

CARRERA: Licenciatura en Ciencias de Datos	CURSO LECTIVO: 2024
CÁTEDRA: Introducción a la Ciencia de Datos	CURSO: 1º año - 1º semestre
DURACIÓN: Semestral	Hs. TOTALES: 48 Hs. Reloj Totales
SEMANAS: 16	Hs. TEÓRICAS: 32 Hs. Reloj Totales Hs. PRÁCTICAS: 16 Hs. Reloj Totales

PROFESORA PROTITULAR: Represa, Soledad
PROFESOR ASISTENTE: Saavedra, Lucas Ariel

1. OBJETIVOS DE LA ASIGNATURA

Que los alumnos logren:

- Familiarizarse con las ciencias de datos.
- Reconocer la importancia de la disciplina en diversos ámbitos: el ámbito social, el empresarial, el profesional, el académico, el gubernamental, etc.
- Incorporar el alcance de las ciencias de datos.
- Comprender la perspectiva histórica, el desarrollo de las ciencias y técnicas asociadas a las ciencias de datos, darle sentido al presente de la práctica y vislumbrar las tendencias a futuro.
- Comprender cómo en ciencias de datos se aplica el método científico.
- Diferenciar correlación de relación causal. Familiarizar el sesgo estadístico de manera de poder llevar este concepto a la cotidianidad: falacias y fake news
- Conocer lo esencial de los principales modelos de ciencias de datos

2. UNIDADES TEMÁTICAS

1. Introducción.

Definición de "Dato". Tipos de Datos. Datos estructurados. Datos semi-estructurados. Datos no estructurados. Alcance de las Ciencias de Datos. Ciencias Asociadas. Aplicaciones en la Industria y en otras Ciencias. Perspectiva histórica.

2. Marco metodológico.

Método científico. El proceso de la ciencia de datos. Perfil del Profesional de los Datos. Equipos de Ciencias de Datos. Multidisciplinariedad. Nociones de Estadística Descriptiva. Muestreo. Variables cuantitativas y cualitativas. Distribución estadística. Sesgo. Eliminación del sesgo. Falacias. Fake news

3. Lenguajes y herramientas de modelado.

Concepto de API. Concepto de Framework. Herramientas para el trabajo en Ciencias de Datos (Python, Jupyter Notebooks, R, C++, Java, Scala, Git, Matlab, SAS, Anylogic,

Optaplanner, Weka). Comunidades. Kaggle.

4. Introducción al modelado.

Estadística inferencial. Modelos determinísticos y estocásticos. Caso práctico de modelado. Tipos de modelos. Machine Learning. Otras técnicas (Metaheurísticas, Algoritmos Genéticos, Sistemas expertos, Nociones de lógica difusa). Análisis e interpretación de datos. Introducción a la visualización de datos.

5. Modelos de Regresión.

Descripción del Problema. Fundamentos estadísticos. Principales APIs. Aplicación de Árboles de Decisión a los modelos de Regresión. Aplicación de Random Forest a los modelos de Regresión. Caso práctico. Métricas de error.

6. Modelos de Clasificación.

Descripción del Problema. Fundamentos estadísticos. Principales APIs. Árboles de Decisión. Random Forest. Caso práctico. Métricas de error. Matriz de confusión.

7. Introducción a la Minería de Datos.

Weka. Descubrimiento de conocimiento en bases de datos (KDD). Reglas de asociación. Algoritmo A-Priori. Análisis de conjuntos de ítems frecuentes. Caso práctico. Uso de metaheurísticas en minería de datos.

8. Redes neuronales.

Concepto de perceptrón. Redes neuronales. Aprendizaje Automático. Aprendizaje Automático reforzado. Aprendizaje Profundo. Concepto de entrenamiento. Tensorflow. Pytorch. Caso práctico.

3. BIBLIOGRAFÍA

3.1 BIBLIOGRAFÍA GENERAL OBLIGATORIA

- *"Spark in Action" 2nd Edition* - J.G. Perrin - Ed. Manning - 2020 - ISBN 9781617295522 .
- *"Big Data: Principles and best practices of scalable realtime data systems"* - N. Marz, J. Warren - Ed. Manning - 2015 - ISBN 9781617290343.
- *"Data Pipelines with Apache Airflow"* - B.P. Harenslak & J. Rutger de Ruiter - Ed. Manning - 2021 - ISBN 9781617296901.
- *"Data Governance: The Definitive Guide"* - E. Eryurek, U. Gilad, V. Lakshmanan, A. Kibunguchy-Grant & J. Ashdown - O'Reilly Media, Inc. - 2021 - ISBN: 9781492063490.

4. METODOLOGÍA

El curso está organizado en 8 unidades temáticas divididas en encuentros de 4 hs cátedras semanales, a realizarse en formato presencial. La modalidad adoptada para el dictado será **teórico-práctica**. En las clases se presentarán los temas de cada unidad, proponiendo espacios de intercambio con el docente y entre los estudiantes a partir de consignas específicas. Se facilitará material de lectura obligatoria y complementaria para complementar la comprensión de las unidades.

5. EVALUACIONES Y CRITERIOS PARA LA APROBACIÓN

La aprobación de la materia estará supeditada al cumplimiento de la condición de asistencia exigida por la Universidad, la aprobación de todas las actividades prácticas y la aprobación del examen integrador.

Los trabajos prácticos podrán ser individuales o grupales, debiéndose cargar a través de la plataforma de Entornos Virtuales de Aprendizaje en tiempo y forma, otorgándose una única instancia de revisión y recuperación. Las actividades prácticas deberán contar con su aprobación para acceder a la instancia de evaluación final.

Para los trabajos prácticos y la evaluación final se realizarán sesiones de consultas individuales y grupales, haciendo además puesta en común general si el caso lo requiera. A los estudiantes que presenten dificultades se les observará y se los guiará para resolver el conflicto.

La instancia de recuperación está prevista para estudiantes que no hayan aprobado el examen integrador o que hayan estado ausentes.

Criterios de Evaluación:

- Respeto de las consignas presentadas.
- Resolución correcta de los problemas planteados.
- Adecuada respuesta a los contenidos teóricos.

6. CRITERIOS y MODALIDAD PARA LA EVALUACIÓN DEL EXAMEN FINAL

El examen final consiste en una evaluación oral y escrita, presencial e individual, donde el alumno deberá demostrar conocimientos teóricos y prácticos. El examen final se diferencia en que abarca todos los temas del programa. Los ejercicios prácticos tendrán un carácter integrador, articulando los distintos contenidos vistos en la materia. Finalmente, en las preguntas teóricas se pretende que el alumno demuestre un conocimiento profundo de los temas, relacionando conceptos entre sí.

Criterios de Evaluación:

- Respeto de las consignas presentadas.
- Adecuada respuesta a los contenidos teóricos.
- Relación de conceptos pertinente.
- Resolución correcta de los problemas planteados.
- Fundamentación bibliográfica de los temas.