



RIDAA
Repositorio Institucional
Digital de Acceso Abierto de la
Universidad Nacional de Quilmes



Universidad
Nacional
de Quilmes

Silva, Germán

Introducción a la ciencia de los datos



Esta obra está bajo una Licencia Creative Commons Argentina.
Atribución - No Comercial - Compartir Igual 2.5
<https://creativecommons.org/licenses/by-nc-sa/2.5/ar/>

Documento descargado de RIDAA-UNQ Repositorio Institucional Digital de Acceso Abierto de la Universidad Nacional de Quilmes de la Universidad Nacional de Quilmes

Cita recomendada:

Silva, G. (2022). *Introducción a la ciencia de los datos. (Programa)*. Bernal, Argentina: Universidad Nacional de Quilmes. Disponible en RIDAA-UNQ Repositorio Institucional Digital de Acceso Abierto de la Universidad Nacional de Quilmes <http://ridaa.unq.edu.ar/handle/20.500.11807/4071>

Puede encontrar éste y otros documentos en: <https://ridaa.unq.edu.ar>

UNIVERSIDAD NACIONAL DE QUILMES
DEPARTAMENTO DE ECONOMÍA Y ADMINISTRACIÓN

Programa Regular – Cursos Presenciales

Carrera: Diplomatura en Economía y Administración

Año: 2022

Curso: Introducción a la ciencia de los datos.

Profesor: Prof. Germán Silva

Carga horaria semanal: 5 horas (4 áulicas y 1 extra áulica)

Horas de consulta extra clase: jueves de 18:00 a 22:00. + 1 hora de trabajo extraúlico

Créditos: 10

Núcleo al que pertenece: Electivo

Tipo de asignatura: Teórico - práctica

Presentación y Objetivos

La asignatura tiene como objetivo la gestión del patrimonio arqueológico para ser ofrecido como un producto turístico, pero con una visión de manejo y conservación del recurso como un bien no renovable.

La presente materia tiene como propósito la presentación y la reflexión crítica sobre algunas de las teorías y herramientas metodológicas disponibles en la actualidad para su uso en la investigación en las ciencias sociales y en las diversas aplicaciones que de ellas derivan. Creemos que la utilización de estas herramientas no debería implementarse sin una discusión epistemológica respecto de los marcos que encuadran sus posibilidades. Nuestra propuesta se desarrolló con este objetivo en mente, ya que confiamos tanto en la importancia del pluralismo metodológico como en la necesidad de suscitar la reflexión sobre perspectivas y herramientas básicas a la hora de gestionar y analizar datos. Dentro de ese pluralismo metodológico destacamos ciertos aspectos de los métodos que consideramos indispensables para poder ser utilizados con cierto rigor. Por un lado, la replicabilidad, es decir la posibilidad de que el método trascienda y pueda ser utilizado por cualquier investigador. Para ello los métodos deben reducir aquellos elementos que se basen en las capacidades y habilidades innatas de cada investigador. Por el otro lado, destacamos la potencia empírica, es decir la capacidad de construir en forma coherente la información necesaria para la interpretación social.

La presente materia tiene como finalidad poder proveer al alumno de las herramientas necesarias para poder analizar e interpretar de manera pertinente todo dato que se encuentre estructurado, lo que resulta de vital trascendencia en la investigación científica aplicada en las ciencias sociales. Históricamente ha sido de gran relevancia el conocimiento pertinente al manejo de bases de datos y la exploración de sus contenidos, pero en el contexto actual de explosión en el almacenamiento masivo de datos, y su habitual liberación de bases por parte de organismos o filtraciones, se vuelve por demás indispensable el manejo de las herramientas que nos permitan acceder analíticamente a la información que allí se encuentra.

La materia comienza con una introducción vinculada a los diferentes tipos de formas que pueden adoptar las bases de datos, seguido por los conceptos básicos y estándares sobre normalización de la información y procesos vinculados al refinamiento de los datos. Para luego abordar contenidos relacionados con la teoría de las bases de datos, como así también procesos de ETL (Extracción, Transformación y carga -Load-). Finalmente, nos introduciremos dentro de las diferentes posibilidades que existen hoy en día a la hora de explorar los datos en la búsqueda de obtener mayor cantidad de información de ellos. Nos centraremos en el modelado de la información, ya sea a través de modelos estadísticos analíticos y predictivos, modelos de aprendizaje automático, como así también modelos basados en agentes -MBA- y, para finalizar, presentar los modelos reticulares con sus estadísticos propios.

En el último segmento de la materia se recorrerán las diferentes posibilidades existentes a la hora de visualizar la información, y la cantidad de herramientas que orientadas a este objetivo se encuentran a disposición.

La asignatura es de tipo teórico-práctico, en la que se considera fundamental la comprensión de los conceptos tanto como la elección de los instrumentos según cada situación, así como, la interpretación de los resultados obtenidos.

En la actualidad, existen distintas aplicaciones de software para la sistematización de la información, preparación, limpieza, normalización, análisis y visualización que utilizaremos para la ejercitación y la aplicación de los conceptos.

Objetivos

Los objetivos que tiene el presente curso es introducir a los alumnos en los conceptos y herramientas básicas de la analítica y ciencias de los datos para ser aplicados en los respectivos campos profesionales. Se espera que al final del curso el alumno:

- Sea capaz de aplicar los conocimientos introductorios a la ciencia de los

datos, como una herramienta para la investigación en su disciplina.

- Adquiera las habilidades para conceptualizar los fundamentos básicos para la comprensión y utilización de las herramientas – software específicos y lenguaje SQL - y técnicas de tamizaje de información que permitan luego generar modelos que den cuenta de las problemáticas analizadas.

- Comprenda las diferencias y los usos de los distintos modos de almacenar la información.

- Alcance un conocimiento elemental de cada etapa del proceso del análisis de los datos (extracción, transformación, modelado, procesamiento y análisis de información)

Objetivos Específicos

- Que los y las estudiantes, obtengan conocimiento sobre la importancia de la conservación e interpretación del patrimonio cultural tangible y su relación con las actividades provenientes del turismo y el hotelería.
- Que los y las estudiantes, se capaciten para intervenir en planes de manejo de recursos culturales ligados a actividades turísticas.
- Que los y las estudiantes, aprendan a usar sus capacidades analíticas y desarrolle una actitud crítica.

Contenidos Mínimos

Bases de datos: tipo de bases, bases de datos relacionales, jerárquicas, orientadas a grafos, tablas, relaciones entre tablas; construcción de una base de datos relacional. Tamizaje: limpieza, normalización. Metodologías de scraping. Consultas a bases de datos en lenguaje SQL. Análisis de datos: epistemología detrás de los modelos, modelos descriptivos, modelos predictivos. Modelos basados en Agentes y Análisis de Redes Sociales. Visualización de la información.

UNIDAD 1

La ciencia de los datos: 1). ¿Qué son los datos?, ¿Qué es la ciencia de los datos? Sistematización de los diferentes modos de abordar los datos. (big data, small data, machine learning, data mining, bases de datos, etc. 2) Repaso sobre Unidad de análisis, Variables y Categorías. Matriz de datos versus una Base de datos

UNIDAD 1 Bibliografía Obligatoria

- **Sosa, E. W., 2019.** Big data: Breve manual para conocer la ciencia de datos que ya invadió nuestras vidas. Buenos Aires: Siglo Veintiuno Editores Argentina. Capítulo 1: Perdidos en el océano de datos. Big data, aprendizaje automático, ciencias de datos, estadística y otras yerbas.

UNIDAD 2

La construcción de modelos en el proceso de investigación científica: 1) ¿Qué es un modelo? La importancia de poseer control sobre el modelo construido. Diferentes tipos de modelados. 2) Los cuatros tipos de modelos: modelos mecánicos, modelos estadísticos, modelos sistémicos, y modelos interpretativos. 3) Los modelos predictivos o las simulaciones como posible superación de la dicotomía entre el modo inductivo y el modo deductivo.

UNIDAD 2 Bibliografía obligatoria

- **Epstein, J. y R. Axtell, 1996,** Growing Artificial Societies: Social Sciences from the Bottom Up, MIT Press, Cambridge (Massachusetts) Cap I Introduction. (Traducción propia).
- **Gómez Bailón, S., 2004,** “Sociedades artificiales, una introducción a la simulación social”, Revista Internacional de Sociología 39 2004: 199-222

UNIDAD 3

Aproximación a la teoría de BBDD: 1) Diferentes tipos de bases de datos. Beneficios y contras de cada modalidad. 2) La base de datos relacional. Matriz, tabla versus BBDD. La importancia del armado de la base de datos. Tipos de conexiones entre tablas. Buenas prácticas en la construcción de una base de datos. 3) Introducción al lenguaje SQL.

UNIDAD 3 Bibliografía obligatoria

- **Russell Bernard, H., 2006.** Métodos de investigación en Antropología. Segunda edición. Capítulo 6. Sampling. The Foundations of Social Research. Altamira Press. Capitulo XVII: Codificación y clave de códigos para datos cuantitativos.
- **Silberschaltz, A., Korth, H. F., & Escalona, G. R., 1987.** Fundamentos de bases de datos. Méjico: McGraw-Hill. Capítulo 1 y 2.

UNIDAD 4

Uso de entorno de almacenamiento de base de datos relacional (mysql). Tipos y estructuras de datos. Procesos de ETL (extracción, transformación y carga) de datos: técnicas de carga, limpieza y normalización de datos, y procesamiento básico. Condiciones lógicas e iteraciones.

UNIDAD 4 Bibliografía obligatoria

- **Silberschaltz, A., Korth, H. F., & Escalona, G. R. (1987).** Fundamentos de bases de datos. México: McGraw-Hill. Capítulo 3 y 4.

UNIDAD 5

Análisis Reticular: 1) Estadísticos propios de la metodología de Análisis de Redes Sociales. Medidas de centralidad. Medidas holísticas de una red. 2) Construcción de una red con datos públicos. Análisis a partir de un software específico (GEPHI). 3) Bases de datos orientadas a grafos.

UNIDAD 5 Bibliografía obligatoria

- **Jariego, Isidro & Holgado, Daniel & Molina, José Luis. (2014).** 7 lecturas para iniciarse en el análisis de redes sociales. 10.13140/2.1.4616.2882.
- **Hanneman, Robert A.** Introducción a los métodos del análisis de redes sociales. Versión en español en <http://revista-redes.rediris.es/webredes/textos> Capítulo Segundo ¿Por qué utilizar métodos formales en el análisis de redes sociales?

UNIDAD 6

Visualización de los datos: 1) Diferentes tipos de visualizaciones, gráficos y grafos. Barras, columnas, Mapas proyectados. Dispersiones. Grafos. Flujos. Diagramas de Gantt. 2) Recursos online vs programación. 3) Presentación de datos. Confección de informes a través de tableros de control. Creación de tableros de control. (Tableau, PowerBi). Periodismo de datos.

UNIDAD 6 Bibliografía obligatoria

- **Axelrod Robert, 2003.** La complejidad de la cooperación. Fondo de Cultura Económica, Buenos Aires.
- **Cruscianelli, S., 2013.** - ¿Qué es el periodismo de datos? En Cuadernos de Periodistas. <http://www.cuadernosdeperiodistas.com/media/2013/12/106-1241.pdf>.
- **Huff, D., 2015.** Cómo mentir con estadísticas. Barcelona: Critica.

Bibliografía de Consulta

- **Borgatti, Steve, 2003** - Conceptos básicos de Redes Sociales. XXIII conferencia Internacional de Análisis de Redes Sociales en Cancún, México. En www.analytictech.com/networks
- **Granovetter, Mark S.** The strength of weak ties, American Journal of Sociology, Vol. 78, N° 6 (Pp.1360 - 1380). Johns Hopkins University. 1973.
- **Holland, J. H., & Torres, A. E.** (2004). El orden oculto: De cómo la adaptación crea la complejidad. México: Fondo de Cultura Económica.

Modalidad de dictado: El desarrollo del curso cuenta con dos modalidades diferentes: Clases teóricas y Clases prácticas. En las clases teóricas se desarrollarán los contenidos del Programa, mientras que en las clases prácticas se realizarán ejercitaciones ya sea a través del uso de aplicaciones de software o de guías de ejercicios

Actividades extra-áulicas obligatorias:

Se realizarán dos de trabajos prácticos a lo largo del cuatrimestre, consistentes en guías preguntas o ejercicios, en función del contenido de cada unidad. La realización de

los mismos es domiciliaria y la entrega dentro de la fecha establecida para cada uno de ellos es obligatoria y condición para la aprobación de la materia.

Evaluación: Dos parciales escritos, una instancia de recuperación (para ausentes o desaprobados) y presentación de un trabajo práctico individual grupal.

Aprobación de la asignatura: Se regirá según el Art. 9no del Régimen de estudios de la UNQ (Res. 201/18), bajo las siguientes condiciones:

a) Asistencia no inferior al 75% de las clases; b) Obtención de un promedio mínimo de 7 puntos en los parciales y de un mínimo de 6 puntos en cada uno de ellos.; c) Obtención de un mínimo de 4 puntos en cada parcial y en el examen integrador que será obligatorio en estos casos. Este examen se tomará dentro de los plazos del curso. Los alumnos que obtengan un mínimo de 4 puntos en cada uno de los parciales y no hubiesen aprobado el examen integrador mencionado en el inciso b) deberán rendir un examen integrador en un lapso que no superará el cierre de actas del siguiente cuatrimestre. El Departamento respectivo designará a un profesor del área, quien integrará con el profesor a cargo del curso, la mesa evaluadora de este nuevo examen integrador. Se garantizará que los alumnos tengan al menos una instancia parcial de recuperación.

A handwritten signature in black ink, appearing to read 'Germán Silva', with a large, sweeping flourish underneath.

Prof. Germán Silva.