

Carrera: INGENIERIA EN INFORMATICA		
Asignatura 3670 - Ciencia de Datos		
Trayecto de Desarrollo de Software		
Año académico: 2023		
Responsable Mg. Abraham Alan Panuccio		
Carga horaria semanal: 4 hs	Carga horaria total: 64 hs	Créditos: -----
Modalidad: Presencial		
Correlativas anteriores: ESTADISTICA APLICADA - INTELIGENCIA ARTIFICIAL	Correlativas posteriores: -----	
Conocimientos necesarios: -----		

Descripción de la asignatura

Los contenidos que el alumno verá en esta asignatura son de la máxima actualidad e interés, ya que se enmarcan en el fenómeno conocido como Ciencia de Datos o Data Science, que está relacionado con la generación, el procesamiento y, sobre todo, la interpretación o búsqueda de información de interés a partir de grandes volúmenes de datos de múltiples orígenes, superando el alcance de las bases de datos tradicionales.

Metodología de enseñanza

La modalidad de enseñanza contempla clases teórico-prácticas con una duración de cuatro horas semanales ofrecidas en dos formatos: presencial o virtual.

La cátedra pone a disposición de los alumnos material teórico-práctico de cada uno de los temas desarrollados en clase, casos de estudio, trabajos de aplicación práctica, bibliografía y todo artículo conexo sobre los cuales los estudiantes deban investigar para facilitar la comprensión de los contenidos.

Los alumnos podrán utilizar Anaconda y Python para resolver problemas de aplicación o utilizar Google Colab.

A modo de complemento, se suman las siguientes herramientas de soporte:

- Plataforma MleL: donde los estudiantes acceden al material didáctico seleccionado o diseñado por los docentes. También se utiliza esta herramienta como vía de comunicación asincrónica entre estudiantes y con los docentes, a partir de foros de consulta o mensajería.
- Plataforma Teams: Entorno donde se desarrollan las clases sincrónicas cuando la modalidad de trabajo sea virtual.

Objetivos de aprendizaje

Objetivo general:

Brindar una visión general y sólida de los fundamentos que engloba la ciencia de datos, la inteligencia de negocios y el Big Data, buscando satisfacer la creciente demanda de profesionales con habilidades y conocimientos en el análisis, gestión y explotación de cantidades masivas de datos. Para ello, es necesaria la aplicación de nuevas soluciones que van más allá de los procesos tradicionales o herramientas de software habituales en la práctica diaria.

Objetivos específicos:

- Comprender la problemática de los sistemas de Soporte a la Toma de Decisiones, en la coyuntura tecnológica y socioeconómica actual de la inmediatez.
- Modelar adecuadamente los datos para lograr resultados satisfactorios, evaluando las técnicas más apropiadas para cada caso concreto.
- Proporcionar los conocimientos para definir métricas y atributos que permitan implementar una solución de BI capaz de mejorar la capacidad de la organización en la toma de decisiones.
- Abordar problemas complejos mediante el análisis y explotación de datos.
- Comprender los algoritmos más importantes en el ámbito de Machine Learning.
- Desarrollar modelos descriptivos y predictivos para tomar mejores decisiones, más rápido y más informadas, anticipándose incluso a la ocurrencia futura de los hechos de incidencia.
- Comprender el ciclo de trabajo de la ciencia de datos y de la gestión del conocimiento (KDD) dentro de las organizaciones y saber gestionarlo con éxito.
- Generar y transmitir conocimientos de manera efectiva.
- Experimentar el uso de herramientas para inteligencia de negocios y ciencia de datos para su aplicación en proyectos de explotación de información de cualquier ámbito.

Contenidos mínimos

- Fundamentos para la Ciencia de Datos.
- Modelos descriptivos y predictivos.
- Regresión lineal simple y múltiple.
- Regresión logística.
- Captura y Almacenamiento de Datos.
- Minería de Datos.
- Análisis del Modelo de Negocios.
- Visualización de Datos.

Competencias a desarrollar

Genéricas

- Generación de desarrollos tecnológicos y/o innovaciones tecnológicas.
- Desempeño en equipos de trabajo.
- Comunicación efectiva.
- Actuación profesional ética y responsable.

