



Ministerio de Cultura y Educación  
Universidad Nacional de San Luis  
Facultad de Ciencias Físico Matemáticas y Naturales  
Departamento: Informatica  
Area: Area II: Sistemas de Computacion

(Programa del año 2024)

### I - Oferta Académica

Materia	Carrera	Plan	Año	Período
() TALLER INTRODUCTORIO A LA CIENCIA DE DATOS	LIC.CS.COMP.	32/12	2024	2° cuatrimestre

### II - Equipo Docente

Docente	Función	Cargo	Dedicación
BARRIONUEVO, MERCEDES DEOLINDA	Prof. Responsable	P.Adj Exc	40 Hs
ARROYUELO BILLIARDI, JORGE A.	Prof. Co-Responsable	P.Adj Exc	40 Hs
TISSERA, PABLO CRISTIAN	Prof. Co-Responsable	P.Adj Exc	40 Hs

### III - Características del Curso

Credito Horario Semanal				
Teórico/Práctico	Teóricas	Prácticas de Aula	Práct. de lab/ camp/ Resid/ PIP, etc.	Total
2 Hs	Hs	Hs	3 Hs	5 Hs

Tipificación	Periodo
B - Teoria con prácticas de aula y laboratorio	2° Cuatrimestre

Duración			
Desde	Hasta	Cantidad de Semanas	Cantidad de Horas
05/08/2024	15/11/2024	15	75

### IV - Fundamentación

Para la correcta toma de decisiones, toda organización tiene una ingente cantidad de datos de todo tipo, muchos de ellos recolectados por sistemas propios, pero muchos más obtenidos a partir de una importante variedad de fuentes como lo son los datos publicados por otras organizaciones públicas y privadas, redes sociales, sensores, etc. Estos datos pueden otorgar beneficios transformadores para las organizaciones y la sociedad en su conjunto, siempre y cuando seamos capaces de interpretarlos.

Obtener conocimiento a partir de grandes volúmenes de datos implica un gran desafío ya que plantea la necesidad de explorar un universo de conocimientos, técnicas y tecnologías provenientes de diferentes áreas como lo son las matemáticas, en particular la estadística, las ciencias de la información y la informática.

El presente curso pretende utilizar aplicaciones y tecnologías para recopilar, almacenar, analizar, compartir y proporcionar acceso a datos para ayudar a los usuarios a tomar mejores decisiones.

### V - Objetivos / Resultados de Aprendizaje

El objetivo general de la asignatura es suministrar al alumno conocimientos sólidos referidos al campo de ciencia de datos, examinando su amplio conjunto de principios fundamentales, tecnologías y aplicaciones para recopilar, almacenar, analizar y visualizar datos para colaborar con los usuarios en la reducción de incertidumbre y toma de decisiones fundamentadas.

Se espera que al finalizar el curso el alumno sea capaz de:

- \* Introducir y utilizar metodologías para el desarrollo de un proyecto de ciencia de datos que incluya el proceso ETL (Extract, transform and load), minería de datos y visualización de datos.
- \* Interpretar de manera correcta los resultados obtenidos luego de la aplicación de técnicas de análisis y visualización de datos.
- \* Desarrollar la capacidad de reconocer la estrategia a utilizar en base al problema planteado.
- \* Promover el uso de nuevas tecnologías para el análisis de datos.
- \* Desarrollar habilidades para trabajo en equipo, comunicación efectiva en forma oral, visual y escrita.
- \* Promover el aprendizaje autónomo.

## VI - Contenidos

### UNIDAD 1: Introducción a la Ciencia de Datos

¿Qué es la ciencia de datos? Justificación. Tipos de problemas que se pueden resolver mediante la ciencia de datos. Marco metodológico para el análisis de datos. Metodología fundamental para la ciencia de datos. Metodología KDD y fases del proceso de extracción del conocimiento.

