

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA

COORDINACIÓN GENERAL DE FORMACIÓN PROFESIONAL

PROGRAMA DE UNIDAD DE APRENDIZAJE

I. DATOS DE IDENTIFICACIÓN

- 1. Unidad Académica:** Facultad de Ciencias Administrativas y Sociales, Ensenada; Facultad de Ciencias Administrativas, Mexicali; Facultad de Ciencias de la Ingeniería, Administrativas y Sociales, Tecate; y Facultad de Contaduría y Administración, Tijuana.
- 2. Programa Educativo:** Licenciado en Inteligencia de Negocios
- 3. Plan de Estudios:** 2021-2
- 4. Nombre de la Unidad de Aprendizaje:** Big Data
- 5. Clave:** 39056
- 6. HC:** 01 **HT:** 00 **HL:** 04 **HPC:** 00 **HCL:** 00 **HE:** 01 **CR:** 06
- 7. Etapa de Formación a la que Pertenece:** Disciplinaria
- 8. Carácter de la Unidad de Aprendizaje:** Obligatoria
- 9. Requisitos para Cursar la Unidad de Aprendizaje:** Ninguno



Equipo de diseño de PUA

Roberto Carlos Valdés Hernández
Myriam Gabriela Aguilera Zertuche
Héctor Zatarain Aceves
Eduardo Ahumada Tello

Vo.Bo. de subdirectores de las Unidades Académicas

Adelaida Figueroa Villanueva
Angélica Reyes Mendoza
Esperanza Manrique Rojas
Jesús Antonio Padilla Sánchez

Fecha: 10 de marzo de 2021

II. PROPÓSITO DE LA UNIDAD DE APRENDIZAJE

La unidad de aprendizaje Big data tiene el propósito de dotar al estudiante con los conocimientos y habilidades para el procesamiento y el análisis de datos masivos, haciendo uso de las herramientas tecnológicas, para la solución de problemas, que apoye en la toma de decisiones de las organizaciones.

Esta asignatura se imparte en la etapa disciplinaria con carácter obligatorio y pertenece al área de conocimiento Ciencia de Datos

III. COMPETENCIA GENERAL DE LA UNIDAD DE APRENDIZAJE

Gestionar datos masivos mediante la aplicación de herramientas y estrategias de Big Data, para agilizar y optimizar el proceso de toma de decisiones en organizaciones, con honestidad, confidencialidad y pensamiento crítico.

IV. EVIDENCIA(S) DE APRENDIZAJE

Portafolio de prácticas sobre la aplicación de las herramientas para el análisis de los datos masivos, que ofrece el Big Data.

V. DESARROLLO POR UNIDADES
UNIDAD I. Fundamentos de Big Data

Competencia:

Identificar los fundamentos del Big Data, mediante la revisión de sus modelos y estructura, para comprender su aplicación en el manejo de datos de las organizaciones, con pensamiento analítico y reflexivo

Contenido:

Duración: 4 horas

- 1.1. Conceptos de Big Data
- 1.2. Características del Big Data
- 1.3. Casos de estudio de Big Data
- 1.2. Introducción al reconocimiento de patrones
- 1.3. Modelos de regresión lineal y regresión logística
- 1.4. Exploración de la estructura de datos
- 1.5. Introducción a Jupyter Notebook y Pyspark

UNIDAD II. Data warehouse.

Competencia:

Analizar datos masivos mediante la aplicación del Data warehouse, con la finalidad de generar información de utilidad para la toma de decisiones estratégica, con pensamiento analítico, reflexivo y honestidad.

Contenido:

Duración: 2 horas

- 2.1. Concepto de Data warehouse
- 2.2. Características de un Data warehouse
 - 2.2.1. Data Marts
- 2.3. Diseño de Data warehouse
- 2.4. Data warehouse basado en la nube
 - 2.4.1. Solución analítica integrada
- 2.5. Premisas del Data warehouse
- 2.6. Inteligencia de negocios (Business Intelligence)
- 2.7. Analítica de datos en tiempo real (Tracking de datos)

UNIDAD III. Herramientas y Arquitecturas de Datawarehouse

Competencia:

Aplicar las herramientas y arquitectura del Data warehouse, para el análisis de datos masivos, que apoye la toma de decisiones estratégica, con pensamiento analítico, reflexivo y honestidad.

