

**DOCTORADO EN INGENIERÍA
UNIVERSIDAD TECNOLÓGICA DE PEREIRA
MICROCURRÍCULO**

INFORMACIÓN DE LA ASIGNATURA

Programa académico	Doctorado en Ingeniería
Nombre de la asignatura	Introducción a la Ciencia de los Datos
Código de asignatura	DV144
Semestre y año en que se imparte la asignatura	Segundo semestre del 2022
Profesor de la asignatura	Julián David Echeverry Correa
Afiliación institucional del profesor	Profesor Asociado - UTP
Correo electrónico del profesor	jde@utp.edu.co

DESCRIPCIÓN

1. Descripción y justificación de la asignatura

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.

2. Objetivo de la asignatura

Al finalizar esta asignatura el estudiante estará en capacidad de:

- Entender los conceptos de la ciencia de los datos y sus aplicaciones.
- Aplicar habilidades prácticas en el manejo, procesamiento, preparación y análisis de datos.
- Estudiar los esquemas y modelos básicos del aprendizaje de máquina (aprendizaje supervisado y no supervisado).
- Estudiar el manejo de herramientas de software para el procesamiento, análisis y toma de decisiones a partir de datos.

3. Contenido de la asignatura

- T1: Introducción y conceptos básicos (6 h)
- T2: Preprocesamiento de datos y análisis descriptivo (8 h)
- T3: Modelos de regresión y ajuste de datos (10 h)
- T4: Técnicas de aprendizaje automático (8 h)
- T5: Transformación y visualización de datos (8 h)
- T6: Herramientas prácticas para la Ciencia de los Datos (8 h)

4. Requisitos

- Competencias en la resolución de problemas que involucren técnicas del álgebra lineal.
- Adecuada capacidad de lecto-escritura.
- Capacidad de llevar a cabo una adecuada metodología de investigación.

5. Evaluación de la asignatura

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). (20%) Entrega de carácter grupal.

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). (20%) Entrega de carácter grupal.

Entrega 3: Anteproyecto en donde se presenten los objetivos del proyecto final, se identifique la problemática a resolver, se plantee la hipótesis de investigación y se enuncien las posibles soluciones metodológicas al problema. (10%) Entrega de carácter individual.

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).

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.

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 computacionales

- o Python
- o Software de simulación y procesamiento de información:
- o Weka
- o RapidMiner
- o Matlab
- o Knime
- o Orange
- o Guías de clase desarrolladas por el docente
- o Plataforma Google Classroom

- Recursos en línea

- o Bases de datos suscritas por la Universidad Tecnológica de Pereira
- o <https://www.ieee.org/>
- o <http://www.kaggle.com>