

Análisis y Ciencia de Big Data Avanzados

Código: ARC-105

Propuesta de Valor: ARCITURA

Duración: 24 Horas



Este curso profundiza en una variedad de prácticas avanzadas de análisis de datos y técnicas de análisis. Que se exploran en el contexto de Big Data. El contenido del curso se centra en temas que permiten, que los participantes desarrollen una comprensión profunda de la estadística, el modelado y el análisis.

Técnicas para patrones de datos, clústeres y análisis de texto, así como la identificación de valores atípicos, errores que afectan la importancia y la precisión de las predicciones realizadas en conjuntos de datos de Big Data



AUDIENCIA

- Cualquier persona que esté interesada en ampliar sus conocimientos en Big Data y deseen cubrir unas necesidades no satisfechas por las tecnologías existentes en el almacenamiento y tratamiento de grandes volúmenes de datos.



PRE REQUISITOS

- No hay requisitos previos.



OBJETIVOS

- Modelado, evaluación de modelos, ajuste de modelos y sobreajuste de modelos
- Modelos estadísticos, medidas de evaluación del modelo
- Validación cruzada, variación de sesgo, matriz de confusión y puntaje F
- Algoritmos de aprendizaje automático e identificación de patrones
- Reglas de asociación y algoritmo a priori
- Reducción de datos, selección de funciones de dimensionalidad
- Extracción de funciones, discretización de datos (agrupamiento y agrupamiento)
- Técnicas estadísticas avanzadas
- Paramétrico frente a no paramétrico, agrupación frente a no agrupación
- Basado en distancia, supervisado vs. semi-supervisado
- Regresión lineal y regresión logística para Big Data

- Reglas de clasificación para Big Data
- Regresión logística, Naïve Bayes, Suavizado de Laplace, etc.
- Árboles de decisión para Big Data
- Poda de árboles, división de características, algoritmo de una regla (1R)
- Identificación de patrones, reglas de asociación, algoritmo a priori
- Análisis de series de tiempo, tendencia, estacionalidad
- K Vecino más cercano (kNN), K-significa
- Análisis de texto para Big Data
- Bolsa de palabras, frecuencia de término, frecuencia de documento inverso, distancia de coseno, etc.
- Detección de valores atípicos para Big Data
- Técnicas estadísticas, basadas en distancia, supervisadas y semi-supervisadas



CERTIFICACIÓN DISPONIBLE

- Certificado oficial **ARCITURA**.
- Este curso es parte de la ruta de certificaciones de:
 - Consultor de Big Data Certificado



CONTENIDO

1. MODELADO, EVALUACIÓN DE MODELOS, AJUSTE DE MODELOS Y SOBREAJUSTE DE MODELOS
2. MODELOS ESTADÍSTICOS, MEDIDAS DE EVALUACIÓN DE MODELOS
3. VALIDACIÓN CRUZADA, VARIANZA DE SESGO, MATRIZ DE CONFUSIÓN Y PUNTUACIÓN
4. ALGORITMOS DE APRENDIZAJE AUTOMÁTICO E IDENTIFICACIÓN DE PATRONES
5. REGLAS DE ASOCIACIÓN Y ALGORITMO A PRIORI
6. REDUCCIÓN DE DATOS, SELECCIÓN DE FUNCIONES DE DIMENSIONALIDAD
7. EXTRACCIÓN DE FUNCIONES, DISCRETIZACIÓN DE DATOS (BINNING Y AGRUPAMIENTO)
8. TÉCNICAS ESTADÍSTICAS AVANZADAS
9. PARAMÉTRICO FRENTE A NO PARAMÉTRICO, AGRUPAMIENTO FRENTE A NO AGRUPAMIENTO
10. BASADO EN LA DISTANCIA, SUPERVISADO FRENTE A SEMISUPERVISADO
11. REGRESIÓN LINEAL Y REGRESIÓN LOGÍSTICA PARA BIG DATA
12. REGLAS DE CLASIFICACIÓN PARA BIG DATA
13. REGRESIÓN LOGÍSTICA, NAÏVE BAYES, SUAIVIZADO DE LAPLACE, ETC.

14. ÁRBOLES DE DECISIÓN PARA BIG DATA

15. PODA DE ÁRBOLES, DIVISIÓN DE CARACTERÍSTICAS, ALGORITMO DE UNA REGLA (1R)

16. IDENTIFICACIÓN DE PATRONES, REGLAS DE ASOCIACIÓN, ALGORITMO A PRIORI

17. ANÁLISIS DE SERIES TEMPORALES, TENDENCIA, ESTACIONALIDAD

18. K VECINO MÁS CERCANO (KNN), K-MEANS

19. ANÁLISIS DE TEXTO PARA BIG DATA

20. BOLSA DE PALABRAS, FRECUENCIA DE TÉRMINOS, FRECUENCIA DE DOCUMENTO INVERSA, DISTANCIA DE COSENO, ETC.

21. DETECCIÓN DE VALORES ATÍPICOS PARA BIG DATA

22. ESTADÍSTICA, BASADA EN LA DISTANCIA, SUPERVISADA Y SEMISUPERVISADA TÉCNICAS

★ BENEFICIOS

- Al finalizar el curso tendrás una comprensión profunda de la estadística, el modelado y el análisis.