



Universidad Central
Vicerrectoría Académica
PROGRAMA PARA EL DESARROLLO DE LA ASIGNATURA
FACULTAD DE INGENIERÍA Y CIENCIAS BÁSICAS

Denominación de la Asignatura: Métodos estadísticos para analítica de datos	Código: 64491090		Créditos: 4
Nombre del Profesor: Luis Andres Campos Maldonado	Grupo: 06	Horario: Jueves: 18:00-22:00. enlace: https://meet.google.com/vgj-dpjr-sto?authuser=0	
Dispositivo Pedagógico Indicado en el Sílabo: <ul style="list-style-type: none">• La Clase expositiva o dialógica• Talleres• Laboratorio de sistemas usando R y Python			
Otros dispositivos auxiliares y actividades complementarias contemplados para trayectos particulares: Trabajos en R y Python			
Formulación del problema general del cual se ocupará el curso: En el curso se construyen los fundamentos teóricos de los principales métodos de la estadística exploratoria multidimensional y se trabajan sus aplicaciones, utilizando software especializado de uso libre y comercial. Las prácticas se orientan a la aplicación de los métodos a situaciones reales que requieren apropiación de los contextos específicos, a partir de la revisión de artículos, para realizar interpretaciones adecuadas de los datos y del fenómeno en general. Para el aprendizaje de la aplicación de los métodos, el estudiante deberá realizar talleres utilizando el paquete estadístico			



Universidad Central
Vicerrectoría Académica
**PROGRAMA PARA EL DESARROLLO DE LA ASIGNATURA
FACULTAD DE INGENIERÍA Y CIENCIAS BÁSICAS**

Criterios para evaluar la adquisición de competencias de los estudiantes:		Medios (acciones y producciones de los estudiantes a los que se aplicarán los criterios):
Finalidad 1:	<i>En diferentes situaciones hacer uso de métodos de reducción de dimensionalidad para facilitar la comprensión y visualización de datos</i>	<i>Talleres que se realizarán en R o en python en los que se deba reducir la dimensionalidad, además que se deban crear visualizaciones de esos datos.</i>
Finalidad 2:	<i>En diferentes situaciones decidir el uso de aprendizaje supervisado y no supervisado y la conveniencia de ambos</i>	<i>Con base en los objetivos que se plantean con bases de datos y problemas de interés, el estudiante deberá aprender a diferenciar entre el uso de aprendizaje no supervisado o supervisado.</i>
Finalidad 3:	<i>Donde se necesite clasificar individuos y predecir probabilidades, el estudiante deberá aprender a usar distintos algoritmos de clasificación.</i>	<i>En diferentes talleres de clasificación, los estudiantes usarán métodos de clasificación y regresión, y además pronosticar comportamientos y tomar decisiones en base a estos.</i>



Universidad Central
Vicerrectoría Académica
PROGRAMA PARA EL DESARROLLO DE LA ASIGNATURA
FACULTAD DE INGENIERÍA Y CIENCIAS BÁSICAS

Trayectos a recorrer en el tratamiento del problema:		Finalidades formativas del trayecto:	Acciones y producciones de los estudiantes: Sección correspondiente en el texto guía.
Semana 1	<i>Aprendizaje no supervisado y sus aplicaciones</i>	<i>Comprender los conceptos de aprendizaje no supervisado, los tipos de aprendizaje no supervisado y los retos que existen.</i>	<i>Lectura del capítulo 1 del libro de Sarkar.</i>
Semana 2	<i>Análisis de componentes principales</i>	<i>Comprender el uso de componentes principales en situaciones de múltiples variables cuantitativas Analiza la selección y uso de PCA</i>	<i>Everitt, Capítulo 3</i>
Semana 3	<i>Análisis de componentes principales Primer avance. Problema, justificación y objetivos.</i>	<i>Comprender el uso de componentes principales en situaciones de múltiples variables cuantitativas Analiza la selección y uso de PCA</i>	<i>Everitt, Capítulo 5</i>
Semana 4	<i>Análisis Factorial</i>	<i>Utilizar el análisis factorial en situaciones de reducción de dimensionalidad. Realizar las fases exploratorias y confirmatorias de este tipo de análisis</i>	<i>Everitt, Capítulo 4</i>
Semana 5	<i>Análisis de correspondencias múltiples</i>	<i>Comprender las técnicas de visualización y correlación en variables categóricas. Aplicar la distancia Ji-cuadrado para analizar similitudes</i>	<i>Everitt, Sección 4.6</i>



