

UNIVERSIDAD NACIONAL DEL SUR						1	5
BAHIA BLANCA				ARGENTINA			
DEPARTAMENTO DE MATEMÁTICA							
PROGRAMA DE: MÉTODOS PARA EL ANÁLISIS DE DATOS I.						CÓDIGO: 8142	
						ÁREA N°: IV	
HORAS DE CLASES:				PROFESOR RESPONSABLE			
TEÓRICAS		PRÁCTICAS		Dr. José BAVIO			
Por semana	Por cuatrim.	Por semana	Por cuatrim.				
6	96	6	96				
ASIGNATURAS CORRELATIVAS PRECEDENTES							
CARRERA			APROBADA			CURSADA	
Licenciatura en Matemática Aplicada			---			Probabilidad y Estadística	
DESCRIPCIÓN							
<p>El curso se centrará en la presentación y desarrollo teórico de las principales técnicas propias del todo el ciclo de Análisis de Datos-. Para cada técnica introducida se proporcionarán una serie de ejercicios y experiencias con datos reales a efectos de que el estudiante pueda introducirse a problemas concretos. Las clases teóricas tendrán seis (6) horas de duración semanal y las prácticas, incluyendo fundamentalmente una guía para manejo de rutinas de Python y R, tendrán seis (6) de duración semanal, de manera que, posteriormente el estudiante cuente con los recursos teóricos de cada técnica y esté capacitado para ingresar y analizar las salidas ofrecidas por las herramientas que utilice.</p>							
OBJETIVOS							
<ol style="list-style-type: none"> 1. Presentar y discutir las técnicas exploratorias propias del Análisis de Datos que conducen a la reducción e interpretación del número de casos y del número de variables. 2. Comprender conceptos claves de las técnicas estadísticas del Análisis de Datos. 3. Ser capaz de distinguir qué técnica es la más adecuada, en función tanto del tipo de datos como del uso que se quiera hacer de ellos. 4. Discutir los Usos Adecuados e Inadecuados de los distintos Procedimientos. 5. Adquirir los conocimientos necesarios para la aplicación mediante el manejo de distintos softwares, en particular Python y R. 							
PROGRAMA SINTÉTICO SEGÚN PLAN DE ESTUDIOS							
<ol style="list-style-type: none"> 1. Introducción al Análisis de Datos 2. Obtención, Preparación y Exploración de Datos 3. Modelado Estadístico y Regresión 4. Aprendizaje Supervisado 5. Aprendizaje No Supervisado 6. Predicción y Comunicación de Resultados 							
AÑO	2026						

BAHIA BLANCA

ARGENTINA

DEPARTAMENTO DE MATEMÁTICA

PROGRAMA DE: MÉTODOS PARA EL ANÁLISIS DE DATOS I

CÓDIGO: 8142

ÁREA N°: IV

PROGRAMA ANALÍTICO Y METODOLOGÍA DE ENSEÑANZA

CAPÍTULO	CONTENIDO TEMÁTICO	METODOLOGÍA
1.-	Introducción al Análisis de Datos: El ciclo de análisis de datos. Tipos de análisis. Configuración del entorno de trabajo (Python, R, VS Code). Introducción a notebooks (Jupyter, R Markdown). Sintaxis básica de Python y R. Buenas prácticas de programación y reproducibilidad. Caso de estudio completo.	Exposición de los contenidos teóricos con apoyo multimedia. - Ejercicios y aplicaciones con datos reales. - Uso integrado de Python y R en entorno VS Code. - Configuración guiada del entorno de desarrollo. - Primer contacto con notebooks Jupyter y R Markdown. - Caso de estudio completo desde datos crudos hasta resultados.
2.-	Obtención y Preparación de Datos: Fuentes de datos. Carga de datos en Python (pandas) y R (readr, tidyverse). Exploración inicial de datos. Calidad de datos: valores faltantes, duplicados, inconsistencias. Estrategias de tratamiento de valores faltantes. Transformación de datos: tipos de datos, variables derivadas, codificación de variables categóricas. Normalización y estandarización. Discretización de variables continuas. Integración de datos: joins y merges. APIs REST y web scraping. Tidy data. Reshape: pivot, melt, stack.	- Exposición de los contenidos teóricos con apoyo multimedia. - Ejercicios y aplicaciones con datos reales. - Uso integrado de Python y R en entorno VS Code. - Práctica con distintas fuentes de datos (CSV, APIs, web scraping). - Ejercicios de limpieza y transformación de datos reales con problemas de calidad. - Trabajo con pandas (Python) y tidyverse (R)
3.-	Modelado Estadístico y Regresión: Correlación vs causalidad. Regresión lineal simple: formulación, mínimos cuadrados ordinarios (OLS), interpretación. Supuestos del modelo lineal. Análisis de residuos. Evaluación de modelos: R^2 , RMSE, MAE. Regresión lineal múltiple. Interpretación de coeficientes parciales. Multicolinealidad (VIF). Selección de variables: forward, backward, stepwise. Tests de hipótesis e intervalos	- Exposición de los contenidos teóricos con apoyo multimedia.

