

1. Datos Generales de la asignatura

Nombre de la asignatura:	PROGRAMACIÓN PARA CIENCIA DE DATOS
Clave de la asignatura:	CDF - 2305
SATCA¹:	3 - 2 - 5
Carrera:	Ingeniería en Tecnologías de la Información y Comunicaciones

2. Presentación

Caracterización de la asignatura

La pirámide DIKW (Datos, Información, Conocimiento y Sabiduría -*Data, Information, Knowledge, and Wisdom*-), identifica cuatro elementos, de entre los cuales, el que se encuentra en el nivel inferior corresponde a los **Datos**, definidos como una colección de símbolos que por sí mismos no pueden dar respuesta a una pregunta. El elemento siguiente, **Información** permite responder a preguntas del tipo: ¿quién?, ¿cuándo?, ¿dónde?, y ¿qué?. Por su parte, el **Conocimiento**, permite responder a preguntas abiertas del tipo ¿cómo?, que no pueden ser contestadas por los elementos anteriores. Por último, la **Sabiduría**, corresponde a una apreciación del ¿por qué? de las cosas, con una visión a futuro, a diferencia de los Datos, Información y Conocimiento que corresponden a eventos pasados.

Tradicionalmente los **Datos** son recopilados a partir de diferentes fuentes, para ser *preprocesados* en conjuntos de datos (datasets) que contienen **Información** relevante. Si a la **Información** se le aplican métodos y técnicas de Ciencia de Datos, entonces es posible descubrir patrones que muestran tendencias o relaciones de interés, y representan un **Conocimiento** extraído a partir de la **Información**. Por último, se *analizan los riesgos* de considerar los patrones extraídos y se *toma la decisión de ejecutar una acción -Sabiduría-*.

Por consiguiente, esta asignatura representa un complemento al perfil del Ingeniero en Tecnologías de la Información y Comunicaciones, que le permita no solo gestionar sistemas de información, sino que además sea capaz de **extraer patrones** a partir de diferentes fuentes de información, que le permitan la **toma de decisiones**. Para el desarrollo de esta asignatura se propone Python tomando en cuenta las siguientes características: a) En un lenguaje de programación de alto nivel, con uso eficiente de la memoria, b) Es multiplataforma, c) Por su simplicidad, es excelente para el desarrollo rápido de aplicaciones, d) Por su madurez, es ampliamente utilizado en aprendizaje automático (ML -*Machine Learning*-, Ciencia de Datos (Data Science) e

¹ Sistema de Asignación y Transferencia de Créditos Académicos

Inteligencia Artificial (AI -*Artificial Intelligence*), contando con una gran cantidad de librerías aplicables a las áreas antes citadas. Para la toma de decisiones, se incluyen modelos predictivos, utilizando regresión, así como modelos de clasificación y agrupamiento.

Intención didáctica

La asignatura es organizada en 3 temas, de los cuales el primero se destina a una revisión del lenguaje de programación Python.

El tema dos, introduce tres módulos ampliamente utilizados para la manipulación de datos y uno para la visualización de datos que permitan resumir o comparar poblaciones, mostrar y/o comparar tendencias, o tal vez descubrir relaciones entre atributos.

En el tema tres se presenta la aplicación del lenguaje de programación python aplicado a técnicas de:

- a) **Regresión** como un método para entender la relación existente entre las variables dependientes e independientes, de manera que se puedan predecir resultados, una vez descubierta dicha relación.
- b) **Clasificación** aplicando el algoritmo KNN y los árboles de decisión, para predecir una categoría; que, a diferencia de los algoritmos de regresión, se utilizan para predecir un número.
- c) **Agrupación**, utilizando técnicas para agrupar elementos de un conjunto de datos, de acuerdo con su similaridad; de manera que los elementos ubicados en un mismo grupo sean más similares con respecto a aquellos elementos que se ubican en otros conjuntos.

3. Participantes en el diseño y seguimiento curricular del programa

Lugar y fecha de elaboración o revisión	Participantes	Observaciones
Instituto Tecnológico de Toluca, septiembre de 2022	M. en T. E. Luis Antonio Estrada Manuel Dr. Federico del Razo López Dr. Roberto Alejo Eleuterio Dra. Eréndira Rendón Lara	Miembros de la Academia de Sistemas Computacionales del Instituto Tecnológico de Toluca

4. Competencias a desarrollar

Competencias específicas de la asignatura
<ul style="list-style-type: none"> Conoce e implementa los principios de descomposición y abstracción utilizando el lenguaje de programación Python. Conoce e implementa los principios de Programación Orientada a Objetos utilizando el lenguaje de programación Python. Conoce y utiliza herramientas para el procesamiento y manipulación de datos. Conoce y utiliza módulos de código abierto escritos en Python, con algoritmos aplicables en técnicas de aprendizaje supervisado y no supervisado. Comprende los principios básicos de Aprendizaje Automático (ML – Machine Learning) Aplica algoritmos de Aprendizaje Automático, para la identificación de patrones y/o realizar análisis predictivos. Aplica técnicas regresión, clasificación y/o agrupamiento para dar solución a problemas reales.