Universidad Central
 Vicerrectoría Académica
PROGRAMA PARA EL DESARROLLO DE LA ASIGNATURA
FACULTAD DE INGENIERÍA Y CIENCIAS BÁSICAS

Semana 6	Agrupamiento Jerárquico Segundo avance. Marco teórico – Previa metodología-	<i>Estudiar el agrupamiento de individuos a través de múltiples atributos. Entender el agrupamiento por medio de métodos jerárquicos</i>	<i>Kassambara, capítulo 7</i>
Semana 7	Agrupamiento de K-medias	<i>Analizar el uso de agrupamiento de K-medias. Combinar el uso de métodos jerárquicos y de k-medias</i>	<i>Kassambara, capítulo 4</i>
Semana 8	Otros agrupamientos	<i>Comprender el uso de agrupamientos basados en funciones de densidad.</i>	<i>Kassambara, capítulo 19</i>
Semana 9	Aprendizaje supervisado Clasificación y regresión	<i>Comprender el uso de regresión para la predicción de eventos.</i> <i>Analizar la clasificación como método de discriminación de individuos</i>	<i>Zumel, sección 7.1</i> <i>Zumel, sección 7.2</i>
Semana 10	Clasificación y regresión Tercer avance – Metodología y primeros resultados.	<i>Entender la función de decisión, Pronosticar probabilidades y manejar incertidumbre de clasificación</i>	<i>Zumel, sección 5.2</i> <i>Zumel, sección 5.3</i>
Semana 11	Estimaciones de incertidumbre de clasificadores	<i>Usar árboles de decisión para clasificar y predecir comportamientos de individuos</i>	<i>Zumel sección 6.3</i> <i>Han, sección 8.2</i>
Semana 12	Árboles de decisión	<i>Comprender el funcionamiento de clasificadores bayesianos “ingenuos” y su aplicación en predicción de eventos.</i>	<i>Zumel, sección 6.3</i> <i>Han, sección 8.3</i>
Semana 13	Clasificadores de Naive Bayes	<i>Analizar el uso de máquinas de soporte vectorial para la clasificación y regresión con alta dimensionalidad</i>	<i>Han, sección 9.3-9.5</i>
Semana 14	Máquinas de soporte vectorial		<i>Han, sección 9.2</i>



Universidad Central
Vicerrectoría Académica
PROGRAMA PARA EL DESARROLLO DE LA ASIGNATURA
FACULTAD DE INGENIERÍA Y CIENCIAS BÁSICAS

Semana 15	Exposiciones Entrega Final	<i>Diagnosticar la efectividad de la clasificación de los métodos anteriores en problemas de data mining.</i>	<i>Han, sección 8.5-8.6</i>
----------------------	---	---	-----------------------------

EVALUACIÓN:

En cada uno de los cortes las notas se distribuirán de la siguiente manera:

ACTIVIDAD	PORCENTAJE
Quices	40%
Proyecto	60% (10 10 20 20)
Total	100%

Bibliografía:

Textos de Referencia

- Müller, A. C., & Guido, S. (2016). Introduction to machine learning with Python: a guide for data scientists. " O'Reilly Media, Inc.".



Universidad Central
Vicerrectoría Académica

**PROGRAMA PARA EL DESARROLLO DE LA ASIGNATURA
FACULTAD DE INGENIERÍA Y CIENCIAS BÁSICAS**

- Sarkar, D., Bali, R., & Sharma, T. (2018). Practical Machine Learning with Python. A Problem-Solver's Guide to Building Real-World Intelligent Systems, New York: Springer-Verlag
- Nina Zumel, John Mount, "Practical Data Science with R", Manning Publications, 2014
- Everitt, B. S. and Hothorn, T. (2011), An Introduction to Applied Multivariate Analysis with R , New York: Springer-Verlag
- Kassambara, A. (2017). *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning* (Vol. 1). STHDA.
- Jiawei Han and Micheline Kamber, "Data Mining: Concepts and Techniques", Morgan Kaufmann Publishers, 3rd ed, 2010.

Firma del Profesor:	Firma del Director del Programa:	Fecha: