

## ACTIVIDAD CURRICULAR 1

# Introducción al Data Science y R Project

Duración: 5 semanas  
Horas Presenciales: 30  
Horas Autónomas: 30  
Créditos: 2 SCT.

### RESULTADOS DE APRENDIZAJE

1) Explicar los conceptos fundacionales del Data

Science, y las tendencias y tecnologías actuales detrás del Big Data.

2) Aplicar el Ciclo de Vida del Data Science en problemas del mundo real y contexto local.

3) Aplicar los principios fundamentales de las tecnologías RStudio y Shiny como apoyo a la programación del proceso del Data Mining.

### BIBLIOGRAFÍA MÍNIMA

- ♦ Shah, C. (2020). A Hands-On Introduction to Data Science. Cambridge University Press.
- ♦ Boehmke, B. C. (2016). Data wrangling with R. Springer.
- ♦ De Brouwer, P. J. S. (2020). The Big R-Book: From Data Science to Learning Machines and Big Data. Wiley

Total de horas: 180

### EJES TEMÁTICOS

- ♦ Conceptos fundacionales del Data Science, Machine y Deep Learning, Data Mining
- ♦ Comprensión del negocio.
- ♦ Enfoque analítico.
- ♦ Recopilación de los datos (requisitos y comprensión).
- ♦ Preparación de los datos
  
- ♦ Modelado.
- ♦ Implementación
- ♦ Visualización y Comunicación.
- ♦ Descarga de R y Rstudio.
- ♦ Gestión de directorios de trabajo.
- ♦ R como Calculadora, ayuda de R.
- ♦ Objetos y estructuras de los datos.
- ♦ Operaciones sobre vectores.
- ♦ Caracteres especiales.
- ♦ Matrices y dataframes.
- ♦ Operadores lógicos, condicionales, ciclos y funciones.
- ♦ Gestión de librerías.

## ACTIVIDAD CURRICULAR 2

# Estadística para Data Science

Duración: 5 semanas  
Horas Presenciales: 30  
Horas Autónomas: 30  
Créditos: 2 SCT.

## RESULTADOS DE APRENDIZAJE

- 1) Aplicar los diferentes Métodos Estadísticos para la preparación de los datos y sus posibles aplicaciones en la industria.
- 2) Aplicar diferentes técnicas de segmentación a problemas reales mediante el estudio de casos.
- 3) Aplicar los diferentes Métodos de Regresión para predicción, clasificación y detección de patrones en proyectos reales.

## BIBLIOGRAFÍA MÍNIMA

- Boehmke, B. C. (2016). Data wrangling with R. Springer.
- Irizar y, R. (2019). Introduction to Data Science-Data Analysis and Prediction Algorithms with R. CRC Press.
- Dobson, A. J., & Barnett, A. G. (2018). An introduction to generalized linear models. CRC press.

## EJES TEMÁTICOS

- ETL: Extract, Transform and Load.
- Data cleaning, tratamiento de outliers y valores missing.
- Normalización de datos.
- Visualización y Análisis Exploratorio de Datos.
- Métodos de reducción de dimensión.
- Visualización interactiva: Dashboards y Shiny Apps.
- Breve repaso de estadística básica.
- El concepto del valor-p.
- Análisis de Varianza (ANOVA).
- Prueba Chi-Cuadrado.
- Árboles de decisión (CHAID).
- Regresión Lineal.
- Modelos Lineales Generalizados: Regresión Logística y otros casos.
- Interpretación de los coeficientes de regresión.
- Modelado con datos desbalanceados
- Regresión PLS.

### ACTIVIDAD CURRICULAR 3

# Elementos de Machine Learning

Duración: 10 semanas  
Horas Presenciales: 60  
Horas Autónomas: 60  
Créditos: 4 SCT.

## RESULTADOS DE APRENDIZAJE

1) Analizar los diferentes tipos de algoritmos y tareas de aprendizaje de machine learning y data mining tradicional y sus posibles aplicaciones en la industria nacional.

2) Aplicar diferentes algoritmos Reglas de

Asociación y Ensamble a problemas reales para desarrollar estrategias tecnológicas para productos y servicios.

3) Aplicar los principios fundamentales de las redes neuronales artificiales y su evolución hacia arquitecturas del tipo Deep learning para la detección de patrones en problemas reales.

## EJES TEMÁTICOS

- Clustering jerárquico.
- Algoritmo k-medias.
- Redes de Kohonen.
- DBSCAN
- Medidas de evaluación deClustering.
- Técnicas de asociación. técnicas de ensamble.
- Regularización.
- Mejorar el rendimiento del modelo.
- Optimización para Machine Learning.
- Conjuntos de entrenamiento, testeo y cross-validation.
- K-vecinos más cercanos.
- Support Vector Machine.
- Redes Neuronales Artificiales. Medidas de Evaluación de clasificación y Regresión.

## BIBLIOGRAFÍA MÍNIMA

- Larose, D. T. (2015). Data mining and predictive analytics. John Wiley&Sons.
- Irizarry, R. (2019). Introduction to Data Science-Data Analysis and Prediction Algorithms with R. CRC Press.

## PROCESO DE ADMISIÓN

Para la postulación al diplomado se requiere completar una ficha de inscripción online que será evaluada por los Coordinadores del programa para su aprobación.

Documentos y Requisitos:

- Poseer título profesional o técnico o grado de licenciatura.  
Currículum vitae  
Fotocopia del título profesional
- Fotocopia de carnet
- 

Posterior a la selección, el participante debe firmar carta de compromiso con las siguientes acciones:

- Permanencia total en el diplomado.
- La asistencia sincrónica por actividad curricular debe ser al menos 75%.
- Las inasistencias deben ser debidamente justificadas.

- El Programa se reserva el derecho de suspender la realización del diplomado/curso si no cuenta con el mínimo de estudiantes requeridos. En tal caso se devuelve a los alumnos matriculados la totalidad del dinero.

Una vez matriculados, no hay devolución de dinero.

- La coordinación del programa se reserva el derecho de suspender o reprogramar la realización de la actividad si no cuenta con el mínimo de estudiantes requeridos o por motivos de fuerza mayor. En tal caso se devuelve a los alumnos matriculados la totalidad del dinero.