

Plan de estudios Diploma de Especialización en “Ciencia de Datos”

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.

1. Objetivos

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.

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.

2. Perfil del Egresado

Un científico de datos es aquella persona que aplica con solvencia las metodologías de la ciencia de datos. Cuenta con 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.
- Presentar sus conclusiones, en forma visual y escrita, a los no científicos de datos.

3. Ordenamiento

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 involucran 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.

4. 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.

5. Admisión y 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.

6. 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.

7. Estructura del plan de estudios

Los cursos que serán ofrecidos por este Diploma se organizan en torno a un conjunto de áreas temáticas, que llamaremos materias. Dentro de estas materias se definen temas a cubrir, y asociadas a cada tema se encuentran los cursos. Las materias son:

- Fundamentos de Aprendizaje Automático
- Fundamentos de Probabilidad y Estadística
- Modelado y procesamiento de grandes volúmenes de datos
- Técnicas Avanzadas de Análisis de Datos

En el Anexo se presenta un ejemplo de plan, incluyendo cursos necesarios y opcionales en las diferentes materias.

A continuación se describe cada materia:

Fundamentos de Aprendizaje Automático:

Esta materia introduce al estudiante en el aprendizaje automático, dándole herramientas para modelar y resolver distintos problemas típicos del área. La materia conjuga teoría y

práctica, siendo necesarios conocimientos de base en matemática y programación de computadoras.

Fundamentos de Probabilidad y Estadística:

En la ciencia de datos, la probabilidad y la estadística son sustento para la interpretación de los datos y para la generación de hipótesis sobre los mismos. Esta área de formación incluye varias de las teorías de base necesarias para entender la mayoría de los métodos aplicados en el análisis de datos.

Modelado y procesamiento de grandes volúmenes de datos:

La manipulación de datos digitales es clave en el análisis de datos. El área de formación de procesamiento de datos incluye técnicas para recolectar, consolidar, limpiar e imputar datos de distintas fuentes.

Técnicas Avanzadas de Análisis de Datos:

Dependiendo de la naturaleza de los datos se ha desarrollado matemática y técnicas específicas para analizar cada caso. Por ejemplo la teoría de grafos para analizar datos interrelacionados; técnicas de procesamiento de lenguaje natural para analizar textos; las series de tiempo para estudiar datos con dinámica temporal; y la optimización combinatoria para refinar los modelos extraídos de los datos. El área de formación de áreas aplicadas incluye esas disciplinas y otras utilizadas de forma específica para analizar ciertos datos.

A modo de resumen no exhaustivo, la siguiente tabla presenta temas incluidos en cada materia.

Materia	Temas incluidos
Fundamentos de Aprendizaje Automático	Aprendizaje supervisado Aprendizaje no supervisado Aprendizaje profundo
Fundamentos de Probabilidad y Estadística	Modelado matemático Probabilidad Estadística Análisis multivariado
Modelado y Procesamiento de grandes volúmenes de datos	Calidad de datos Integración de datos de múltiples fuentes Modelos de datos y herramientas Cómputo y tratamiento de la información Arquitecturas distribuidas
Técnicas Avanzadas de Análisis de Datos	Análisis de lenguaje natural Análisis de series de tiempo Análisis de grafos y redes Análisis de señales, imágenes, audio, etc. Robótica Optimización

8. 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.

Anexo: Ejemplo de plan

En este Anexo se incluye un ejemplo de curricula de al menos 60 créditos, con cursos opcionales para el estudiante. Esta lista será actualizada por la SCAPA-informática.

La SCAPA-informática, en conjunto con el estudiante, definirá entre las cursos opcionales. Esto dependerá de la disponibilidad de los cursos, los intereses del estudiante y su formación previa. Por ejemplo para egresados de Ingeniería en Computación se recomiendan cursos avanzados en "Modelado y Procesamiento de grandes volúmenes de datos"; y para los egresados de otras carreras se recomienda mayor profundidad en "Técnicas Avanzadas de Análisis de Datos".

#	Curso	instituto	créditos	materia
1er semestre			20	
1	Técnicas de aprendizaje estadístico	IMERL	6	Fundamentos de Aprendizaje Automático
2	Introducción a la estadística usando software	IMERL	4	Fundamentos de Probabilidad y Estadística
3	Calidad e integración de datos	InCo	6	Modelado y procesamiento de grandes volúmenes de datos
4	Visualización científica	InCo	4	Modelado y procesamiento de grandes volúmenes de datos
2do semestre			20	
5	Aprendizaje automático aplicado	InCo	6	Fundamentos de Aprendizaje Automático
6	Estadística multivariada computacional	IMERL	6	Fundamentos de Probabilidad y Estadística
7	Análisis de textos	InCo	4	Técnicas Avanzadas de Análisis de Datos
8	Análisis de datos en redes	InCo	4	Técnicas Avanzadas de Análisis de Datos
3er semestre			20	
9	Modelado y procesamiento de datos no relacionales	InCo	6	Modelado y procesamiento de grandes volúmenes de datos
10	Estadísticas en series temporales de memoria larga	IMERL	4	Técnicas Avanzadas de Análisis de Datos
11	Un curso en "Modelado y procesamiento de grandes volúmenes de datos"		6	Modelado y procesamiento de grandes volúmenes de datos
12	Un curso en "Técnicas Avanzadas de Análisis de Datos"		4	Técnicas Avanzadas de Análisis de Datos

Ejemplos de cursos opcionales en Modelado y procesamiento de grandes volúmenes de datos:

<i>11a. Recuperación de información y recomendaciones en la Web</i>	<i>InCo</i>
<i>11b. Computación de alta performance</i>	<i>InCo</i>
<i>11c. Data mining y toma de decisiones</i>	<i>InCo</i>
<i>11d. Procesamiento analítico en línea en big data</i>	<i>InCo</i>

Ejemplos de cursos opcionales en Técnicas Avanzadas de Análisis de Datos:

<i>12a. Algoritmos Evolutivos</i>	<i>InCo</i>
<i>12b. Metaheurísticas y Optimización sobre Redes</i>	<i>InCo</i>
<i>12c Fundamentos de Informática Urbana: Análisis y Procesamiento de Datos</i>	<i>InCo</i>