



2019 Data Scientist

CURSO SEMIAVANZADO



+54 9 11 63761130

| www.alexandria.com.ar

| C.A.B.A., Argentina

QUIENES SOMOS

Somos un equipo de especialistas en Data Mining y Business Intelligence. Contamos con Ingenieros, Economistas y Estadísticos.

Tenemos experiencia en la implementación de este tipo de herramientas en el sector bancario, seguros, retail, oil & gas y otras.

Nuestro objetivo principal es generar una cultura de datos en las empresas. Para ello, consideramos primordial capacitar al estrato gerencial en cuestiones de datos, donde las decisiones de implementaciones se hacen más accesibles y se potencian los resultados.

Queremos ayudar en el cómo plantear las necesidades, cómo enfocar los proyectos, y cómo puede aplicarse la tecnología en todas las actividades del mercado.

Para ello brindamos capacitaciones y consultoría especializada en Machine Learning.

APLICACIONES

Generación de hipótesis, preparación de datos, así como mejora de la calidad de estos. Análisis exploratorio con métodos estadísticos. Modelado, que implica algoritmos de aprendizaje supervisado y no supervisado.

DESCRIPCIÓN DEL CURSO

Este curso pretende otorgar al participante conocimientos básicos de datos y la utilidad que brinda a la toma de decisiones el conocer su potencial. Para ello, se brindará una introducción teórica a las bases de datos, el modo en que se relacionan las tablas y como se llega a explotar esos datos de manera accesible a personas que no son especialistas en Business Intelligence ni Bases de Datos.

Se utilizarán herramientas de software libre, de manera tutorada, que permitirán generar modelos de predictivos sobre un conjunto de datos sencillos. Pero, sobre todo, donde podremos trabajar de manera conjunta en el análisis exploratorio de datos, previo a la implementación de los algoritmos.

Dado que se trata de un ámbito que utiliza especialmente la estadística, y que este curso no requiere conocimientos previos en este ámbito, se brindarán también herramientas para poder comprender porque ciertas variables deben transformarse y cómo es que generan mejores resultados.

DETALLES

Modalidad: In-Company

Tiempo: 3 días, 24 hs.

Costos: (Incluye material didáctico del curso)

Prerrequisito: Instalación de entorno Anaconda.

TESTES
TESTES
TESTES
TESTES
TESTES
TESTES
TESTES
TESTES
TESTES
TESTES

PERFIL DEL PARTICIPANTE

Profesionales de todas las áreas con predisposición de utilizar herramientas de análisis de datos para generar proyectos, o comprender como se generan los proyectos relacionados con Analytics.

Directivos, conducción intermedia y técnico-profesionales que participen de la generación de información.

Se recomienda conocimientos de estadística, programación y nociones de análisis de datos.

OBJETIVOS

Al finalizar el curso, el participante, será capaz de discernir si se trata de un proyecto a ser resuelto por medio de un algoritmo de aprendizaje supervisado o no supervisado.

Analizar el problema para identificar la variable target, tener los conocimientos básicos para saber que tratamiento aplica.

Podrá también, visualizar variables, realizar análisis estadísticos básicos de variables, aplicar transformaciones, trabajar con distintos tipos de variables, y habrá aplicado al menos un algoritmo de aprendizaje supervisado por su cuenta.

Se le brindará al participante herramientas para poder interactuar con profesionales de Analytics, generar solicitudes y proponer proyectos que mejoren la performance de su propia área.



Programa

CURSO SEMIAVANZADO

MODULO 1

Preprocesamiento de datos. Conceptos.

Limpieza de Datos.

Integración y transformación de datos.

Reducción de la dimensionalidad.

Discretización de variables.

Implementación de conocimientos adquiridos

MODULO 2

Introducción estadística.

Análisis exploratorio de datos.

Características de la distribución de las variables.

Análisis de las relaciones entre las variables.

Transformación y análisis de las mejoras.

Exploración de resultados y aplicaciones prácticas.

MODULO 3

Market Basket Analysis:

Problemas de overfitting y underfitting.

Problemas de Generalización.

Estimación de error (train-test split. cross-validation).

Ajuste de parámetros e hiperparámetros.

Práctica.

MODULO 4

XGBoost:

Problemas de overfitting y underfitting.

Problemas de Generalización.

Estimación de error (train-test split. cross-validation).

Ajuste de parámetros e hiperparámetros.

Práctica.

MODULO 5

Support Vector Machine:

Problemas de overfitting y underfitting.

Problemas de Generalización.

Estimación de error (train-test split. cross-validation).

Ajuste de parámetros e hiperparámetros.

Práctica.

MODULO 6

Deep Learning: Redes Neuronales

Red Neuronal Backpropagation.

Problemas de Generalización y de Mínimos locales.

Práctica.

PROGRAMA