### UNIDAD II: Preprocesamiento de los datos.

Datos. Tipos y características de datos. Limpieza de datos (data cleaning). Análisis exploratorio de datos (Análisis univariado, bivariado). Ruidos y anomalías (Outliers). Normalización de datos. Discretización y Binarización de variables. Selección de características. Correlación de datos.

### Unidad V: Análisis de datos.

Análisis Exploratorio de los Datos. Aprendizaje de modelos predictivos. Predicción con datos de entrenamiento reducidos. Métodos semi-supervisados. Validación de modelos. Fusión de información. Aprendizaje de modelos descriptivos. Agrupamiento. Cohesión y separación de grupos. Autocodificadores. Reglas de asociación. Análisis de datos secuenciales. Codificadores y decodificadores.

### Unidad VI: Visualización.

El proceso de la Visualización. Estados y transformaciones de los datos. Variables Visuales. Interacciones. Herramientas de visualización. Dashboard y reportes.

## VII - Plan de Trabajos Prácticos

Laboratorio 1: Pre-procesamiento de datos

Laboratorio 2: Aprendizaje supervisado y no supervisado

Laboratorio 3: Utilización de herramientas de análisis y modelado de datos.

Laboratorio 4: Utilización de herramientas de visualización para la muestra de resultados.

## VIII - Regimen de Aprobación

Régimen para Regularizar el curso:

1. Tener un mínimo de 60% de asistencia a las clases teóricas y prácticas.
2. Aprobar los laboratorios con al menos el 60%.
3. Aprobar el trabajo final integrador con al menos el 60%.

Si cualquier punto no fuese cumplimentado, implicará que el alumno pase a condición de libre.

Régimen para promocionar el curso:

1. Tener un mínimo de 80% de asistencia a las clases teóricas y prácticas.
2. Aprobar los laboratorios con al menos el 70%.
3. Aprobar el trabajo final integrador con al menos el 70%.

La nota final se computará promediando las notas obtenidas en cada uno de los puntos mencionados previamente.

El curso no admite rendir el examen final en condición de Libre por las características inherentemente prácticas de la misma.

### **IX - Bibliografía Básica**

- [1] Data Science Concepts and Techniques with Applications. Usman Qamar, Muhammad Summair Raza. Springer 2020, ISBN: 978-981-15-6132-0.
- [2] Probabilidad y Estadística para Ingeniería y Ciencias. 9th Edición. Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers and Keying Ye. Pearson Education, 2012. ISBN: 978-607-32-1417-9.
- [3] Hernández Orallo, Ramírez Quintana, Ferri Ramírez. Introducción a la Minería de Datos. Prentice Hall. 2004. ISBN 84-205-4091-9.
- [4] Tamara Munzner and Eamonn Maguire. Visualization analysis and design. A K Peters visualization series. CRC Press, Boca Raton, FL, 2015.
- [5] Wolfgang Aigner, Silvia Miksch, Heidrun Schumann, and Christian Tominski. Visualization of Time-Oriented Data. Springer Publishing Company, Incorporated, 1st edition, 2011.

### **X - Bibliografía Complementaria**

### **XI - Resumen de Objetivos**

Mostrar la diferencia entre el análisis de datos y la ciencia de datos  
Identificar las fuentes de obtención y almacenamiento de los datos  
Mejorar la calidad y la utilidad de un conjunto de datos.  
Realizar una correcta representación del conjunto de datos a trabajar.  
Reconocer el mejor algoritmo de predicción o descripción para un conjunto de datos en particular.  
Realizar representaciones visuales significativas y acordes al tipo de dato ingresado.

### **XII - Resumen del Programa**

El Proceso de Extracción de Conocimiento. Recolección, almacenamiento y acceso de datos. Preprocesamiento de los datos.  
Representación de datos. Análisis de datos. Proceso de Visualización de datos.

### **XIII - Imprevistos**

### **XIV - Otros**

Contacto  
laboratoriodedatos.unsl@gmail.com  
Box 8. 1° Piso. Bloque 2.