

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:** Minería de Datos
- 5. Clave:** 36325
- 6. HC: 01 HL: 03 HT: 00 HPC: 00 HCL: 00 HE: 01 CR: 05**
- 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

Félix Fernando González Navarro
Adolfo Heriberto Ruelas Puente

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

El curso de Minería de Datos tiene como propósito integrar conocimientos de programación de computadoras, teoría de algoritmos, matemáticas y estadística en el análisis inteligente de datos. El análisis inteligente de datos consiste en aplicar algoritmos matemáticos y técnicas estadísticas en la búsqueda de patrones de información en grandes bases de datos. Estos patrones de información, en función del contexto en que se apliquen, son de gran importancia en aspectos de toma de decisiones, modelación de procesos, diseño de sistemas eficientes, entre otros.

Durante el curso, se enseña a los estudiantes diversos algoritmos y técnicas y sus fundamentos matemáticos. Además, se refuerzan los conocimientos teóricos con la implementación mediante lenguajes de programación de propósito general y especializados. Esta asignatura se ubica en la etapa terminal optativa y pertenece al área de Ciencias de la Ingeniería.

III. COMPETENCIA DE LA UNIDAD DE APRENDIZAJE

Implementar algoritmos computacionales, para el análisis de datos inteligentes, empleando lenguajes de propósito general y programas de modelación matemática, con una actitud crítica ante los distintos modelos matemáticos y su aplicabilidad a escenarios de la vida real.

IV. EVIDENCIA(S) DE DESEMPEÑO

Implementa diversos algoritmos en computadora, realiza reportes escritos de análisis de resultados y llevar a cabo un estudio de minería de datos en un caso real.

V. DESARROLLO POR UNIDADES

UNIDAD I. Descubrimiento de conocimiento y minería de datos

Competencia:

Interpretar el concepto de minería de datos y el proceso de descubrimiento de conocimiento, en el marco del análisis de datos e información, para valorar los distintos conceptos y modelos de minería de datos, con actitud crítica, metódica y analítica.

Contenidos:**Duración:** 4 horas

- 1.1 Definición de Minería de Datos (MD)
- 1.2 Definición del proceso de descubrimiento del conocimiento (KDP)
- 1.3 Modelos del KDP y características particulares
- 1.4 Algoritmos de Aprendizaje, de Clasificación y de Regresión
- 1.5 El proceso de KDP en Big Data

UNIDAD II. Preprocesamiento de datos

Competencia:

Diferenciar distintas técnicas de pre-procesado de datos, como etapa previa a la implementación de algoritmos de computación inteligente en el planteamiento de modelos, para solucionar problemas de Ingeniería y Ciencias, con actitud creativa y de manera eficiente.

Contenidos:

Duración: 4 horas

- 2.1 Definición de atributos, datos y almacenamiento
- 2.2 Normalización y Estandarización de datos
- 2.3 Manejo de Outliers
- 2.4 Datos faltantes y ruido
- 2.5 Discretización de datos
- 2.6 Extracción y selección de atributos

UNIDAD III. Construcción de modelos para la Minería de Datos

Competencia:

Aplicar distintos modelos de minería de datos en el análisis inteligente de datos, utilizando lenguajes de programación, para la solución de problemas de ciencias e ingeniería, con una actitud innovadora y creativa.

Contenidos:

Duración: 4 horas

- 3.1 Modelos Supervizados
 - 3.1.1 Funciones de distancia y matrices de proximidad
 - 3.1.2 K-means
 - 3.1.3 Clustering Jerárquico
- 3.2 Modelos No Supervizados
 - 3.2.1 Métodos Bayesianos
 - 3.2.2 Algoritmos Basados en Instancias
 - 3.2.3 Redes Neuronales y Deep Learning
- 3.3 Manejo de Modelos de Visualización
 - 3.3.1 Análisis de Componentes Principales
 - 3.3.2 Discriminante Lineal de Fisher
 - 3.3.3 Modelos Estadísticos de visualización

UNIDAD IV. Selección y validación de modelos

Competencia:

Evaluar los diferentes métodos de selección de modelos, para implementar soluciones a problemas de ingeniería y ciencias de la vida real, utilizando diversos métodos estadísticos, con actitud analítica, creativa y proactiva.

Contenidos:

- 4.1 Conceptos principales y aplicación en Modelos de Clasificación y Regresión.
- 4.2 Validación Cruzada y Matrices de Confusión
- 4.3 Métodos Estadísticos para Comparación de Algoritmos

Duración: 4 horas

VI. ESTRUCTURA DE LAS PRÁCTICAS DE LABORATORIO

No. de Práctica	Competencia	Descripción	Material de Apoyo	Duración
1	Aplicar distintos Modelos de Minería de datos para realizar análisis inteligente de información en la solución de problemas de ciencias e ingeniería, mediante la identificación de aspectos particulares de dichos problemas y su adecuación a los modelos, de manera ordenada y metódica.	<p>El docente presenta los aspectos básicos de los distintos modelos de minería de datos, haciendo énfasis en la aplicación a problemas de la vida real.</p> <p>El estudiante:</p> <ol style="list-style-type: none"> 1. Investiga en internet un caso práctico de aplicación de modelos de minería de datos 2. Realiza y entrega un reporte ejecutivo con los conceptos aprendidos y la aplicación práctica a un caso de la vida real. 3. Expone frente a grupo el caso aplicación encontrado en internet 	Pintarrón, plumones, cañón de proyección.	8 horas
2	Aplicar métodos de preprocesamiento de datos, como etapa previa a la implementación de modelos de minería de datos, mediante la implementación de algoritmos en computadora, de forma eficiente y metódica,	<p>El docente presenta las operaciones de preprocesado de datos</p> <p>El estudiante:</p> <ol style="list-style-type: none"> 1. Implementa códigos de computadora sobre las principales operaciones de preprocesado de datos: Normalización y Estandarización de datos, Manejo de Outliers, Datos faltantes y ruido, Discretización de datos, Extracción y selección de atributos; su fundamento matemático y pseudocódigo. 2. Entrega un documento que consiste en explicar los códigos de las operaciones de preprocesado. 	Pintarrón, plumones, cañón de proyección computadora.	12 horas

