

**UNIVERSIDAD DE LA REPÚBLICA
FACULTAD DE INGENIERÍA
COMISIÓN ACADÉMICA DE POSGRADO**

DESCRIPCIÓN DEL PROGRAMA DE POSGRADO

Nombre del Programa: *Diploma de Especialización en Ciencia de Datos*

Montevideo – 2018

1. IDENTIFICACIÓN:

DE LA CARRERA

Nombre del Programa: **Diploma de Especialización en Ciencia de Datos**

Programa (especialización, maestría académica o profesional, o doctorado): Programa de especialización Profesional, Posgrado

ÁREA ACADÉMICA

Área (Instituto/ Grupo/ Núcleo, etc.): Instituto de Computación

Institutos vinculados al Área: InCo, IMERL

Contacto institucional del Programa

Nombre: Pablo Rodríguez Bocca

Teléfono: 2 7114244

E-mail: prbocca@fing.edu.uy

Programa compartido con otra Institución:

Nombre de la Institución:

En caso afirmativo adjuntar copia del acuerdo establecido.

ANTECEDENTES Y FUNDAMENTACIÓN

La Ciencia de datos estudia la forma de extraer conocimiento a partir de los datos. Como todas las ciencias, aplica el método científico a sus objetos de estudio, que en este caso son los datos. En la actualidad, en todas las áreas de desempeño profesional, los datos se presentan en grandes volúmenes, y en diferentes formas y formatos. Esto transforma a la ciencia de datos en una disciplina de creciente importancia y de naturaleza interdisciplinaria, que logra de forma exitosa utilizar la misma metodología en las más variadas áreas de aplicación. Dicha metodología usualmente incluye una etapa inicial de acceso a los datos, que implica la obtención, limpieza y muestreo de los datos en forma eficaz y eficiente, adaptable para cada caso; seguido por una etapa de extracción de conocimiento en base al planteo de hipótesis científicas, utilizando modelado estadístico, programación y aprendizaje automático; terminando en una etapa de documentación y visualización de los resultados generados.

Las carreras de grado que incluyen formación en matemática e informática, como las ingenierías, ofrecen una formación metodológica de base, que combinada con los conocimientos técnicos del área de aplicación de cada variante particular, constituyen los requisitos necesarios para formar científicos de datos. Por tanto, este diploma busca ofrecer una formación de posgrado orientada a profesionales egresados de carreras de carácter cuantitativo, que aplican (o desean aplicar) el método científico en el análisis de datos.

La formación asume que el profesional cuenta con una base de conocimientos en modelado matemático (probabilidad y estadística) y en informática (lenguajes de programación), además de entender cómo se aplica el método científico. De este modo, la propuesta de formación apunta a un nicho definido por profesionales con una formación de base cuantitativa, y una fuerte motivación por las actividades del análisis de datos. El diploma plantea un plan de cursos adaptable a distintas formaciones de base de los profesionales; logrando en cada caso un balance desafiante entre la matemática y la computación.

2. UBICACIÓN FÍSICA DEL PROGRAMA

Lugar y dirección completa de la sede del programa:

Instituto de Computación
Facultad de Ingeniería - Universidad de la República
Herrera y Reissig 565
Montevideo / Uruguay

Nombre y teléfono de un contacto en la Institución Sede:

Nombre: Pablo Rodríguez Bocca
Teléfono: 27114244

Personal, instalaciones, y materiales disponibles para la realización del programa:

El Instituto de Computación cuenta con más de 150 docentes y el Instituto de Matemática con aproximadamente 100 docentes, de los cuales un alto porcentaje posee título de posgrado (doctorado o maestría). Debido a las temáticas abordadas en esta especialización se espera contar con el apoyo de todos ellos para participar en el programa tanto de forma estable como ocasionalmente a través de actividades puntuales.

Entre ambos institutos se cuenta con un importante plantel docente en los temas de esta especialización. Son decenas los docentes involucrados en el programa propuesto, todos contando con formación de maestría o doctorado. Estos docentes dictarán de forma regular el conjunto de cursos de posgrado que integrarán los cursos obligatorios y opcionales del programa propuesto.

Se cuenta con aulas y salas de computadoras de uso compartido de la Facultad de Ingeniería, así como de equipos especializados (hardware y software) disponible en los diversos grupos de investigación del InCo, que permiten dar los recursos necesarios para la realización de los trabajos de los estudiantes del programa. En particular, se cuenta con infraestructura de hardware de alto poder de cómputo, así como software específico en las áreas de análisis de datos, simulación y optimización. Asimismo, la Facultad cuenta con varias salas de posgrados equipadas.

Se cuenta con acceso electrónico a la mayoría de las revistas científicas relevantes del área a través del portal Timbó. Asimismo se cuenta con acceso a ediciones impresas actualizadas de algunas de estas revistas y libros especializados, a través de la biblioteca del Instituto de Computación.

3. OBJETIVOS DEL PROGRAMA

FINALIDAD:

El Diploma de especialización en "Ciencia de Datos" se dirige a profesionales interesados en la temática, que deseen especializarse en el análisis de datos y la inteligencia artificial.

La formación de grado en general encara aspectos técnicos y la solución de problemas de cada área específica. En este diploma se busca complementar esta formación básica, brindando los elementos necesarios para analizar cualquier tipo de dato; poder extraer conocimiento de los datos; modelar la realidad y predecir su comportamiento en forma fundamentada.

OBJETIVOS ESPECÍFICOS:

Este diploma tiene como objetivos brindar una formación específica en el análisis de datos y la inteligencia artificial, basándose en áreas de conocimiento como el Aprendizaje Automático, el Modelado y Procesamiento de Datos, la Probabilidad y Estadística, la Optimización Combinatoria, y la Teoría de Grafos entre otras.

PERFIL DEL EGRESADO:

Un científico de datos es aquella persona que aplica con solvencia las metodologías de la ciencia de datos. Cuenta con muy buena formación en estadística y computación, siendo capaz de modelar problemas reales en forma creativa y abstracta.

El egresado adquirirá la capacidad de aplicar los temas de estudio incluidos en el Diploma en un amplio rango de áreas de aplicación, tanto en el campo profesional como académico, con profundidad y solvencia. Asimismo, adquirirá los elementos metodológicos que junto con la capacidad de abordar bibliografía especializada, le permitan comprender y emplear las nuevas tecnologías para el análisis de datos en su actividad profesional.

Se espera que el egresado de este diploma tenga la capacidad para:

- Recopilar, procesar y extraer valor de fuentes de datos heterogéneas.
- Abstractar y formular hipótesis de forma creativa para extraer conocimiento, que puedan responderse mediante el análisis de datos.
- Identificar, seleccionar y aplicar los métodos más apropiados de análisis de datos y aprendizaje automático al problema específico.
- Dominar la tecnología necesaria para analizar datos mediante paquetes de software especializados y procesar los datos que los mismos requieren.
- Mantenerse actualizado en la bibliografía referente a los puntos anteriores.

TÍTULO:

Cuando el aspirante haya completado los requisitos del programa, la SCAPA-informática notificará a la CAP, quien propondrá al Consejo de la Facultad el otorgamiento del Título "Especialista en Ciencia de Datos". Este diploma será firmado por el Decano de la Facultad de Ingeniería y el Rector de la Universidad de la República.

4. ORGANIZACIÓN Y NORMAS DE FUNCIONAMIENTO

La Sub Comisión Académica de Posgrado de Informática (SCAPA-informática) supervisará las actividades ligadas al desarrollo del Diploma de Especialización en “Ciencia de Datos”; sin perjuicio de las competencias que correspondan a la Comisión Académica de Posgrado (CAP) y al Consejo de la Facultad de Ingeniería.

La SCAPA-informática nombrará para cada estudiante un Director Académico, responsable de la organización de las actividades y de la orientación del mismo. La SCAPA-informática dispondrá del apoyo de un responsable para las decisiones que involucren al diploma.

Los aspectos reglamentarios no mencionados explícitamente se ajustan a lo establecido por los documentos: Ordenanza de las Carreras de Posgrado de la Universidad de la República, aprobado en fecha 25/09/01 por el Consejo Directivo Central y el Reglamento General de las Actividades de Posgrado y Educación Permanente de la Facultad de Ingeniería (RGP-FING), 2003.

Formación:

Por cada estudiante, la SCAPA-informática formulará una propuesta de plan de formación, que será aprobada por la CAP. Los planes de formación se integrarán con actividades programadas (cursos de actualización y/o posgrado, seminarios, etc.), de manera de cumplir un mínimo de 60 créditos (el crédito es la unidad de medida de la carga de trabajo en los planes de estudio de la Facultad de Ingeniería, y equivale a quince horas de dedicación por parte del estudiante).

Todas las actividades programadas deberán contar con alguna forma de evaluación de los conocimientos adquiridos.

La duración prevista para la realización del Diploma de Especialización en “Ciencia de Datos” es de 18 meses con una dedicación estimada entre 15 y 20 horas semanales.

Número de plazas previstas (incluyendo becas si es aplicable): Se prevén entre 15 y 20 alumnos

Número mínimo de alumnos para realizar el programa: El número mínimo de alumnos es de 5

Requisitos para obtener el título:

Número de créditos mínimos de Actividad Programada: 60 Créditos

Horas presenciales mínimas de Actividad Programada: No corresponde

Estructura de la Actividad Programada (fundamentales y técnicas): Cursos de entre 2 y 10 créditos cada uno.

INGRESO

Perfil de ingreso

El programa está dirigido a profesionales en general; con interés en profundizar sus capacidades científicas, y motivación por aplicar este conocimiento al análisis de datos.

Requisitos de Ingreso

Podrán ingresar al Diploma de Especialización "Ciencia de Datos" quienes cumplan con al menos una de las siguientes condiciones:

- **Condición 1:** Contar con un título de grado en carreras que incluyan formación en matemática e informática, otorgado por la Universidad de la República de al menos 360 créditos.
- **Condición 2:** Contar con formación equivalente que, a juicio de la Comisión de Posgrado, permita la realización y aprovechamiento del Plan de Estudios del Diploma de Especialización en "Ciencia de Datos".

En cualquiera de las condiciones anteriores, la SCAPA-informática podrá proponer la realización de cursos de nivelación en caso de considerarlo necesario para la correcta realización del plan de estudio.

En todos los casos se requiere el dominio del idioma inglés, a nivel de lectura de documentos y comprensión oral.

Criterios de selección de los candidatos

Las candidaturas deberán ser presentadas a la SCAPA-informática, quien deberá elevar un informe a la CAP sugiriendo la aprobación o no de la candidatura. La admisión tendrá en cuenta los antecedentes del candidato, pudiéndose realizar una entrevista a los aspirantes para complementar la información presentada. La CAP resolverá la admisión de los candidatos en base a sus antecedentes y al informe de la SCAPA-informática.

5. CUERPO DOCENTE Y SUS ACTIVIDADES

Titulación/nombre/cargo/instituto	Horas aula anuales dedicadas al programa	Nº previsto de candidatos a orientar	Nº previsto de estudiantes a orientar en otros programas	Horas anuales de otras actividades vinculadas al programa
01. Dr Mathías Bourel, gr. 3, DT./IMERL	60			
02. Dra. Ing. Andrea Delgado, gr. 4, DT/InCo	-			20
03. Dra. Ing. Lorena Etcheverry, gr. 3, DT./InCo	60			
04. Msc. Mathias Etcheverry, gr. 2, InCo	20			
05. Dr. Ing. Eduardo Fernández, gr. 4, DT/InCo	33			
06. Msc. Diego Garat, gr. 3, InCo	-			20
07. Msc. Jorge Graneri, gr. 2, DT./IMERL	20			
08. Dr. Gustavo Guerberoff, gr. 4, DT./IMERL	60			
09. Dr. Ing. Santiago Iturriaga, gr. 2, DT./InCo	20			20
10. Dr. Juan Kalemkerian, gr. 3, 20 hrs./IMERL	24			
11. Msc. Luis Chiruzzo, gr. 3, DT/InCo	20			
12. Dr. Ing. Sergio Nesmachnow, gr. 5, DT./InCo	20			20
13. Dra. Ing. Adriana Marotta, gr. 4, DT/InCo	50			
14. Dr. Ing. Guillermo Moncecchi, gr. 3, InCo	15			
15. Msc. Lic. Fabio Pazos, gr. 2, 20 hrs./IMERL	-			20
16. Dr. Gonzalo Perera, gr. 5, DT./IMERL	-			20
17. Dr. Juan Piccini, gr 3, 35 hrs./IMERL	20			
18. Msc. Juan José Prada, gr. 3, InCo	-			20
19. Dr. Ing. Claudio Risso, gr. 3, DT/InCo	-			20
20. Dr. Ing. Franco Robledo, gr. 5, DT./InCo	-			20
21. Dr. Ing. Pablo Rodríguez-Bocca, gr. 4, 35 hrs./InCo	42			60
22. Dra. Ing. Aiala Rosá, gr. 3, DT./InCo	20			
23. Dr. Ing. Raúl Ruggia, gr. 5, 6 hrs./InCo	-			20
24. Dr. Marco Scavino, gr. 4, 40 hrs./FCEA, InEst	-			20
25. Dra. Ing. Libertad Tansini, gr. 3, DT./InCo	40			
26. Dra. Ing. Dina Wonsever, gr. 5, DT/InCo	-			20

6. CURRÍCULA

asignatura n.º 01 : **Técnicas de Aprendizaje Estadístico**

Responsable de la unidad curricular (docente): Dr. Gustavo Guerberoff

Instituto: IMERL

Departamento: LPE

Arancel:

Nº de Créditos: 6

Cupos:

Horas Presenciales: 60

Objetivos: Introducir los aspectos metodológicos de algunas técnicas modernas de aprendizaje automático. Aplicar dichas técnicas a conjuntos de datos reales e interpretar los resultados obtenidos. Acercar el estudiante al empleo de los paquetes para el análisis estadístico de datos disponibles en el ambiente de desarrollo de software libre R (<http://www.r-project.org/>). .

Conocimientos previos exigidos: Introducción a la probabilidad y estadística

Metodología de enseñanza:

- ! Horas clase (teórico): 22
- ! Horas clase (práctico): 6
- ! Horas clase (laboratorio): 20
- ! Horas consulta: 10
- ! Horas evaluación: 2
 - o Subtotal horas presenciales: 60
- ! Horas estudio: 12
- ! Horas resolución ejercicios/prácticos: 8
- ! Horas proyecto final/monografía: 10
 - o Total de horas de dedicación del estudiante: 90

Forma de evaluación

Resolución de ejercicios propuestos durante el curso y trabajo final a entregar con defensa oral.

Temario:

1. Análisis de datos. Problemas de clasificación.
2. Clasificación Bayesiana.
3. Regresión Logística.
4. Introducción a las técnicas de Aprendizaje Automático. Principios básicos.
5. Árboles de clasificación y regresión (CART).
6. k vecinos más cercanos.
7. Support Vector Machines.
8. Métodos de agregación de modelos: Bagging, Boosting, Random Forest.
9. Aprendizaje no supervisado.

Bibliografía:

- Bourel, M. (2012) Model aggregation methods and applications. Memorias de trabajos de difusión científica y técnica, Vol. 10, p. 19-32, 2012.
- Breiman L. (2001). Random forests. Machine Learning 45 (1): pp 5–32.
- Breiman L., Friedman J, Stone CJ & RA Olshen (1984) Classification and Regression Trees. Wadsworth Internacional Group, Belmont, CA.
- Hastie T., Tibshirani R and Friedman J (2011). The Elements of Statistical Learning: Data Mining, Inference and Prediction. (5th. Edition). Springer Seris in Statistics.
- James, G., Witten, D., Hastie, T., Tibshirani, R. (2013) An introduction to Statistical Learning with Applications in R, Springer.
- Peña, D. (2002). Análisis de datos multivariantes, Daniel Peña, Mac Graw Hill

asignatura nº 02 : Introducción a la Estadística usando software.

Responsable de la unidad curricular (docente): Msc. Jorge Graneri / Dr. Juan Piccini

Instituto: IMERL

Departamento: LPE

Arancel:

Nº de Créditos: 4

Cupos:

Horas Presenciales: 40

Objetivos: Introducir los aspectos básicos de la inferencia estadística clásica. Aplicar dichas técnicas a conjuntos de datos reales e interpretar los resultados obtenidos. Acercar el estudiante al empleo de los paquetes de software para el análisis estadístico de datos disponibles.

Conocimientos previos exigidos:

Introducción a la probabilidad y estadística.

Metodología de enseñanza:

Horas clase (teórico): 20

Horas clase (práctico): 18

Horas clase (laboratorio):

Horas consulta:

Horas evaluación: 2

Subtotal horas presenciales: 40

Horas estudio: 8

Horas resolución ejercicios/prácticos: 12

Horas proyecto final/monografía: 0

Total de horas de dedicación del estudiante: **60**

Forma de evaluación

Resolución de ejercicios propuestos durante el curso y examen final.

