

Formulario de aprobación de curso de posgrado/educación permanente

Asignatura: Introducción a la Ciencia de Datos

(Si el nombre contiene siglas deberán ser aclaradas)

Modalidad:

(posgrado, educación permanente o ambas)

Posgrado

Educación permanente

Profesor de la asignatura ¹: Dr. Marcelo Fiori (Gr 3, DT, IMERL)

(título, nombre, grado o cargo, instituto o institución)

Profesor Responsable Local ¹:

(título, nombre, grado, instituto)

Otros docentes de la Facultad: Dra. Lorena Etcheverry (Gr 3, DT, INCO)
Dr. Ignacio Ramírez (Gr 3, DT, IIE)
Dr. Mathias Bourel (Gr 3, DT, IMERL)
Dra. María Inés Fariello (Gr 2, DT, IMERL)
Dr. Guillermo Moncecchi (Gr 3, INCO)

Docentes fuera de Facultad:

(título, nombre, cargo, institución, país)

¹ Agregar CV si el curso se dicta por primera vez.

(Si el profesor de la asignatura no es docente de la Facultad se deberá designar un responsable local)

Programa(s) de posgrado: Maestría en Ingeniería Matemática, Maestría en Ingeniería Eléctrica, Maestría en Informática.

Instituto o unidad: IMERL/INCO/IIE

Departamento o área:

Horas Presenciales: 60

(se deberán discriminar las horas en el ítem Metodología de enseñanza)

Nº de Créditos: 10

[Exclusivamente para curso de posgrado]

(de acuerdo a la definición de la UdelaR, un crédito equivale a 15 horas de dedicación del estudiante según se detalla en el ítem Metodología de enseñanza)

Público objetivo: Estudiantes de posgrado, egresados de carreras de ingeniería o licenciaturas en matemática o física, con interés en realizar estudios en ciencia de datos.

Cupos: 25 estudiantes

(si corresponde, se indicará el número de plazas, mínimo y máximo y los criterios de selección. Asimismo, se adjuntará en nota aparte los fundamentos de los cupos propuestos. Si no existe indicación particular para el cupo máximo, el criterio general será el orden de inscripción, hasta completar el cupo asignado)

Objetivos:

El objetivo central es proveer de una base común a los estudiantes que ingresan a una carrera de posgrado para trabajar con problemas de Ciencia de Datos. En particular, introducir los conceptos principales y reforzar los conocimientos fundamentales necesarios para facilitar el desempeño en el resto de la carrera.

Al culminar el curso, el estudiante tendrá un panorama global de lo que es la Ciencia de Datos y ramas afines, un manejo del lenguaje y terminología general aplicada en el área, y elementos prácticos, teóricos y metodológicos como para encarar un problema de Ciencia de Datos.

En particular, el objetivo es que al final del curso el estudiante sea capaz de:

- Formular una estrategia completa de solución para problemas de ciencia de datos, incluyendo procesado y gestión de datos, formulación del problema, exploración de los datos, modelado, implementación, evaluación, y análisis crítico de los resultados.
- Identificar problemas en los datos como ser datos faltantes, anomalías, redundancia.
- Comprender los conceptos básicos de aprendizaje supervisado y no supervisado, incluyendo conceptos como sobreajuste, poder de generalización, separación de datos en entrenamiento, validación y test, la importancia de las características.
- Identificar problemas éticos, incluyendo implicancias de privacidad, sesgo, y toma de decisiones automáticas, entre otros.

Conocimientos previos exigidos: Conocimientos básicos de cálculo diferencial e integral, probabilidad y estadística, álgebra lineal, y programación.

Conocimientos previos recomendados:

Metodología de enseñanza:

Descripción de la metodología:

Se hará énfasis en los aspectos formativos y metodológicos de la práctica en Ciencias de Datos a través de la aplicación de ciertos principios generales y métodos básicos a uno o más problemas prácticos del área. Dichos principios y métodos serán profundizados en otras asignaturas.

A su vez, se introducirá al estudiante a un conjunto de herramientas de uso general. Dicho material será impartido mediante tutoriales escritos y puestos en práctica en las instancias de laboratorio.

El curso se estructura en 12 semanas compuestas cada una por dos clases teórico-prácticas de dos horas cada una.

Detalle de horas:

- Horas de clase (teórico): 36
- Horas de clase (práctico): 12
- Horas de clase (laboratorio):
- Horas de consulta: 12
- Horas de evaluación:
 - Subtotal de horas presenciales: 60
- Horas de estudio: 40
- Horas de resolución de ejercicios/prácticos: 50

- Total de horas de dedicación del estudiante: 150

Forma de evaluación:

Se hará sobre la base de 100 puntos repartidos en cuatro instancias (pruebas escritas presenciales, entrega de ejercicios, ejercicios de programación, u otras) de 15 puntos cada una (no obligatorias ni eliminatorias), más un trabajo final de 40 puntos (obligatorio). Para aprobar, el estudiante deberá entregar el trabajo final y reunir, en la totalidad de los puntos, al menos 60 de los 100 totales.

Temario:

1. **Introducción y conceptos generales de la Ciencia de Datos.** Tipos de datos, tipos de modelos. Componentes y etapas de un sistema completo, desde la recolección de datos a los resultados y su análisis.
2. **Manejo y gestión de datos.** Modelos de datos y modelado. Calidad de datos y limpieza. Recolección e integración de datos. Interpretación y visualización. Ética y privacidad.
3. **Fundamentos matemáticos.** Conceptos de Probabilidad y Álgebra Lineal asociados a métodos de aprendizaje automático. Por ejemplo: Regla de Bayes, distribuciones de probabilidad, correlación, valores y vectores propios, descomposición de matrices, mínimos cuadrados y optimización.
4. **Métodos de aprendizaje automático.** Tipos de métodos: supervisado, no-supervisado, por recompensa. Evaluación y sobreajuste. Clasificación y regresión. Vecinos más cercanos. Clasificadores lineales. K-means. Análisis de Componentes Principales.

Bibliografía:

(título del libro-nombre del autor-editorial-ISBN-fecha de edición)

Pattern recognition and machine learning. Bishop, Christopher M. Springer, 2006.

The elements of statistical learning. Friedman, Jerome, Trevor Hastie, and Robert Tibshirani.. Vol. 1. No. 10. New York: Springer series in statistics, 2001.

Data and Information Quality: Dimensions, Principles and Techniques. Carlo Batini, Monica Scannapieca, Springer, 978-3-319-24106-7, 2016. (Disponible en Portal Timbó <https://doi-org.proxy.timbo.org.uy/10.1007/978-3-319-24106-7>)

The DataOps Cookbook: Methodologies and Tools That Reduce Analytics Cycle Time While Improving Quality, Christopher Bergh, Gil Benghiat, and Eran Strod, DataKitchen, Inc, 2019.

Practical DataOps: Delivering Agile Data Science at Scale, Harvinder Atwal, Apress, 9781484251041, 2019