



UNIVERSIDAD CENTRAL VICERRECTORÍA ACADÉMICA

PROGRAMA DE DESARROLLO DE LA ASIGNATURA

Denominación de la Asignatura: Métodos estadísticos para analítica de datos	Código: 64491090	Créditos: 4
Nombre del Profesor: Luis Andres Campos Maldonado	Grupo: 1 y 2	Período Académico: 2024 - 2
Dispositivo Pedagógico Indicado en el Sílabo: <ul style="list-style-type: none">• Clase magistral, expositiva o dialógica en mediaciones tecnológicas• Laboratorio.• Taller. Recursos educativos y herramientas de apoyo: <ul style="list-style-type: none">• Grabaciones de clases.• Notebooks de Google Colab.		Horario Grupo 1: Viernes de 18:00 a 22:00 - Remoto, enlace de conexion: https://meet.google.com/bhq-nhiy-gnj Grupo 2: Jueves de 18:00 a 22:00 - Sala de cómputo 104
Formulación del problema general del cual se ocupará el curso: <p>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 software libre python, en la versión proporcionada por Google Collaboratory (actualmente 3.10.12).</p>		
Método de evaluación y recomendaciones: <p>La evaluación de la disciplina se dividirá en dos bloques principales que serán evaluados y ponderados a través de todo el semestre. Un primer bloque corresponde a actividades en clase y talleres, en ambos casos corresponden a réplicas y prácticas de lo presentado en las sesiones de clase. El segundo bloque corresponde a la formulación, construcción y desarrollo de un proyecto sobre cualquiera de las grandes temáticas del curso. Este proyecto debe incluir un documento técnico, así como todo el código en python que llega a la consecución de los objetivos tanto general como específicos. La ponderación de las actividades para la definición de la nota definitiva del curso es la siguiente:</p>		

	Peso porcentual sobre la nota definitiva	Comentarios
Actividades en clase y talleres	50%	Mínimo 5 (cinco) actividades durante el semestre académico. En caso de desarrollarse más de 5 (cinco) o no, todas las actividades tendrán el mismo peso porcentual.
Proyecto de la asignatura.	50%	Tres momentos en el proyecto: 1. Selección del conjunto de datos, planteamiento del problema, justificación y objetivos. Peso porcentual 10% 2. Marco teórico, metodología, EDA. Peso porcentual 20% 3. Exposición y entrega final de documento con resultados, código en python y conclusiones. Peso porcentual 20%
Total	100%	

Dado que la asignatura tiene 4 créditos académicos y que cada crédito corresponde a 48 horas de trabajo, el tiempo total que se debe dedicar a esta asignatura durante el semestre es 192 horas. Por otro lado, como cada semana se realiza 1 sesión de 4 horas y el semestre consta de 16 semanas, de las 192 horas, 64 horas son de clase. Así, las 128 horas restantes corresponden a trabajo autónomo de cada estudiante, es decir, cada estudiante debe dedicar durante el semestre 128 horas de estudio por fuera de las clases. En base a esto, y considerando que la apropiación de los conceptos estudiados en esta asignatura depende fuertemente de la realización de ejercicios, se recomienda distribuir esas 128 horas de la forma más equitativa posible a lo largo del semestre. Así, para ayudarle a que gestione su tiempo, considere las siguientes sugerencias:

1. Cada semana intente dedicar por lo menos 8 horas extra clase para desarrollar actividades asociadas a esta asignatura.
2. Toda sesión de clase tendrá el desarrollo de un método estadístico o profundización y/o refuerzo de uno ya presentado, estos temas se encuentran en los libros de texto. Así, para cada una de esas sesiones de clase se sugiere la lectura de algunas secciones de libros que son

indicadas más adelante en este documento (columna llamada acciones y producciones de los estudiantes). De esta manera, dedique por lo menos 1 hora semanal a la lectura de las secciones sugeridas para las sesiones de clase de esa semana.

3. Toda semana que no se haya programado un taller dedique tiempo a la elaboración del proyecto y/o revisión de los temas presentados en clase.
4. Las semanas en las que haya programado un taller de esta asignatura no se asignará trabajo extra. Durante esas semanas dé prioridad a la preparación y desarrollo de este taller y/o práctica.
5. Antes de entregar un taller y/o práctica siempre es conveniente leerlo de nuevo, pues a veces se pasan errores que son fáciles de corregir. Así, reserve unos minutos para la lectura final de sus resultados..
6. Finalmente, nunca deje sus tareas para último momento. Las cosas que se hacen con afán tienen buena probabilidad de salir mal.

