

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA

COORDINACIÓN GENERAL DE FORMACIÓN BÁSICA
COORDINACIÓN GENERAL DE FORMACIÓN PROFESIONAL Y VINCULACIÓN UNIVERSITARIA
PROGRAMA DE UNIDAD DE APRENDIZAJE

I. DATOS DE IDENTIFICACIÓN

- 1. Unidad Académica:** Facultad de Ingeniería, Mexicali; Facultad de Ingeniería, Arquitectura y Diseño, Ensenada y Facultad de Ciencias Químicas e Ingeniería, Tijuana.
- 2. Programa Educativo:** Ingeniero en Computación
- 3. Plan de Estudios:** 2020-1
- 4. Nombre de la Unidad de Aprendizaje:** Ciencia de los Datos
- 5. Clave:** 36323
- 6. HC:** 01 **HL:** 02 **HT:** 02 **HPC:** 00 **HCL:** 00 **HE:** 01 **CR:** 06
- 7. Etapa de Formación a la que Pertenece:** Terminal
- 8. Carácter de la Unidad de Aprendizaje:** Optativa
- 9. Requisitos para Cursar la Unidad de Aprendizaje:** Ninguno



Equipo de diseño de PUA

Mauricio Alonso Sánchez Herrera
Félix Fernando González Navarro
Mabel Vázquez Briseño

Vo.Bo. de Subdirectores de Unidades Académicas

Alejandro Mungaray Moctezuma
Humberto Cervantes de Ávila
Rocío Alejandra Chávez Santoscoy

Fecha: 17 de octubre de 2019

II. PROPÓSITO DE LA UNIDAD DE APRENDIZAJE

La finalidad de esta asignatura es brindar las habilidades técnicas y metodológicas para el desarrollo de la ciencia de datos al integrar diferentes campos de conocimiento.

Su utilidad radica en que les permite a los alumnos la aplicación de teorías y metodologías de la ciencia de los datos en el desarrollo de soluciones de análisis y modelado de diversos fenómenos de naturaleza ingenieril.

Esta unidad de aprendizaje es de carácter optativa y se imparte en la etapa terminal. Pertenece al área de conocimiento Ciencias de la Ingeniería. Se recomienda tener conocimientos previos en las áreas de inteligencia artificial, algebra lineal, probabilidad y estadística, y cálculo.

III. COMPETENCIA DE LA UNIDAD DE APRENDIZAJE

Aplicar las teorías, metodologías y técnicas de la ciencia de los datos, a través de modelos de estimación, clasificación y predicción a partir de datos para el desarrollo de soluciones de análisis y modelado de fenómenos de naturaleza ingenieril, con responsabilidad social y objetividad.

IV. EVIDENCIA(S) DE DESEMPEÑO

Presenta un proyecto que debe contener: integración de datos, limpieza y preprocesamiento de datos, minado y validación de algún problema establecido a resolver. También debe presentarse un reporte técnico como soporte donde incluya el plan de administración, arquitectura, desarrollo, y prototipo final.

V. DESARROLLO POR UNIDADES

UNIDAD I. Ciencia de datos

Competencia:

Describir el uso de la ciencia de datos, a través de su revisión histórica, analítica y aplicación, para identificar su utilidad y limitantes, de manera analítica y reflexiva.

Contenido:**Duración:** 2 horas

- 1.1 Definición de Ciencia de Datos
- 1.2 Breve historia de ciencia de datos y conceptos relacionados
- 1.3 Hacia la era de la información
- 1.4 Información que puede extraerse a través de la Ciencia de Datos
- 1.5 Tecnologías utilizadas para llevar a cabo Ciencia de Datos
- 1.6 Áreas de aplicación de la Ciencia de Datos
- 1.7 Problemas con la Ciencia de Datos
- 1.8 Descubrimiento de conocimiento en bases de datos (KDD)

UNIDAD II. Análisis de datos

Competencia:

Describir los datos, mediante la identificación de sus atributos y visualización, para entender su comportamiento, de forma inquisitiva y exacta.