AÑO

2026

BAHIA BLANCA

ARGENTINA

DEPARTAMENTO DE MATEMÁTICA

PROGRAMA DE: MÉTODOS PARA EL ANÁLISIS DE DATOS I

CÓDIGO: 8142

ÁREA N°: IV

de confianza. Overfitting vs underfitting. Bias-variance tradeoff. Regularización: Ridge, LASSO, Elastic Net. Regresión logística: clasificación binaria, función logística, odds ratios, máxima verosimilitud. Validación de modelos: conjuntos train/validation/test. Validación cruzada (k-fold, stratified). Curvas de aprendizaje.

- Ejercicios y aplicaciones con datos reales.
- Uso integrado de Python y R en entorno VS Code.
- Construcción paso a paso de visualizaciones con matplotlib, seaborn y ggplot2.
- Análisis exploratorio de datasets multivariados.
- Énfasis en principios de visualización efectiva y comunicación visual

4.-

Aprendizaje Supervisado: Clasificación: conceptos fundamentales. Árboles de decisión: construcción, criterios de splitting, poda. Random Forest: bagging, bootstrap, importancia de variables, hiperparámetros. Support Vector Machines (SVM): hiperplano óptimo, vectores de soporte, kernels. K-Nearest Neighbors (KNN): algoritmo, selección de k. Redes Neuronales, mlps, algoritmos de entrenamiento, otras arquitecturas. Métricas de clasificación: matriz de confusión, accuracy, precision, recall, F1-score. Curvas ROC y AUC. Curvas precision-recall. Ingeniería de features. Clases desbalanceadas: técnicas de resampling (SMOTE, undersampling), ajuste de pesos. Comparación de modelos: grid search, pipelines.

- Exposición de los contenidos teóricos con apoyo multimedia.
- Ejercicios y aplicaciones con datos reales.
- Uso integrado de Python y R en entorno VS Code.
- Implementación de modelos de regresión desde cero y con librerías (statsmodels, scikit-learn, lm).
- Diagnóstico exhaustivo de modelos: residuos, supuestos, métricas.
- Casos prácticos de selección de variables y regularización

5.-

Aprendizaje No Supervisado: Clustering: conceptos fundamentales. Métricas de similitud y distancia. K-means: algoritmo, selección de k (método del codo, silueta), limitaciones. Clustering jerárquico: aglomerativo y divisivo, dendrogramas, métodos de enlace. DBSCAN: clustering basado en densidad, parámetros eps y minPts, detección de ruido. Gaussian Mixture Models (GMM): modelos probabilísticos, algoritmo EM, comparación con k-means.

- Exposición de los contenidos teóricos con apoyo multimedia.
- Ejercicios y aplicaciones con datos reales.

AÑO

2026

	<p>Evaluación de clustering: índice de silueta, Davies-Bouldin, Calinski-Harabasz.</p> <p>Reducción de dimensionalidad: Análisis de Componentes Principales (PCA): formulación matemática, interpretación de componentes, varianza explicada, visualización. t-SNE (t-Distributed Stochastic Neighbor Embedding): preservación de estructura local, parámetros, aplicaciones en visualización. UMAP (Uniform Manifold Approximation and Projection): fundamentos, ventajas sobre t-SNE. Linear Discriminant Analysis (LDA): maximización de separabilidad entre clases, reducción supervisada vs no supervisada.</p> <p>Detección de anomalías: Isolation Forest: construcción de árboles de aislamiento, score de anomalía. Local Outlier Factor (LOF): densidad local, detección de outliers contextuales. One-Class SVM: frontera de decisión, aplicaciones.</p> <p>Estimación de densidad: Kernel Density Estimation (KDE): selección de kernel, bandwidth, aplicaciones en análisis exploratorio. Aprendizaje profundo para datos no etiquetados: Autoencoders: arquitectura encoder-decoder, función de pérdida, reconstrucción. Autoencoders para reducción de dimensionalidad y detección de anomalías.</p>	<ul style="list-style-type: none"> - Uso integrado de Python y R en entorno VS Code. - Aplicación de técnicas de clustering y reducción de dimensionalidad a problemas reales. - Visualización de estructuras latentes en datos de alta dimensionalidad. - Implementación de autoencoders con Keras/TensorFlow (Python) para detección de anomalías.
6.-	<p>Predicción y Comunicación de Resultados: Ciclo de vida de modelos en producción. Serialización de modelos. Interpretabilidad de modelos: SHAP, LIME, partial dependence plots.</p> <p>Comunicación efectiva de resultados estadísticos. Reportes reproducibles: Jupyter Notebooks, R Markdown. Ética en análisis de datos.</p>	<ul style="list-style-type: none"> - Exposición de los contenidos teóricos con apoyo multimedia. - Ejercicios y aplicaciones con datos reales. - Uso integrado de Python y R en entorno VS Code. - Generación de reportes reproducibles con Jupyter y R Markdown. - Interpretación de modelos black-box con SHAP y LIME. - Discusión de casos éticos en análisis de datos y sesgos algorítmicos

SISTEMA DE EVALUACIÓN

Se resolverán problemas y ejercicios durante el desarrollo de la cursada. Periódicamente, se requerirá la entrega de soluciones a problemas teórico prácticos y con datos reales previamente asignados y que formarán parte de la evaluación final.

Se tomarán por lo menos dos (2) exámenes parciales. La nota final del curso se calculará como el promedio simple de las notas obtenidas por cada instancia de evaluación.

UNIVERSIDAD NACIONAL DEL SUR		5	5
BAHIA BLANCA		ARGENTINA	
DEPARTAMENTO DE MATEMÁTICA			
PROGRAMA DE: MÉTODOS PARA EL ANÁLISIS DE DATOS I		CÓDIGO: 8142	
		ÁREA N°: IV	

BIBLIOGRAFÍA


Bibliografía Básica

- [1] DRAPER, N.R. & SMITH, H. (1981): *"Applied Regression Analysis"*. John Wiley & Sons, New York, 2nd edition.
- [2] de OLIVEIRA BUSSAB, W. (1988): *"Análise de Variância e de Regressão"*. Atual Editora Ltda. Sao Paulo. Brasil. pp.147.
- [3] ELIAN, S.N. (1988): *"Análise de Regressão"*. Editora da Universidade de Sao Paulo. Brasil. pp. 230.
- [4] HAERDLE, W. (1990): *"Applied nonparametric Regression"*. Cambridge University Press. New York.
- [5] HOSMER, D.W. & LEMESHOW, S. (2000): *"Applied Logistic Regression"*. 2^o Edition. Wiley Series in Probability and Statistics. Pp. 376.
- [6] MONTGOMERY, D.C. and PECK, E.A. (1992): *"Introduction to Linear Regression Analysis"*. John Wiley & Sons, Inc. New York. 2nd edition. pp. 527.
- [7] MONTGOMERY, D.C., PECK, E.A. & VINING, G.G. (2006): *"Introducción al Análisis de Regresión Lineal"*. Compañía Editorial Continental. 3^o Reimpresión. México. Pp. 608
- [8] SEBER, G.A.F & LEE, A. (2003). *"Linear Regression Analysis"*. Second Edition. John Wiley, New York.
- [9] NETER, J.; NACHTSHEIM, C. J.; KUTNER, M.H. & LI, W. (2013): *"Applied Linear Statistical Models"*. McGraw-Hill.
- [10] R CORE TEAM. (2016): *"R: A language and environment for statistical Computing"*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>

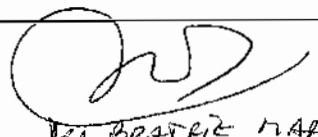


Bibliografía Complementaria

- [10] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer
- [11] Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (3rd ed.). O'Reilly Media.

VIGENCIA DE ESTE PROGRAMA

AÑO	PROFESOR RESPONSABLE (Firma aclarada)	AÑO	PROFESOR RESPONSABLE (Firma aclarada)
2026	 Dr. José BAVIO		

VISADO

COORDINADORA DE ÁREA	SECRETARÍA ACADÉMICA	DIRECTOR DEL DEPARTAMENTO
 Dra. BEATRICE NAZZARI	 Dra. DIANA JORCZULINA RECCHI SECRETARÍA ACADÉMICA DEPARTAMENTO DE MATEMÁTICA UNIVERSIDAD NACIONAL DEL SUR	 Dr. MARTIN D. SAFE DIRECTOR DECANO DEPARTAMENTO DE MATEMÁTICA UNIVERSIDAD NACIONAL DEL SUR
FECHA:	FECHA:	FECHA:

AÑO	2026						
-----	------	--	--	--	--	--	--