesos de preprocesamiento. Mencionar cómo errores no detectados podrían alterar la interpretación de resultados.

ACTIVIDADES COMPLEMENTARIAS

Repetir el proceso con otro dataset descargado por el alumno.
Realizar una limpieza automática con un script y luego justificar su lógica.

EVALUACIÓN Y EVIDENCIAS DE APRENDIZAJE

Criterios de evaluación	
Rúbricas o listas de cotejo para valorar desempeño	Rúbrica Práctica de Laboratorio
Formatos de reporte de prácticas	

FUENTES DE INFORMACIÓN

1. McKinney, W. (2022). Python for data analysis: Data wrangling with pandas, NumPy, and Jupyter (3ra ed.). O'Reilly Media. <https://wesmckinney.com/book/>
2. VanderPlas, J. (2016). Python data science handbook: Essential tools for working with data. O'Reilly Media. <https://jakevdp.github.io/PythonDataScienceHandbook/>
3. Han, J., Kamber, M., & Pei, J. (2011). Data mining: Concepts and techniques (3ra ed.). Morgan Kaufmann. <https://www.sciencedirect.com/book/9780123814791/data-mining>
4. Kaggle. (s.f.). Titanic - Machine Learning from Disaster. Kaggle. <https://www.kaggle.com/competitions/titanic/data>
5. Waskom, M. (2023). seaborn: Statistical data visualization. <https://seaborn.pydata.org/>
6. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830. <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>
7. Pandas development team. (2024). pandas-dev/pandas: Powerful data structures for data analysis, time series, and statistics. GitHub. <https://pandas.pydata.org/>
8. NumPy Developers. (2024). NumPy. <https://numpy.org/doc/>

NOMBRE DE LA PRÁCTICA	Aplicación de la Minería de Datos para Análisis Predictivo.
COMPETENCIA DE LA PRÁCTICA	Analizar tendencias y relaciones significativas en grandes volúmenes de datos mediante técnicas avanzadas de minería, bajo criterios de precisión y relevancia, para respaldar la toma de decisiones estratégicas en entornos empresariales complejos y desarrollar pensamiento crítico y habilidades de resolución de problemas.

FUNDAMENTO TEÓRICO
<p>El análisis predictivo es una técnica avanzada de minería de datos que busca prever comportamientos o eventos futuros mediante el análisis de datos históricos y el uso de algoritmos estadísticos y de aprendizaje automático. Se basa en la construcción de modelos matemáticos que identifican patrones dentro de conjuntos de datos complejos y voluminosos.</p> <p>Entre las aplicaciones más comunes del análisis predictivo se encuentran la predicción de ventas, el abandono de clientes, la detección de fraude, el mantenimiento predictivo de maquinaria y la segmentación de usuarios. Estos modelos utilizan variables independientes para predecir el valor de una variable objetivo o dependiente.</p> <p>Python ofrece una amplia gama de bibliotecas para este propósito, como pandas para manipulación de datos, scikit-learn para modelado predictivo, y matplotlib o seaborn para visualización, lo cual permite una integración efectiva de todo el proceso de análisis.</p>

MATERIALES, EQUIPAMIENTO Y/O REACTIVOS
<ul style="list-style-type: none"> • Computadora con acceso a Jupyter Notebook o Google Colab • Paquetes instalados: pandas, numpy, matplotlib, seaborn, scikit-learn • Dataset proporcionado por el facilitador o seleccionado de fuentes confiables como Kaggle • Conexión a internet para descarga de datos o bibliotecas

PROCEDIMIENTO O METODOLOGÍA
<p>1. Formación de equipos: Se formarán equipos de hasta tres integrantes para trabajar de manera colaborativa, favoreciendo la discusión de ideas y el análisis conjunto de resultados.</p> <p>2. Selección y comprensión del dataset: El facilitador proporcionará un conjunto de datos estructurados que incluya una variable objetivo. Cada equipo deberá realizar una descripción inicial del contenido del dataset, identificando variables numéricas, categóricas y posibles valores ausentes.</p> <p>Exploración y limpieza de los datos: Se realizará una exploración estadística básica utilizando funciones como <code>.describe()</code>, <code>.info()</code> y <code>.value_counts()</code>. Posteriormente, se aplicarán técnicas de limpieza: imputación de valores nulos, codificación de variables categóricas, normalización o estandarización según corresponda.</p> <p>3. División de los datos:</p>

Se dividirá el dataset en dos subconjuntos: entrenamiento (70%) y prueba (30%), utilizando `train_test_split` de `scikit-learn`.

4. Entrenamiento del modelo:

Se aplicará un modelo de regresión o clasificación, según la naturaleza de la variable objetivo. Podrá ser Regresión Lineal, Regresión Logística, Árboles de Decisión o K-Vecinos más Cercanos.

5. Evaluación del modelo:

Se utilizarán métricas apropiadas como precisión, exactitud, matriz de confusión, MAE, RMSE o R^2 , para evaluar el rendimiento del modelo.

6. Visualización de resultados:

Los equipos generarán gráficos comparativos antes y después de la limpieza de datos y presentarán visualizaciones del desempeño del modelo predictivo

RESULTADOS ESPERADOS

Se espera que los estudiantes sean capaces de implementar un modelo de análisis predictivo a partir de un conjunto de datos reales, desarrollando habilidades para la preparación del dataset, la selección de variables significativas, la codificación de variables categóricas y la validación de resultados. El producto final incluirá un reporte técnico que documente el proceso completo de minería predictiva, desde la carga de los datos hasta la interpretación de los resultados, destacando gráficamente la mejora en la calidad del dataset y el rendimiento del modelo.

ANÁLISIS DE RESULTADOS

Los alumnos deberán analizar cómo influyen la calidad del dataset y la preparación de las variables en el desempeño del modelo predictivo. Mediante la evaluación de métricas como el error cuadrático medio, la precisión o la exactitud, los equipos identificarán si el modelo predice adecuadamente el comportamiento esperado de la variable objetivo.

Se deberán responder las siguientes preguntas:

- ¿Qué variables mostraron mayor influencia sobre la predicción?

Se deberá justificar la relevancia de las variables seleccionadas, con base en su relación estadística o impacto en el modelo.

- ¿Cómo varió el desempeño del modelo antes y después de la limpieza del dataset?

Los equipos deberán presentar evidencias gráficas o numéricas que comparen la calidad del modelo utilizando datos sin limpiar y datos preparados adecuadamente.

- ¿Cuál fue el modelo más eficiente para el problema analizado?

En caso de que se hayan probado diferentes algoritmos, se deberá justificar cuál mostró mejor balance entre simplicidad, rendimiento y capacidad predictiva.

Este análisis permitirá valorar el impacto de la preparación de los datos en la calidad del análisis, y la utilidad real de los modelos en contextos empresariales.

CONCLUSIONES Y REFLEXIONES

Cada equipo redactará una conclusión fundamentada sobre el papel del análisis predictivo en el ámbito empresarial, destacando cómo el modelo generado podría apoyar decisiones estratégicas. Se reflexionará sobre la importancia de contar con datos de calidad, así como la necesidad de comprender el comportamiento de las variables involucradas. También se mencionará el potencial y los límites del modelo empleado, y cómo puede mejorarse su desempeño mediante ajustes técnicos o refinamiento del dataset.

ACTIVIDADES COMPLEMENTARIAS

Repetir el ejercicio utilizando otro algoritmo de predicción.
 Evaluar el modelo con nuevos datos (dataset distinto o modificado).
 Aplicar validación cruzada para comparar el rendimiento con otras particiones.

EVALUACIÓN Y EVIDENCIAS DE APRENDIZAJE

Criterios de evaluación	
Rúbricas o listas de cotejo para valorar desempeño	Rúbrica Práctica de Laboratorio
Formatos de reporte de prácticas	

FUENTES DE INFORMACIÓN

1. VanderPlas, J. (2016). Python Data Science Handbook: Essential tools for working with data. O'Reilly Media. <https://jakevdp.github.io/PythonDataScienceHandbook/>
2. McKinney, W. (2022). Python for Data Analysis (3ra ed.). O'Reilly Media. <https://wesmckinney.com/book/>
3. Han, J., Kamber, M., & Pei, J. (2011). Data mining: Concepts and techniques. Morgan Kaufmann. <https://www.sciencedirect.com/book/9780123814791/data-mining>
4. Scikit-learn Developers. (2024). scikit-learn: Machine learning in Python. <https://scikit-learn.org/stable/index.html>
5. Kaggle. (s.f.). Titanic - Machine Learning from Disaster. <https://www.kaggle.com/competitions/titanic/data>

NOMBRE DE LA PRÁCTICA	Técnicas de Minería de Datos
COMPETENCIA DE LA PRÁCTICA	Analizar tendencias y relaciones significativas en grandes volúmenes de datos mediante técnicas avanzadas de minería, bajo criterios de precisión y relevancia, para respaldar la toma de decisiones estratégicas en entornos empresariales complejos y desarrollar pensamiento crítico y habilidades de resolución de problemas.

FUNDAMENTO TEÓRICO

La minería de datos es el proceso de descubrir patrones útiles, relaciones ocultas y conocimientos valiosos a partir de grandes conjuntos de datos. Sus técnicas pueden clasificarse en dos grandes categorías: técnicas supervisadas (cuando existe una variable objetivo, como clasificación o regresión) y técnicas no supervisadas (cuando se busca descubrir estructuras subyacentes, como clústeres o asociaciones).

Algunas de las técnicas más utilizadas incluyen:

- Clasificación (por ejemplo, árboles de decisión, regresión logística, KNN)
- Regresión (como regresión lineal o polinómica)
- Agrupamiento (K-means, DBSCAN, jerárquico)
- Reglas de asociación (Apriori, FP-Growth)
- Detección de anomalías (Isolation Forest, Local Outlier Factor)

Estas técnicas permiten resolver problemas empresariales como segmentación de clientes, detección de fraudes, predicción de ventas y análisis de comportamiento. Python, a través de bibliotecas como scikit-learn, mlxtend o statsmodels, brinda las herramientas necesarias para implementarlas eficientemente.

MATERIALES, EQUIPAMIENTO Y/O REACTIVOS

- Estación de cómputo con entorno Jupyter o Google Colab
- Bibliotecas Python: pandas, numpy, matplotlib, scikit-learn, mlxtend, seaborn
- Dataset seleccionado por el docente con variables categóricas y numéricas
- Conexión a internet

PROCEDIMIENTO O METODOLOGÍA

1. Formación de equipos:
Se conformarán equipos de hasta tres alumnos para trabajar colaborativamente durante la práctica.
2. Exploración inicial del dataset:

Cada equipo analizará la estructura del conjunto de datos, identificando tipos de variables, presencia de valores nulos, distribución de clases (si aplica), valores atípicos, etc.

3. Aplicación de técnica supervisada:

Se seleccionará una técnica de clasificación (por ejemplo, árbol de decisión o KNN) para predecir una variable objetivo. El dataset será dividido y se entrenará el modelo con scikit-learn.

4. Aplicación de técnica no supervisada:

En paralelo, se aplicará una técnica de agrupamiento como K-means para descubrir clústeres dentro de los datos. Se visualizarán los grupos generados y se compararán con etiquetas reales (si existen).

5. Comparación de resultados:

Se analizará la efectividad de cada técnica aplicada y su utilidad para extraer conocimiento relevante a partir de los datos.

6. Presentación de hallazgos:

Los equipos presentarán los resultados obtenidos, visualizaciones clave, métricas de evaluación y reflexiones sobre las ventajas y limitaciones de cada técnica.

RESULTADOS ESPERADOS

Al finalizar la práctica, los estudiantes habrán aplicado al menos dos técnicas de minería de datos (una supervisada y una no supervisada), interpretado sus resultados y evaluado su aplicabilidad en un contexto organizacional. Deberán ser capaces de justificar por qué se seleccionó cada técnica, cómo se prepararon los datos, qué patrones se identificaron y qué utilidad estratégica puede derivarse de dichos hallazgos.

ANÁLISIS DE RESULTADOS

Los estudiantes analizarán cómo distintas técnicas de minería permiten responder a diferentes tipos de preguntas de negocio. Se valorará la capacidad del equipo para elegir la técnica adecuada según el tipo de problema (predicción, segmentación, clasificación).

Se deberán responder las siguientes preguntas:

- ¿Qué técnica supervisada resultó más eficaz para el problema planteado?

Se evaluará con base en métricas como exactitud, precisión, recall o matriz de confusión.

- ¿Qué grupos o patrones fueron detectados mediante técnicas no supervisadas?

Se valorará la coherencia de los clústeres con el conocimiento previo del dominio.

- ¿Qué diferencias clave existen entre los enfoques supervisado y no supervisado?

