

**Código de asignatura: 47AA14**

<b>Nombre del programa académico</b>	Maestría en Ingeniería Eléctrica		
<b>Nombre completo de la asignatura</b>	Introducción a la ciencia de los datos		
<b>Número de créditos ECTS por categoría</b>	Ciencias naturales y matemáticas	Módulos profesionales y especiales	Humanidades y ciencias sociales y económicas
	3	3	1
<b>Semestre y año de actualización</b>	Semestre 2 – 2023		
<b>Semestre y año en que se imparte</b>	Semestre 2 – Año 1		
<b>Tipo de asignatura</b>	[ ] Obligatoria [X] Electiva		
<b>Director o contacto del programa</b>	Julián David Echeverry Correa		
<b>Coordinador o contacto de la</b>	Andrés Escobar Mejía		
<b>Descripción y contenidos</b>			
<b>1. Breve descripción</b> La asignatura de Introducción a la Ciencia de los Datos es de naturaleza teórico práctica, el propósito que tiene es el estudio de las técnicas matemáticas, probabilísticas, estadísticas y de aprendizaje de máquina relacionadas con el análisis, presentación, gestión y visualización de datos, extracción de información, el descubrimiento de patrones ocultos y correlaciones desconocidas en grandes bases de datos; y el manejo de herramientas de software para el procesamiento, análisis y toma de decisiones a partir de los datos.			
<b>2. Objetivos. Se espera que al finalizar este curso el estudiante esté en la capacidad de:</b> - OA1: Introducir al estudiante a los conceptos de la ciencia de los datos y sus aplicaciones. Se corresponde con los RAP: 3, 4, 9, 11 y 12  -OA2: Formar al estudiante en habilidades prácticas en el manejo, procesamiento, preparación y análisis de datos. Se corresponde con los RAP: 3, 4, 6, 7 y 13  - OA3: Estudiar los esquemas y modelos básicos del aprendizaje de máquina (aprendizaje supervisado y no supervisado). Se corresponde con los RAP: 6, 7, 8, 10 y 13  - OA4: Estudiar el manejo de herramientas de software para el procesamiento, análisis y toma de decisiones a partir de datos. Se corresponde con los RAP: 6, 7, 8, 10 y 13			
<b>3. Resultados de aprendizaje. Los propósitos de formación en el estudiante de posgrado son:</b> -RAA-1. Definir los conceptos fundamentales de las técnicas de reprocesamiento y análisis de datos. Se corresponde con los RAP: 3, 4, 6, 9, 10 y 13 -RAA-2. Distinguir entre enfoques supervisados y no supervisados para el entrenamiento de modelos de análisis de datos. Se corresponde con los RAP: 10, 11 y 13 -RAA-3. Identificar los algoritmos óptimos para una mejor eficiencia en el procesamiento de grandes volúmenes de información. Se corresponde con los RAP: 6, 7, 9, 11 y 13 -RAA-4. Aplicar modelos de extracción de información y de descubrimiento de patrones ocultos en bases de datos. Se corresponde con los RAP: 6, 7, 9, 12 y 13 -RAA-5. Resolver problemas de análisis de datos mediante algoritmos computacionales. Se corresponde con los RAP: 9, 10, 11, 12 y 13 -RAA-6. Aplicar técnicas de análisis estadístico y probabilístico en el desarrollo de sistemas automáticos de análisis de datos. Se corresponde con los RAP: 6, 7, 9 y 12 -RAA-7. Capacidad de trabajo en equipo. Se corresponde con los RAP: 3, 6, 7, 9, 10, 13			
<b>4. Contenido</b> - T-1. Introducción y conceptos básicos (6 h). - T-2. Reprocesamiento de datos y análisis descriptivo (8 h).			

- T-3. Modelos de regresión y ajuste de datos (10 h).
- T-4. Técnicas de aprendizaje automático (8 h)
- T-5. Transformación y visualización de datos (8 h)
- T-6: Herramientas prácticas para la Ciencia de los Datos (8 h)

**5. Requisitos.** Los definidos en requisito de admisión de la IES.

## **6. Recursos**

### **Bibliografía**

#### 6. Recursos

Libros de texto:

- [1] Witten et al. Data Mining. Morgan Kaufmann. 3rd edition. 2011.
- [2] Bell, J. (2015) Machine Learning - Hands-on for developers and Technical Professionals. Ed. Wiley
- [3] Murphy K.P. (2012) Machine Learning: A Probabilistic Perspective. The MIT Press, 1st Edition.
- [4] Duda R.O., Hart P.E. and Stork D.G. (2001), Pattern Classification, John Wiley & Sons, 2nd Edition.
- [5] Hastie T., Tibshirani R. and Friedman J.H. (2009), The Elements of Statistical Learning, Springer, 2nd Edition.
- [6] Schölkopf B., Smola A. (2001), Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, The MIT Press, 1st Edition.
- [7] Schutt R. and O'Neil C. (2013), Doing Data Science, O'Reilly, 1st Edition.
- [8] Vercellis C. (2009), Business Intelligence: Data mining and optimization for decision making, John Wiley & Sons, 1st Edition.

Herramientas informáticas

- Software de simulación y procesamiento de información:

o Weka

o RapidMiner

o Matlab

o Knime

- Guías de clase desarrolladas por el docente

- Plataforma Google Classroom

Recursos de internet:

- <https://www.ieee.org/>
- <http://www.kaggle.com>

## **7. Herramientas técnicas de soporte para la enseñanza**

- Tareas enfocadas al desarrollo de algoritmos para la solución de problemas en el área de la ciencia de los datos.
- Desarrollo de sistemas automáticos de procesamiento de información.
- Ejercicios individuales, grupales y tareas propuestas.
- Otras herramientas técnicas presentadas previamente en el ítem 6 (Recursos).

## **8. Trabajos en laboratorio y proyectos**

- Se proponen, en el desarrollo del curso, proyectos de laboratorio y trabajos que relacionen las temáticas tratadas en él.

## **9. Métodos de aprendizaje**

- Clases magistrales apoyadas por tecnologías de la información para la presentación de las clases y como apoyo al proceso de aprendizaje del estudiante.
- Lecturas de artículos especializados y de material adicional propuesto por el profesor como parte del trabajo individual semanal del alumno.
- Tutorías.

## **10. Métodos de evaluación**

Para la obtención de la nota final se realizan distintas entregas escritas individuales y grupales durante el semestre. Están previstas las siguientes entregas:

- Entrega 1: Taller propuesto para evaluar los objetivos de los temas Introducción y conceptos básicos (T1) y Preprocesamiento de datos y análisis descriptivo (T2). (RA1 y RA2). Entrega de carácter individual.
- Entrega 2: Taller propuesto para evaluar los objetivos de los temas Modelos de regresión y ajuste de datos (T3), Técnicas de aprendizaje automático (T4) y Transformación y visualización de datos (T5). (RA1, RA2, RA3, RA4, RA5, RA6). Entrega de carácter individual.
- Entrega 3: Taller propuesto para desarrollar un sistema de procesamiento y análisis de datos. Se evalúan todos los temas incluyendo el tema Herramientas prácticas para la Ciencia de los Datos (T6). (RA1, RA2, RA3, RA4, RA5, RA6 y RA7). Entrega de carácter individual.
- Las entregas tienen un valor del 40% sobre la nota final
- Presentación del anteproyecto de curso (10%)
- Presentación del proyecto de curso (50%). Proyecto propuesto para desarrollar un sistema de procesamiento y análisis de datos. Se evalúan todos los temas incluyendo el tema Herramientas prácticas para la Ciencia de los Datos (T6). (RA1, RA2, RA3, RA4, RA5, RA6 y RA7).

El trabajo debe ser presentado en la fecha establecida, no se admiten entregas posteriores a la fecha establecida. El trabajo debe ser sustentado y su exposición hará parte de la evaluación del trabajo.

En esta asignatura se evalúa la competencia transversal\* de trabajo en equipo (RA7). La evaluación se realiza en los talleres y tiene un peso de 10%.