

Plan de estudios

Maestría en Ciencia de Datos Aplicada

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 profesionales de la ciencia de datos. Por tanto, esta maestría 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 (principalmente en probabilidad y estadística) y en informática (principalmente en 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. La maestría 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 informática.

1. Objetivos

Esta maestría en Ciencia de Datos Aplicada 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 esta maestría 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.

Esta maestría 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 profesional de ciencia 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 e integrar los temas de estudio incluidos en la Maestría en un amplio rango de áreas de aplicación, 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 esta maestría 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 de la Maestría en Ciencia de Datos Aplicada; 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 a la maestría.

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 a la Maestría en Ciencia de Datos Aplicada 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 de la Maestría en Ciencia de Datos Aplicada.

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.), y con la realización de una Tesis, de manera de cumplir:

- Un mínimo de 70 créditos en actividades programadas (el crédito es la unidad de medida de la carga de trabajo en la Universidad de la República, y un crédito equivale a quince horas de dedicación por parte del estudiante).
- Las actividades programadas deberán cubrir el mínimo de horas presenciales exigido por la Ordenanza de las Carreras de Posgrado de la Universidad de la República.
- La realización de una Tesis, la cual deberá ser defendida en una exposición oral y pública ante un tribunal designado por la CAP a propuesta de la SCAPA-Informática.
- Todas las actividades programadas deberán contar con alguna forma de evaluación de los conocimientos adquiridos.

La duración prevista para la realización de la Maestría en Ciencia de Datos Aplicada es de 24 meses, con una dedicación estimada entre 15 y 20 horas semanales.

7. Estructura del plan de estudios

Los cursos que serán ofrecidos por esta maestría 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. Tesis

El estudiante deberá contar con un Director de Tesis y, eventualmente, un co-director, que deberá ser nombrado por la CAP a propuesta de la SCAPA-Infomática a lo sumo un año después del ingreso a la Maestría.

El tema de la tesis se definirá en mutuo acuerdo entre el estudiante y el Director de Tesis, debiendo ser comunicado a la SCAPA-Infomática para su validación. El trabajo de tesis tendrá un valor de 40 créditos.

En la elaboración de la Tesis, el estudiante deberá alcanzar el estado del arte y desarrollar un análisis en profundidad del tema elegido. El manuscrito de la Tesis deberá contener una correcta exposición del tema encarado, una discusión de la bibliografía actualizada (preferentemente aquella publicada en conferencias y/o revistas arbitradas y reconocidas en el área del tema de tesis) y deberá incluir los detalles necesarios para la comprensión de los objetivos perseguidos y de los resultados y conclusiones alcanzadas durante el trabajo.

Cuando el Director de Tesis considere que el trabajo ha sido culminado, el mismo se presentará por escrito y se defenderá en una exposición oral y pública ante un tribunal designado por la CAP a propuesta de la SCAPA-Infomática, según lo establecido en el Reglamento General de las Actividades de Posgrado y Educación Permanente de la Facultad de Ingeniería.

9. Título

Cuando el aspirante haya completado los requisitos del programa, iniciará los trámites pertinentes para la expedición del Título "Magíster en Ciencia de Datos Aplicada". 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 70 créditos, con cursos opcionales para el estudiante. Esta lista será actualizada por la SCAPA-informática, y adecuada a las necesidades de formación de cada estudiante.

La SCAPA-informática, en conjunto con el estudiante, seleccionará los cursos opcionales, dependiendo del tema de tesis definido, de los intereses del estudiante y su formación previa, y de la disponibilidad de los cursos. En todos los casos se recomendará al menos un curso de los opcionales en fundamentos relacionado al tema de tesis del estudiante, que aporte en el conocimiento del estado del arte en el tema y a la mejor ejecución del trabajo de tesis.

#	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 y 4to semestre		30	
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” orientado al tema de tesis		6	Modelado y procesamiento de grandes volúmenes de datos
12	Un curso en “Técnicas Avanzadas de Análisis de Datos” orientado al tema de tesis		4	Técnicas Avanzadas de Análisis de Datos
13	Un curso en fundamentos orientado al tema de tesis		10	Depende del tema de tesis

Ejemplos de cursos opcionales en “Modelado y procesamiento de grandes volúmenes de datos” (curso nro. 11):

<i>Recuperación de información y recomendaciones en la Web</i>	<i>InCo</i>
<i>Computación de alta performance</i>	<i>InCo</i>
<i>Datamining y toma de decisiones</i>	<i>InCo</i>
<i>Procesamiento analítico en línea en big data</i>	<i>InCo</i>

Ejemplos de cursos opcionales en “Técnicas Avanzadas de Análisis de Datos” (curso nro. 12):

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

Ejemplos de cursos opcionales en fundamentos orientado al temas de tesis (curso nro. 13):

<i>Modelado y Optimización</i>	<i>InCo</i>
<i>Optimización Bajo Incertidumbre</i>	<i>InCo</i>
<i>Estimación Numérica Monte Carlo</i>	<i>InCo</i>
<i>Gramáticas Formales para el Lenguaje Natural</i>	<i>InCo</i>
<i>Introducción al Procesamiento de Lenguaje Natural</i>	<i>InCo</i>
<i>Análisis Semántico de Lenguaje Natural</i>	<i>InCo</i>
<i>Álgebra lineal numérica</i>	<i>InCo</i>
<i>Computación de propósito general en unidades de procesamiento gráfico</i>	<i>InCo</i>
<i>Simulación a eventos discretos</i>	<i>InCo</i>
<i>Modelos de espacio de estado con aplicaciones a series de tiempo</i>	<i>IMERL</i>

Se consideraran también otros cursos de posgrado, pertinentes al tema de tesis del estudiante, en coordinación con las SCAPAs de referencia de cada caso.