Contenido:

Duración: 4 horas

- 3.1. Redes sociales (facebook graph)
- 3.2. Google trends
- 3.3. Operational Data Store
- 3.4. Sistemas de procesamiento analítico en línea
 - 3.4.1 OLAP (Online Analytical Processing)
 - 3.4.2 OLTP (Online Transaction Processing)
 - 3.4.3 ROLAP (Relational Online Analytical Processing)
 - 3.4.4 HOLAP (Hybrid Online Analytical Processing)
- 3.5. Selección de plataformas analíticas.
- 3.6. HDFS (Hadoop Distributed File System) Architecture
 - 3.6.1. HDFS Usage Examples

UNIDAD IV. Minería de datos

Competencia:

Aplicar la Minería de datos, mediante el uso de herramientas y algoritmos, para generar información objetiva, que apoye la toma de decisiones estratégica, con pensamiento analítico, reflexivo y honestidad.

Contenido:

Duración: 4 horas

- 4.1. Introducción a minería de datos
- 4.2. Herramientas de minería de datos
- 4.3. Exploración de datos
- 4.4. Preprocesamiento de datos
- 4.5. Clasificación de datos
- 4.6. Evaluación de datos
- 4.6. Técnicas de minería de datos
 - 4.6.1. Naive Bayes
 - 4.6.2. K-medias
 - 4.6.3. Redes Neuronales
 - 4.6.4. Modelización con redes neuronales
 - 4.6.5. Clasificación automática
 - 4.6.6. Reducción dimensión
- 4.7. Análisis de componentes principales
- 4.8. Análisis exploratorio

UNIDAD V. Visualización de datos

Competencia:

Representar el comportamiento de los datos masivos, mediante la comparación de diferentes gráficas y diagramas, que permita su visualización e interpretación para el apoyo de la toma de decisiones estratégica, con pensamiento analítico, reflexivo y honestidad.

Contenido:

Duración: 2 horas

- 5.1. Gráfica de Barras
- 5.2. Gráfica de pastel
- 5.3. Diagrama de Caja
- 5.4. Diagrama de dispersión
- 5.5. Gráfica de Mapa
- 5.6. Diagrama de Grafo
- 5.7. Mapa de calor con agrupamiento
- 5.8. Diagramas de cuadrícula
 - 5.8.1. Cuadrícula de Par (Pair Grid)
 - 5.8.2. Cuadrícula de Facetas (Facet Grid)

VI. ESTRUCTURA DE LAS PRÁCTICAS DE LABORATORIO

No.	Nombre de la Práctica	Procedimiento	Recursos de Apoyo	Duración
UNIDAD I				
1	Exploración de Jupyter Notebook y Pyspark	<ol style="list-style-type: none"> 1. Atiende las indicaciones del docente para realizar la práctica del correcto manejo de las herramientas de Jupyter Notebook y Pyspark. 2. Navega por las diferentes herramientas de Jupyter Notebook y Pyspark. 3. Elabora un informe del uso de las herramientas utilizadas. 4. Entrega al docente. 5. Integra al portafolio de evidencias 	<ul style="list-style-type: none"> • Computadora • Internet. • Herramientas digitales • Referencias 	4 horas
UNIDAD II				
2	Diseño de Data warehouse	<ol style="list-style-type: none"> 1. Atiende las indicaciones del docente para realizar la práctica del diseño del Data warehouse. 2. Identifica las fuentes de datos. 3. Elabora el modelo de Data warehouse. 4. Selecciona herramienta ETL adecuada. 5. Elabora reporte de práctica 6. Entrega al docente para su revisión y retroalimentación. 7. Corrige e Integra al portafolio de evidencias. 	<ul style="list-style-type: none"> • Computadora • Internet. • Herramientas digitales • Referencias 	6 horas
3	Analítica de datos en tiempo real	<ol style="list-style-type: none"> 1. Atiende las indicaciones del docente para realizar la práctica analítica de datos en tiempo real. 	<ul style="list-style-type: none"> • Computadora • Internet. • Herramientas digitales • Referencias 	

		<ol style="list-style-type: none"> 2. Identifica las fuentes de datos. 3. Extrae los datos. 4. Analiza el comportamiento. 5. Elabora reporte de práctica 6. Entrega al docente para su revisión y retroalimentación. 7. Corrige e Integra al portafolio de evidencias. 		
UNIDAD III				
4	Analítica de datos en redes sociales	<ol style="list-style-type: none"> 1. Atiende las indicaciones del docente para realizar la práctica Analítica de datos en redes sociales. 2. Identifica los tipos de herramientas para analizar los datos en redes sociales. 3. Identifica las fuentes de datos. 4. Extrae los datos. 5. Procesa y analiza patrones de comportamiento. 6. Elabora reporte de práctica 7. Entrega al docente para su revisión y retroalimentación. 8. Corrige e Integra al portafolio de evidencias. 	<ul style="list-style-type: none"> • Computadora • Internet. • Herramientas digitales • Referencias 	6 horas
5	Procesamiento analítico en línea	<ol style="list-style-type: none"> 1. Atiende las indicaciones del docente para realizar la práctica procesamiento analítico en línea. 2. Identifica los tipos de sistemas de procesamiento de datos en línea. 3. Identifica las fuentes de datos. 4. Extrae los datos. 5. Procesa y analiza patrones de comportamiento. 6. Elabora reporte de práctica 	<ul style="list-style-type: none"> • Computadora • Internet. • Herramientas digitales • Referencias 	8 horas

		<ol style="list-style-type: none"> 7. Entrega al docente para su revisión y retroalimentación. 8. Corrige e Integra al portafolio de evidencias. 		
6	Arquitectura HDFS	<ol style="list-style-type: none"> 1. Atiende las indicaciones del docente para utilizar la arquitectura HDFS. 2. Identifica la arquitectura HDFS para el procesamiento de datos masivos. 3. Identifica las fuentes de datos. 4. Extrae los datos. 5. Procesa y analiza patrones de comportamiento. 6. Elabora reporte de práctica 7. Entrega al docente para su revisión y retroalimentación. 8. Corrige e Integra al portafolio de evidencias. 	<ul style="list-style-type: none"> • Computadora • Internet. • Herramientas digitales • Referencias 	4 horas
UNIDAD IV				
7	Proceso de minería de datos	<ol style="list-style-type: none"> 1. Atiende las indicaciones del docente para utilizar la minería de datos. 2. Identifica las etapas de exploración, preprocesamiento, evaluación, y clasificación de los datos masivos. 3. Identifica las fuentes de datos. 4. Extrae los datos. 5. Procesa y analiza patrones de comportamiento. 6. Elabora reporte de práctica 7. Entrega al docente para su revisión y retroalimentación. 8. Corrige e Integra al portafolio de evidencias. 	<ul style="list-style-type: none"> • Computadora • Internet. • Herramientas digitales • Referencias 	8 horas

8	Técnicas de Minería de datos	<ol style="list-style-type: none"> 1. Atiende las indicaciones del docente para utilizar la Minería de datos. 2. Identifica las técnicas más utilizadas de Minería de datos (redes neuronales, k-medias y naive bayes). 3. Identifica las fuentes de datos. 4. Extrae los datos. 5. Procesa y analiza patrones de comportamiento. 6. Elabora reporte de práctica 7. Entrega al docente para su revisión y retroalimentación. 8. Corrige e Integra al portafolio de evidencias. 	<ul style="list-style-type: none"> • Computadora • Internet. • Herramientas digitales • Referencias 	10 horas
9	Análisis de componentes principales	<ol style="list-style-type: none"> 1. Atiende las indicaciones del docente para realizar análisis de componentes principales 2. Identifica la técnica de componentes principales. 3. Proyectar los datos en términos de mínimos cuadrados. 4. Analiza las variables correlacionadas. 5. Elabora reporte de práctica 6. Entrega al docente para su revisión y retroalimentación. 7. Corrige e Integra al portafolio de evidencias. 	<ul style="list-style-type: none"> • Computadora • Internet. • Herramientas digitales • Referencias 	6 horas
UNIDAD V				
10	Presentación de datos	<ol style="list-style-type: none"> 1. Atiende las indicaciones del docente para realizar la visualización de datos. 2. Identifica los requerimientos para utilizar los datos y determina los diferentes tipos de 	<ul style="list-style-type: none"> • Computadora • Internet. • Herramientas digitales • Referencias 	6 horas

		<p>gráficos disponibles para su visualización.</p> <ol style="list-style-type: none">3. Identifica las fuentes de datos.4. Extrae los datos.5. Procesa los datos utilizando los gráficos aplicables dependiendo de los requerimientos del punto 2.6. Visualiza de manera gráfica cada dato requerido para el apoyo del proceso de toma de decisiones.7. Entrega al docente para su revisión y retroalimentación.8. Corrige e Integra al portafolio de evidencias.		
--	--	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--	--

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):

- Presenta información sobre los conceptos básicos de Big data
- Presenta ejercicios prácticos relacionados con las temáticas
- Proporciona información para las prácticas de laboratorio
- Resuelve y ejemplifica con casos prácticos
- Dirige, supervisa y retroalimenta las prácticas de laboratorio
- Propicia la participación activa de los estudiantes
- Revisa y evalúa reportes de prácticas y actividades
- Diseña y aplica evaluaciones
- Muestra el uso de software especializado

Estrategia de aprendizaje (alumno):

- Indaga y analiza información sobre conceptos básicos de Big data
- Resuelve ejercicios prácticos proporcionados por el profesor
- Realiza las prácticas de laboratorio
- Participa activamente en clase
- Elabora y entrega reportes de prácticas
- Trabaja de manera individual y en equipo
- Elabora y entrega actividades y prácticas en tiempo y forma
- Utiliza software especializado
- Atiende puntualmente las indicaciones del 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:

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.
- Por ser una unidad de aprendizaje eminentemente práctica no se evalúa en extraordinario, según lo especifica Estatuto Escolar vigente.

Criterios de evaluación

- Evaluaciones parciales..... 20%
- Tareas e investigaciones15%
- Prácticas de laboratorio..... 45%
- Portafolio de prácticas de laboratorio..... 20%
- Total.....100%**

IX. REFERENCIAS

Básicas

Lee, W., Leung, C., Nasridinov, A. (2021). *Big Data Analyses, Services, and Smart Data*. Singapore: Springer Singapore.

Luengo, J., García-Gil, D., Ramírez-Gallego, S., García López, S., Herrera, F. (2020). *Big Data Preprocessing. Enabling Smart Data*. Switzerland: Springer International Publishing

Quinto, B. (2018) *Next-Generation Big Data. A Practical Guide to Apache Kudu, Impala, and Spark*. USA: Apress

Ríos, D., & Gómez, D. (2019) *Big data: Conceptos, tecnologías y aplicaciones*. España: Catarata

Sakr, S. (2020) *Big Data 2.0 Processing Systems. A Systems Overview*. USA: Springer International Publishing

Complementarias

Marr. B., (2015). *Big Data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance*. USA: John Wiley & Sons, Ltd.

Pérez, M.,(2015). *Big Data – Técnicas, herramientas y aplicaciones*. México D.F., México: AlfaOmega

Warren, J., & Marz, N. (2015). *Big Data Principles and Best Practices of Scalable Realtime Data Systems*. United Kingdom Manning Publications.

X. PERFIL DEL DOCENTE

El docente que imparta la unidad de aprendizaje Big Data, deberá ser Licenciado en Informática, Licenciado en Inteligencia de Negocios, Ingeniero en Computación, Licenciado en Sistemas Computacionales o área a fin. De preferencia con experiencia mínima de tres años en la docencia y profesional. Ser proactivo, analítico y que fomente el trabajo en equipo.