

Plan de estudios

Maestría en Ciencia de Datos y Aprendizaje Automático

La Ciencia de Datos es la disciplina que busca extraer conocimiento de forma sistemática y computacionalmente eficiente a partir de los datos de un dominio. Para esto utiliza principalmente métodos y técnicas de la matemática y la estadística, la computación, y la visualización de datos. Entre las disciplinas relacionadas con la Ciencia de Datos y áreas de aplicación se destacan el aprendizaje automático, el modelado estadístico, el procesamiento de señales, el análisis de datos de alta dimensionalidad, el procesamiento de datos multimedia, el almacenamiento y la recuperación de información en bases de datos distribuidas, su aplicación a la visión artificial y el procesamiento de lenguaje natural, entre otras.

En la Universidad de la República, y en Facultad de Ingeniería en particular, existe una larga trayectoria y amplios antecedentes de diferentes e importantes líneas de investigación vinculadas a la Ciencia de Datos y el Aprendizaje Automático. El Instituto de Matemática y Estadística (IMERL, FIng) tiene en su plantel docentes con dedicación total que trabajan en aspectos fundamentales de la ciencia de datos: por un lado, en estadística -en particular para datos de altas dimensiones- y análisis multivariado, prestando especial atención a los problemas relacionados con el aprendizaje automático y con el análisis de datos en general; por otro lado, en optimización y otros métodos de análisis numérico, que se encuentran presentes en cualquier algoritmo de aprendizaje automático. El Instituto de Ingeniería Eléctrica (IIE, FIng), a través del Departamento de Procesamiento de Señales, tiene una amplia trayectoria de más de dos décadas trabajando en análisis de datos, en particular en el área de procesamiento y análisis de señales, tanto con enfoques basados en modelos como basados en datos. Las aplicaciones tradicionales han sido la transmisión y comunicación de datos, la biología, la biomedicina, la biometría y el agro, entre otras, tanto a partir de señales unidimensionales como de imágenes (microscopía, cámaras, satelitales, radar, etc) o datos multidimensionales. Más recientemente, con el aumento vertiginoso de los volúmenes de datos y de la potencia de cálculo, el aprendizaje automático y los enfoques data driven ha venido cobrando un protagonismo cada vez mayor entre las metodologías para resolver este tipo de problemas. La investigación y producción científica del Instituto de Ingeniería Eléctrica ha ido acompañando esta tendencia. En el Instituto de Computación (INCO, FIng), diferentes grupos de investigación, han estudiado algoritmos de aprendizaje supervisado y no supervisado como parte de los métodos principales para sus dominios, vinculados a diferentes aspectos de la inteligencia artificial como, por ejemplo, robótica, procesamiento de lenguaje natural, o investigación de operaciones. La larga tradición de investigación en Sistema de Información ha generado asimismo conocimiento en el análisis y organización de grandes volúmenes de datos. La experiencia generada en estos temas ha posibilitado que exista ya un perfil "Inteligencia Artificial" en la Licenciatura en Computación, que incluye materias vinculadas a temas de Ciencia de Datos.

1. Objetivos

Los objetivos de la Maestría en Ciencia de Datos y Aprendizaje Automático son promover la profundización y actualización de los conocimientos en las diferentes disciplinas relacionadas con la Ciencia de Datos y el Aprendizaje Automático. Se busca formar recursos humanos altamente capacitados para la investigación y el desarrollo científico-tecnológico, así como para el ejercicio innovador en la profesión y la resolución de problemas específicos de la industria nacional,

promoviendo el desarrollo de la misma. Asimismo se busca formar profesionales con una actitud crítica ante las nuevas posibilidades y tecnologías, y con la capacidad de aplicarlas a la resolución de las necesidades de la sociedad uruguaya en el área.

2. Perfil del Egresado

El egresado adquiere formación en diferentes disciplinas de la Ciencia de Datos y el Aprendizaje Automático. Esto lo capacita en la recolección, depurado, modelado, representación y análisis de datos para resolver problemas como los que se presentan en la industria o la academia. De igual modo adquiere capacidades para interpretar, visualizar y comunicar los resultados obtenidos. A la vez, es capaz de entender los fundamentos matemáticos y algorítmicos detrás de los métodos de la ciencia de datos y del aprendizaje automático, y comprender la literatura especializada de la disciplina. Por lo tanto, es capaz de proponer y desarrollar nuevos métodos y algoritmos adaptados al tipo de datos y al problema que tenga que resolver.

3. Marco del Plan de Estudios

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), 2006.