Bibliografía:

- Díaz L & Morales M, *Análisis estadístico de datos multivariados*. Publicaciones Facultad de ciencias, UNAL, 2012.
[Disponible aquí](#)
- Kassambara, A, *Practical Guide To Principal Component Methods in R*. STHDA, 2017
- Géron, A, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition*, O'Reilly, 2019
- Müller A & , Guido S, *Introduction to Machine Learning with Python*, O'Reilly, 2016
- Sarkar, D., Bali, R., & Sharma, T. *Practical Machine Learning with Python. A Problem-Solver's Guide to Building Real-World Intelligent Systems*, New York: Springer-Verlag, 2018
- Everitt, B. S. and Hothorn, T. (2011), *An Introduction to Applied Multivariate Analysis with R* , New York: Springer-Verlag
- Han J & Kamber M, *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, 3rd ed, 2010.

Trayectos a recorrer en el tratamiento del problema	Finalidades formativas del trayecto	Acciones y producciones de los estudiantes
Semana 1: Clase dialógica:	Contextualizar los temas que se trabajarán durante las sesiones del curso, el método de evaluación, la bibliografía del curso.	Fuera de clase:

Introducción al curso. Introducción - Aprendizaje automático	Comprender los conceptos de aprendizaje automático, los tipos de aprendizaje supervisado y no supervisado.	Lectura (3 horas): Revisión de conceptos del curso previo de Fundamentos de estadística. Taller (5 horas): Desarrollo del taller número 1 de la asignatura, ejercicios de código en python, estadísticas descriptivas, implementación y análisis.
Semana 2 Clase dialógica: Análisis de componentes principales (PCA)	Comprender el uso de componentes principales en situaciones de múltiples variables cuantitativas. Analizar la selección, uso e implementación en Python del PCA.	Fuera de clase: Lectura (5 horas): Capítulo 7, del libro: <i>Análisis estadístico de datos multivariados</i> . Proyecto (3 horas): Avance del proyecto.
Semana 3 Clase dialógica: Análisis de componentes principales - Parte II	Múltiples ejemplos de análisis e implementación de PCA en Python.	Fuera de clase: Taller (5 horas): Desarrollo de taller práctico de PCA en Python. Proyecto (3 horas): Avance del proyecto.
Semana 4 Clase dialógica: Análisis de correspondencias simples.	Comprender las técnicas de visualización y correlación en variables categóricas. Aplicar la distancia Ji-cuadrado para analizar similitudes	Fuera de clase: Lectura (3 horas): Capítulo 8, del libro: <i>Análisis estadístico de datos multivariados</i> . Taller (3 horas): Avance del proyecto. Taller (2 horas): Solución del taller práctico de análisis de correspondencias simples.
Semana 5 Clase dialógica: Análisis de correspondencias múltiples. Análisis Factorial.	Comprender las técnicas de visualización y correlación en variables categóricas. Aplicar la distancia Ji-cuadrado para analizar similitudes.	Fuera de clase: Proyecto (7 horas): Avance del proyecto. Lectura (1 hora): Capítulo 6, del libro: <i>Practical Guide To Principal Component Methods in R</i>

Primer avance del proyecto: Problema, justificación y objetivos	Utilizar el análisis factorial en situaciones de reducción de dimensionalidad. Realizar la fase exploratoria.	
Semana 6 Clase dialógica: Agrupamiento Agrupamiento Jerárquico	Estudiar el agrupamiento de individuos a través de múltiples atributos. Entender el agrupamiento por medio de métodos jerárquicos.	Fuera de clase: Lectura (3 horas): Capítulo 9, del libro: <i>Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow</i> . Proyecto (5 horas): Avance del proyecto.
Semana 7 Clase dialógica: Agrupamiento de K-means	Analizar el uso de agrupamiento de K-medias. Combinar el uso de métodos jerárquicos y de k-means.	Fuera de clase: Lectura (5 horas): Capítulo 9, del libro: <i>Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow</i> Proyecto (3 horas): Avance del proyecto.
Semana 8 Clase dialógica: Otros agrupamientos: <ul style="list-style-type: none"> - DSCAN - Agrupamientos basados en funciones de densidad. 	Hacer uso de agrupamiento para la detección de anomalías. Comprender el uso de agrupamientos basados en funciones de densidad.	Fuera de clase: Taller (8 horas): Taller sobre métodos de clustering en Python
Semana 9 Clase dialógica:	Comprender el uso de regresión para la predicción de eventos.	Lectura (3 horas): Capítulo 4, el libro: <i>Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow</i>

