



PROGRAMA ANALÍTICO DE ASIGNATURA

DEPARTAMENTO: Ingeniería en Sistemas de Información

CARRERA: Ingeniería en Sistemas de Información

NOMBRE DE LA ACTIVIDAD CURRICULAR: Ciencia de Datos

Año Académico: Plan 2023

Área: Sistemas Inteligentes

Bloque: Tecnologías Aplicadas

Nivel: 5º

Tipo: Obligatoria

Modalidad: Cuatrimestral

Cargas horarias totales:

<i>Horas reloj</i>	<i>Horas cátedra</i>	<i>Horas cátedra semanales</i>
72	96	3

OBJETIVOS

- Distinguir estrategias para gestionar un proyecto de ciencia de datos.
- Aplicar estrategias de extracción, visualización y transformación de datos en forma de patrones útiles y aplicables en el desarrollo de sistemas inteligentes.
- Crear modelos que resuelvan situaciones problemáticas particulares en el contexto del negocio.

CONTENIDOS

Contenidos mínimos

- Gestión de proyectos de ciencia de datos
- Análisis exploratorio de los datos
- Visualización de datos
- Extracción y Transformación de datos
- Ingeniería de datos
- Algoritmos de Clasificación y Agrupamiento
- Minería de Texto

Contenidos analíticos



Unidad 1: Concepto de Ciencia de Datos. Importancia de su aplicación. Historia y orígenes. Para que se utiliza. Relación con otras Ciencias. Perfiles en la industria.

Gestión de Proyectos de Ciencia de Datos. Conformación de Equipos. Perfiles involucrados.

Logros pedagógicos: interiorizar al alumno en los conceptos y las áreas vinculadas a la Ciencia de Datos. Comprender el proceso completo de la Gestión de Proyectos de Datos y la conformación de equipos y perfiles.

Unidad 2: Proceso de Ciencia de Datos. Obtención. Depuración. Exploración. Modelado e Interpretación. Sesgo de los Datos. Objetivos de su utilización. Metodologías del Proceso de Datos. Tecnologías y Técnicas en la Ciencia de Datos. Minería de texto.

Logros pedagógicos: Aprender el proceso completo llevado adelante aplicando metodologías. Distinguir y conocer las diferentes Tecnologías y técnicas aplicadas para la manipulación de los datos.

Unidad 3: Concepto de Ingeniería de Datos. Conceptos y Técnicas de Extracción, Transformación, Homogeneización, limpieza y visualización de datos.

Logros pedagógicos: Conceptualizar al alumno en las técnicas de manejo de los datos dentro de la ingeniería de datos. Formalizar el proceso llevado adelante por la ingeniería de datos.

Unidad 4: Algoritmos de Regresión, Clasificación y Agrupamiento aplicados en la Ciencia de Datos. Aprendizaje supervisado, no supervisado, semi supervisado y aprendizaje reforzado. Algoritmos basados en instancias o basados en modelos.

Logros pedagógicos: Aprender los diferentes algoritmos de clasificación y agrupamientos aplicados. Seleccionar su utilización en función de diferentes problemas y complejidades requeridas.

Unidad 5: Evaluación de Resultados obtenidos. Aplicación de los resultados en el mundo real. Aplicaciones en diferentes áreas de negocio. Casos de Estudio.

Logros pedagógicos: Clarificar el concepto de la aplicación y evaluación de resultados obtenidos. Conocer las diferentes áreas donde se aplica desde el análisis de casos de estudio.

BIBLIOGRAFÍA OBLIGATORIA

- Agrawal, T. (2020). Hyperparameter Optimization in Machine Learning: Make Your Machine Learning and Deep Learning Models More Efficient. Ed. Apress.
- Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn. Ed. Keras, and TensorFlow.



- Hurley, R. (2020). A Comprehensive Guide to Data Science, Data Analytics, Data Mining, Artificial Intelligence, Machine Learning, and big Data. Ed. Ationa Publications.
- McGregor, S. (2021) Practical Python Data Wrangling and Data Quality. Ed. O'Reilly.
- Paper, D. (2019) Hands-on Scikit-Learn for Machine Learning Applications: Data Science Fundamentals with Python. Ed. O'Reilly.
- Raschka, S. Mirjalili, V. (2019). Python Machine Learning. Ed. Packt Publishing.
- Raschka, S. (2019). Python Machine Learning and Deep learning. Ed. Packt Publishing.
- Rupam, N., Gypsy Dr., Sharma, K. (2020). Data Science Fundamentals and Practical Approaches. Ed. BPB Publications.

PÁGINAS WEB

- Python <https://www.python.org/>
- Scikit Learn. <https://scikit-learn.org/stable/>
- Anaconda <https://www.anaconda.com/products/individual>
- Stack overflow. <https://stackoverflow.com/>
- GitHub. <https://github.com/>
- Towards data Science. <https://towardsdatascience.com/>
- LightGBM <https://lightgbm.readthedocs.io/en/latest/>
- Apache Spark <https://spark.apache.org/>
- Databricks <https://databricks.com/documentation>
- Kaggle <https://www.kaggle.com/>

CORRELATIVAS

Para cursar y rendir

- Cursadas:
 - Simulación
- Aprobadas:
 - Probabilidad y Estadística
 - Bases de Datos