

**ESPECIALIZACIÓN EN INGENIERÍA EN SISTEMAS DE INFORMACIÓN**  
**MAESTRÍA EN INGENIERÍA EN SISTEMAS DE INFORMACIÓN**

**1. Nombre de la actividad curricular:** Minería en grandes volúmenes de datos

**2. Año Académico:** 2020

**3. Docente:** Dr. Mariano Rubiolo

**4. Fundamentación**

El avance y la popularidad de los sistemas de información y las tecnologías de la comunicación trajeron como consecuencia la producción diaria de “Grandes Volúmenes de Datos”. Estos datos no sólo se producen en gran cantidad, sino que de forma variada y a gran velocidad, llegando a ser necesario analizar la veracidad de los mismos. Esta era de los grandes volúmenes de datos plantea un nuevo desafío a las empresas, industrias, organizaciones y comunidad científica del mundo entero, las que diariamente intentan analizar y aprender, no sólo de la gran cantidad de datos que ellas mismas producen, sino también de los datos globales de los que disponen y consumen.

La Minería de Datos es un proceso que utiliza un conjunto de técnicas y tecnologías que permiten explorar base de datos, de manera automática o semiautomática, y analizar los datos contenidos en las mismas con el objetivo de descubrir patrones, tendencias o reglas que permitan explicar su comportamiento. En la era de los grandes volúmenes de datos, el proceso de minería de datos se ve afectado por las características de volumen, variedad, velocidad y veracidad de los datos. Por lo tanto, las técnicas y tecnologías de exploración y análisis deben adaptarse al nuevo proceso de minería de grandes volúmenes de datos.

Para responder a este nuevo desafío, los profesionales del área de sistemas de información, informática e industriales, en su instancia de formación a nivel de posgrado, deben adquirir conocimiento específico de máximo nivel que les permita llevar a cabo de manera eficiente procesos de minería de grandes volúmenes de datos en diferentes tipos de organizaciones. El fundamento de este curso de posgrado es el estudio de los últimos avances teórico-prácticos de las técnicas para realizar minería en grandes volúmenes de datos, adaptarlas e integrarlas a los sistemas existentes o desarrollar nuevos sistemas y evaluar su desempeño, acorde a las necesidades de cada organización.

**5. Objetivos**

El objetivo de este curso es el estudio y uso de modelos de minería de datos aplicados a grandes volúmenes de datos. Estos datos pueden poseer variada procedencia y formato. Se estudian modelos, métodos y herramientas para el análisis de un conjunto de datos, y cómo se pueden utilizar para el contexto en que estos datos existan en gran cantidad.

Al finalizar el curso se pretende que el alumno adquiriera los conocimientos en el tema y sea capaz de:

- comprender el proceso de minería de grandes volúmenes de datos cuyas actividades comprenden: la obtención de los datos, el preprocesamiento y análisis de los mismos, y la generación del resultado final.
- comprender los fundamentos teóricos básicos de los modelos de minería de datos presentados.
- utilizar y evaluar diferentes herramientas tradicionales y no tradicionales para el análisis y modelado de problemas reales de minería de datos.
- desarrollar, implementar y evaluar herramientas de minería de datos en el campo de la Ingeniería en

Sistemas de Información.

## 6. Contenidos

### Contenidos analíticos:

El contenido del curso se organiza en las siguientes unidades:

#### *Unidad I*

Proceso de Minería de datos. Adaptación del proceso para el contexto de grandes volúmenes de datos. Ciencia de datos.

#### *Unidad II*

Introducción a los grandes volúmenes de datos. Características: volumen, velocidad, variedad y veracidad. Ecosistema de grandes volúmenes de datos. Técnicas y algoritmos de procesamiento de grandes volúmenes de datos. Herramientas de aplicación. Gestión de datos en gran escala. Tecnologías en la nube.

#### *Unidad III*

Análisis Exploratorio de Datos. Problemáticas fundamentales de datos: Consistencia, Valores atípicos, ruido y ausencia de valor. Pre-procesamiento de datos. Reducción de la dimensionalidad. Datos faltantes. Visualización de datos. Técnicas, herramientas y algoritmos.

#### *Unidad IV*

Clasificación. Regresión. Predicción. Técnicas, herramientas y algoritmos.

#### *Unidad V*

Segmentación. Agrupamiento. Ensamble de modelos. Técnicas, herramientas y algoritmos.

## 7. Metodología de Enseñanza y Formación práctica

Las clases serán del tipo teórico-práctico, en donde se expondrán los conceptos teóricos y se realizará la discusión de problemas típicos, casos de estudios presentados por el docente y propuestos por los alumnos, con el propósito de afianzar los conocimientos. El resto serán clases prácticas, donde se trabajará en aplicaciones concretas de los conceptos presentados, y se favorecerá la discusión entre alumnos a través de la presentación de casos de estudio basados en situaciones reales, con consignas que ellos deberán resolver.

## 8. Carga horaria total

Carga horaria teórica	Carga horaria práctica	Carga horaria total
24	36	60

## 9. Modalidad de Evaluación

En primera instancia, presentar y defender un trabajo práctico final en el que se aplique los conceptos

analizados en el curso a un caso de estudio real, preferentemente del ámbito laboral del alumno, para favorecer la transferencia del conocimiento adquirido. En segunda instancia, aprobar un examen final escrito individual que se tomará después de finalizado el curso, para acreditar los conocimientos vertidos en el curso.

### **10. Requisitos de aprobación y promoción**

Los alumnos deberán registrar un 80% de asistencia y aprobar las instancias de evaluación anteriormente indicadas. La calificación se expresará en escala numérica de cero (0) a diez (10) sin decimales. Para la promoción se requerirá la norma mínima de siete (7). (Extraído de la Ordenanza N° 1313)

### **11. Infraestructura y equipamiento**

La infraestructura y ámbitos a utilizar en el dictado son los siguientes:

1. Campus virtual: el material bibliográfico del curso, las presentaciones y los enunciados de las ejercitaciones y trabajos prácticos se encuentran disponibles en el campus virtual de la Facultad Regional Santa Fe.
2. Aulas: las clases teóricas se desarrollan en un aula con capacidad para 30 estudiantes, equipo de proyección y acceso a internet mediante conexión WiFi. Todo el equipamiento mencionado es empleado en el dictado de las clases teóricas.
3. Laboratorio: las clases prácticas se desarrollan en un laboratorio móvil compuesto por notebooks de la Facultad Regional Santa Fe, o por las notebooks personales de los alumnos.

### **12. Bibliografía**

DEAN, J. Big data, Data mining and Machine Learning. (2014). Wiley.

LESKOVEC, J., RAJARAMAN, A., ULLMAN, J.D. Mining of Massive Datasets (2015). 3rd Edition. Cambridge University Press.

WITTEN, I.H., FRANK, E. Data mining. Practical machine learning tools and techniques (2016). 4th edition. Elsevier.

OLSON, D., DELEN, D. Advanced data mining techniques (2008). Springer.

WANG, I., XIUJU F. Data mining with computational intelligence (2005). Springer.