



Formato para registro de Unidades de aprendizaje 2021

I.- Datos de identificación de la unidad de aprendizaje

| | | | | | | | | | | |
|---|---|--------------|-----------------|-------------------------------------|------------------------------|------------------------|---------------------------------|--|--------------------|--|
| Unidad académica: | Multisede (CIC, CIDETEC, ESCOM, ESFM, UPIITA) | | | | | | | | | |
| Programa académico: | Doctorado en Ciencia y Tecnología de Inteligencia Artificial y Ciencia de Datos | | | | | | | | | |
| | X | Doctorado | | | Orientación profesional | | | | | |
| | | Maestría | X | | Orientado a la investigación | | | | | |
| | | Especialidad | | | Con la industria | | | | | |
| | | | | | Especialidad médica | | | | | |
| Nombre de unidad de aprendizaje: | Sesión de colegio donde se propuso: | | | | | | Fecha de propuesta: | | | |
| Tipo de unidad de aprendizaje: | Datos masivos y minería de datos | | | | | | | | | |
| | Clave de la unidad de aprendizaje: | | | | | | Créditos: | | 5 <i>REP 2017</i> | |
| | Semanas del semestre | | 18 | Horas a la semana: | | 4 | Horas totales: | | 72 | |
| | Obligatoria: | | Optativa: | | X | Observaciones: | | | | |
| | Semestre: | 1 - 4 | | | | | | | | |
| | Teórica (%): | 50 | Práctica (%): | | 50 | Teórico-prácticas (%): | | | | |
| Área del conocimiento: | Ingeniería y Ciencias Fisicomatemáticas | | X | Ciencias Sociales y Administrativas | | | Ciencias Médico Biológicas | | Interdisciplinario | |
| Modalidad no escolarizada: | No escolarizada | | | Nombre de la Plataforma: | | | | | | |
| | Mixta | | Presencial (%): | | | | En plataforma (%): | | | |
| Horas establecidas en el programa de estudios: | Presenciales (si procede) (horas x semana) | | | | | | En plataforma (horas x semana): | | | |



Formato para registro de Unidades de aprendizaje 2021

I. Aprendizajes que el estudiante deberá demostrar al finalizar

| Conocimientos | Habilidades y destrezas | Actitudes y valores |
|---|---|---|
| <ul style="list-style-type: none"> • Diferenciar preguntas de negocios que para su respuesta requieren generar modelos de Minería de Datos a simples consultas a bases de datos. • Saber usar organizaciones o modelos de datos orientados al análisis de datos en grandes almacenes o bodegas de datos (Data Warehouse). • Conocer los algoritmos para generar diferentes modelos de Minería de Datos para responder preguntas de negocios. • Saber probar los modelos de minería con diferentes arquitecturas de datos de entrenamiento, de prueba y validación con métodos formales. | <ul style="list-style-type: none"> • Identificar cuando es una pregunta compleja de negocios o minería de datos o Inteligencia de Negocios de una simple consulta. • Identificar el modelo de datos a usar para contestar preguntas de negocios. • Saber usar con solvencia las técnicas de minería de datos. • Describir con facilidad las características importantes de los conjuntos de datos a analizar. • identificar el conocimiento que es relevante o interesante para el conjunto de datos con ayuda del dueño de los datos. | <ul style="list-style-type: none"> • Innovación y creatividad para aprovechar la información que se genera día con día a través de diferentes medios. • Respetar la información confidencial y el conocimiento derivado de los conjuntos de datos de los dueños de ella, sin importar si son, personas, empresas y competidores. • Utilizar únicamente información disponible y legal. |

Resolución que aborda la propuesta con su enfoque disciplinar

La informatización de nuestra sociedad ha mejorado sustancialmente nuestras capacidades tanto para generar como para recopilar datos de diversas fuentes. Una enorme cantidad de datos ha inundado casi todos los aspectos de nuestras vidas. Este crecimiento explosivo de datos almacenados o transitorios ha generado una necesidad urgente de nuevas técnicas y herramientas automatizadas que puedan ayudarnos de manera inteligente a transformar grandes cantidades de datos en información y conocimiento útiles. Esto ha llevado a la generación de una frontera prometedora y floreciente en la informática llamada minería de datos y sus diversas aplicaciones.

La minería de datos, también conocida popularmente como descubrimiento de conocimiento a partir de datos (KDD), es la extracción automatizada o conveniente de patrones que representan conocimiento implícitamente almacenado o capturado en grandes bases de datos, almacenes de datos, la Web, otros repositorios masivos de información o flujos de datos. Referencias

Jiawei Han and Micheline Kamber , “Data Mining: Concepts and Techniques” , Third edition The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor , [Morgan Kaufmann Publishers](http://www.mkp.com), March 2011. ISBN 1-55860-901-6



Formato para registro de Unidades de aprendizaje 2021

II. Proximidad formativa

Áreas multi, inter y transdisciplinarias

Líneas de Generación y Aplicación de Conocimiento

Sectores sociales

| | | |
|---|--|---|
| <ul style="list-style-type: none"> • Lógica matemática • Probabilidad y estadísticas • Matemáticas para análisis • Modelado e inferencia • Informática • Algoritmos | <ul style="list-style-type: none"> • Aprendizaje automático • Redes neuronales y aprendizaje profundo • Reconocimiento de patrones • Generación de soluciones y decisiones inteligentes: sistemas híbridos, búsqueda y optimización • Minería de datos, descubrimiento de conocimiento y analítica avanzada • Visión computacional • Robótica inteligente • Sistemas difusos y razonamiento bajo incertidumbre • Minería de texto y procesamiento de lenguaje natural | <ul style="list-style-type: none"> • Los sectores que se establecen en las LGAC, agua, cultura, educación, energía y cambio climático, salud, seguridad, soberanía alimentaria y vivienda, |
|---|--|---|

Estrategia de asociación:

Trabajar con datos del libro de texto, con datos de bases de datos de ejemplos que acompañan a los sistemas administradores de datos y con datos abiertos que se publican en las páginas de organizaciones del gobierno, empresas privadas y organizaciones no gubernamentales (ONGs), las cuales integran y sensibilizan al estudiante de maestría Inteligencia Artificial y Ciencia de Datos en los problemas relacionados tanto a la sociedad como de empresas, y así aplicar los conocimientos impartidas en la U. de A.

III Metodología de enseñanza – aprendizaje

Descripción



Formato para registro de Unidades de aprendizaje 2021

Se dará un seguimiento personalizado y de trabajo colaborativo mediante las diversas actividades que se realizan durante la unidad de aprendizaje con el fin de que el estudiante:

- El trabajo se desarrolla en las 18 semanas, en las 72 horas de estudio a distancia y/o presencial
- Las sesiones están previstas para el seguimiento del trabajo colaborativo de los estudiantes, así como para ofrecerles apoyo en todas las dudas que tengan.
- En el aspecto teórico, se desarrollarán parte de los ejercicios de los primeros 10 capítulos del libro de texto [1]
- En el aspecto práctico, se revisará la aplicación de las técnicas a los datos que permitan abordar los problemas que estén relacionados con los sectores sociales ya enumerados en el punto anterior.

Las evaluaciones se realizarán con base en la rúbrica correspondiente.

La calificación mínima aprobatoria para la Unidad de aprendizaje será ocho (8). Siendo la escala de notas y valores numéricos la siguiente:

Menor a 8: No aprobado

8: Bueno

9: Muy bueno

10: Excelente

Evidencias como proceso de aprendizaje

- Lista de asistencia en las 18 semanas
 - Lista de dudas registradas en las asesorías solicitadas
 - Entrega y discusión de los ejercicios de los primeros 10 capítulos del libro de texto [1]
- Entregas revisadas de las presentaciones y documentos relacionadas a las actividades descritas en el punto V.
Secuencia programática y que cubren al contenido temático

Evidencias integradoras (resultados que contribuyen al curriculum) Ponderación

| | |
|---|--|
| Está centrado en el producto del portafolio de evidencias de aprendizaje (bitácora) | |
|---|--|



Formato para registro de Unidades de aprendizaje 2021

IV. Descripción de la participación esperada en el estudiante

| Receptiva | Resolutiva | Autónoma | Estratégica |
|-----------|------------|----------|---|
| | | | Los proyectos están abiertos a lo que proponga el alumno. |

Contenido temático

1. INTRODUCCIÓN. (2 horas)
 - a. ¿Qué motiva la minería? ¿Por qué es importante?
 - b. ¿Qué es la minería de datos?
 - c. Minería de datos—¿Sobre qué tipos de datos?
 - d. Funciones de minería de datos ¿Qué tipo de patrones pueden ser minados?