Contenido:

- 2.1 Objetos de datos y tipos de atributos
- 2.2 Descripciones estadísticas básicas de los datos

Duración: 2 horas

UNIDAD III. Preprocesamiento de datos

Competencia:

Manipular datos en preparación a su minado, mediante el uso de técnicas diversas de limpieza, integración, reducción y transformación de datos, para obtener modelos de mejor calidad durante el proceso de minado, de forma lógica y analítica.

Contenido:

Duración: 4 horas

- 3.1 Necesidad de pre-procesar los datos
- 3.2 Limpieza de datos
- 3.3 Integración de datos
- 3.4 Reducción de datos
- 3.5 Transformación de datos
- 3.6. Manejo de datos atípicos

UNIDAD IV. Minería de datos y validación

Competencia:

Aplicar minado de datos, mediante el uso de diversas técnicas estadísticas, probabilísticas, bioinspiradas, entre otras; para extraer conocimiento a partir de estos y evaluar sus resultados, con actitud planificadora y sistemática.

Contenido:

Duración: 8 horas

- 4.1 Separación de datos
- 4.2 Agrupación
 - 4.2.1 Visualización
 - 4.2.2 Validación de agrupación
- 4.3 Clasificación
 - 4.3.1 Validación de clasificación
- 4.4 Regresión
 - 4.4.1 Validación de regresión
- 4.5 Pronóstico
 - 4.5.1 Validación de pronóstico
- 4.6 Métodos estadísticos de comparación
 - 4.6.1 Métodos paramétricos
 - 4.6.2 Métodos no paramétricos

VI. ESTRUCTURA DE LAS PRÁCTICAS DE TALLER

No. de Práctica	Competencia	Descripción	Material de Apoyo	Duración
UNIDAD I				
1	Identificar los beneficios y alcances de la ciencia de datos, para concientizar al alumno de la utilidad de este, mediante una investigación en internet, de manera ordenada y sistemática.	Realiza una investigación a través de internet sobre los alcances de la ciencia de datos, y crea una mapa conceptual que ligue todas las ideas como producto entregable.	PC, internet	2 horas
2	Analizar las herramientas disponibles para llevar a cabo ciencia de datos, mediante la búsqueda, análisis, e instalación de herramientas, para instalar un ambiente que propicie la ciencia de datos, con disciplina e interés.	Se familiariza con las herramientas disponibles para realizar ciencia de datos a través de la instalación y configuración de herramientas diversas. El alumno decide cuál ambiente utilizar en la instalación.	PC, internet	2 horas
UNIDAD II				
3	Analizar los tipos de atributos existentes dentro de una base de datos, a través de la revisión detallada con múltiples bases de datos de naturalezas diferentes, para inferir el tipo de procesamiento requerido, con objetividad y eficacia.	Realiza un análisis de los atributos existentes en una base de datos, y genera un mapa conceptual que describa cada detalle de todos los tipos de atributos existentes.	PC, internet	2 horas
4	Describir los datos dentro de bases de datos diversas, mediante el uso de estadística, para entender el comportamiento de éstos, con exactitud y congruencia.	Utiliza estadística para describir la información en diferentes bases de datos, y comprende los comportamientos y sus implicaciones posibles a partir de las descripciones.	PC, internet	2 horas
UNIDAD III				

5	Aplicar técnicas de limpieza a datos, a través del uso de las diversas metodologías existentes, para reducir los errores y datos faltantes en las bases de datos, con paciencia y coherencia.	Usa diferentes técnicas para limpieza y con ello homogeniza la información eliminando ruido y datos faltantes, normalizando valores. Lleva a cabo esto en diversas bases de datos.	PC, internet	2 horas
6	Construir bases de datos a partir de múltiples fuentes de información, mediante el análisis de la compatibilidad entre fuentes, para compilar bases de datos procesables durante el minado, con organización y lógica.	Usa diferentes técnicas para integrar los contenidos de múltiples fuentes de datos. Analiza la compatibilidad entre fuentes para crear una sola base de datos compuesta de información relevante.	PC, internet	2 horas
7	Aplicar reducción de datos, mediante la utilización de diversas técnicas en reducción, para crear bases de datos más compactas y con datos significativos, con interés y observación.	Usa diferentes técnicas de reducción de datos para reducir numerosidad y cantidad de atributos en diversas bases de datos. Genera bases de datos más compactas y con datos significativos	PC, internet	2 horas
8	Aplicar transformación de datos, mediante la utilización de metodologías y herramientas en transformación, para crear bases de datos con datos significativos, con interés y observación.	Usa diferentes técnicas para transformar de datos en diferentes bases de datos y mejora la calidad de la información previo a su análisis.	PC, internet	2 horas
UNIDAD IV				
9	Experimentar con diversas técnicas de separación de datos previo a su análisis, mediante el uso de herramientas computacionales, para llevar a cabo separación de datos, para mejorar el rendimiento de	Aplica diferentes técnicas para la separación de datos para eliminar o reducir el sobreentrenamiento y sobreajuste en diferentes bases de datos.	PC, internet	2 horas

	creación de modelos durante el minado, con exactitud y cuidado.			
10	Aplicar técnicas de visualización en datos, mediante el uso herramientas computacionales, para entender los datos previos a su análisis, con creatividad y lógica.	Usa técnicas de visualización para analizar los datos y comprender sus comportamientos y posibles acciones a tomar en la etapa de modelado. Utiliza diversas bases de datos para practicar estas técnicas.	PC, internet	2 horas
11	Aplicar técnicas de análisis de grupos, mediante el uso de herramientas computacionales, para llevar a cabo análisis de grupos, para encontrar agrupaciones entre los datos, con eficacia y reflexión.	Aplica técnicas de minado para análisis de grupos y valida el resultado. Utiliza diversas bases de datos para practicar estas técnicas.	PC, internet	2 horas
12	Implementar técnicas de clasificación, regresión y pronóstico, mediante el uso de herramientas estadísticas, probabilísticas, bioinspiradas, entre otras, para generar modelos de clasificación, estimación y pronóstico, con disposición y exactitud.	Selecciona una variedad de bases de datos. Aplica técnicas de minado para clasificar patrones y valida el resultado. Genera un modelo de clasificación.	PC, internet	2 horas
13		Selecciona una variedad de bases de datos Aplica técnicas de minado en las diferentes bases de datos para estimar patrones y valida el resultado. Crea un modelo de estimación	PC, internet	2 horas
14		Selecciona un conjunto de bases de datos Aplica técnicas de minado para pronosticar en series de tiempo y valida el resultado. Genera un modelo de estimación	PC, internet	2 horas
15	Distinguir las técnicas estadísticas de comparación de modelos, mediante la	Identifica los diferentes métodos estadísticos paramétricos y no	PC, internet	4 horas

	identificación de sus características particulares, para conocer su confiabilidad, con actitud crítica y analítica.	paramétricos de comparación mediante el diseño de la herramienta para la validación de modelos.		
--	---	---	--	--

VI. ESTRUCTURA DE LAS PRÁCTICAS DE LABORATORIO

No. de Práctica	Competencia	Descripción	Material de Apoyo	Duración
UNIDAD I				
1	Identificar múltiples fuentes de datos, mediante la investigación en internet, para localizar sitios de descarga de fuentes de información de diferentes naturalezas, de manera ordenada y sistemática.	Realiza investigación en internet e identifica fuentes de datos que permitan descarga de fuentes de información. Compara las fuentes de información y las ordena de acuerdo a su naturaleza.	PC, internet, software especializado.	2 horas
2	Analizar las herramientas disponibles para llevar a cabo ciencia de datos, mediante la búsqueda, análisis, e instalación de herramientas, para instalar un ambiente que propicie la ciencia de datos, con disciplina e interés.	Instala y configura diferentes plataformas para llevar a cabo el proceso de ciencia de datos. Analiza las ventajas y desventajas de cada ambiente.	PC, internet, software especializado.	2 horas
UNIDAD II				
3	Analizar los tipos de atributos existentes dentro de una base de datos, a través de la revisión detallada con múltiples bases de datos de naturalezas diferentes, para inferir el tipo de procesamiento requerido, con objetividad y eficacia.	Analiza y describe los diferentes tipos de atributos existentes en distintas bases de datos. Compara los tipos de atributos y concluye sobre las implicaciones de las diferencias.	PC, internet, software especializado.	2 horas
4	Describir los datos dentro de bases de datos diversas, mediante el uso de estadística, para entender su comportamiento, con exactitud y congruencia.	Utiliza estadísticas para describir los comportamientos de los datos respecto sus atributos en diversas bases de datos. Analiza las descripciones obtenidas y genera estrategias de procesamiento.	PC, internet, software especializado.	2 horas

UNIDAD III				
5	Aplicar diversas técnicas de limpieza a datos, a través del uso de las metodologías existentes, para reducir los errores y datos faltantes en las bases de datos, con paciencia y coherencia.	Usa diferentes técnicas de limpieza para homogeneizar los datos y eliminar ruido en diferentes bases de datos. Reduce los errores y datos faltantes en las bases de datos.	PC, internet, software especializado.	2 horas
6	Construir bases de datos, a partir de múltiples fuentes de información, mediante el análisis de la compatibilidad entre fuentes, para compilar bases de datos procesables durante el minado, con organización y lógica.	Usa diferentes técnicas para integrar los contenidos de múltiples fuentes de datos. Analiza la compatibilidad entre estas fuentes. Crea una sola base de datos compuesta de información relevante.	PC, internet, software especializado.	2 horas
7	Aplicar reducción de datos, a través del uso de diversas metodologías de reducción de numerosidad y atributos, para crear bases de datos más compactas y con datos significativos, con interés y observación.	Usa diferentes técnicas para reducir numerosidad y cantidad de atributos en diversas bases de datos. Genera una bases de datos más compacta con datos significativos.	PC, internet, software especializado.	2 horas
8	Realizar transformación de datos, mediante la utilización de diversas técnicas en transformación, para crear bases de datos con datos significativos, con interés y observación.	Usa diferentes técnicas para transformar de datos en diferentes bases de datos y mejora la calidad de la información previo a su análisis. Crea bases de datos con datos significativos..	PC, internet, software especializado.	2 horas
UNIDAD IV				
9	Experimentar con diversas técnicas de separación de datos previo a su análisis, mediante el uso de herramientas existentes, para mejorar	Aplica diferentes técnicas para la separación de datos para eliminar o reducir el sobreentrenamiento y sobreajuste en diferentes bases	PC, internet, software especializado.	2 horas

	el rendimiento de creación de modelos durante el minado, con exactitud y cuidado.	de datos. Realiza mejoras en el rendimiento de creación de modelos durante el minado.		
10	Aplicar técnicas de visualización en datos, mediante el uso de las diversas herramientas computacionales, para comprender los datos previos a su análisis, con creatividad y lógica.	Usa técnicas de visualización para analizar los datos y comprender sus comportamientos y posibles acciones a tomar en la etapa de modelado. Utiliza diversas bases de datos para practicar estas técnicas.	PC, internet, software especializado.	2 horas
11	Implementar técnicas de análisis de grupos, mediante el uso de herramientas computacionales, para encontrar agrupaciones entre los datos, con eficacia y reflexión.	Aplica técnicas de minado para análisis de grupos y valida el resultado. Utiliza diversas bases de datos para practicar estas técnicas.	PC, internet, software especializado.	2 horas
12	Aplicar técnicas de clasificación, mediante el uso de herramientas computacionales, para generar modelos de clasificación, con disposición y exactitud.	Aplica técnicas de minado para clasificar patrones y valida el resultado. Utiliza diversas bases de datos para practicar estas técnicas.	PC, internet, software especializado.	2 horas
13	Utilizar técnicas de regresión, mediante el uso de herramientas computacionales, para generar modelos de estimación, con disposición y exactitud.	Aplica técnicas de minado para estimar patrones y valida el resultado. Utiliza diversas bases de datos para practicar estas técnicas.	PC, internet, software especializado.	2 horas
14	Aplicar diversas técnicas de pronóstico, mediante el uso de herramientas computacionales, para generar modelos de pronóstico, con disposición y exactitud.	Aplica técnicas de minado para pronosticar en series de tiempo y valida el resultado. Utiliza diversas bases de datos para practicar estas técnicas.	PC, internet, software especializado.	2 horas
15	Aplicar distintas técnicas estadísticas de comparación de modelos, mediante el uso de técnicas paramétricas y no paramétricas, para validar la	Aplica diferentes métodos estadísticos paramétricos y no paramétricos de comparación para validar modelos diferentes.	PC, internet, software especializado.	4 horas

	confiabilidad de dichos modelos, con exactitud y conciencia.			
--	--	--	--	--

VII. MÉTODO DE TRABAJO

Encuadre: El primer día de clase el docente debe establecer la forma de trabajo, criterios de evaluación, calidad de los trabajos académicos, derechos y obligaciones docente-alumno.

Estrategia de enseñanza (docente)

Presentaciones y demostraciones de técnicas, metodologías y plataformas diversas.

- Mediante la exposición por parte del maestro de forma ordenada y consistente, explica la teoría.
- En sesiones de taller se desarrollarán ejercicios por parte de los alumnos, en los que identifique y explore la aplicación de la ciencia de datos;
- siguiendo con dinámicas en grupos de trabajo para la solución de ejercicios, siendo el maestro un monitor y guía de éstos
- propicia la participación activa del estudiante
- elabora y aplica evaluaciones parciales

Estrategia de aprendizaje (alumno)

Investigar, presentar, interactuar con plataformas, redactar reportes técnicos y desarrollar actividades de talleres, así como un proyecto final que integre tanto conocimiento como habilidades adquiridas durante el curso.

- En sesiones de taller, aplicará los conceptos de ciencia de datos.
- Realiza las prácticas de laboratorio y taller.
- Aplica las herramientas computacionales en las sesiones de laboratorio, para llevar a cabo un análisis de las teorías, metodologías y técnicas de la ciencia de datos.
- Se recomienda los ejercicios de tarea en su modalidad individual y por equipos.
- Resuelve evaluaciones parciales
- Entrega el proyecto final.

VIII. CRITERIOS DE EVALUACIÓN

La evaluación será llevada a cabo de forma permanente durante el desarrollo de la unidad de aprendizaje de la siguiente manera:

Criterios de acreditación

- Para tener derecho a examen ordinario y extraordinario, el estudiante debe cumplir con los porcentajes de asistencia que establece el Estatuto Escolar vigente.
- Calificación en escala del 0 al 100, con un mínimo aprobatorio de 60.

Criterios de evaluación

- Evaluaciones parciales.....	20%
- Exposiciones	10%
- Prácticas de taller y laboratorio.....	30%
- Evidencia de desempeño.....	40%
(Proyecto final y reporte técnico)	
Total.....	100%

IX. REFERENCIAS

Básicas

Gutiérrez, J. A. G. (2018). *Fundamentos de Minería de Datos en Ciencias e Ingeniería: a través de ejemplos*. España: Editorial Académica Española.

Ian, H., Eibe, F., Hall M. y Pal C. (2016). *Data Mining: Practical Machine Learning Tools and Techniques* (4^a ed.). Estados Unidos: Morgan Kaufmann.

Morgan, P. (2018). *Data Science from Scratch with Python: Step-by-Step Guide* (2^a ed.). Estados Unidos: AI SCIENCES

Tan, P.N., Steinbach, M., Karpatne, A., & Kumar, V. (2018). *Introduction to Data Mining* (2^a ed.). Estados Unidos: Pearson

Complementarias

Aurelien, G. (2017). *Hands-On Machine Learning with Scikit-Learn and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems*. Estados Unidos: O'Reilly Media.

Healy, K. (2018). *Data Visualization: A Practical Introduction*. Estados Unidos: Princeton University Press.

McKinney, W. (2015). *Data Science from Scratch: First Principles with Python*. Estados Unidos: O'Reilly Media.

X. PERFIL DEL DOCENTE

El profesor debe poseer título de licenciatura en Ingeniería en Computación, o afín. Contar con posgrado y preferentemente tener grado de doctor en ciencias de la computación o afín, así como experiencia comprobable de trabajo con ciencia de datos de 2 años y 2 años de experiencia docente. Además, deberá ser una persona proactiva, responsable y comprometida con el aprendizaje significativo de los alumnos.