Aprendizaje supervisado Regresión Lineal múltiple.		Proyecto (5 horas): Avance del proyecto.
Semana 10 Clase dialógica: Regresión Logística. Clasificación, matriz de confusión, métricas Segundo avance. Marco teórico – EDA, Pruebas del transformador.	Analizar la clasificación como método de discriminación de individuos.	Fuera de clase: Lectura (2 horas): Capítulo 4, el libro: <i>Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow - Section: Logistic Regression</i> . Taller (6 horas): Taller en Python sobre regresión múltiple.
Semana 11 Clase dialógica: Clasificadores de Naive Bayes	Comprender el funcionamiento de clasificadores bayesianos “ingenuos” y su aplicación en predicción de eventos	Fuera de clase: Lectura (2 horas): Secciones 9.3-9.5, del libro: <i>Data Mining: Concepts and Techniques</i> Proyecto (6 horas): Avance del proyecto.
Semana 12 Clase dialógica: Árboles de decisión	Usar árboles de decisión para clasificar y predecir comportamientos de individuos.	Fuera de clase: Lectura (6 horas): Capítulo 6, del libro: <i>Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow</i> Proyecto (2 horas): Avance del proyecto.
		Fuera de clase:

Semana 13 Clase dialógica: Random Forest	Usar el algoritmo para decisión para clasificar y predecir comportamientos de individuos	Lectura (5 horas): Capítulo 7, del libro: <i>Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow</i> Proyecto (3 horas): Avance del proyecto.
Semana 14 Clase dialógica: Random Forest	Usar el algoritmo para decisión para clasificar y predecir comportamientos de individuos	Fuera de clase: Lectura (7 horas): Taller sobre métodos de clasificación usando diferentes clasificadores en Python. Proyecto (1 hora): Avance proyecto.
Semana 15 Clase dialógica: Extreme Gradient Boosting(XGBoost), LightGBM	Usar el algoritmo para decisión para clasificar y predecir comportamientos de individuos Diagnosticar la efectividad de la clasificación de los métodos anteriores en problemas de data mining.	Fuera de clase: Taller (2 horas): Taller sobre métodos de clasificación en Python, usando este par de librerías y sus algoritmos como clasificadores. Proyecto (5 horas): Avance proyecto.
Semana 16 Exposiciones: Entrega Final-Resultados Conclusiones	Socialización y presentación de resultados del proyecto del curso, así como documento y código	En clase: Socialización y presentación de resultados

**UNIVERSIDAD CENTRAL
VICERRECTORÍA ACADÉMICA**

PROGRAMA PARA EL DESARROLLO DE LA ASIGNATURA

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:	En diferentes situaciones hacer uso de métodos de reducción de dimensionalidad para facilitar la comprensión y visualización de datos	Talleres que se realizarán en Python en los que se deba reducir la dimensionalidad, además que se deban crear visualizaciones de esos datos.
Finalidad 2:	En diferentes situaciones decidir el uso de aprendizaje supervisado y no supervisado y la conveniencia de ambos	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
Finalidad 3:	Donde se necesite clasificar individuos y predecir probabilidades, el estudiante deberá aprender a usar distintos algoritmos de clasificación	En diferentes talleres de clasificación, los estudiantes usarán métodos de clasificación y regresión, y además pronosticar comportamientos y tomar decisiones en base a estos
Finalidad 4:	Gestionar y administrar el esfuerzo colaborativo al enfrentar un proyecto de mediano aliento. Este se sustenta en una división racional y eficiente del trabajo, identificando fortalezas y debilidades de su equipo de trabajo, buscando obtener los mejores réditos que cada integrante en la consecución del mismo.	Desarrollar un proyecto de analítica de datos, comprendiendo la adecuación de los datos para lograr el objetivo principal, así como el desarrollo de todas las fases del mismo, incluyendo su documentación y construcción de un documento técnico incluyendo resultados.

Firma del Profesor

Luis Campos

Firma del Director del Programa

Fecha: