



FaCENA - UNNE



UNNE

**CURSO DE POSGRADO**

# **CIENCIA DE DATOS**

**INFORMACIÓN AMPLIADA**

**Tipo de actividad:** Curso de posgrado.

**Denominación:** Ciencia de Datos.

**Destinatarios:** Profesionales con título de grado vinculados con la Informática Los cursantes deberán contar con equipamiento informático y conectividad para realizar las actividades en la plataforma virtual.

**Carga horaria:** 45 horas.

**Fecha de inicio y finalización del curso:** 9 de agosto al 7 de septiembre de 2024.

**Inscripción:** Hasta el 2 de agosto por SIU Guaraní.

**Cupos:** Mínimo 15 personas – Máximo 40 personas.

**Modalidad:** Virtual.

**Aranceles:**

- Arancel General de \$40.000
- Docentes de la UNNE \$35.000
- Docentes y estudiantes de posgrado de FaCENA \$30.000

**Fundamentación**

El curso de Ciencia de Datos responde a la demanda de una propuesta específica en el área que se dirija al desarrollo integral de habilidades en relación a la gestión de datos y su procesamiento para dar soporte a propuestas de investigación con datos.

La Ciencia de Datos propone la integración de estrategias para el abordaje de problemas específicos a partir de un conjunto de habilidades, incluyendo la visualización y el procesamiento de los datos, el entrenamiento de modelos de aprendizaje automático, y estrategias para la gestión de Grandes Datos. Este curso incluye estos elementos, así como la presentación de enfoques para la gestión de proyectos de Ciencia de Datos.

Una vez concluido el curso, el estudiante será capaz de diseñar, implementar y evaluar proyectos de Ciencia de Datos.

**Objetivos generales**

Que el estudiante logre:

- Apropiarse de los conceptos fundamentales de la Ciencia de Datos, sus antecedentes en el ámbito de la estadística y su rol actual como herramienta para asistir en la toma de decisiones en las organizaciones.
- Desarrollar estrategias representativas de la Ciencia de Datos que incluyen la visualización de datos en distintas etapas de los proyectos; el procesamiento de los datos y la preparación para su uso en productos de datos; una introducción al entrenamiento de modelos de aprendizaje automático, su evaluación y optimización; y estrategias para la gestión de Grandes Datos.
- Desarrollar las habilidades necesarias para la gestión de proyectos de Ciencia de Datos con la presentación de enfoques de desarrollo y metodologías específicas.

### **Contenidos**

- Concepto de Ciencia de Datos. Relación con otras áreas de conocimiento. Herramientas de software para Ciencia de Datos.
- Visualización de datos. Criterios de visualización de datos, Distribuciones de datos, Reportes, Estadística aplicada a los datos. Visualizaciones de acuerdo al dominio de aplicación (texto, análisis de redes sociales).
- Preparación y análisis de datos. Acceso y orígenes de datos. Limpieza de datos, Integración de datos. Análisis exploratorio de datos.
- Aprendizaje de Máquina. El flujo de trabajo del Aprendizaje de Máquina. Ingeniería de características (features), Algoritmos de Supervisado y No supervisado. Entrenamiento de modelos y optimización,
- Evaluación y optimización de modelos.
- Gestión de Grandes Datos. Aplicaciones del Big Data. Modelado y gestión de soluciones Big Data. Aprendizaje automático con Big Data. Servicios cloud como soporte a soluciones de Ciencia de datos.
- Gestión de proyectos de Ciencia de Datos. Conceptos de Gubernanza de la Tecnología de la Información y de Datos. Definición de Proyectos. Metodologías Ágiles en proyectos de Ciencia de Datos. Metodología CRISP- DM. ASUM-DM. TDSP.

### **Metodología de enseñanza**

El curso prevé el desarrollo de los contenidos teóricos organizados de manera que el estudiante se vaya familiarizando con las distintas áreas abordadas por la Ciencia de Datos, partiendo de los conceptos fundamentales y dirigiéndose hacia los aspectos específicos de la gestión de proyectos. Adicionalmente, el curso prevé el uso de herramientas software de acceso abierto y de uso extendido en la gestión de proyectos de Ciencia de Datos.

Actividades prácticas: Planteo y resolución de problemas. Utilización de herramientas y entornos de desarrollo. Se propondrán actividades prácticas, articulando entre modalidades síncronas y asincrónicas. Se establecerán consignas de actividades prácticas y resúmenes teóricos en formato de videos explicativos utilizando plataformas basadas en internet. Los objetivos de cada actividad práctica serán presentados en forma de documento web, utilizando la plataforma de educación a distancia de la institución. Cada contenido estará soportado por recursos de evaluación en línea (cuestionarios) y foros de consultas específicos.

Las actividades de desarrollo de software se realizarán mediante plataformas basadas en internet mediante herramientas software de código abierto (ej: Jupyter- notebooks).

Modalidad de supervisión: Presencial (remoto), en clases. No presencial, mediante evaluaciones en proceso teórico/prácticos.

Modalidad de evaluación: Participación activa de los estudiantes en clases teóricas y prácticas. Resolución de problemas teórico/prácticos.

### **Instancias de evaluación y aprobación**

Se prevén dos seminarios a realizarse en las semanas 3 y 5 del curso. Los estudiantes deberán participar en el foro donde deberán exponer sobre cada uno de los seminarios brindados por los expertos. La participación se podrá realizar en grupos.

### **Requisitos de aprobación del curso**

Participación de los estudiantes en clases teóricas y prácticas. Resolución y aprobación de las actividades prácticas. Aprobación del Trabajo Integrador. La calificación respetará la Escala de Calificación aprobada en cada institución. Para la emisión del certificado de aprobación del curso el estudiante debe cumplir con el pago de la totalidad del arancel estipulado.

### **Equipo Docente**

Profesor Responsable: Dr. Horacio Kuna (UNaM).

Dr. Eduardo Zamudio (UNaM), Dr. Hemán Merlino (UNLa), Dra. María Inés Pisarello (UNNE)

Coordinación: Mgter. Gladys Dapozo. (UNNE)

### **Cronograma estimativo:**

El curso iniciará el viernes 9 de agosto y se extenderá por cinco semanas y la entrega del Trabajo Integrador.

Se prevén los siguientes encuentros sincrónicos:

<b>Día</b>	<b>Horario</b>
Viernes 9 de agosto	18h a 21h
Sábado 10 de agosto	09h a 12h
Viernes 23 de agosto	18h a 21h
Sábado 24 de agosto	09h a 12h
Viernes 6 de septiembre	18h a 21h
Sábado 7 de septiembre	09h a 12h

### **Bibliografía básica**

- Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., & Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0 Step-by-step data mining guide.
- Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., & Kem, J. (2001). Manifiesto far agile software development.

- Fowler, M., & Highsmith, J. (2001). The agile manifesto. *Software Development*, 9(8), 28-35.
- García Martínez, R. Britos, P. Martins, S. Baldizzoni, E. (2013). *Explotación de Información. Ingeniería de Proyectos*. Nueva Librería.
- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- Larose, D. T. (2006). *Data mining methods & models*. John Wiley & Sons.
- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- Cam Davidson-Pilon. *GitHub - Bayesian Methods for Hackers: An introduction to Bayesian methods + probabilistic programming with a computation/understanding- first, mathematics second point of view*. Retrieved November 28, 2018,
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. Retrieved from <https://www.deeplearningbook.org/>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning (Vol.103)*. New York, NY: Springer New York. <https://doi.org/10.1007/978-1-4614-7138-7>
- Leskovec, J., Rajaraman, A., & Ullman, J. D. (2014). *Mining of Massive Datasets*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139924801>
- Marsland, S. (2009). *Machine learning: an algorithmic perspective*. CRC Press.
- Ng, A. (2018). *Machine Learning Yearning. Draft Version*. Retrieved from [https://gallery.mailchimp.com/dc3a7ef4d750c0abfc19202a3/files/5dd91615-3b3f-4f5d-bbfb-4ebd8608d330/Ng\\_MLY01\\_13.pdf](https://gallery.mailchimp.com/dc3a7ef4d750c0abfc19202a3/files/5dd91615-3b3f-4f5d-bbfb-4ebd8608d330/Ng_MLY01_13.pdf)
- Richert, W., & Coelho, L. P. (2013). *Building Machine Learning Systems with Python*. Packt Publishing.
- Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding Machine Learning*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781107298019>
- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: an introduction*. MIT Press. Retrieved from <https://mitpress.mit.edu/books/reinforcement-learning>