

## SÍLABO 2025-2

### I. INFORMACIÓN GENERAL

Nombre del curso:	TALLER DE HERRAMIENTAS CUANTITATIVAS PARA EL ANÁLISIS SOCIAL 3
Clave del curso:	1INT47
Créditos:	2
Número de horas de teoría:	1
Número de horas de práctica:	2
Clave del horario:	0801
Profesor/a:	Christian Chiroque Ruiz
Correo electrónico PUCP:	christian.chiroque@pucp.edu.pe
Horario de sesiones Teóricas:	Miércoles 5:00 p.m. a 6:00 p.m.
Horario de sesiones prácticas:	Miércoles 6:00 p.m. a 8:00 p.m.

### II. SUMILLA

Taller de contenido variable sobre herramientas de análisis cuantitativo vinculadas a las ciencias sociales. El objetivo de este taller es ofrecer a los y las estudiantes de la Facultad espacios para la adquisición y perfeccionamiento del uso de herramientas útiles para la sistematización y análisis de información cuantitativa en los procesos de investigación social.

### III. PRESENTACIÓN

El curso ***Introducción al Machine Learning aplicado a las ciencias sociales*** busca introducir a los estudiantes de ciencias sociales en el campo del aprendizaje automático, presentando sus fundamentos, principales aplicaciones y limitaciones en el análisis de fenómenos sociales. Se abordarán casos de uso relevantes como la detección de fraude financiero, la identificación de patrones delictivos, predicción de conflictos sociales, análisis socioeconómico, entre otros.

El curso combina una base teórica con ejercicios prácticos basados en diversas bases de datos, lo que permite a los estudiantes desarrollar competencias analíticas, críticas y técnicas para enfrentar desafíos contemporáneos desde una



**PUCP**

Facultad de  
Ciencias Sociales

perspectiva basada en datos.

Este curso es de naturaleza interdisciplinaria porque integra los marcos analíticos y críticos de las ciencias sociales con herramientas computacionales propias de la estadística, la ciencia de datos y la informática. Lejos de abordar estas herramientas de forma puramente técnica, se explora cómo los modelos pueden aportar a la comprensión de patrones de comportamiento, dinámicas estructurales y procesos sociales, manteniendo una mirada crítica sobre su uso, sus límites y sus implicancias éticas. Esta aproximación permite tender puentes entre el análisis computacional y la reflexión social, promoviendo una formación integral y contextualizada.

#### **IV. RESULTADOS DE APRENDIZAJE**

Al finalizar el curso se espera que el estudiante sea capaz de:

- Explicar los conceptos básicos del aprendizaje automático, tanto para los métodos supervisados como los no supervisados.
- Formular problemas de predicción o clasificación con diversos datasets, seleccionar algoritmos adecuados y evaluar su desempeño utilizando métricas apropiadas.
- Utilizar el lenguaje de programación R (V4.5.1), a través del IDE R Studio (3.6.0+), para desarrollar flujos de trabajo completos de ML: desde la limpieza de datos hasta la validación de modelos.
- Evaluar críticamente el uso de algoritmos predictivos en contextos donde el objetivo sea explicar fenómenos sociales o tomar decisiones políticas.
- Elaborar informes, visualizaciones y presentaciones que integren resultados cuantitativos de modelos de ML con narrativas comprensibles para audiencias no técnicas.

#### **V. CONTENIDO DEL CURSO**

1. Fundamentos de Aprendizaje automático
2. Tipos de aprendizaje automático
3. Preprocesamiento para ML
4. Métodos supervisados
5. Métodos no supervisados
6. Métricas de evaluación
7. Hiperparámetros y *model tuning*
8. Introducción al Deep Learning y la visión artificial
9. Presentación de proyectos integrador



**PUCP**

Facultad de  
Ciencias Sociales

## VI. METODOLOGÍA

El curso contempla las siguientes consideraciones metodológicas:

- **Sesiones teóricas:** Durante la sesión se expondrá algunos puntos teóricos necesarios para el uso de las herramientas a analizar. Así también, comprenderá la realización de ejercicios por parte de los alumnos con la dirección del docente.

La dinámica del curso parte del supuesto de que los estudiantes llegan a las clases teóricas habiendo revisado con detenimiento las lecturas o material audiovisual asignado para cada semana.

- **Sesiones prácticas:** Las sesiones tienen como principal objetivo la resolución de ejercicios dirigidos y resolver las dudas de los estudiantes sobre su aprendizaje del software estadístico.
- **Software estadístico:** Se utilizará principalmente el software R, a través del entorno de desarrollo R Studio. No es necesario tener experiencia previa con el programa, ya que se enseñará desde los conceptos básicos.

La última versión de R (V4.5.1) puede ser descargada gratuitamente para Windows o macOS:

<https://cran.r-project.org/bin/windows/base/>  
<https://cran.r-project.org/bin/macosx/>

La última versión de Rstudio Desktop (versión RStudio 3.6.0+) puede ser descargada aquí para Windows o macOS.

<https://posit.co/download/rstudio-desktop/>

De forma opcional podrán utilizar la plataforma web basado en la nube R Studio Cloud. Los recursos mencionados se encuentran dentro de la categoría de software libre, por lo que no tienen costo.

Asimismo, de forma complementaria, para las clases de deep learning, se utilizará el lenguaje de programación Python a través de la plataforma Google Colab:

<https://colab.research.google.com/>



**PUCP**

Facultad de  
Ciencias Sociales

- **Tareas calificadas:** Se desarrollan de forma individual. Estas tareas buscan consolidar progresivamente las habilidades en el uso de herramientas de aprendizaje automático aplicadas a problemas sociales reales. Cada tarea incluirá:
  - Un dataset asignado o propuesto por el estudiante.
  - Aplicación de uno o dos algoritmos vistos en clase.
  - Interpretación crítica de los resultados.
  - Se espera la entrega en formato HTML (Quarto o R Markdown), PDF o notebook de Google Colab, según corresponda al lenguaje utilizado.
- **Proyecto integrador:**
  - Se realizará de forma grupal (2-3 estudiantes).
  - Tiene como finalidad que los estudiantes apliquen de manera autónoma y crítica los conocimientos adquiridos durante el curso, desarrollando un proyecto de machine learning desde la exploración y preparación de datos hasta la evaluación de modelos y la presentación de resultados.
  - El proyecto puede centrarse en datasets vinculados a cualquier especialidad de las ciencias sociales: economía, ciencia política, finanzas, antropología, etc.
  - Se valorará la creatividad, el uso correcto de las herramientas vistas en clase, y la claridad en la explicación de decisiones técnicas.
  - Incluye exposiciones tanto en la entrega parcial como final del proyecto.

## VII. EVALUACIÓN

TIPO DE EVALUACIÓN	CANTIDAD	FECHA(S)	PESO TOTAL (EN %)
Tareas calificadas	3	19 de setiembre 24 de octubre 21 de noviembre	40%
Proyecto integrador: Entrega parcial y exposición.	1		30%
Proyecto integrador: Entrega final y exposición.	1		30%

**Fórmula de calificación:**

$$(40Tc + 30PI\_Ep + 30PI\_Ef) / 100$$

**Observación:**

No se elimina ninguna tarea calificada. Los estudiantes podrán realizar sus solicitudes de recalificación a través del Campus Virtual (INTRANET), dentro de los 5 días calendario posteriores al registro de la nota.

**VIII. BIBLIOGRAFÍA OBLIGATORIA**

AGRESTI, Alan and Barbara Finlay (2009). *Statistical Methods for the Social Sciences*. Upper Saddle River, New Jersey: Pearson Prentice Hall

BOEHMKE, Bradley & Brandon Greenwell (2020). *Hands-On Machine Learning with R*. The R Series. CRC Press. Taylor & Francis Group.

<https://bradleyboehmke.github.io/HOML/>

CHIROQUE, Christian (2024). *Estadística para el Análisis Político 1*. Estadística descriptiva e inferencial en R.

CHOLLET, François (2017). *Deep Learning with Python*. Manning Publications.

FRIEDMAN, JEROME, Trevor Hastie, and Robert Tibshirani (2001). *The Elements of Statistical Learning*. Vol. 1. Springer Series in Statistics New York, NY, USA:

GARETH, JAMES, Daniela Witten, Trevor Hastie and Robert Tibshirami (2023). *An Introduction to Statistical Learning with Applications in R*. Second Edition. Springer.