3	Identificar los principales modelos de minería de datos, para su implementación en entornos requeridos, a través del uso modelos matemáticos y algoritmos, con actitud metódica, pensamiento analítico y orden.	El docente presenta los principales modelos de Minería de Datos y de la Inteligencia Artificial: El estudiante: 1. Implementa y realiza un análisis comparativo de cada uno de los modelos supervisados, modelos no supervisados y de visualización. 2. Elabora un reporte escrito de los modelos implementados que incluya el código en computadora.	Pintarrón, plumones, cañón de proyección computadora.	18 horas
4	Aplicar distintos procedimientos para la selección y validación de modelos, mediante la aplicación de distintos métodos estadísticos con actitud metodológica y analítica.	El docente presenta los conceptos estadísticos para la selección y validación de modelos de minería de datos. El estudiante: 1. Implementa cada uno de los modelos de selección y validación. 2. Elabora un reporte escrito de la implementación de los modelos que incluya el código en computadora.	Pintarrón, plumones, cañón de proyección computadora.	10 horas

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)

- El docente funge como orientador en comprensión de conceptos matemáticos, la implementación de éstos en algoritmos de computación y la interpretación de resultados
- Presenta estudios de caso
- Realiza ejercicios y resuelve con el grupo
- Supervisa prácticas
- Propicia la participación activa del estudiante

Estrategia de aprendizaje (estudiante)

- Participa de forma activa y autónoma en la construcción de sus conocimientos
- Integra las competencias del plan de estudios
- Trabaja de forma crítica y analítica en la revisión el análisis de los fundamentos matemáticos y las características de los algoritmos computacionales para la implementación
- Realiza investigaciones documentales sobre la minería de datos
- Resuelve ejercicios sobre estudios de caso planteados por el docente

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:

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 (2).....40%
- Evidencia de desempeño.....60%

(Implementa diversos algoritmos en computadora, realizar reportes escritos de análisis de resultados y llevar a cabo un estudio de minería de datos en un caso real.)

Total.....100%

IX. REFERENCIAS

Básicas	Complementarias
Doug, A. (2019). <i>Data mining</i> . Recuperado de http://www.laits.utexas.edu/~anorman/BUS.FOR/course.mat/Alex/ .	Balakrishnan, N., Koutras, M. y Politis, K. (2019). <i>Introduction to Probability: Models and Applications</i> . (5th ed.). Estados Unidos: Wiley
Hastie, T. y Tibshirani R. (2016). <i>The Elements of Statistical Learning: Data Mining, Inference, and Prediction</i> (2ª ed.). Estados Unidos: Springer Series. [clásica]	D. Smith, P. (2018). <i>Hands-On Artificial Intelligence for Beginners: An introduction to AI concepts, algorithms, and their implementation</i> . (7th ed.). Estados Unidos: Packt Publishing.
Ian, H. y Eibe, F., Hall, M. y Pal, C. (2016). <i>Data Mining: Practical Machine Learning Tools and Techniques</i> (4ª ed.). Estados Unidos: Morgan Kaufmann.	Edwards, C., Penney, D. y Calvis, D. (2017). <i>Differential Equations and Linear Algebra</i> (4a ed.). Estados Unidos: Pearson.
Kelleher, J. y Tierney, B. (2018). <i>Data Science</i> . Estados Unidos: MIT Press.	Lay, D., Lay, S., y Mcdonald, J. (2015). <i>Linear Algebra and its Applications</i> (5ª ed.). Estados Unidos: Pearson.
Lanantz, B. (2019). <i>Machine Learning with R: Expert techniques for predictive modeling</i> . (3a ed.). Estados Unidos: Packt Publishing.	McClave, B.S. (2017). <i>Statistics for Business and Economics</i> (13ª ed.). Estados Unidos: Pearson.
Mitchel, T. (2019). <i>Machine Learning</i> . Estados Unidos. Recuperado de http://www.cs.cmu.edu/~tom/mlbook-chapter-slides.html	Stewart, J. (2015). <i>Calculus</i> . (7 ^{ma} ed.). Estados Unidos: Brooks Cole.
Tan, P., Steinbach, M., Karpatne, A., y Kumar, V. (2014). <i>Introduction to Data Mining</i> . Estados Unidos: Addison-Wesley. [clásica]	
Witten, I. H., Frank, E., Hall, M.A., y Pal, C.J. (2016). <i>Data Mining: Practical Machine Learning Tools and Techniques</i> (4ª ed.). Estados Unidos: Morgan Kaufmann. [clásica]	

X. PERFIL DEL DOCENTE

El docente que imparta esta asignatura debe contar con título de Licenciatura en Ingeniería en Computación o área afín, se sugiere contar con maestría y/o doctorado en Ciencias o en Ingeniería, con especialización en el área de computación, minería de datos e inteligencia artificial. Asimismo, deberá contar con experiencia docente en educación superior y ser una persona responsable con disposición para promover el aprendizaje significativo en los estudiantes.