Los estudiantes reflexionarán sobre los beneficios y limitaciones de cada uno en distintos escenarios empresariales.

Este análisis permitirá comprender la complementariedad entre técnicas, así como su utilidad en contextos reales de minería de datos para la toma de decisiones.

CONCLUSIONES Y REFLEXIONES

Cada equipo elaborará una conclusión donde reflexione sobre la importancia de seleccionar adecuadamente la técnica de minería de datos más apropiada según el problema. Se deberá destacar el valor estratégico de transformar datos en conocimiento accionable, así como el aprendizaje obtenido respecto a la preparación del dataset, la implementación de modelos y la interpretación de resultados. También se invitará a considerar escenarios donde sería conveniente combinar técnicas supervisadas y no supervisadas.

ACTIVIDADES COMPLEMENTARIAS

Repetir la práctica con otro dataset del área de interés del estudiante.
Investigar e implementar otra técnica no supervisada como DBSCAN o jerárquico.
Generar un reporte de comparación entre modelos con validación cruzada.

EVALUACIÓN Y EVIDENCIAS DE APRENDIZAJE

Criterios de evaluación	
Rúbricas o listas de cotejo para valorar desempeño	Rúbrica Práctica de Laboratorio
Formatos de reporte de prácticas	

FUENTES DE INFORMACIÓN

1. Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques. Morgan Kaufmann. <https://www.sciencedirect.com/book/9780123814791/data-mining>
2. VanderPlas, J. (2016). Python Data Science Handbook. O'Reilly Media. <https://jakevdp.github.io/PythonDataScienceHandbook/>
3. McKinney, W. (2022). Python for Data Analysis (3ra ed.). O'Reilly Media. <https://wesmckinney.com/book/>
4. Scikit-learn Developers. (2024). Scikit-learn documentation. <https://scikit-learn.org/stable/index.html>
5. Mlxtend Developers. (2023). MLxtend: Machine Learning Extensions. <http://rasbt.github.io/mlxtend/>

NOMBRE DE LA PRÁCTICA	Aplicación de Técnica de Minería con Python (Clúster)
COMPETENCIA DE LA PRÁCTICA	Analizar tendencias y relaciones significativas en grandes volúmenes de datos mediante técnicas avanzadas de minería, bajo criterios de precisión y relevancia, para respaldar la toma de decisiones estratégicas en entornos empresariales complejos y desarrollar pensamiento crítico y habilidades de resolución de problemas.

FUNDAMENTO TEÓRICO
<p>El agrupamiento (clustering) es una técnica de minería de datos no supervisada que permite identificar grupos homogéneos dentro de grandes volúmenes de datos, basándose en la similitud entre observaciones. A diferencia de los algoritmos supervisados, el clustering no utiliza una variable objetivo, sino que busca patrones subyacentes que agrupen a los datos de forma natural.</p> <p>Uno de los algoritmos más utilizados es K-Means, que divide un conjunto de datos en k clústeres, minimizando la distancia entre los puntos y el centroide del grupo. Otros algoritmos incluyen DBSCAN, clustering jerárquico y Mean Shift. Estas técnicas son especialmente útiles en aplicaciones como segmentación de clientes, análisis de comportamiento, reducción de dimensionalidad y descubrimiento de conocimiento.</p> <p>Python ofrece herramientas potentes para implementar técnicas de clustering mediante bibliotecas como scikit-learn, seaborn, matplotlib y pandas, permitiendo una visualización clara de los grupos formados y facilitando la toma de decisiones basada en datos.</p>

MATERIALES, EQUIPAMIENTO Y/O REACTIVOS
<ul style="list-style-type: none"> • Computadora con acceso a Jupyter Notebook o Google Colab • Paquetes Python: pandas, numpy, scikit-learn, matplotlib, seaborn • Dataset multivariable (sin variable objetivo) proporcionado por el docente o descargado de fuentes abiertas • Conexión a internet

PROCEDIMIENTO O METODOLOGÍA
<ol style="list-style-type: none"> 1. Formación de equipos: Los alumnos trabajarán en equipos de hasta tres integrantes, fomentando la discusión analítica y la colaboración técnica. 2. Selección y preparación del dataset: El facilitador proporcionará un conjunto de datos con variables cuantitativas. Los equipos deberán realizar una limpieza básica: eliminación de duplicados, manejo de valores nulos y normalización de variables. 3. Aplicación del algoritmo K-Means: Determinar el número óptimo de clústeres usando el método del codo.

