

1. Datos Generales de la asignatura

Nombre de la asignatura:	Ciencia de Datos.
Clave de la asignatura:	DSD-2304
SATCA¹:	2-3-5
Carrera:	Ingeniería en Informática.

2. Presentación

Caracterización de la asignatura
<p>La Ciencia de Datos tiene como propósito hacer que los datos que nos rodean tengan un sentido. El uso de la Ciencia de Datos ayuda a tomar decisiones informadas, visualizar datos e incluso predecir eventos futuros mediante el uso del aprendizaje Máquina.</p> <p>Esta asignatura aporta al perfil del Ingeniero Informático las habilidades prácticas para analizar datos, visualizarlos y, a través del análisis de casos prácticos, analizar las decisiones a tomar apoyándose en la información generada.</p> <p>Se sugiere el uso del lenguaje de programación Python para desarrollar los temas de esta asignatura. Este lenguaje es libre, relativamente simple para codificar y cuenta con un soporte amplio para realizar tareas de Ciencia de Datos por medio de sus librerías, lo que permite demostrar los conceptos de esta ciencia de una manera clara y fácil de entender.</p> <p>Al egresar, el estudiante contará con las herramientas y conocimientos necesarios que le permitan adquirir, transformar y procesar datos con la finalidad de generar información útil para la toma de decisiones en las empresas o entidades gubernamentales.</p> <p>Se relaciona con la materia de inteligencia de negocios, en donde se realiza el proceso de ETL y de presentación de reportes principalmente, complementando el proceso de análisis de datos. También se relaciona con la materia de aprendizaje máquina en la cual se estudia a profundidad los algoritmos que se emplean en la unidad 3 de esta asignatura.</p>

1 Sistema de Asignación y Transferencia de Créditos Académicos

Intención didáctica

Esta asignatura comprende 4 unidades. En la primera unidad se analizan los conceptos generales de la Ciencia de Datos y se revisan las herramientas utilizadas para el procesamiento básico de datos, a los cuales se les aplicarán técnicas estadísticas básicas. Se sugiere el uso de las librerías NumPy y pandas de Python.

En la segunda unidad el estudiante graficará datos provenientes de fuentes de acceso libre, la idea central es poder visualizar estos datos de una manera eficaz, que permita entender de una manera simple la información que estos datos contienen. Se sugiere el uso de las librerías matplotlib y seaborn.

En la tercera unidad se abordará la aplicación de algoritmos de aprendizaje máquina para las tareas de clasificación supervisada, no supervisada, regresión y reglas de asociación, a problemas reales para generar y evaluar modelos que permitan automatizar la inferencia lógica y la toma de decisiones. Se sugiere el uso de las librerías Scikit-Learn y TensorFlow.

Finalmente, en la unidad 4, se revisarán conceptos de minería de texto para aplicarlos en un proyecto de reconocimiento de lenguaje natural, el cual permitirá la aplicación de técnicas de minería de datos y el reforzamiento de los temas realizados en las unidades 1, 2 y 3.

3. Participantes en el diseño y seguimiento curricular del programa

Lugar y fecha de elaboración o revisión	Participantes	Observaciones
--	----------------------	----------------------

Instituto Tecnológico del Valle de Oaxaca, 16 de octubre de 2019.	Academia de Ingeniería Informática e Ingeniería en Tecnologías de la Información y Comunicaciones del ITVO	Taller para generar la especialidad de la Ingeniería en Tecnologías de la Información y Comunicaciones.
---	--	---

4. Competencia(s) a desarrollar

Competencia específica de la asignatura
Extraer y analizar datos procedentes de múltiples fuentes para transformarlos en información de valor útiles para las empresas o entidades gubernamentales.

5. Competencias previas

<p>Conocimientos matemáticos y estadísticos.</p> <p>Lógica de programación.</p> <p>Habilidad en el empleo de un lenguaje de programación.</p> <p>Realizar operaciones con matrices.</p> <p>Uso de bases de datos relacionales.</p>
--

6. Temario

No.	Temas	Subtemas
1	Introducción a la Ciencia de Datos	1.1 Conceptos generales. 1.2 Herramientas utilizadas en la Ciencia de Datos. 1.3. Procesamiento básico de datos. 1.4 Técnicas estadísticas para el análisis de datos.

2	Visualización de la información	2.1 Principios de visualización de la información. 2.2 Graficación de datos. 2.3 Aplicación de visualizaciones.
3	Herramientas de Aprendizaje Máquina	3.1 Algoritmos de clasificación supervisada. 3.2 Algoritmos de clasificación no supervisada. 3.4 Algoritmos de regresión. 3.5 Algoritmos de reglas de asociación.
4	Minería de texto.	4.1 Manejo de texto. 4.2 Estructura de texto. 4.3 Manipulación de texto. 4.4 Expresiones regulares. 4.5 Limpieza de texto. 4.6 Métodos de procesamiento natural de lenguaje. 4.7 Clasificación de texto.

7. Actividades de aprendizaje de los temas

1. Introducción a la Ciencia de Datos	
Competencias	Actividades de aprendizaje
<p>Competencia Específica:</p> <p>Maneja datos tabulares.</p> <p>Aplica algoritmos para la limpieza de datos.</p> <p>Emplea herramientas computacionales para la visualización de datos.</p> <p>Realiza análisis estadístico inferencial.</p> <p>Competencias Genéricas:</p> <p>Propiciar actividades de búsqueda, selección y análisis de información en distintas fuentes.</p>	<p>Realiza investigación documental de aplicaciones de Data Science.</p> <p>Realiza ejercicios de carga, limpieza y procesamiento básico de datos utilizando la librería Pandas.</p> <p>Realiza operaciones con Dataframes.</p> <p>Aplica técnicas estadísticas (distribuciones, muestreo y t-tests).</p>

<p>Llevar a cabo actividades prácticas que promuevan el desarrollo de habilidades para la experimentación, tales como: observación, identificación manejo y control de variables y datos relevantes, planteamiento de hipótesis, de trabajo en equipo.</p>	
<p>2. Visualización de la información</p>	
<p>Competencias</p>	<p>Actividades de aprendizaje</p>
<p>Competencia Específica:</p> <p>Identifica el método de graficación adecuado para analizar los datos de un tipo determinado de problema.</p> <p>Competencias Genéricas:</p> <p>Capacidad creativa, capacidad para identificar, planear y resolver problemas. Toma de decisiones. Trabajo en equipo.</p>	<p>Realiza una investigación bibliográfica acerca de la creación de visualizaciones efectivas.</p> <p>Grafica con datos reales utilizando Matplotlib.</p> <p>Realiza gráficas compuestas.</p> <p>Investiga un problema y lo analiza con gráficas interactivas utilizando conjuntos de datos públicos.</p>
<p>3. Herramientas de Aprendizaje Máquina</p>	
<p>Competencias</p>	<p>Actividades de aprendizaje</p>
<p>Competencia Específica:</p> <p>Aplicará técnicas de aprendizaje máquina en tareas de clasificación supervisada, no supervisada, regresión y reglas de asociación.</p> <p>Competencias Genéricas:</p> <p>Aplica actividades de búsqueda, selección y análisis de información en distintas fuentes.</p>	<p>Resuelve problemas de clasificación supervisada utilizando algoritmos de Scikit Learn. (Redes Neuronales, Máquinas de soporte vectorial, árboles de decisión).</p> <p>Evalua modelos utilizando cross-validation).</p> <p>Resuelve problemas de clasificación no supervisada utilizando el algoritmo k-nearest neighbors con la librería Scikit Learn.</p>

Resolución de problemas. Presentación de resultados.	Resuelve problemas de reglas de asociación utilizando el algoritmo aprioricon la librería Scikit Learn.
--	---

4. Minería de texto.

Competencias	Actividades de aprendizaje
<p>Competencia Específica:</p> <p>Aplicará técnicas de procesamiento de lenguaje natural para realizar tareas de clasificación automática de texto.</p> <p>Competencias Genéricas:</p> <p>Desarrolla la capacidad para coordinar y trabajar en equipo; orientar el trabajo y potenciar la autonomía, el trabajo cooperativo y la toma de decisiones.</p>	<p>Investigación bibliográfica sobre minería de texto.</p> <p>Manejo de texto con Python.</p> <p>Práctica de expresiones regulares con pandas.</p> <p>Realiza tareas básicas de Procesamiento natural de lenguaje.</p> <p>Utilización de técnicas de identificación de características y aprendizaje máquina para clasificación de texto.</p>

8. Práctica(s)

<p>Análisis de datos utilizando técnicas estadísticas.</p> <p>Visualización de datos efectivas utilizando fuentes de datos públicas.</p> <p>Desarrollo de modelos inteligentes para la resolución de problemas.</p> <p>Minería de texto para el procesamiento natural del lenguaje.</p> <p>Desarrolla una aplicación móvil que consuma un servicio web de terceros.</p> <p>Realiza una aplicación móvil multiplataforma con el uso de formularios.</p> <p>Realiza una aplicación móvil multiplataforma con acceso a datos.</p>
--

9. Proyecto de asignatura

La intención del proyecto que plantee el docente que imparta esta asignatura, es demostrar el desarrollo y alcance de la(s) competencia(s) de la asignatura, considerando las siguientes fases:

- **Fundamentación:** marco referencial (teórico, conceptual, contextual y legal) en el cual se fundamenta el proyecto de acuerdo con un diagnóstico realizado, mismo que permite a los estudiantes lograr la comprensión de la realidad o situación objeto de estudio para definir un proceso de intervención o hacer el diseño de un modelo.
- **Planeación:** con base en el diagnóstico en esta fase se realiza el diseño del proyecto por parte de los estudiantes con asesoría del docente; implica planificar un proceso: de intervención empresarial, social o comunitario, el diseño de un modelo, entre otros, según el tipo de proyecto, las actividades a realizar los recursos requeridos y el cronograma de trabajo.
- **Ejecución:** consiste en el desarrollo de la planeación del proyecto realizada por parte de los estudiantes con asesoría del docente, es decir en la intervención (social, empresarial), o construcción del modelo propuesto según el tipo de proyecto, es la fase de mayor duración que implica el desempeño de las competencias genéricas y específicas a desarrollar.
- **Evaluación:** es la fase final que aplica un juicio de valor en el contexto laboral-profesión, social e investigativo, ésta se debe realizar a través del reconocimiento de logros y aspectos a mejorar se estará promoviendo el concepto de “evaluación para la mejora continua”, la metacognición, el desarrollo del pensamiento crítico y reflexivo en los estudiantes.

10. Evaluación por competencias

Realizar prácticas con herramientas de limpieza de datos.

Realizar prácticas con herramientas de extracción de características.

Examen escrito.

Proyecto integrador que ponga en práctica los conocimientos obtenidos en esta asignatura.

Participación en eventos académicos, tales como Congresos o concursos donde se traten los avances en el área de análisis inteligente de datos.

Investigación documental.

Rúbricas para la evaluación de las prácticas.

Utilizar herramientas de desarrollo y programación.

11. Fuentes de información

- Charles R. Severance. Charles Severance, (2016). Python for Everybody. Exploring Data Using Python 3.
- Allen B. Downey. Elements of Data Science... Green Tea Press.
- Joel Grus. Data Science from Scratch: First Principles with Python... Springer International Publisher.
- Alberto Boschetti, Luca Massaron. Python Data Science Essentials - Learn the fundamentals of Data Science with Python. Packt Publishing.
- Allen B. Downey. Think Python 2e: How to Think Like a Computer Scientist. Green Tea Press.

Referencias Virtuales:

- <https://www.python.org>
- <https://numpy.org/>
- <https://pandas.pydata.org/>
- <https://matplotlib.org/>
- <https://seaborn.pydata.org>
- <https://scikit-learn.org/>