Formato para registro de Unidades de aprendizaje 2021

- e. ¿Qué tecnologías se utilizan?
- f. ¿En dónde se aplica?
- g. Los grandes desafíos de minería de datos

2. Conociendo los datos (6 horas)

- a. Objetos, datos y tipos de datos
- b. Estadística básica y descripción de los datos
- c. Visualización de datos
- d. Medidas de similitud y no similitud de datos

3. Preprocesamiento de datos (8 horas)

- a. Resumen del preprocesamiento
- b. Limpieza de datos
- c. Integración de datos
- d. Reducción de Datos
- e. Transformación de datos y discretización de datos

4. Bodega de Datos y OLAP (8 horas)

- a. Conceptos básicos de una bodega de datos
- b. Modelado de bodega de datos: Cubo de datos y OLAP
- c. Diseño de una Bodega de Datos y su utilización
- d. Implementación de bodegas de datos
- e. Generalización de datos por inducción orientado a atributos

5. Tecnología de cubo de datos (8 horas)

- a. Conceptos preliminares de cálculo de los cubos de datos
- b. Métodos de cálculo de los cubos de datos
- c. Tipos avanzados de procesamiento de consultas y tecnología de exploración de cubos
- d. Análisis multidimensional de datos en espacios de cubos

6. Minado avanzado de patrones frecuentes, asociaciones, y correlaciones: Conceptos básicos (8 horas)



Formato para registro de Unidades de aprendizaje 2021

- a. Conceptos básicos y una revisión rápida
 - b. Métodos de minado de conjuntos frecuentes
 - c. ¿Cuáles patrones son interesantes? Métodos de evaluación de patrones
7. Minado avanzado de patrones (8 horas)
- a. Minado de patrones; Una visión rápida
 - b. Minado patrones en espacios de multinivel y multi-dimensionales
 - c. Minado de patrones frecuentes en base a restricciones
 - d. Minado de datos de alta dimensionalidad y patrones colosales
 - e. Minado de patrones compactados o aproximados
 - f. Exploración de patrones y su aplicación
8. Clasificación: Conceptos básicos (8 horas)
- a. Conceptos básicos
 - b. Clasificación por inducción de un árbol de decisión
 - c. Métodos de clasificación bayesiana
 - d. Clasificación basada en reglas
 - e. Modelo de evaluación y selección
 - f. Técnicas para mejorar la exactitud de la clasificación
9. Métodos avanzados de clasificación (8 horas)
- a. Redes Bayesianas
 - b. Clasificación por Retropropagación
 - c. Máquinas de soporte vectorial
 - d. Clasificación usando patrones frecuentes
 - e. Aprendizaje Lazy (o el aprendizaje a partir de sus vecinos)
 - f. Otros métodos de clasificación
 - g. Tópicos adicionales
10. Análisis de conglomerados: Conceptos básicos y métodos (8 horas)
- a. Análisis de conglomerados
 - b. Métodos de particionamiento



Formato para registro de Unidades de aprendizaje 2021

- c. Métodos jerárquicos
- d. Métodos basados en densidad
- e. Métodos basados en Grid
- f. Evaluación de conglomerados

V. Secuencia programática

| No. | Tema | Objetivo de aprendizaje / competencia específica | Tiempo/Horas/Semanas | |
|----------------|------|--|--------------------------|--|
| 1 | | | | |
| Actividad(es): | | | Tipo de interacción(es): | |
| Evidencia(s): | | | Referencias (s): | |

Tipo de interacción: ID–Instrucción directa, TC–Trabajo colaborativo, AC–Análisis en campo, RP–Reflexión personal, PE–Presentación expositiva

Nota: *Replique esta sección las veces que sea necesario para cubrir toda la secuencia programática*

Indicar solo el número de las *Referencias* indizadas en la sección VII de este documento.

VI. Habilitadores tecnológicos

| Disposiciones | Especificaciones / descripción de efectos |
|-----------------------|---|
| Conectividad | Acceder a diferentes fuentes de datos con diferentes formatos |
| Habilidades digitales | Internet para complementar la información del libro de texto, acceder a documentación de herramientas |
| Interoperabilidad | Tema de la unidad de aprendizaje, utilizado en la compartición de datos abiertos |
| Datos abiertos | Utilizados, para plantear diferentes análisis a las bases de datos abiertas seleccionadas |



Formato para registro de Unidades de aprendizaje 2021

| | |
|-------------------------|--|
| <i>Big Data</i> | Se usan los datos, pero no se generan más datos |
| <i>Machine Learning</i> | Descubrir patrones interesantes en datos no estudiados |
| Simulación | |
| Realidad aumentada | |
| Otro... | |

Conferencias magistrales

| |
|----|
| 1. |
| 2. |
| 3. |

Notas complementarias

| |
|--|
| |
| |
| |

VII. Referencias

Documentales / electrónicas

| |
|---|
| 4. Jiawei Han and Micheline Kamber , “Data Mining: Concepts and Techniques” , Third edition The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor , Morgan Kaufmann Publishers, March 2011. ISBN 1-55860-901-6 |
| 5. Jure Leskovec. Anand Rajaraman, Jeffrey D. Ullman, “The Mining of Massive Datasets book”, 3th Edition, 2020 |
| 6. Fayyad, U.M., Piatetsky-Shapiro, G. “Advanced in Knowledge Discovery and Data Mining”, AAAI Press, 1996 |
| 7. Chen, Z. “Intelligent Data Warehousing”, CRC Press, Boca Raton, FL, 2001 |
| 8. Witten, Ian H., Frank, Kaufmann., Kaufmann, Morgan. “Data Mining: Practical Machine Learning Tools and Techniques“. (Second Edition), June 2005. |

VIII. Créditos y responsabilidades

| Responsabilidad | Nombre completo | Clave de nombramiento /No. de empleado |
|---------------------|--------------------------------|--|
| Coordinador (Autor) | Gilberto Lorenzo Martínez Luna | 16300-EH-22/9601000 |



Formato para registro de Unidades de aprendizaje 2021

| | | |
|---|---|--------------------------------------|
| Participante (Coautor) | Adolfo Guzmán Arenas Elsa Rubio Espino | 13702-EG-18/6/9500093 17133-EI-23 |
| Asesor didáctico / Diseñador Instruccional | | |
| Tecnólogo educativo / Comunicólogo | | |
| Corrector de estilo | | |
| Programador multimedia / Diseñador gráfico | | |
| Otro... | | |

VERIFICACIÓN GENERAL DE LA PLANEACIÓN DIDÁCTICA

Por la División de Operación y Promoción al Posgrado de la SIP

Nombre _____

FIRMA _____

REVISIÓN DE LA PLANEACIÓN DIDÁCTICA (VIABILIDAD)

Por la Subdirección de Diseño y Desarrollo de la DEV

Nombre _____

FIRMA _____

VERIFICACIÓN PARA SU PUESTA EN OPERACIÓN

Por la Dirección de Posgrado

Nombre _____

REVISIÓN TÉCNICO-PEDAGÓGICA PARA LA MODALIDAD

Por la Dirección para la Educación Virtual

Nombre _____



Instituto Politécnico Nacional

Secretaría Académica
Dirección de Educación Virtual

Secretaría de Investigación y Posgrado
Dirección de Posgrado

SIP-30

Formato para registro de Unidades de aprendizaje 2021

| | |
|--|--------------------|
| <p>FIRMA _____</p> <p>SELLO DE VALIDACIÓN</p> | <p>FIRMA _____</p> |
|--|--------------------|