- Aprendizaje continuo.
- Desarrollo de una actitud profesional emprendedora.
- Identificación, formulación y resolución de problemas de ingeniería en sistemas de información/informática.
- Gestión, planificación, ejecución y control de proyectos de ingeniería en sistemas de información / informática.
- Utilización de técnicas y herramientas de aplicación en la ingeniería en sistemas de información / informática.

Específicas

- Establecimiento de métricas y normas de calidad de software.
- Procedimientos y certificaciones del funcionamiento, condición de uso o estado de sistemas de información, sistemas de comunicación de datos, software, seguridad informática y calidad de software.
- Dirección y control de la implementación, operación y mantenimiento de sistemas de información, sistemas de comunicación de datos, software, seguridad informática y calidad de software.

Programa analítico	
Unidad 1	Introducción a la Inteligencia de Negocios, Analítica y Ciencia de Datos
Unidad 2	Introducción a Data Mining y Machine Learning
Unidad 3	Introducción a Python / R para ciencia de datos
Unidad 4	Métodos de Clasificación. Generación y evaluación de modelos
Unidad 5	Regresión Lineal. Modelos Descriptivos

Planificación de actividades T: Teoría P: Práctica E: Evaluación				
Semana	Actividad	Tipo	Duración estimada	Unidad/des
1	Introducción. Fundamentos para la Ciencia de Datos	T	4	1,2
2	Introducción a Python / R para ciencia de datos	P	4	3
3	Métodos de Clasificación - Árboles de Decisión	T	4	4
4	Métodos de Clasificación - Clasificadores Bayesianos	T	4	4
5	Árboles de Decisión y Clasificadores Bayesianos	P	4	4
6	Métodos de Clasificación - KNN y Regresión Logística	TyP	3	4
7	Evaluación de modelos. Overfitting/Underfitting. ROC	T	4	4

8	Evaluación de modelos. Lineamientos del trabajo a presentar	PyE	4	4
9	Regresión Lineal	TyP	4	5
10	Regresión Lineal	TyP	4	5
11	Modelos descriptivos. Clustering. KMeans y DBSCAN	T	4	5
12	Modelos descriptivos. Clustering, KMeans y DBSCAN	PyE	4	5
13	Modelos descriptivos. Clustering jerárquico. Reglas de Asociación	T	4	5
14	Modelos descriptivos. Clustering Jerárquico y Reglas de Asociación	P	4	5
15	Workshop de temas nuevos.	E	4	
16	Entrega de Notas finales	E	4	

Evaluación

El alumno deberá desarrollar un trabajo integrador aplicando la metodología CRISP-DM, procesos y algoritmos de machine Learning supervisado / no supervisado sobre un caso propuesto, utilizando un producto comercial o de código abierto.

Los lineamientos y las consignas del trabajo serán entregadas al alumno, aproximadamente en la clase 8, ya que, a partir de la misma, contarán con las herramientas necesarias para su realización.

A partir de dicha fecha contará con 4 semanas para la entrega y a la semana siguiente tendrá el resultado de la evaluación por parte del docente, pudiendo haber aprobado o desaprobado en dicha instancia.

Si el alumno no aprobó el trabajo integrador en la primera entrega realizada, tendrá 2 semanas más para realizar una segunda entrega del trabajo considerando las correcciones indicadas por el docente.

Si el alumno aprobó el trabajo integrador, ya sea en la primera o segunda entrega, habrá aprobado la materia.

Si el alumno no logró completar los conocimientos necesarios en las instancias antes mencionadas, deberá rendir una evaluación (recuperatorio) y en caso de aprobarla, habrá aprobado la materia.

Primera evaluación	12	Entrega 1 del TP	1 hora reloj
Segunda evaluación	15	Entrega 2 del TP	1 hora reloj
Recuperatorio	16	Evaluación Final	2 horas reloj

Bibliografía obligatoria				
Título	Autor	Editorial	Edición	Año
The Data Warehouse Toolkit	Kimball, R., Ross M.	John Wiley and Sons	2da.	2002
Building a data warehouse for decision support	Poe, Vidette; Reeves, Laura L.	Prentice Hall	1ra.	1996

Bibliografía complementaria recomendada				
Título	Autor	Editorial	Edición	Año

Otros recursos obligatorios	
Nombre	

Otros recursos complementarios: Videos, enlaces	
Nombre	Material didáctico multimedia diseñado por el equipo docente.