5. Competencias previas

Las adquiridas en las asignaturas siguientes:

- Álgebra Lineal
- Estructuras y Organización de Datos
- Matemáticas para la Toma de Decisiones
- Probabilidad y Estadística
- Programación Orientada a Objetos
- Tópicos de Desarrollo de Sistemas

6. Temario

No.	Temas	Subtemas
1	Introducción - Python	1.1 Tipos de Datos 1.2 Condiciones y Control de Flujo 1.3 Funciones 1.4 Colecciones 1.5 Recursión 1.6 POO
2	Procesamiento y Manipulación de Datos	2.1 NumPy 2.2 SciPy 2.3 Pandas 2.4 Mathplotlib

3	Machine Learning in Python (Scikit-learn)	3.1 Regresión - Regresión Lineal con Una Variable - Regresión Lineal con Múltiples Variables 3.2 Clasificación - KNN - Árboles de Decisión 3.3 Agrupamiento - Modelos de Agrupación - K-Means
---	--	---

7. Actividades de aprendizaje de los temas

Tema 1: Introducción - Python	
Competencias	Actividades de aprendizaje
Específicas: <ul style="list-style-type: none"> • Conoce e implementa los principios de descomposición y abstracción utilizando el lenguaje de programación Python. • Conoce e implementa los principios de Programación Orientada a Objetos utilizando el lenguaje de programación Python. Genéricas: <ul style="list-style-type: none"> • Habilidades de investigación. • Capacidad de análisis y de síntesis. • Habilidad para trabajar en forma autónoma y en equipo. • Comunicación oral y escrita. • Flexibilidad. • Gestión de tiempo. • Colaboración. 	<ul style="list-style-type: none"> • Configuración del entorno de trabajo: <ul style="list-style-type: none"> ◦ Sistema Operativo (Windows, Linux), ◦ Virtualización (VirtualBox, Vagrant) ◦ Editor/Depurador (Code Visual Studio, Sublime Text) • Instalación de la plataforma para Ciencia de Datos “Anaconda” • Creación de Programas individuales en Python usando los principios de programación orientada. • Creación de programas en equipo

Tema 2: Procesamiento y Manipulación de Datos	
Competencias	Actividades de aprendizaje
<p>Específicas:</p> <ul style="list-style-type: none"> Conoce y utiliza herramientas para el procesamiento y manipulación de datos. <p>Genéricas:</p> <ul style="list-style-type: none"> Habilidades de investigación. Capacidad de análisis y de síntesis. Habilidad para trabajar en forma autónoma y en equipo. Comunicación oral y escrita. Flexibilidad. Gestión de tiempo. Colaboración. 	<ul style="list-style-type: none"> Uso de herramientas para manipulación de datos usando ejemplos reales. Uso de herramientas para conectarse y extraer datos de bases de datos relacionales, mediante el uso de APIs Realizar ejercicios de visualización de datos Por equipos, identificar fuentes reales de datos estructurados, semi-estructurados o no estructurados, para su procesamiento en Python.
Tema 3: Machine Learning in Python (Scikit-learn)	
Competencias	Actividades de aprendizaje
<p>Específicas:</p> <ul style="list-style-type: none"> Conoce y utiliza módulos de código abierto escritos en Python, con algoritmos aplicables en técnicas de aprendizaje supervisado y no supervisado. Comprende los principios básicos de Aprendizaje Automático (ML – Machine Learning) Aplica algoritmos de Aprendizaje Automático, para la identificación de patrones y/o realizar análisis predictivos. Aplica técnicas regresión, clasificación y/o agrupamiento para dar solución a problemas reales. <p>Genéricas:</p> <ul style="list-style-type: none"> Habilidades de investigación. Capacidad de análisis y de síntesis. Habilidad para trabajar en forma autónoma y en equipo. Comunicación oral y escrita. Flexibilidad. Gestión de tiempo. Colaboración. 	<ul style="list-style-type: none"> Conocer y aplicar la herramienta “scikit-learn” para predecir valores (regresión), identificar categorías (clasificación) o realizar agrupamientos. Identificar un conjunto de datos asociado a un problema real, para aplicarle una técnica de regresión, clasificación o agrupamiento, según su dominio.

8. Práctica(s)

- **Práctica 1.** Instalación y configuración del ambiente de trabajo (Sistema Operativo, Python, IDE, APIs)
- **Práctica 2.** Desarrollo de programas en Python, aplicando principios de descomposición y abstracción.
- **Práctica 3.** Desarrollo de programas en Python, aplicando principios de programación orientada a objetos.
- **Práctica 4.** Desarrollo de programas en Python, para manipulación de datos: SQL, XML, CSV, ...
- **Práctica 5.** Desarrollo de programas en Python, para procesar datos con NumPy, Pandas, y Matplotlib.
- **Práctica 6.** Desarrollo de programas usando “**scikit-learn**”, para realizar regresión, clasificación y agrupamiento.

9. Proyecto de asignatura

El objetivo del proyecto que planteé el docente que imparte 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, 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, esta 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

La evaluación debe ser continua y formativa por lo que se debe considerar el desempeño en cada una de las actividades de aprendizaje, haciendo especial énfasis en:

- Uso de rúbricas de evaluación y tablas de cotejo
- Aplicación de exámenes teóricos y prácticos
- Entrega a tiempo de productos de aprendizaje
- Realizar una retroalimentación efectiva de los diferentes productos de aprendizaje
- Calendarización congruente de la entrega de los productos de aprendizaje para los diversos momentos de las evaluaciones diagnóstica, formativa y sumativa

11. Fuentes de información

1. Ozdemir, S. (2016). *Principles of Data Science* (1st ed.). Packt Publishing. Retrieved
2. Nagy, Z. (2018). *Artificial Intelligence and Machine Learning Fundamentals* (1st ed.). Packt Publishing. Fuente APA
3. Danesi, I. L., Perugini, F. (2019). *Data Exploration for Data Science*. EDUCatt.
4. Jafari, R. (2022). *Hands-On Data Preprocessing in Python* (1st ed.). Packt Publishing.
5. Harrington, P. (2012). *Machine Learning in Action*. Manning Publications.
6. Hastie, T., Hastie, T., Tibshirani, R., & Friedman, J. H. (2001). *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer.