



Diplomado en

APPLIED DATA SCIENCE

CURSO 1
INTRODUCCIÓN
A DATA SCIENCE

CURSO 2
VISUALIZACIÓN
DE DATOS

CURSO 3
MACHINE LEARNING
PARA CIENCIA DE DATOS



Curso

INTRODUCCION A DATA SCIENCE

Los datos impulsan todo lo que hacemos, saber gestionarlos y extraer insights relevantes es la clave para el éxito del negocio. La ciencia de datos es un campo interdisciplinario que involucra métodos científicos, procesos y sistemas para extraer conocimiento o un mejor entendimiento de los datos en sus diferentes formas, ya sean estructurados o no estructurados.

La ciencia de datos juega un rol en el apoyo de las estrategias de los responsables a la hora de fundamentar las decisiones de negocios, es decir, la ciencia de datos permite conseguir los objetivos empresariales, a partir del análisis de datos.

OBJETIVOS

Este curso busca introducir a los participantes de forma simple y didáctica en la ciencia de datos y sus procesos para comprender globalmente lo que hay detrás y cómo convertir los datos en información y la información en conocimiento.

METODOLOGÍA

La metodología de UAI Online contempla clases en línea con diferentes elementos de aprendizaje individual que incluye controles por clase, lecturas, tareas y videos multimedia con el contenido del programa de estudio, por su parte, el aprendizaje colaborativo se fomenta a través de Foro de la Clases de discusión entre los participantes del curso.

El curso tiene una duración de 5 semanas para las clases online, que incluyen material de lectura obligatoria y lecturas complementarias, la participación en los Foro de la Clases y las actividades evaluativas a fin de desarrollar los conocimientos contemplados por el mismo. Trabajo semanal de 12 horas.

CONTENIDOS

- Introducción. ¿Qué es data science?
- Metodologías para desarrollar proyectos de data science. Técnicas
- Modelos Predictivos
- Preprocesamiento y Exploración de Datos. Relación entre Atributos
- Modelos de regresión lineal
- Modelo de regresión logística

PROFESOR



Rolando De la Cruz

Magíster y Doctorado en Estadística, Pontificia Universidad Católica de Chile. Profesor regular de la Facultad de Ingeniería y Ciencias e investigador del Smart Center de la UAI. Director Académico del Magíster en Data Science.

Su investigación se enfoca en el desarrollo de metodologías para el análisis de datos. Cuenta con más de 30 artículos científicos publicados en revistas ISI. Actualmente es editor asociado del ISI journal Statistical Modelling y evaluador de los programas FONIS y FONDECYT, para la National Security Agency Mathematical Sciences Program, EE.UU, y es par evaluador de la CNA para programas de postgrado. Además brinda asesoría en temas de Business Analytics, Big Data, Data Science y Machine Learning.



Curso

VISUALIZACIÓN Y ANÁLISIS EXPLORATORIO DE DATOS

Una imagen vale más que mil palabras, es por ello que un gráfico es una herramienta muy poderosa para transmitir un mensaje. Específicamente, a través del correcto estímulo en la zona visual del cerebro, un gráfico puede mostrar claramente el mensaje deseado.

En este curso aprenderá una serie de métodos de visualización de información y a ilustrar su aplicación en el análisis exploratorio de datos.

OBJETIVOS

- Comprender rigurosamente conceptos claves de visualización.
- Criticar y redefinir un gráfico usando los conceptos aprendidos.
- Crear un gráfico con las características deseadas a través del uso de GGLOT.
- Observar patrones básicos en distintos tipos de gráficos.

METODOLOGÍA

La metodología de UAI Online contempla clases en línea con diferentes elementos de aprendizaje individual que incluye controles por clase, lecturas, tareas y videos multimedia con el contenido del programa de estudio, por su parte, el aprendizaje colaborativo se fomenta a través de Foro de la Clases de discusión entre los participantes del curso.

El curso tiene una duración de 5 semanas para las clases online, que incluyen material de lectura obligatoria y lecturas complementarias, la participación en los Foro de la Clases y las actividades evaluativas a fin de desarrollar los conocimientos contemplados por el mismo. Trabajo semanal de 12 horas.

CONTENIDOS

- Introducción y percepción humana
- Principios básicos de la visualización
- Gráficos, imágenes, y la gramática de los gráficos
- GGLOT, modificando características
- GGLOT, gráficos para datos categóricos, dispersión y línea
- GGLOT, distribuciones
- GGLOT, distribuciones
- Multidimensionalidad

PROFESOR



Sebastián Moreno

Doctor en Ciencias de la Computación y Master of Science en Ciencias de la Computación y Estadística, Purdue University. Magíster en Ciencias de la Computación, Universidad Técnica Federico Santa María. Profesor regular de la Facultad de Ingeniería y Ciencias e investigador del Smart Center de la UAI. Director del Magíster en Ciencias de la Ingeniería mención Tecnologías de la Información, UAI. Su investigación se enfoca en el área de machine learning desarrollando nuevos modelos para el análisis de datos, específicamente en el área de datos relacionales, como por ejemplo redes sociales.



Curso

MACHINE LEARNING PARA CIENCIA DE DATOS

La transformación digital está generando una ingente cantidad de datos que pueden ser utilizados para mejorar los procesos productivos, el desarrollo y venta de productos y servicios, entre otros. Las empresas y organizaciones tienen en los datos a uno de sus principales activos y requieren de un buen análisis para una óptima toma de decisiones para así mantener u obtener ventajas competitivas.

El machine learning o aprendizaje automático es un conjunto de algoritmos que se entrena en un conjunto de datos para hacer predicciones o tomar acciones para optimizar algunos sistemas. Por ejemplo, los algoritmos de clasificación supervisados se utilizan para clasificar a los posibles clientes en prospectos buenos o malos, para fines de préstamo, con base en datos históricos. Las técnicas del machine learning, junto a reglas de negocio se pueden utilizar para reducir el riesgo y calificar objetivamente la calidad de un cliente.

OBJETIVO

Este curso presentará algoritmos de aprendizaje supervisados y no supervisados tales como: métodos de regularización, naive Bayes, k-vecinos más cercanos, árboles de clasificación/regresión, algoritmos ensamblados, Support Vector Machine, Algoritmos jerárquicos, K-medias, DBScan y reglas de asociación. Se explica cómo validar modelos supervisados de machine learning para evitar sobreajustes, así como evaluar su performance mediante métricas de calidad predictiva. Para cada uno de los temas revisados se muestran ejemplos de cómo usar los algoritmos en el lenguaje de programación R/Python.

METODOLOGÍA

La metodología de UAI Online contempla clases en línea con diferentes elementos de aprendizaje individual que incluye controles por clase, lecturas, tareas y videos multimedia con el contenido del programa de estudio, por su parte, el aprendizaje colaborativo se fomenta a través de Foro de la Clases de discusión entre los participantes del curso.

El curso tiene una duración de 4 a 8 semanas para las clases online, que incluyen material de lectura obligatoria y lecturas complementarias, la participación en los Foro de la Clases y las actividades evaluativas a fin de desarrollar los conocimientos contemplados por el mismo. Trabajo semanal de 12 horas.

PROFESOR



Rolando De la Cruz

Magíster y Doctorado en Estadística, Pontificia Universidad Católica de Chile. Profesor regular de la Facultad de Ingeniería y Ciencias e investigador del Smart Center de la UAI. Director Académico del Magíster en Data Science. Su investigación se enfoca en el desarrollo de metodologías para el análisis de datos. Cuenta con más de 30 artículos científicos publicados en revistas ISI. Actualmente es editor asociado del ISI journal Statistical Modelling y evaluador de los programas FONIS y FONDECYT, para la National Security Agency Mathematical Sciences Program, EE.UU, y es par evaluador de la CNA para programas de postgrado. Además brinda asesoría en temas de Business Analytics, Big Data, Data Science y Machine Learning.

CONTENIDOS

Métodos de Regularización

- Introducción
- Ridge, Lasso y Elastic Net en regresión lineal y logística
- Elección del parámetro de regularización
- Aplicaciones con R/Python
-

Modelos supervisados: Naive Bayes y K-vecinos más cercanos

- Teorema de Bayes. Ejemplo de aplicación
- Naive Bayes. Características
- K-vecinos más cercanos. Características y consideraciones
- Elección del parámetros K.

Modelos supervisados: Árboles de clasificación/regresión

- Introducción
- Algoritmos para construir árboles
- Poda del árbol
- Predicción. Evaluación de calidad predictiva
- Aplicaciones con R/Python

Modelos supervisados: Algoritmos ensamblados – bagging/boosting

- Random Forest. Génesis. Bootstrapping. Bagging (Bootstrap Aggregating).
- Número de árboles. Out of bag error
- Importancia de las variables
- Algoritmos boosting: AdaBoost, XGBoost
- Aplicaciones con R/Python

Support Vector Machine

- Problemas de regresión y clasificación con SVM
- Aprendizaje con Kerneks
- Selección de kernels por validación cruzada
- Implementación de SVM para resolver problemas de clasificación y regresión
- Predicción
- Aplicaciones con R/Python

Algoritmos de Aprendizaje No Supervisados: Algoritmos jerárquicos

- Algoritmos jerárquicos divisivos
- Algoritmos jerárquicos aglomerativos
- Dendograma
- Elección del número de grupos
- Aplicaciones con R/Python

Modelos no supervisados: Algoritmos no jerárquicos

- Algoritmo K-medias
- Elección del número de grupos
- DBScan. Descripción del algoritmo
- Selección de parámetros
- Pros y contras de DBScan
- Evaluación de resultados
- Aplicaciones con R/Python

Modelos no supervisados: Reglas de Asociación

- Introducción. Definiciones
- Construcción de las reglas. Algoritmo A priori
- Evaluación (Sorporte, Confianza, lift, ...)
- Aplicaciones con R/Python