Aplicar el algoritmo KMeans de scikit-learn para dividir los datos en grupos.

Agregar la etiqueta de clúster al dataframe original.

4. Visualización de clústeres:

Utilizar gráficos de dispersión, diagramas de radar o mapas de calor para visualizar los grupos formados.

Si se trabaja con más de dos dimensiones, aplicar PCA para reducir la dimensionalidad y facilitar la representación visual.

5. Análisis del perfil de cada clúster:

Describir las características predominantes de cada grupo, identificando similitudes y diferencias.

Interpretar el significado estratégico de los clústeres generados.

6. Documentación del proceso:

Registrar los pasos metodológicos seguidos, incluir capturas de código y gráficas.

Justificar las decisiones tomadas (número de clústeres, normalización, etc.).

RESULTADOS ESPERADOS

Se espera que los estudiantes sean capaces de aplicar con éxito una técnica de agrupamiento sobre un conjunto de datos real, obteniendo como resultado clústeres interpretables que puedan apoyar decisiones estratégicas. Los equipos deberán entregar un informe técnico que contenga el dataset procesado, el número óptimo de clústeres, visualizaciones gráficas claras de los grupos generados, y una interpretación empresarial o contextual de cada clúster.

ANÁLISIS DE RESULTADOS

Los alumnos deberán analizar cómo el uso de técnicas de agrupamiento permite descubrir segmentos o patrones no visibles en el análisis tradicional. A través de los clústeres generados, podrán identificar grupos con comportamientos similares, lo cual es útil para personalización de estrategias, mejora de procesos o diseño de nuevos productos.

Se deberán responder las siguientes preguntas:

- ¿Cuántos clústeres fueron identificados como óptimos y por qué?

El equipo deberá justificar esta decisión con el gráfico del método del codo u otra técnica.

- ¿Qué características comparten los elementos dentro de cada clúster?

Los estudiantes deberán describir el perfil de cada grupo, explicando qué variables los agrupan.

- ¿Qué implicaciones prácticas podrían tener los clústeres identificados?

Se valorará si los estudiantes son capaces de traducir los hallazgos técnicos en posibles estrategias o decisiones organizacionales.

Este análisis contribuirá a la comprensión del valor del clustering como herramienta de descubrimiento y segmentación avanzada.

CONCLUSIONES Y REFLEXIONES

Cada equipo elaborará una conclusión reflexiva sobre la utilidad del análisis de clústeres en el contexto de la minería de datos. Se valorará la capacidad para identificar patrones ocultos y convertir los resultados técnicos en acciones concretas. Asimismo, se espera que los estudiantes reconozcan los desafíos técnicos del agrupamiento, como la selección de k , la interpretación de grupos y la necesidad de datos bien estructurados.

ACTIVIDADES COMPLEMENTARIAS

Aplicar otro algoritmo de clustering como DBSCAN o clustering jerárquico sobre el mismo dataset y comparar los resultados.

Ejecutar el análisis con un dataset del área de interés profesional del estudiante.

Utilizar métricas como Silhouette Score para validar la calidad de los clústeres.

EVALUACIÓN Y EVIDENCIAS DE APRENDIZAJE

Criterios de evaluación

Rúbricas o listas de cotejo para valorar desempeño

Formatos de reporte de prácticas

[Rúbrica Práctica de Laboratorio](#)

FUENTES DE INFORMACIÓN

1. Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques. Morgan Kaufmann. <https://www.sciencedirect.com/book/9780123814791/data-mining>
2. VanderPlas, J. (2016). Python Data Science Handbook. O'Reilly Media. <https://jakevdp.github.io/PythonDataScienceHandbook/>
3. Scikit-learn Developers. (2024). Clustering — scikit-learn documentation. <https://scikit-learn.org/stable/modules/clustering.html>
4. McKinney, W. (2022). Python for Data Analysis (3ra ed.). O'Reilly Media. <https://wesmckinney.com/book/>
5. Waskom, M. (2023). seaborn: Statistical Data Visualization. <https://seaborn.pydata.org/>



ANEXOS



UES

Universidad Estatal de Sonora
La Fuerza del Saber Estimulará mi Espíritu