

CARTA AL ESTUDIANTE
XS0126 Minería de Datos y Análisis Multivariado
I Semestre 2025

	GRUPO 01	GRUPO 02
Docente:	Ricardo Alvarado Barrantes	Shirley Rojas Salazar
Correo:	estad.ucr@gmail.com	shirleyelena.rojas@ucr.ac.cr
Teléfono:	84021263	87737058 / 25119172
Clases:	L: 1:00-2:50 pm (108 CE) J: 1:00-2:50 pm (240 CE)	L: 1:00-2:50 pm (240 CE) J: 1:00-2:50 pm (014 CE)
Consulta:	M: 1:00-3:00 pm	M: 11:00am-1:00 pm
Zoom:	https://udecr.zoom.us/j/86862167699	https://udecr.zoom.us/j/89709609762
Materiales:	https://6f33fa7f78ea46e2aaca-my.sharepoint.com/:f/g/personal/ricardo_alvarado_ucr_ac_cr/EmBNliWB3e9NrFUia28TPYEBW1MAX_K6HFJuKh41vtXbA?e=CDLyt5	

Características del curso

Horas:	2 teoría y 2 práctica
Créditos:	4
Requisitos:	XS2130 Modelos de Regresión Aplicados, XS0130 Programación para Estadística II (Equiv. XS2230), XS0123 Modelos Probabilísticos II (Equiv.XS2330)
Correquisitos:	ninguno
Ciclo:	VII (Plan 1), VI (Plan 2)
Clasificación:	propio

Descripción

En este curso se presentan diversas técnicas estadísticas multivariadas para minería de datos con un enfoque teórico-práctico. Se estudia el análisis de componentes principales, el análisis de agrupamiento, el escalamiento multidimensional, técnicas de clasificación y técnicas de clasificación basadas en árboles. Además de los conocimientos teóricos, el estudiante debe aplicar las técnicas usando lenguajes de programación en prácticas y trabajos de investigación.

Objetivo general

Aplicar las técnicas básicas de minería de datos, gráficas y cuantitativas, para contribuir en la resolución de problemas que se presenten en distintos contextos mediante la interpretación adecuada de los resultados.

Objetivos específicos

Al finalizar el curso el/la estudiante estará en la capacidad de:

1. Discriminar entre técnicas de minería de datos para aplicar el análisis apropiado a una situación específica.
2. Comprender los resultados de los análisis realizados con cada técnica de minería de datos para su correcta interpretación.
3. Aplicar técnicas de minería de datos para la obtención y caracterización de conglomerados.
4. Aplicar distintas técnicas de minería de datos para establecer un procedimiento adecuado a un problema de clasificación.

Habilidades y conocimientos (perfil de salida)

Habilidades	Conocimientos
HM02 - Emplear lenguaje matemático para expresar propiedades estadísticas	CM04 - Conocimientos avanzados de aspectos teórico-matemáticos que dan fundamento al uso de las técnicas de análisis estadístico
HE01 - Identificar y aplicar modelos estadísticos apropiados según el problema de investigación	CE01 - Conocimientos avanzados de técnicas clásicas y modernas de análisis de datos univariados y multivariados para comprender los fenómenos en diferentes áreas del conocimiento CE02 - Aplicación de modelos estadísticos a problemas de diversas áreas del conocimiento
HE04 - Aprender de forma autónoma metodología estadística	CE06 - Conocimientos básicos en la exploración y actualización de metodologías estadísticas
HT02 - Crear, interpretar y modificar programas de código escrito en lenguajes de programación (ej: R, Python y SQL)	CT08 - Conocimientos intermedios en lenguajes de programación para análisis estadístico (ej: R, Python)
HC04 - Comunicar asertivamente mediante la manipulación creativa de datos pero manteniendo su integridad	CC04 - Conocimientos avanzados en construcción de cuadros y gráficos
HC05 - Comunicar adecuadamente de forma escrita diferentes elementos del quehacer estadístico	CC06 - Conocimientos intermedios de técnicas de escritura de artículos e informes técnicos CC08 - Conocimientos en técnicas de organizar y sintetizar la información

Contenidos

1. Análisis de componentes principales (PCA):
 - a) Características de los componentes principales.
 - b) Construcción: valores y vectores propios, variancia explicada, cálculo de los puntajes en los componentes principales, uso de covariancias o correlaciones, número de componentes principales.
 - c) Representación gráfica: biplot.
 - d) Evaluación de resultados: reproducción de matriz de variancias, correlación entre componentes y variables originales.
2. Análisis de agrupamientos y escalamiento multidimensional (MDS):
 - a) Distancias entre individuos (variables continuas, nominales, mezclas).
 - b) Distancias entre grupos (vecino más cercano, vecino más lejano, salto promedio).
 - c) Selección de variables para el análisis / estandarización.
 - d) Agrupamientos jerárquicos: algoritmo, representación (dendograma).
 - e) Métodos de k-medias y k-medoides: algoritmo. selección del número de clusters.
 - f) Validación: número de clústers.
 - g) Presentación de resultados: mapas de calor.
 - h) Escalamiento multidimensional.
3. Técnicas de clasificación:
 - a) Modelos de clasificación:
 - Regresión logística binomial y multinomial
 - Análisis discriminante
 - K-vecinos más cercanos
 - b) Métodos basados en árboles:
 - Árboles de decisión
 - Agregación de Bootstrap (bagging), bosques aleatorios (random forest)
 - Boosting
 - BART
 - c) Evaluación del modelo:
 - Métricas de desempeño.
 - Validación cruzada.

Metodología

El curso es teórico-práctico y exige el uso frecuente de la computadora. Se espera que el estudiante aprenda los fundamentos teóricos de las técnicas de minería de datos y que aplique las técnicas a archivos de datos utilizando lenguajes de programación estadística. Se propone una combinación de actividades, tales como:

1. Presentaciones teóricas: lecciones por parte del docente donde se explican los conceptos y sus aplicaciones.
2. Ejercicios en clase para que las sesiones sean activas.
3. Laboratorios: sesiones estructuradas con ejercicios sobre los contenidos desarrollados en las clases teóricas con solución disponible. Durante las sesiones de laboratorio se utilizan lenguajes de programación (por ejemplo, R y Python).
4. Prácticas: ejercicios fuera de clase que incluyen aplicaciones con datos para ser analizados, así como interpretaciones de los resultados.
5. Trabajos de investigación: con el objetivo de poner en práctica los conocimientos, el estudiantado debe enfrentar problemas reales que debe analizar y presentar con el formato de un artículo científico.
6. Revisión bibliográfica: los estudiantes deben seleccionar un artículo de una revista científica donde se aplique el PCA, exponerlo y entregar un reporte sobre el artículo con énfasis en la metodología, incluyendo una crítica sobre el mismo.

Evaluación

Primer examen	25%
Segundo examen	25%
Primer artículo	20%
Segundo artículo	20%
Trabajos adicionales	10%

Cronograma

Mes	Módulo	L	J	Actividad
MAR	Análisis de componentes principales	10		
			13	
		17		
			20	
ABRIL	Análisis de agrupamiento y MDS	24		
			27	
		31		
			3	Presentación aplicaciones PCA
		7		
			10	Anteproyecto
		14		SEMANA SANTA
			17	SEMANA SANTA
MAYO	Técnicas de clasificación	21		
			24	
		28		
			1	FERIADO
		5		
			8	
		12		Artículo 1 (escrito y oral)
			15	
JUNIO	Técnicas de clasificación	19		Examen No.1
			22	
		26		
			29	
		2		
			5	
		9		
			12	
JULIO	Técnicas de clasificación	16		
			19	
		23		
			26	Artículo 2 (escrito y oral)
JULIO	Técnicas de clasificación	30		
			3	
		7		Examen No.2
			10	

Bibliografía

- Cichosz, Pawel. (2015). Data Mining Algorithms: Explained Using R. Wiley.
- Everitt, B. y Hothorn, T. (2011). An Introduction to Applied Multivariate Analysis with R. Springer.

BIBLIOTECA LUIS DEMETRIO TINOCO 519.535.028.5 E93i

- Hair, J.F. et al (2014). Multivariate Data Analysis. Pearson Education Limited.

BIBLIOTECA LUIS DEMETRIO TINOCO 519.535 M958m7 2015

- James, G., Witten, D., Hastie, T. y Tibshirani. R. (2021). An Introduction to Statistical Learning: with Applications in R (2da. Ed). Springer.
- Johnson, R. A. y Wichern, D. W. (2007). Applied Multivariate Statistical Analysis. Prentice-Hall International, Inc.

BIBLIOTECA LUIS DEMETRIO TINOCO 519.535 J68a6

- Mirkin, B (2005). Clustering for Data Mining: A Data Recovery Approach. Chapman & Hall.
- Mishra, P. (2016). R Data Mining Blueprints. Packt Publishing.
- Olson et al. (2017). Predictive Data Mining Models. Springer.
- Pan et al. (2013). Introduction to Data Mining. Pearson.
- Ramasubramanian, K y Singh, A (2017). Machine Learning Using R. Apress.
- Sarkar, D (2008). Lattice: Multivariate Data Visualization with R. Springer.

BIBLIOTECA LUIS DEMETRIO TINOCO 006.6 S245L

Reglamentación

- La reglamentación sobre sus deberes y derechos como estudiante se encuentra en el **Reglamento de Régimen Académico Estudiantil**
https://www.cu.ucr.ac.cr/normativ/regimen_academico_estudiantil.pdf
- La reglamentación y sanciones ante fraudes en las evaluaciones o comportamientos anómalos por parte de los y las estudiantes, la pueden encontrar en **Reglamento de Orden y Disciplina de los Estudiantes de la Universidad de Costa Rica** https://www.cu.ucr.ac.cr/normativ/orden_y_disciplina.pdf
- Con el fin de garantizar un espacio libre de violencia y sexismo en el desarrollo de este curso, les recomiendo que revisen **el Reglamento de la Universidad de Costa Rica contra el Hostigamiento Sexual**
https://www.cu.ucr.ac.cr/normativ/hostigamiento_sexual.pdf