

MÓDULO 1: DATA SCIENCE Y BIG DATA. LA NUEVA REALIDAD

CONTENIDO

1. Métodos Estadísticos

- 1.1. Definición y clasificación de la Estadística
- 1.2. Conceptos estadísticos fundamentales

2. Data Science y Big Data. La Nueva Realidad

- 2.1. Conceptos clave
- 2.2. Minería de Datos o Data Mining
- 2.3. Modelos SEMMA Y CRISP-DM
- 2.4. Principales Métodos y Algoritmos en la Minería de Datos

3. Introducción a las bases de datos relacionales

- 3.1. Definición de base de datos
- 3.2. Bases de datos relacional y no relacional
- 3.3. Lenguaje SQL

4. Big Data

- 4.1. Acceso a fuente de datos
- 4.2. Consideraciones legales básicas
- 4.3. Aplicaciones del Big Data
- 4.4. Herramientas Big Data - Spark
- 4.5. Big Data en la Nube y Servidores cloud - Microsoft Azure

5. Introducción a la programación en Data Science

- 5.1. R y R-Studio - Librerías para Data Science
- 5.2. Python (Notebooks y Pycharm) - Librerías para Data Science
- 5.3. Otro Software utilizado en Data

6. Gestión de proyectos de Software aplicados al Data Science

- 6.1. Metodología SCRUM y Agile
- 6.2. Repositorio de Código. Introducción a Git
- 6.3. Plataformas de desarrollo colaborativo - GitHub
- 6.4. Estructura de proyectos de Data Science

PREPROCESADO Y LIMPIEZA DE DATOS

1. Introducción

- 1.1. Formulación del problema. Integración de la información

2. Análisis gráfico y descriptivo de la información

- 2.1. Visualización de datos para extracción de información

3. Detección e imputación de datos ausentes

- 3.1 Métodos univariantes y multivariantes

4. Tratamiento de valores extremos (outliers)

- 4.1 Métodos univariantes y multivariantes

5. Transformación de la base de datos

- 5.1 Tipificación y discretización de variables

6. Equilibrado de la muestra

PROFESORES RESPONSABLES

- Dr. Julio González Arias
- D. Pablo Sánchez Cabrera
- D. Juan José García Muñoz
- D. José Manuel González Fidalgo

(Para mayor información de los profesores, puede consultar el menú “Equipo Docente” de la Web)