4. Ordenamiento

Los posgrados en esta área se encuentran orientados por la Sub-Comisión Académica de Posgrado en el área de Ciencia de Datos y Aprendizaje Automático (SCAPA-CDAA), asesora de la Comisión Académica de Posgrado de la Facultad de Ingeniería (CAP), de acuerdo a las ordenanzas vigentes de la Universidad de la República y la Facultad de Ingeniería.

5. Requisitos de Ingreso

Pueden ingresar a la Maestría en Ciencia de Datos y Aprendizaje Automático quienes posean antecedentes académicos de acuerdo a lo expresado en el Artículo 19 del Reglamento General de las Actividades de Posgrado y Educación Permanente de la Facultad de Ingeniería (RGP-FING).

Los aspirantes deben tener conocimientos sólidos en computación, cálculo, álgebra lineal, probabilidad y estadística. Aquellos aspirantes que a juicio de la SCAPA-CDAA así lo requieran, previo a su admisión al programa, deberán completar actividades de nivelación para asegurar el completo aprovechamiento de las actividades de posgrado.

6. Admisión y selección de los candidatos

El aspirante debe solicitar su ingreso a la SCAPA de acuerdo a los requisitos dispuestos en el Artículo 21 de la RGP-FING, la documentación que permita analizar sus antecedentes académicos y profesionales. La SCAPA-CDAA recomendará a la CAP-FING la aceptación de la solicitud si a su juicio el aspirante presenta méritos suficientes. De ser aceptado en el programa, la SCAPA-CDAA le asignará al postulante un Director Académico, con quien deberá proponer un plan de actividades primario en un plazo no superior a seis meses desde el momento de su inscripción.

7. Organización de la formación

La formación se organiza en un conjunto de actividades definidas en la Actividad Programada y la realización de una Tesis. El estudiante debe reunir un mínimo de 65 créditos en la Actividad Programada, de los cuales 57 créditos deben corresponderse a cursos o estudios dirigidos, cumpliendo con los mínimos de cada una de las áreas de formación definidas más adelante. Un crédito equivale a 15 horas de trabajo del estudiante, tal como se define en la Ordenanza de las Carreras de Posgrado de la Universidad de la República.

En un plazo no mayor a un año, el estudiante, en acuerdo con su Director Académico, debe presentar a la SCAPA un tema de Tesis, y un candidato para desempeñarse como Director de Tesis. Este Director de Tesis debe acreditar una trayectoria relevante en el tema de tesis propuesto por el estudiante. La tesis de maestría corresponde a un trabajo de 45 créditos.

Actividad programada

La actividad programada se encuentra constituida por cursos, pasantías en el país o en el extranjero, en instituciones universitarias, centros de investigación o empresas representativas del área, así como por otras actividades aprobadas por la el Director Académico, la SCAPA-CDAA y la CAP. El contenido de la actividad programada debe equilibrarse de manera que el maestrando profundice su formación básica y adquiera conocimientos acordes con la tesis a realizar.

La actividad programada se divide en dos Grupos de Áreas de Formación: el Grupo de Áreas de Formación Básica y el Grupo de Áreas de Formación Específica. Es esperable que el estudiante realice al menos un curso en cada una de las áreas de formación básica y al menos dos cursos en alguna de las áreas de formación específica, cumpliendo con los créditos mínimos de estas áreas.

El número de créditos mínimos de la actividad programada es de 65 créditos.

Grupo de Áreas de Formación Básica

El número de créditos mínimos por concepto de cursos en este grupo de áreas es 42. Las Áreas de Formación Básica son:

- **Ciencia de Datos.** Con esta área se busca presentar un panorama de los métodos, aplicaciones y herramientas que se utilizan en la Ciencia de Datos y el Aprendizaje Automático, así como de los aspectos éticos y de privacidad de los datos. También se abordan aspectos vinculados a la comunicación de resultados. El mínimo de créditos en esta área es 10.
- **Estadística.** En esta área se busca desarrollar el pensamiento crítico sobre el análisis estadístico de datos en general así como la comprensión de modelos actuales particulares para la realización de dicho análisis. Para ello se desarrollarán las nociones básicas para la exploración inicial de los datos a través de cálculo de estadísticos y su visualización que permitan reconocer patrones y detectar anomalías. Además se brindarán las herramientas necesarias para construir, ajustar y validar modelos acordes a los datos. Éstos temas se podrán abordar en asignaturas que incluyan temas de estadística paramétrica y no paramétrica, estadística descriptiva, estadística inferencial y estadística en espacios de alta dimensión, entre otros. El mínimo de créditos en esta área es 8.
- **Aprendizaje Automático.** En esta área se cubre las técnicas y algoritmos que permiten que las computadoras *aprendan*, esto es, cómo crear programas capaces de mejorar su performance a partir de información dada en forma de ejemplos. Entre los temas a abordar

se encuentran: el aprendizaje supervisado, no supervisado, por refuerzos, etc. El mínimo de créditos en esta área es 8.

- **Gestión de datos e información.** Esta área cubre conceptos y herramientas para el manejo eficiente de datos complejos, incluyendo su almacenamiento, búsqueda, limpieza e integración. La complejidad de los datos no sólo refiere a su volumen, sino que puede responder a otros factores como su naturaleza, su estructura, o su dimensionalidad, por mencionar algunos. El mínimo de créditos en esta área es 8.
- **Optimización.** Prácticamente la totalidad de los problemas que se presentan en ciencia de datos y aprendizaje automático se traducen en la resolución de un problema de optimización. Esta área cubre las técnicas que permiten encontrar o aproximar la solución a estos problemas planteados en diferentes ámbitos, como por ejemplo métodos de descenso por gradiente, métodos de Lagrangeano y otros métodos de optimización con restricciones, así como métodos estocásticos y aproximados. Pueden incluirse problemas de optimización lineal y no lineal, convexa y no-convexa, discreta y continua. El mínimo de créditos en esta área es 8.

Grupo de Áreas de Formación Específica

El Grupo de Áreas de Formación Específica permite que el estudiante profundice en alguna de las disciplinas comprendidas en la ciencia de datos y el aprendizaje automático, o en algún dominio de aplicación. El mínimo de créditos en el Área de Formación Específica es 15.

Algunas de estas disciplinas o dominios de aplicación son: Procesamiento de Lenguaje Natural, Visión Artificial, Procesamiento de señales (audio, imágenes, video, ...), Sensado Remoto, Sistemas de comunicación, Genómica, entre otros. Esta enumeración no restringe la existencia de nuevos dominios que serán definidos por la coherencia de los cursos realizados; esta coherencia será revisada por el Director Académico y la SCAPA-CDAA.

Tesis

El trabajo de tesis tendrá una dedicación correspondiente a 45 créditos. El tema de tesis se definirá de mutuo acuerdo entre el estudiante y el Director de Tesis, y estará sujeto a la aprobación de la SCAPA-CDAA. Los aspectos vinculados a la culminación y defensa de la tesis se regirán por la RGP-FING.

8. Título

Una vez que el aspirante haya completado su plan de trabajo, incluyendo la actividad programada oportunamente aprobada, y haya defendido su tesis con aprobación del tribunal, la SCAPA-CDAA notificará a la CAP, quien propondrá al Consejo de la Facultad el otorgamiento del título de “Magíster en Ciencia de Datos y Aprendizaje Automático”.

Anexo 1. Posibles implementaciones y perfiles

El curso de *Introducción a la Ciencia de Datos* no existe en la oferta actual de cursos de posgrado de Facultad de Ingeniería. Este curso busca cubrir la formación en el área de formación básica de Ciencia de Datos para todos los estudiantes del programa. Se busca presentar un panorama general de los métodos, aplicaciones y herramientas que se utilizan en la Ciencia de Datos y el Aprendizaje Automático; también tiene el objetivo de crear un lenguaje común entre estudiantes provenientes de diferentes disciplinas.

Semestre	Asignatura	Créditos	Área
1	Introducción a la Ciencia de Datos (a crear)	10	Ciencia Datos
1	Estadística Multivariada Computacional	10	Estadística
1	Bases de Datos No Relacionales	10	GDI
2	Tratamiento de Imágenes por Computadora	10	Específica
2	Aprendizaje Profundo para Visión Artificial	8	Específica
2	Reconocimiento de Patrones	12	Aprendizaje
3	Teoría y Algoritmia de Optimización	10	Optimización
Total		70	

Semestre	Asignatura	Créditos	Área
1	Introducción a la Ciencia de Datos (a crear)	10	Ciencia Datos
1	Estadística Multivariada Computacional	10	Estadística
1	Bases de Datos No Relacionales	10	GDI
2	Introducción al Procesamiento de Lenguaje Natural	12	Específica
2	Procesamiento Semántico del Lenguaje Natural	8	Específica
2	Reconocimiento de Patrones	12	Aprendizaje
3	Teoría y Algoritmia de Optimización	10	Optimización
Total		72	

Semestre	Asignatura	Créditos	Área
1	Introducción a la Ciencia de Datos (a crear)	10	Ciencia Datos
1	Estadística Multivariada Computacional	10	Estadística
1	Computación Distribuida y Procesamiento de Grandes Volúmenes de Datos	10	GDI
2	Estimación y Predicción en Series Temporales	10	Específica
2	Aprendizaje Automático	12	AA
2	Análisis de Datos en Redes	5	Específica
3	Teoría y Algoritmia de Optimización	10	Optimización
Total		67	