



Título

CG6: Fundamentos de la ciencia de datos con R

I. DATOS GENERALES

1. Horas totales: 30 horas
10 h grabadas, el cual se divide en cinco sesiones
20 h de lecturas, prácticas y realización del trabajo final
2. Modalidad: **No presencial (virtual)**
3. Aula virtual: Los matriculados tendrán acceso al Google Classroom
4. Docente: Mg. Sthip Blas Oyola
<https://drive.google.com/file/d/16uX4nUCbRNwvZHdRDIZTHAJDuhbRTFuq/view?usp=sharing>
Magister en Estadística (PUCP). Experto en implementar soluciones analíticas y desarrollo de metodologías para el análisis de datos. Con competencias en tecnologías y el mundo digital.

II. PRESENTACIÓN Y DESCRIPCIÓN DEL CURSO

Actualmente las empresas están enfocadas en la generación de estrategias e implementación de procesos de transformación digital; produciendo muchos datos que podrían transformarse en información de valor. Esta situación ha visibilizado la importancia de contar con profesionales que puedan analizar estos datos y explotarlos para mejorar el proceso de toma de decisiones en las empresas, incrementando la productividad y rentabilidad de estas. El curso fundamentos de la ciencia de datos con R te permitirá aportar los conocimientos y capacidades necesarias para proponer proyectos analíticos en tu organización y, de esta forma, generar valor a las organizaciones a través del análisis de datos. Todo ello con el uso de R, un entorno de software libre y muy utilizado en la actualidad a nivel mundial.

III. CONTENIDOS

Sesión 1: Introducción a la ciencia de datos

1. Conceptos: ciencia de datos, machine learning, deep learning, big data
2. Metodología de la ciencia de datos
3. Conociendo RStudio
4. Taller de la sesión 1

Sesión 2: Análisis exploratorio de datos con RStudio

1. Carga de datos, importación, exportación y limpieza de datos, filtros.
2. Imputación datos, eliminación datos.
3. Estadísticos, histogramas, boxplot
4. Taller de la sesión 2

Sesión 3: Métodos no supervisados con RStudio

1. Reducción de dimensiones: PCA
2. Cluster: Kmeans – Kmodes - Kprototypes

3. Detección outlier: DBScan – Isolation Forest
4. Taller de la sesión 3

Sesión 4: Métodos supervisados con RStudio

1. Regresión lineal simple y múltiple
2. Regresión no lineal - Logístico
3. Clasificaciones: Random Forest – Algoritmos gradient boosting
4. Taller de la sesión 4

Sesión 5: Temas varios con RStudio

1. Estadística: Muestreo, series de tiempo, análisis de experimentos
2. Optimización: Programación Lineal
3. Aplicaciones en los negocios
4. Taller final y explicación del trabajo práctico

IV. APORTES

Al término del curso-taller, con el soporte de R, los estudiantes:

Analizan los datos cuantitativos y cualitativos del entorno digital para organizar información de valor que ayude a las empresas a tomar decisiones.

Tienen las competencias para la elaboración de aplicaciones usando técnicas estadísticas. Comprenden los fundamentos estadísticos y de ciencia de datos para la aplicación en sus instituciones.

Analizan datos para predecir acciones o comportamientos de empresas o instituciones a partir del análisis de grandes volúmenes de datos históricos.

V. PRODUCTO FINAL Y CERTIFICACIÓN

Para la emisión del certificado a nombre de la Sociedad Científica de Bibliometría y Cienciometría, cada participante elaborará un producto final (El trabajo práctico es el desarrollo de un modelo analítico enfocado a un problema de su ambiente laboral) que será realizado en las horas prácticas.

Puede descargar el instructivo [AQUÍ](#)

V. METODOLOGÍA

La metodología de trabajo consiste en observar minuciosamente las grabaciones de las cinco sesiones.

Luego de cada sesión debe avanzar el producto final, el cual presentará al término de la observación de las cinco grabaciones.

Los participantes tendrán acceso al Classroom (materiales, grabaciones y envío del producto final) durante tres meses desde el momento de la matrícula.

Contacto: Andrea Rojas
a.rojas@ssbibliometrics.com

VI. BIBLIOGRAFÍA

<http://www.datascienceassn.org/sites/default/files/Building%20Data%20Science%20Teams.pdf>

<https://rstudio-education.github.io/hopr/r-notation.html>