



DIPLOMA EN DATA SCIENCE

I. Descripción

En la actualidad es de gran interés en los ambientes industriales, organizacionales y académicos el uso de herramientas interdisciplinarias para el tratamiento de datos de forma tal que permita descubrir la información valiosa que allí pueda ser conseguida y que apoye a la toma de decisiones de forma inteligente. Así, la Ciencia de datos (Data Science) emerge como esa combinación exitosa de métodos y técnicas que incluyen herramientas matemáticas, estadísticas, computacionales, inteligencia artificial y comportamientos sociales y organizacionales que han brindado soluciones de gran valor en los entornos desarrollados.

El Diploma en Data Science, que ofrece la Universidad Técnica Federico Santa María (UTFSM) a través de su Departamento de Informática, es un programa de continuidad de estudios para profesionales, enfocado en brindar conocimientos sobre la ciencia de datos, sus técnicas, metodologías, herramientas y aplicaciones, habilitando a los participantes para reconocer las oportunidades de incorporar el uso de herramientas de Data Science en sus áreas de interés.

II. Objetivo del Programa

Proveer una formación práctica, actualizada y especializada en ciencia de datos, dirigida a profesionales interesados en perfeccionar sus habilidades para mejorar los procesos decisionales de una organización o crear oportunidades de negocios a partir de fuentes heterogéneas de datos, usando para esto avances recientes en Data Science y gestión de datos masivos.

El diplomado hace un fuerte énfasis en la aplicación de los métodos estudiados, a través de demostraciones prácticas usando Python y jupyter notebooks, y a través de dos módulos de integración de contenidos.

III. Dirigido a

Podrán postular al programa profesionales de cualquier área, que buscan formarse adquiriendo conocimientos sobre la ciencia de datos, sus técnicas, herramientas y aplicaciones, habilitándolos para reconocer las oportunidades de incorporar estas herramientas en sus áreas de interés. Entre los posibles participantes se pueden destacar:

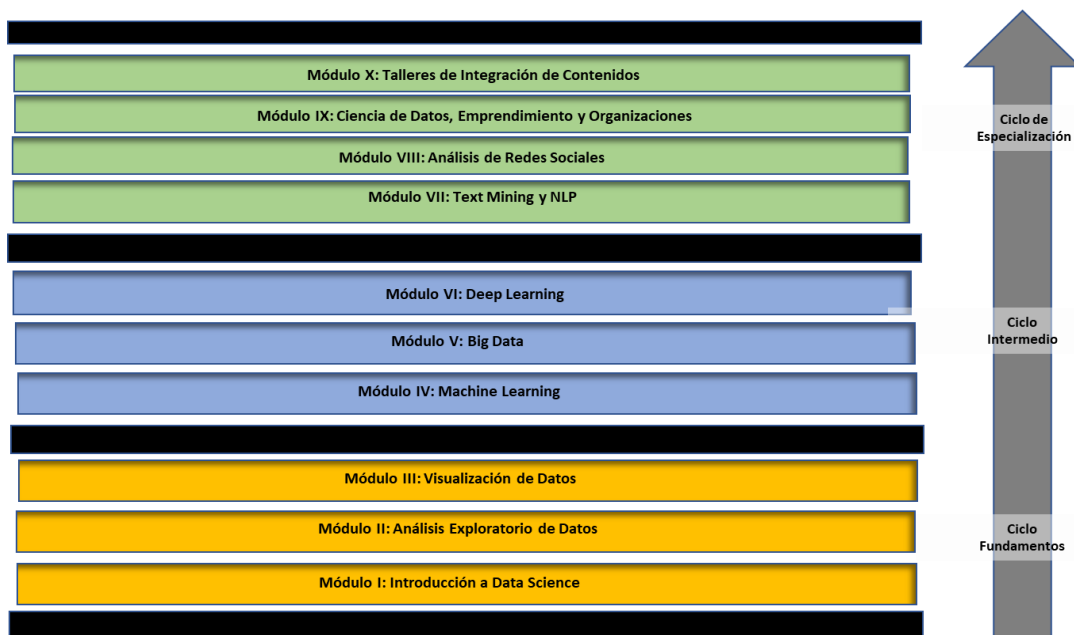


- Gerentes o directivos de empresas, organismos públicos o privados que deseen conocer sobre ciencias de datos y evaluar la posibilidad de integrar su utilización para mejorar sus procesos.
- Analistas de negocio y de información.
- Arquitectos y desarrolladores de sistemas de información y de software.
- Responsables de infraestructura y servicios TIC.
- Consultores y auditores TIC.
- Profesionales en general interesados en conocer las potencialidades y oportunidades de uso de la ciencia de datos.

IV. Plan de Estudios

El Diploma estructura tres ciclos, con un total de 120 horas, a través de 10 módulos de 12 horas, estructurados en un ciclo de fundamentos, uno intermedio y un ciclo de especialización.

El Ciclo de fundamentos está constituido por tres módulos (Introducción a Data Science, Análisis Exploratorio de Datos y Visualización de Datos). El Ciclo Intermedio incluye tres módulos (Machine Learning, Big Data y Deep Learning) El Ciclo de Especialización considera cuatro módulos adicionales (Text Mining y NLP, Análisis de Redes Sociales, Ciencia de Datos, Emprendimiento y Organizaciones y Talleres de Integración de Contenidos). Los talleres se irán modificando en cada oportunidad que se realice el Diploma.





1. Ciclo de Fundamentos

Módulo I: Introducción a Data Science (12 horas)

- Introducción a la Ciencia de Datos, Problemática y Perspectivas.
- Tipos de Datos. Datos estructurados y no estructurados.
- Veracidad de datos, privacidad y seguridad.
- Aspectos éticos y sociales.
- Discusión de casos.
- Ejemplos en Jupyter Notebooks.

Módulo II: Análisis Exploratorio de Datos (12 horas)

- Medidas de Similaridad y Distancia.
- Elementos de Teoría de Probabilidades.
- Análisis de Datos Univariados. Medidas de Tendencia y Dispersión.
- Detección de Datos Atípicos.
- Análisis de Datos Multivariados. Matriz de Correlaciones.
- Análisis de Varianza Intra e Varianza Inter.
- Estimación Paramétrica Clásica.
- Regresión Lineal.

Módulo III: Visualización de Datos (12 horas)

- Visualización de datos y su relación a la cognición humana.
- Qué, por qué y cómo visualizar.
- Elementos gráficos y canales visuales.
- Herramientas de visualización de datos.
- Métodos centrados en usuarios para diseño, prototipado y evaluación.
- Revisión de casos prácticos.

2. Ciclo Intermedio

Módulo IV: Machine Learning (12 horas)

- Introducción al Aprendizaje Supervisado.
- Árboles de Clasificación y Regresión
- Máquinas de Vectores de Soporte (SVMs).
- Ensamblados: Bagging, Boosting, Random Forest.
- Métodos de Evaluación y Selección de Modelos.
- Introducción al Aprendizaje No-Supervisado.

Módulo V: Big Data (12 horas)

- Introducción a la computación distribuida y sistemas para Big Data
- Sistemas de archivos distribuidos para Big Data (BigTable, Cassandra, MongoDB).
- Sistemas de procesamiento para Big Data (Map-reduce, Storm, Spark).
- Sistemas de comunicación para Big Data (Kafka).



Módulo VI: Deep Learning (12 horas)

- Introducción y Redes Feed-forward.
- Entrenamiento de Redes Profundas: Backpropagation y Dropout.
- Redes Neuronales Convolucionales.
- Redes Neuronales Recurrentes.
- Entrenamiento No-supervisado de Redes Neuronales.

3. Ciclo de especialización

Módulo VII: Text Mining y NLP (12 horas)

- Extracción de Información (Keywords, Collocations, Entidades Nombradas).
- Recuperación de Información Clásica (TF-IDF, BM25).
- Representaciones de Texto (BOW, Word2Vec, CBOW, GloVE, Doc2Vec).
- Aplicaciones: Neural IR, Clasificación de Contenidos, Detección de Tópicos, Sentiment Analysis, Traducción.

Módulo VIII: Análisis de Redes Sociales (12 horas)

- Modelos de Redes Sociales
- Medidas de Centralidad, Influencia y Robustez.
- Scrapping de Redes Sociales
- Evolución de Redes Sociales

Módulo IX: Ciencia de Datos, Emprendimiento y Organizaciones (12 horas)

- Rol de la Ciencia de Datos para las Organizaciones.
- Ciencia de Datos y Emprendimiento.
- Discusión de Casos de éxito.

Módulo X: Talleres de Integración de Contenidos (12 horas)

- Taller de Aplicaciones I. Ejemplo: Análisis de Sentimiento en Tweets.
Taller de Aplicaciones II. Ejemplo: Sistema Recomendador en Facebook.

V. Metodología

El programa combina horas presenciales teóricas y prácticas, realización de trabajos grupales y desarrollo de un proyecto final de integración de los contenidos consistente en el diseño conceptual o un caso de estudio con implementación de una aplicación de Data Science en su área de interés.

El equipo de profesores está compuesto por un seleccionado grupo de académicos y profesionales, con sólidos conocimientos y una amplia experiencia en el campo de la ciencia de datos, lo que permitirá brindar a los participantes diferentes visiones y aportes en Data Science y formar sus propios criterios sobre las potencialidades y oportunidades de su utilización en sus áreas de interés.



VI. Organización

Dirección Académica del Programa

- Dr. Ricardo Ñanculef Alegría, UTFSM, Chile (jnancu@inf.utfsm.cl)

Dirección Ejecutiva del Programa

- Dr. Francklin Rivas Echeverría, UTFSM, Chile (firivas@inf.utfsm.cl)

Relatores asociados al Programa

- Héctor Allende
- José Luis Martí
- Marcelo Mendoza
- Andrés Moreira
- Ricardo Ñanculef
- Claudia López
- Erika Rosas
- Carlos Valle

NOTA: La dirección del Programa se reserva el derecho de cambiar algún académico en caso de fuerza mayor.

VII. Información general

Horario: viernes 15:30-21:30 horas; sábado 08:30-14:30 horas.

Lugar: Campus San Joaquín, Universidad Técnica Federico Santa María. Vicuña Mackenna 3939, Santiago.
Campus Casa Central Valparaíso, Universidad Técnica Federico Santa María. Avenida España 1680, Valparaíso.

Inicio: abril 2020

Término: julio 2020

Costo del programa

Derecho a reserva: 2 UF

Valor diploma: 105 UF (103 UF considerando la reserva)

NOTA: El programa se realizará siempre y cuando se complete con el mínimo de participantes, lo que aplica para cada uno de los lugares propuestos de manera independiente. Derecho a reserva sólo se reembolsa en caso de no realizarse el programa.



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA
Departamento de Informática

VIII. Contacto y consultas

Ivonne Barra

Departamento de Informática
Universidad Técnica Federico Santa María
Fono: 2232028200
Email: sec-postitulos@inf.utfsm.cl

Carolina Leal

Departamento de Informática
Universidad Técnica Federico Santa María
Fono: 322654424
Email: vinculacion@inf.utfsm.cl

Departamento de Informática
Universidad Técnica Federico Santa María
www.inf.utfsm.cl