<https://www.statlearning.com/>

GÉRON, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow (2ª ed.)*. O'Reilly Media.

KUHN, MAX and Kjell Johnson (2013). *Applied Predictive Modeling*. Springer.

KUHN, MAX and Kjell Johnson (2019). *Feature Engineering and Selection: A Practical Approach for Predictive Models* .

<http://www.feat.engineering/>

KUHN, MAX and Julia Silge (2023). *Tidy Modeling with R. A framework for Modeling in the Tidyverse*.

<https://www.tmwr.org/>

RASCHKA, S., & Mirjalili, V. (2022). *Python machine learning (4th ed.)*. Packt Publishing.

STARMER, Josh (2022). *The StatQuest Illustrated Guide to Machine Learning*.

URDINEZ, F. y Cruz, A. (Eds.) (2020). *Political Data Science Using R: A Practical Guide*. CRC Press.

WICKHAM H. (2014). "Tidy Data". *The Journal of Statistical Software*, 59.

<https://vita.had.co.nz/papers/tidy-data.pdf>

WICKHAM H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. Springer.

<https://ggplot2-book.org/index.html>

WICKHAM, HADLEY, Mine Cetinkaya-Rundel and Garret Golemund (2017) *R for Data Science*. Second Edition. O'Reilly Media.

<https://r4ds.hadley.nz/>

WICKHAM H. (2019) *The Tidyverse style guide*.

<https://style.tidyverse.org/>

## IX. CRONOGRAMA

SEMANA		CONTENIDO SESIÓN TEÓRICA	CONTENIDO SESIÓN PRÁCTICA	OBSERVACIÓN
1	20 de agosto	Introducción y panorama general del machine learning.	<ul style="list-style-type: none"> <li>- Presentación de R y RStudio: interfaz, paquetes y conceptos básicos de programación.</li> <li>- Instalación y manejo de paquetes básicos y funciones.</li> <li>- Diferencias entre Script (.r), R Markdown (.rmd) y Quarto (.qmd).</li> </ul>	

2	27 de agosto	Análisis Exploratorio de Datos I	<ul style="list-style-type: none"> <li>- Importación y exportación de datos: lectura de CSV, Excel, y otros formatos.</li> <li>- Introducción al tidyverse.</li> <li>- Manipulación de datos con <i>dplyr</i>: selección (<i>select</i>), filtrado (<i>filter</i>), ordenamiento (<i>arrange</i>), creación de nuevas variables (<i>mutate</i>) y resúmenes (<i>summarise</i>).</li> </ul>	
3	3 de septiembre	Análisis Exploratorio de Datos II. Preparación de data para aplicación de algoritmos de ML.	<ul style="list-style-type: none"> <li>- Agrupación de datos y operaciones por grupo (<i>group_by</i>).</li> <li>- Uso de pipes (<i> &gt;</i>) para crear flujos de trabajo claros y legibles.</li> <li>- Creación de gráficos básicos con <i>ggplot2</i>: histogramas, gráficos de barras, diagramas de cajas, gráficos de dispersión.</li> <li>- <b>Aplicación</b>: Descripción de datos socioeconómicos.</li> </ul>	
4	10 de septiembre	Modelos supervisados regresores: Regresión Lineal I	<ul style="list-style-type: none"> <li>- Introducción al <i>tidy models</i>.</li> <li>- <i>rsample</i>: data de entrenamiento y test.</li> <li>- <i>recipes</i> y <i>parsnip</i> (<i>linear_reg()</i>)</li> <li>- <b>Aplicación</b>: Predicción del Basel AML Index (Índice de riesgo de lavado de activos).</li> </ul>	
5	17 de septiembre	<ul style="list-style-type: none"> <li>- Modelos supervisados regresores: Regresión Lineal II</li> <li>- Predicción y evaluación de modelos.</li> </ul>	<ul style="list-style-type: none"> <li>- Utilización de la data de test.</li> <li>- Uso del <i>RMSE</i>.</li> <li>- <b>Aplicación</b>: Predicción del Global Organized Crime Index.</li> </ul>	<b>Deadline entrega de Tarea 1: 19 de septiembre</b>
6	24 de septiembre	Árboles de decisión y Random Forest	<ul style="list-style-type: none"> <li>- Uso de <i>rand_forest()</i>.</li> <li>- Hiperparámetros y <i>model tuning</i>.</li> <li>- Comparación de resultados con regresión lineal.</li> </ul>	



**PUCP**

Facultad de  
Ciencias Sociales

7	01 de octubre	Modelos supervisados para clasificación: Regresión Logística	- Evaluación de desempeño con curvas ROC y matriz de confusión. - <b>Aplicación:</b> Predicción de la conflictividad social en industrias extractivas.	
8	8 de octubre	Feriado		
9	15 de octubre	<b>Proyecto integrador: Entrega parcial y exposición.</b>		
10	22 de octubre	Evaluación de modelos de clasificación y validación cruzada	- Uso del paquete <i>yardstick</i> para evaluar las métricas de los modelos: accuracy, precisión, recall, F1. - <code>vfold_cv()</code>	<b>Deadline entrega de Tarea 2:</b> 24 de octubre
11	29 de octubre	Modelos no supervisados: Clustering	- Algoritmos de análisis clúster basados en partición: k- means (centroides); PAM y CLARA (medoides). - Algoritmos basados en densidad: DBSCAN. - <b>Aplicación:</b> Tipologías de regiones según nivel de desarrollo.	
12	05 de noviembre	Reducción de dimensionalidad	- Análisis de componentes principales - Análisis factorial exploratorio - <b>Aplicación:</b> Indicadores criminológicos.	
13	12 de noviembre	Algoritmos de detección de anomalías	- Isolation Trees e Isolation Forest - <b>Aplicación:</b> Detección de fraude financiero.	
14	19 de noviembre	Introducción al aprendizaje profundo I: Redes neuronales	- Introducción a keras en R (también Google Colab con Python). - Construcción de redes neuronales.	<b>Deadline entrega de Tarea 3:</b> 21 de noviembre



15	26 de noviembre	Introducción al aprendizaje profundo II: Introducción a la visión por computadora	- Uso de redes convolucionales simples (CNN). - Discusión sobre sus aplicaciones.	
16	3 de diciembre	<b>Proyecto integrador: Entrega final y exposición.</b>		

La evaluación de todos los trabajos contemplará el respeto de los derechos de autor. En este marco, cualquier indicio de plagio tendrá como consecuencia la nota cero. Esta medida es independiente del proceso disciplinario que la Secretaría Académica de la facultad estime iniciar según cada caso. Para obtener más información sobre el citado visitar el siguiente sitio web: [www.pucp.edu.pe/documento/pucp/plagio.pdf](http://www.pucp.edu.pe/documento/pucp/plagio.pdf)

**Atención a la diversidad y la equidad:** Nuestra universidad y sus aulas de clase son espacios seguros para todos/as los/as estudiantes. En ese sentido, si algún/a estudiante tiene alguna necesidad o inquietud, comuníquese lo antes posible al/la docente y/o al área de bienestar de la facultad a través del correo [bienestarsociales@pucp.edu.pe](mailto:bienestarsociales@pucp.edu.pe).

Si algún/a estudiante es portador de una discapacidad, sea visible, no visible o neurodiversa, y por ello requiere algún ajuste razonable para su desarrollo académico, puede comunicarse con el Programa de Apoyo en la Inclusión a Estudiantes (PAIE) de la universidad al siguiente correo: [paie@pucp.edu.pe](mailto:paie@pucp.edu.pe). La información que brinde será confidencial.

Si el nombre consignado en el Campus Virtual no corresponde con su identidad de género, puede comunicarse con el/la docente para que utilice su nombre social. También puede comunicarse con la Secretaría General ([secgen@pucp.edu.pe](mailto:secgen@pucp.edu.pe)) para que la Oficina Central de Registro (OCR) incluya el nombre social en el sistema informático de la PUCP. Esto permite que el nombre social sea considerado en las Listas de asistencia, las comunicaciones internas, las ceremonias de entrega de premios y reconocimientos, así como en todos los ámbitos y servicios de la PUCP.



**PUCP**

Facultad de  
Ciencias Sociales

Video institucional sobre prevención del hostigamiento sexual:

<https://www.youtube.com/watch?v=gZx7651ugwg>

Site Bienestar Sociales:

<https://sites.google.com/pucp.edu.pe/bienestarsociales/sociales-contr-el-acoso-y-hostigamiento-sexual?authuser=0>