Temario:

1. Introducción. Tipo de variables. Definiciones básicas (3h)
2. Repaso de Estadística Descriptiva uni y multidimensional. Histograma. Indicadores de media y de dispersión. (4h)
3. Relaciones entre variables cualitativas, variables cualitativa y cuantitativas, variables cuantitativas. Regresión lineal simple. (7h)
4. Repaso de probabilidad. Variable aleatoria. Independencia. Momentos. Distribuciones usuales. Convergencia de variables aleatorias (7h)
5. Muestreo. Simulaciones (5h)
6. Estimación puntual. Métodos de estimación. (7h)
7. Test de hipótesis. Nociones generales. Test sobre la media en una población normal. (7h)

Bibliografía:

- James, G., Witten, D., Hastie, T., Tibshirani, R. (2013) An introduction to Statistical Learning with Applications in R, Springer.
- Peña, D. "Fundamentos de Estadística", Alianza Editorial.
- Feller, William. "Introducción a la Teoría de Probabilidades y sus aplicaciones". Limusa-Wiley S.A. 1973.
- Perera, G. "Probabilidad y Estadística Matemática". Editorial Fin de Siglo, en venta en la librería del CEI.
- Everitt, B.S y Hothorn T. (2010) A Handbook of Statistical Analyses Using R, 2nd edition, Chapman & Hall/CRC.

asignatura nº 03 : Calidad e Integración de Datos

Responsable de la unidad curricular (docente): Dra. Adriana Marotta,
Instituto: InCo

Departamento: Concepción de Sistemas de Información Arancel:

Nº de Créditos: 6

Cupos:

Horas Presenciales: 50

Objetivos:

Presentar a los estudiantes fundamentos sobre calidad de datos e integración de datos. Proveerles conocimientos, herramientas prácticas y tecnologías que les permitan enfrentar el problema de la evaluación de calidad de datos, limpieza y prevención de errores. Presentar los problemas y soluciones posibles relativos a la integración de datos.

Conocimientos previos exigidos:

Conocimientos de Bases de Datos y Programación

Metodología de enseñanza:

Se presentarán en clase los conceptos teóricos y sus aplicaciones prácticas mediante ejemplos. Además, se estudiarán trabajos recientes de investigación en el área. Promediando el curso se presentará a los estudiantes un trabajo obligatorio, el cual irán resolviendo en clase con apoyo de los docentes, y en parte como trabajo domiciliario.

- Horas clase (teórico-práctico): 30
- Horas clase (práctico):
- Horas clase (laboratorio): 10
- Horas consulta: 10
- Horas evaluación:
 - Subtotal horas presenciales: 50
- Horas estudio:
- Horas resolución ejercicios/prácticos:
- Horas proyecto final/monografía: 40
 - Total de horas de dedicación del estudiante: 90

Forma de evaluación:

- Realización de un proyecto donde se aplican las técnicas vistas en clase.
- Defensa oral del trabajo realizado.

Temario:

- Introducción
 - Conceptos de calidad de datos
 - Problemas y consecuencias de la mala calidad
 - Necesidades en distintas áreas de aplicación
 - Gestión de calidad de datos
- Dimensiones y métricas de calidad
 - Multi-dimensionalidad de la calidad
 - Metamodelo de calidad
 - Estudio de un conjunto de dimensiones y métricas
 - Agregaciones según datos y según aspectos de calidad
- Evaluación de calidad de datos
 - Data Profiling
 - Modelo de calidad
 - Requerimientos de calidad
 - Metadatos de calidad
- Mantenimiento de la calidad de datos
 - Integración de datos
 - Limpieza de datos
 - Prevención
 - Monitoreo y roles

Bibliografía:

- Carlo Batini, Monica Scannapieco. Data and Information Quality. Springer, 2016. ISBN: 978-3-319-24104-3.
- Peter Christen. Data Matching - Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection. Springer, 2012. ISBN: 978-3-642-31163-5.
- X. L. Dong y D. Srivastava. Big Data Integration. Morgan & Claypool Publishers, 2015. ISBN: 978-1-62705-224-5

asignatura nº 04 : Visualización Científica

Responsable de la unidad curricular (docente): Dr. Eduardo Fernández

Instituto: InCo

Departamento: Centro de Cálculo

Arancel:

Nº de Créditos: 4

Cupos:

Horas Presenciales: 33

Objetivos:

Introducir en el área de la visualización de información, en especial información cuantitativa. Brindar un recorrido por los principales elementos a considerar al realizar visualización científica, así como una práctica básica en el uso de algunas herramientas informáticas de visualización.

Conocimientos previos exigidos:

Conocimientos de programación

Metodología de enseñanza:

Se presentarán en clase los conceptos teóricos del curso, se estudiarán y presentarán trabajos recientes de investigación en el área, y ejemplos prácticos. Se realizará experiencias de laboratorio en el manejo de Octave ParaView y otros paquetes de visualización científica. Se entregarán ejercicios de visualización utilizando los paquetes mencionados, como trabajo domiciliario. La resolución de dichos ejercicios se realizará de forma individual y se presentará un informe con los resultados obtenidos. A final del curso se realiza una prueba teórica. La nota final surge de la evaluación de ambas actividades.

- ! Horas clase (teórico): 20
- ! Horas clase (práctico): 4
- ! Horas clase (laboratorio): 3
- ! Horas consulta:
- ! Horas evaluación: 3
 - o Subtotal horas presenciales: 30
- ! Horas estudio: 10
- ! Horas resolución ejercicios/prácticos: 20
- ! Horas proyecto final/monografía:
 - o Total de horas de dedicación del estudiante: 60

Forma de evaluación

Realización de un informe final con los resultados obtenidos de los ejercicios prácticos.
Prueba escrita final.

Temario:

- Introducción
 - Objetivos del curso
 - Motivaciones para la visualización de la información.
 - Etapas de la visualización.
 - Símbolos sensoriales vs arbitrarios.
 - Modelo de Procesamiento Perceptual.
 - Tipos de datos a visualizar.
 - Bases Naturales de la percepción visual.
 - Memoria.
 - Propiedades preatentivas.
 - Procesamiento Postatentivo.
 - Principios de la Gestalt.
 - Tipos de visualización de datos.
 - Visualización de datos univariados y bivariados y multivariados.
 - Visualización de datos jerárquicos.
 - Elementos del proceso de visualización.
 - Producción de datos de entrada.
 - Preparar datos a visualizar (Análizar/filtrar/reformatear).
 - Técnicas de modelado/computación gráfica (modelado geométrico/rendering).
 - Visualizar e interpretar.
 - Herramientas de visualización.
 - Octave, Matlab.
 - ParaView.
 - Gephi.
-
- Preparar datos a visualizar (Análizar/filtrar/reformatear).
 - Técnicas de modelado/computación gráfica (modelado geométrico/rendering).
 - Visualizar e interpretar.
-
- Herramientas de visualización.
 - Octave, Matlab.
 - ParaView.
 - Gephi.

Bibliografía:

- Information Visualization, Perception for design, Colin Ware, Morgan Kaufmann, ISBN 978-0-12-381464-7, 2012.
- The Visualization Handbook, Christopher Johnson and Charles Hansen (eds.), Elsevier, ISBN 0-12-387582-X, 2005.
- Introduction to Information Visualization, Ricardo Mazza, Springer, ISBN: 978-1-84800-218-0, 2009.
- Visual Computing, Scientific Visualization and Imaging Systems, Fabiana Rodrigues Leta, Springer, ISBN: 978-3-642-55130-7,

asignatura nº 05 : Aprendizaje automático aplicado

Responsable de la unidad curricular (docente): Dr. Inq. Guillermo Moncecchi.

Otros docentes: Msc. Luis Chiruzzo, Msc. Diego Garat, Msc. Mathias Etcheverry

Instituto: InCo

Departamento: Grupo Procesamiento de Lenguaje Natural

Arancel:

Nº de Créditos: 6

Cupos:

Horas Presenciales: 45

Objetivos:

El objetivo del curso es mostrar una metodología para la aplicación efectiva de diferentes métodos de aprendizaje automático, incluyendo métodos de gran desarrollo en la actualidad, como redes neuronales profundas. Se introducirán las etapas típicas de modelado, entrenamiento, y evaluación. Se utilizarán bibliotecas de código abierto para el lenguaje de programación Python a los efectos de aplicar el conocimiento adquirido sobre diferentes conjuntos de datos disponibles públicamente.

Conocimientos previos exigidos: conocimientos sobre métodos de Aprendizaje Automático

Metodología de enseñanza:

- ! Horas clase (teórico): 21
- ! Horas clase (práctico):
- ! Horas clase (laboratorio): 21
- ! Horas consulta:
- ! Horas evaluación: 3
 - o Subtotal horas presenciales: 45
- ! Horas estudio: 15
- ! Horas resolución ejercicios/prácticos:
- ! Horas proyecto final/monografía: 30
 - o Total de horas de dedicación del estudiante: 90

Forma de evaluación:

Informe de proyecto final (en grupo).

Prueba escrita individual presencial.

Temario:

- Conceptos básicos de Aprendizaje Automático. Tipos de aprendizaje. Sesgo y varianza.
- Técnicas para preprocesamiento de datos y extracción de atributos. Atributos discretos y continuos. Atributos faltantes. Experimentación con bibliotecas para análisis y procesamiento de datos en Python (por ejemplo, Pandas, Scikit-learn).
- Aprendizaje supervisado aplicado: Metodología para el aprendizaje. Medidas de evaluación: accuracy, precision, recall, medida-F. Validación. Evaluación cruzada. Experimentación con métodos básicos para clasificación y regresión. Se utilizarán bibliotecas de Python como Scikit-learn.
- Aprendizaje no supervisado aplicado. Clustering. Reducción de dimensiones. Medidas de performance. Se utilizarán bibliotecas de Python como Scikit-learn.
- Modelado y aplicación de redes neuronales profundas con Python. Se trabajará con bibliotecas como Tensorflow y Keras.

Bibliografía:

- Learning scikit-learn: Machine Learning in Python. Raúl Garreta y Guillermo Moncecchi. Packt Publishing. ISBN-10: 1783281936. 2013.
- Documentación de las bibliotecas de código abierto en Python: Scikit-learn, Pandas, Tensorflow y Keras.
- Machine Learning Yearning. Andrew Ng (draft versions).

Bibliografía complementaria:

- Machine Learning. Tom Mitchell. McGraw-Hill. ISBN 0070428077. 1997.
- Deep Learning. Ian Goodfellow, Yoshua Bengio, y Aaron Courville. Deep Learning. MIT Press. ISBN-10: 0262035618. 2016.

asignatura nº 06 : Estadística Multivariada Computacional

Responsable de la unidad curricular (docente): Dr. Mathias Bourel

Instituto: IMERL

Departamento: LPE

Arancel:

Nº de Créditos: 6

Cupos:

Horas Presenciales: 60

Objetivos:

Introducir los aspectos metodológicos de algunas técnicas de la estadística multivariada y sus modernas variantes. Aplicar dichas técnicas a conjuntos de datos reales e interpretar los resultados obtenidos.

Acercar el estudiante al empleo de los paquetes para el análisis estadístico de datos disponibles en el ambiente de desarrollo de software libre R (<http://www.r-project.org/>).

Conocimientos previos exigidos: introducción a la probabilidad y estadística, álgebra lineal, cálculo matricial.

Metodología de enseñanza:

- ! Horas clase (teórico): 22
- ! Horas clase (práctico): 6
- ! Horas clase (laboratorio): 20
- ! Horas consulta: 10
- ! Horas evaluación: 2
 - o Subtotal horas presenciales: 60
- ! Horas estudio: 12
- ! Horas resolución ejercicios/prácticos: 8
- ! Horas proyecto final/monografía: 10
 - o Total de horas de dedicación del estudiante: 90

Forma de evaluación:

Resolución de ejercicios propuestos durante el curso y trabajo final a entregar con defensa oral.

Temario:

- Introducción al análisis estadístico de datos multivariados con R.
- Regresión lineal simple y múltiple.
- Análisis de componentes principales. Análisis factorial.
- Análisis discriminante.
- Técnicas de segmentación de datos (cluster análisis).
- Escalamiento multidimensional.

Bibliografía:

- Modern Multivariate Statistical Techniques, Alan Julian Izenman, Springer, ISBN: 978-0-387-78188-4, 2008.
- Analisis de datos multivariantes, Daniel Peña, Mac Graw Hill, 2002
- Multivariate Statistics: High-Dimensional and Large-Sample Approximations, Yasunori Fujikoshi, Vladimir V. Ulyanov, Ryoichi Shimizu, Wiley, ISBN: 978-0-470-41169-8, 2010.
- Multivariate Statistical Inference and Applications, Alvin C. Rencher, Wiley, ISBN: 0-471-57151-2, 1998.
- A First Course in Multivariate Statistics, Bernhard Flury, Springer, ISBN: 978-0-387-98206-9, 1997.
- A Handbook of Statistical Analyses Using R, 2nd edition, Brian S. Everitt y Torsten Hothorn, Chapman & Hall/CRC, ISBN: 978-1-4200-7933-3, 2010.
- An introduction to Statistical Learning with Applications in R, G. James, D. Witten, T. Hastie, R. Tibshirani, Springer, 2013.

asignatura nº 07 : Análisis de Textos

Responsable de la unidad curricular (docente): Dra. Ing. Aiala Rosá

Otros docentes: Dra. Ing. Dina Wionsever, Msc Mathias Etcheverry, Msc. Juan José Prada

Instituto: InCo

Departamento: Grupo Procesamiento de Lenguaje Natural Arancel:

Nº de Créditos: 4

Cupos:

Horas Presenciales: 30

Objetivos:

El objetivo del curso es introducir los conceptos, los métodos y las aplicaciones principales del área Procesamiento de Lenguaje Natural, haciendo énfasis en problemas y enfoques actuales, basados principalmente en técnicas de Aprendizaje Automático. Se presentará un panorama completo de los desafíos que involucra el área y se experimentará con las herramientas y los recursos habituales, incluyendo alguna práctica orientada a problemas de dimensiones reales de gran desarrollo en la actualidad, como Análisis de Sentimientos y Extracción de Información en Redes Sociales.

Conocimientos previos exigidos: Programación,

Conocimientos previos recomendados: Lenguajes Formales, Aprendizaje Automático, Probabilidad y Estadística, Álgebra.

Metodología de enseñanza:

- Horas clase (teórico): 15
- Horas clase (práctico): -
- Horas clase (laboratorio): 12
- Horas consulta:-
- Horas evaluación: 3
 - Subtotal horas presenciales: 30
- Horas estudio: 10
- Horas resolución ejercicios/prácticos: -
- Horas proyecto final/monografía: 20
 - Total de horas de dedicación del estudiante: 60

Forma de evaluación:

- Informe sobre el proyecto final (grupal).
- Prueba escrita individual presencial.

Temario:

- Introducción al Procesamiento de Lenguaje Natural
 - Conceptos generales, niveles clásicos de análisis.
 - Aplicaciones más comunes del área: traducción automática, resumen automático, extracción de información, análisis de textos de prensa y de redes sociales.
 - Herramientas y recursos usuales: analizadores lingüísticos, corpus.
 - Evaluación.
- Conceptos generales de gramática del español:
 - Definiciones básicas: oración y sintagma, categorías gramaticales (léxicas y sintagmáticas), funciones sintácticas y semánticas.
 - Análisis gramatical básico de oraciones.
- Tratamiento previo de los textos:
 - Obtención de textos.
 - Pre-procesamiento: limpieza, tokenización, lematización, stemming, pos-tagging.
 - Análisis sintáctico y semántico: recursos disponibles, aplicaciones y problemas.
 - Recursos léxicos.
 - Corpus anotados.
- Métodos clásicos y actuales para el PLN:
 - Métodos simbólicos (reglas).
 - Métodos de aprendizaje automático basados en atributos.
 - Modelos de datos de desarrollo reciente: Representaciones distribuidas de palabras y aprendizaje profundo.
- Problemas actuales del área:
 - Análisis de sentimientos.
 - Extracción de información.

Bibliografía:

Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics, 2nd edition. Daniel Jurafsky, and James H. Martin. Prentice-Hall. ISBN-10: 0131873210. 2009.

Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics, 3rd edition draft. Daniel Jurafsky, and James H. Martin. <https://web.stanford.edu/~jurafsky/slp3/>. 2017.

Bibliografía complementaria

Gramática del español para maestros y profesores del Uruguay. Ángela Di Tullio y Marisa Malcuori. ANEP, Prolee. ISBN 978-9974-688-79-7. 2012.

Machine Learning. Tom Mitchell. McGraw-Hill. ISBN 0070428077. 1997.

Learning scikit-learn: Machine Learning in Python. Raúl Garreta y Guillermo Moncecchi. Packt Publishing. ISBN-10: 1783281936. 2013. ISBN-13: 978-1-78328-193-6. 2016.

asignatura nº 08 : Análisis de datos en Redes

Responsable de la unidad curricular (docente): Dr. Pablo Rodríguez Bocca

Instituto: InCo

Departamento: Investigación Operativa

Arancel:

Nº de Créditos: 4

Cupos:

Horas Presenciales: 42

Objetivos:

Una red es un concepto más amplio que un grafo, y en general se utiliza para modelar sistemas complejos como las interacciones entre personas, proteínas, neuronas, computadoras, elementos de comunicación, sistemas de transporte, etc. El análisis de una red incluye el modelado de su estructura, la visualización, y el cálculo de medidas, todos muy útiles para extraer nuevo conocimiento de estos sistemas.

El objetivo de esta unidad curricular es brindar una visión general del análisis de redes, describiendo sus fundamentos, técnicas, y herramientas para su aplicación práctica.

La unidad curricular presenta un equilibrio entre las visiones teórica y práctica. En lo teórico, se incluyen los principales conceptos del análisis de redes, a saber: ¿Qué es el análisis de datos? Repaso de los principales conceptos de la matemática en teoría de grafos. Modelado de grafos (redes aleatorias). Análisis desde los nodos (centralidad, redes egocéntricas,...). Análisis desde los enlaces (detección de comunidades). Dinámica y difusión en grafos.

En la parte práctica: se brindan razones para analizar redes, y los principales usos de la disciplina. ¿Cómo se analizan las redes?, ¿Qué herramientas se utilizan? Casos de aplicaciones reales.

Al finalizar el curso, el estudiante podrá diseñar y realizar investigación relacionada al análisis en redes, incluyendo: la recolección de datos, el análisis, y la documentación de resultados en formato científico (publicación académica).

Conocimientos previos exigidos: introducción a la probabilidad y estadística, matemática discreta, cálculo matricial.

Metodología de enseñanza:

- ! Horas clase (teórico): 25
- ! Horas clase (práctico):
- ! Horas clase (laboratorio): 15
- ! Horas consulta:
- ! Horas evaluación: 2
 - o Subtotal horas presenciales: 42
- ! Horas estudio: 8
- ! Horas resolución ejercicios/prácticos: 10
- ! Horas proyecto final/monografía:
 - o Total de horas de dedicación del estudiante: 18

Forma de evaluación:

Para la evaluación se tendrá en cuenta:

- 40% entrega de trabajos domiciliarios.
- 60% examen final

Temario:

- 1. Motivación
- 2. Matemática de las redes: teoría de grafos
- 3. Medidas en redes
- 4. Modelos de redes
- 5. Procesos sobre redes
- 6. Tendencias en el análisis de redes

Bibliografía:

- [NE] Mark Newman. Networks: An Introduction. Oxford University Press, Inc. New York, NY, USA 2010. ISBN:0199206651 9780199206650
- [SANDR] Kolaczyk, E.D. and Csardi, G. "Statistical Analysis of Network Data with R". Use R!, Springer New York, 2014. ISBN 9781493909834. (disponible en portal Timbó)

Bibliografía Recomendada:

- [SAND] Kolaczyk, E.D. "Statistical Analysis of Network Data: Methods and Models". Springer Series in Statistics. Springer New York, 2009. ISBN 9780387881461. (disponible en portal Timbó)
- [BA] Albert-László Barabási, Network Science. (disponible online <http://barabasilab.com/networksciencebook/>)
- [NCM] David Easley and Jon Kleinberg. Networks, Crowds, and Markets, Reasoning About a Highly Connected World. Cornell University, New York. Date Published: July 2010. ISBN: 9780521195331 (disponible online)
- [ASN] Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). Analyzing Social Networks. SAGE Publications Limited. ISBN: 9781446247419.

asignatura nº 09 : Modelado y procesamiento de datos no relacionales

Responsable de la unidad curricular (docente): Dra. Lorena Etcheverry

Instituto: InCo

Departamento: Concepción de Sistemas de Información

Arancel:

Nº de Créditos: 6

Cupos:

Horas Presenciales: 60

Objetivos: Durante más de 30 años, los sistemas de bases de datos basados en el modelo relacional dominaron el almacenamiento y la gestión de datos. En la última década, y debido a múltiples razones, esta hegemonía se ha visto desafiada. En particular, las aplicaciones sobre la Internet, con billones de usuarios distribuidos mundialmente generando datos constantemente, configuran un nuevo escenario y plantean problemas que los sistemas de bases de datos relacionales no logran resolver adecuadamente. Como consecuencia, en la actualidad coexisten nuevos modelos y estrategias de gestión de datos. A la hora de diseñar una solución informática resulta imprescindible contar con información y hacer un análisis crítico de cada una de estas estrategias, que permita tomar decisiones fundadas más allá de las modas o tendencias.

El objetivo de esta asignatura es brindar una visión general sobre estos nuevos modelos de gestión de datos, analizando las fortalezas y debilidades de cada uno, y buscando así enriquecer la mirada y el espíritu crítico sobre estas estrategias que llegaron para quedarse. Durante el curso se hará énfasis en los modelos de datos subyacentes y los lenguajes de consulta asociados. Al final del curso el estudiante contará con elementos para decidir cuál es el modelo y la estrategia de gestión de datos que mejor se ajusta a su problema.

Conocimientos previos exigidos:

Conocimientos de Bases de Datos y Programación

Metodología de enseñanza:

- ! Horas clase (teórico): 30
- ! Horas clase (práctico): 20
- ! Horas clase (laboratorio):
- ! Horas consulta: 10
- ! Horas evaluación:
 - o Subtotal horas presenciales: 60
- ! Horas estudio:
- ! Horas resolución ejercicios/prácticos: 30
- ! Horas proyecto final/monografía:
 - o Total de horas de dedicación del estudiante: 90

Forma de evaluación:

Entrega de ejercicios prácticos que podrán ser resueltos en grupo.

Prueba final individual.

Temario:

- ! Introducción:
 - Perspectiva histórica y evolución de los sistemas de bases de datos.
 - El modelo relacional y las propiedades ACID.
- ! Nuevas estrategias de modelado y procesamiento de datos:
 - Bases de datos de documentos
 - Bases de datos de grafos
 - BigData y el ecosistema Hadoop
 - El movimiento NoSQL y las key-value stores
- ! Aspectos técnicos
 - Distribución de datos
 - Modelos de consistencia
 - Modelos de datos y almacenamiento
 - Lenguajes de consulta

Bibliografía:

Next Generation Databases: NoSQL, NewSQL, and Big Data. Guy Harrison, Apress, ISBN 978-1-4842-1330-8, 2015. (disponible en el Portal Timbó)

Concise Guide to Databases: A Practical Introduction. Peter Lake and Paul Crowther, In Undergraduate Topics in Computer Science, Springer-Verlag London, ISBN 978-1-4471-5600-0, 2013. (disponible en el Portal Timbó)

Lectura de artículos seleccionados.

asignatura nº 10 : Estadísticas en series temporales de memoria larga

Responsable de la unidad curricular (docente): Dr. Juan Kalemkerian

Instituto: IMERL

Departamento: LPE

Arancel:

Nº de Créditos: 4

Cupos:

Horas Presenciales: 24

Objetivos:

Los procesos de Ornstein Uhlenbeck Fraccionarios pueden ser usados para modelar series de tiempo de "memoria larga" y han tenido un amplio desarrollo en los últimos 15 años, tanto del punto de vista teórico como del práctico mediante la modelación de diversos fenómenos. Recientemente se han encontrado aplicaciones de estos modelos en finanzas, hidrología, tráfico de internet, genética, geofísica y medicina.

El objetivo del curso es el de introducir al estudiante en el estudio de la modelación práctica de series de tiempo de memoria larga, particularmente mediante los procesos de Ornstein Uhlenbeck Fraccionarios.

Conocimientos previos exigidos: introducción a la probabilidad y estadística.

Metodología de enseñanza:

- ! Horas clase (teórico): 16
- ! Horas clase (práctico): 8
- ! Horas clase (laboratorio):
- ! Horas consulta: 4
- ! Horas evaluación:
 - o Subtotal horas presenciales: 24
- ! Horas estudio: 16
- ! Horas resolución ejercicios/prácticos: 20
- ! Horas proyecto final/monografía:
 - o Total de horas de dedicación del estudiante: 60

Forma de evaluación:

Exposición oral de algún tema planteado en clase, entrega de ejercicios.

Examen oral al finalizar el curso.

Temario:

- Nociones básicas de series de tiempo de memoria larga y memoria corta
- El proceso de Wiener y el puente browniano
- Nociones de series de tiempo ARMA, ARIMA y ARFIMA.
- Procesos de Ornstein Uhlenbeck y procesos de Ornstein Uhlenbeck Fraccionarios (definición, propiedades básicas y modelación).
- Estimación de parámetros de un FOU.
- Procesos de Ornstein Uhlenbeck Fraccionarios de orden p .

Bibliografía:

- Hipel, K. W. & McLeod, A. I. Time Series Modelling of Water Resources and Environmental Systems. Elsevier, Developments in Water Sciences 45, 1994.
- Brockwell, P. J. & Davis, R. A. Introduction to Time Series and Forecasting, Springer Texts in Statistics, 2002.
- Karatzas, I. Shreve, S.E. Brownian Motion and Stochastic Calculus. Springer New York, 1991.

asignatura nº 11 : CURSO DE LA MATERIA “Modelado y procesamiento de grandes volúmenes de datos” DE AL MENOS 6 CRÉDITOS.

ES ELEGIDO POR EL DIRECTOR ACADÉMICO SEGÚN DISPONIBILIDAD DE CURSOS DE POSGRADO y SEGÚN NECESIDAD/INTERÉS DEL ESTUDIANTE.

asignatura nº 12 : CURSO DE LA MATERIA “Técnicas Avanzadas de Análisis de Datos” DE AL MENOS 4 CRÉDITOS.

ES ELEGIDO POR EL DIRECTOR ACADÉMICO SEGÚN DISPONIBILIDAD DE CURSOS DE POSGRADO y SEGÚN NECESIDAD/INTERÉS DEL ESTUDIANTE.

7. INFORMACIONES COMPLEMENTARIAS

Antecedentes del Programa

Año de comienzo de actividades:

Este programa no tiene antecedentes.
Comenzará a funcionar en el primer semestre de 2019.

Otras informaciones pertinentes:

8. SUB-COMISIÓN ACADÉMICA DEL ÁREA

Integrantes: Andrea Delgado, Aiala Rosá, Adriana Marotta, Alberto Pardo, Diego Vallespir, María Eugenia Corti.

Firmas:

Lugar y fecha:

Firmas

9. APROBACIONES PARTICULARES

Fecha de aprobación Comisión/es Instituto/s del Área (o sector equivalente) :

(Nº de expediente y anexar resolución)

Fecha de aprobación Consejo de Facultad de Ingeniería

(Nº de expediente y anexar resolución)

Homologación Comisión Académica Posgrado UdelaR

(Nº de expediente y anexar resolución)

Aprobación por el Consejo Directivo Central

(Nº de expediente y anexar resolución)

10. ANEXOS

Curriculum vitae actualizado de cada docente participante del programa incluyendo:

Cargo docente actual

Estudios y títulos

Experiencia docente universitaria:

cursos de actualización y posgrado, orientación de alumnos, dirección de tesis y título de la misma.

Producción académica: publicaciones, etc. (hasta 5)

Producción profesional creativa: (hasta 5)

Otros méritos de valor académico. (hasta 5)