



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Departamento de Informática

Diploma en Data Science

1. Antecedentes Generales

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.

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.





UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Departamento de Informática

Audiencia

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 y con dominio del uso de herramientas computacionales de programación.

2. Plan de Estudios

El Diploma se estructura tres ciclos (Fundamentos, Intermedio y Especialización) con un total de 120 horas, a través de 10 módulos de 12 horas cada uno.

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.

Ciclo 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.





UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Departamento de Informática

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.

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.





UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Departamento de Informática

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.

Ciclo 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





UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Departamento de Informática

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: Seminario de ejemplos y aplicaciones de Inteligencia Artificial (12 horas)

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

Modalidad de Clases

El Diploma en Data Science será con modalidad online basado en la interacción directa (presencial-remota) con los relatores del Diploma y trabajo a distancia.

La duración del Diploma es de 120 horas, compuesto de 10 módulos de 12 horas cada uno.

Formato: tres sesiones presenciales-remotas (2 horas cada una) de interacción entre los relatores y los participantes (estas sesiones serán grabadas para que la puedan ver posteriormente (forma asíncrona) aquellos que no pudieron participar, y 4 horas de supervisión de trabajo a distancia.

Se utilizará la plataforma Moodle que posee el Departamento de Informática.

Las clases presenciales-remotas se realizarán la semana indicada en el cronograma anexo (lunes, miércoles y viernes de 18:00 a 20:00 horas).

Igualmente, las actividades que serán evaluadas en cada módulo la tendrán disponibles los participantes a través de la plataforma Moodle y la deberán entregar a más tardar la semana posterior a la interacción con el relator.





UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Departamento de Informática

3. 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)

- **Académicos asociados al Programa**

Ricardo Ñanculef

Doctor en Ciencias de la Ingeniería Informática.

Magíster en Ciencias de la Ingeniería Informática.

Ingeniero Civil Informático.

Francklin Rivas

Doctor en Ciencias Aplicadas.

Magíster Sc. en Ingeniería de Control.

Ingeniero de Sistemas.

Abogado.

Héctor Allende

Doctor Rer. Nat. Statistik.

Master en Estadística Matemática.

Magíster en Matemática.

Ingeniero Civil Electricista.





UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Departamento de Informática

José Luis Martí

Magíster en Ingeniería Informática.

Ingeniero Civil Informático.

Andrés Moreira

Doctor en Modelamiento Matemático.

Postdoc en el laboratorio TIMC.

Postdoc en el Centro de Modelamiento Matemático.

Erika Rosas

Doctora en Informática.

Magíster en Ciencias de la Ingeniería Informática.

Ingeniera Civil Informática.

Carlos Valle

Doctor en Ingeniería Informática.

Magíster en Ciencias de la Ingeniería Informática.

Ingeniero Civil Informático.

Cecilia Reyes

Magíster en Ingeniería Industrial.

Ingeniera de Ejecución en Informática.

Ana Muñoz

Doctora en Ciencias Aplicadas.

Master en Ingeniería de Control y Automatización.

Ingeniera de Sistemas.





UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Departamento de Informática

Ignacio Tampe

Ingeniero Civil Informático.

Data & Applied Scientist II en Microsoft.

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

Costo del Programa

Valor total diploma: \$2.700.000

NOTA: El programa se realizará siempre y cuando se complete con el mínimo de participantes.

Consultas

Ivonne Barra

Departamento de Informática - Campus Vitacura

Universidad Técnica Federico Santa María

Fonos: 2232028200 | (+569) 44054309

Email: sec-postitulos@inf.utfsm.cl

Carolina Leal

Departamento de Informática - Campus Casa Central Valparaíso

Universidad Técnica Federico Santa María

Fonos: 322654424 | (+569) 4405 3325

Email: vinculacion@inf.utfsm.cl

