

## PROGRAMA DEL CURSO

### Introducción al *Machine Learning* aplicado a la Ciencia Política

Docente	Jesús Sáenz Solís
Datos de contacto	<a href="mailto:jesus.saenz@ucr.ac.cr">jesus.saenz@ucr.ac.cr</a>
Duración del curso	21 horas
Fecha de inicio	24 de mayo
Fecha de finalización	5 de julio
Horario	Miércoles de 5:00pm a 8:00pm
Modalidad	Presencial
Requisitos y público meta	<ul style="list-style-type: none"><li>-Personas profesionales y/o estudiantes en Ciencia Política, Sociología, Estadística u otras disciplinas relacionadas relacionados con la investigación social con interés en profundizar en métodos estadísticos de aprendizaje supervisado.</li><li>-Conocimientos básicos de estadística descriptiva e inferencial.</li><li>-Conocimiento básico de programación en R y del programa Rstudio.</li><li>-Computadora portátil propia con acceso a internet.</li></ul>

### Introducción

La Ciencia Política en años recientes ha experimentado una serie de cambios para constituirse como una disciplina de carácter cada vez más interdisciplinario, con el fin de profundizar el conocimiento existente de los diferentes fenómenos políticos. Para esto, es fundamental echar mano de las diferentes herramientas desarrolladas desde otras disciplinas como la sociología, la estadística, economía, informática, entre otros.

Dentro de las nuevas perspectivas desarrolladas, tiene un rol fundamental por su carácter innovador las relacionadas a la ciencia de datos. Estas permiten generar resultados a partir de grandes conjuntos de información y con una potencia explicativa superior a los métodos tradicionales utilizados en la estadística inferencial aplicada a las ciencias sociales.

Dentro de estos métodos tienen un rol fundamental el denominado *machine learning*, los cuales suponen el aprendizaje de algoritmos a partir de un conjunto de datos para obtener resultados predictivos sobre el comportamiento de una variable dependiente. Estas herramientas, han sido utilizadas en la investigación en una gran variedad de disciplinas, sin embargo, en las ciencias sociales su aparición es aún incipiente a pesar de su gran potencial. Se entiende por este tipo de herramientas cuando se entrena un algoritmo mediante determinadas características y unas respuestas ya establecidas, esto con el fin de que el modelo pueda predecir una característica previamente definida (Sandoval, 2018).

El curso pretende desarrollar capacidades introductorias para las personas interesadas en algunas aplicaciones de estos métodos para la ciencia política y profundizar en su capacidad explicativa. Para esto, se requiere brindar las herramientas conceptuales elementales sobre estos modelos

## Objetivos

- Introducir a las personas profesionales a los principales conceptos de los métodos de aprendizaje supervisado, aplicados a la ciencia política.
- Desarrollar las capacidades básicas de los principales modelos de aprendizaje supervisado.
- Evaluar los alcances y limitaciones de los modelos de aprendizaje supervisado aplicados a la ciencia política.

## Ejes temáticos

**Conceptos básicos:** en esta primera unidad, las personas estudiantes deberán obtener los conocimientos técnicos y conceptuales básicos de previo a la implementación de los métodos de aprendizaje supervisado. Se abordará en primer lugar, algunos elementos básicos de programación en R necesarios para el desarrollo del curso. Si bien es cierto, se espera que las personas tengan un conocimiento básico de R, se realizará un repaso con los elementos que son requeridos para el desarrollo del curso. En un segundo momento, se realizará una introducción general al *machine learning* y a los métodos de aprendizaje supervisado, esto con el fin de contar con las herramientas conceptuales necesarias para comprender la lógica y naturaleza de los modelos utilizados.

**Desarrollo de modelos:** en esta segunda unidad, se realizará un desarrollo de cinco modelos de aprendizaje supervisado como lo son la regresión logística, los árboles de clasificación, los bosques aleatorios, las máquinas de soporte vectorial y el Ada-Boost. En cada sesión se explicarán los elementos básicos y alcances de cada uno de las técnicas. Además de esto, se realizarán ejemplos prácticos aplicados a la ciencia política en el programa Rstudio. Además, se compararán las diferentes técnicas y los hallazgos generados mediante cada una.

## Metodología

El curso se basa en una metodología participativa. Las personas estudiantes deberán realizar las lecturas asignadas de previo, y participar de manera activa durante las clases que serán de tipo presencial, con una duración de 3 horas, con un descanso de 10 minutos por cada hora de clase. En cada una de las sesiones se realizarán preguntas generadoras a partir de los textos asignados, con el fin de promover la participación y el intercambio con los temas analizados y su aplicación a diferentes coyunturas a nivel local y global.

A partir de las discusiones en clase, las personas estudiantes realizarán de manera opcional cinco fichas de lectura. En estas deberán desarrollar un breve resumen sobre cada uno de los métodos vistos en clase y los principales hallazgos que consideran es posible obtener mediante los ejemplos prácticos vistos en clase.

## Evaluación

La participación en la clase tendrá valor del 50% de la nota final, en donde se espera una presencia activa de los y las estudiantes en las discusiones y desarrollo de las técnicas. Asimismo, el otro 50% corresponde a la entrega correspondiente a las 5 fichas de lectura, que deberán

entregarse posterior a cada una de las sesiones destinadas al desarrollo y análisis de los modelos.

### Certificado

En caso de asistir a todas las sesiones y participar activamente, se entregará un certificado de participación, que será suscrito por la Vicerrectoría de Acción Social y la Escuela de Ciencias Políticas de la Universidad de Costa Rica por 21 horas efectivas.

### Perfil docente

Msc. Jesús Sáenz Solís. Politólogo y analista de datos. Licenciado en Ciencias Políticas por la Universidad de Costa Rica. Máster en Análisis Multivariante de Datos y Big Data por la Universidad de Salamanca. Doctorando en Estadística Multivariante Aplicada por la Universidad de Salamanca. Especialista en Evaluación, Monitoreo y Rendición de Cuentas en Hábitat para la Humanidad Internacional, con experiencia docente en la Escuela de Ciencias Políticas de la Universidad de Costa Rica, consultor independiente en sistemas de justicia, política judicial, violencia y migración con énfasis en técnicas de datos multivariantes.

### Cronograma

Unidad	Sesión	Fecha	Material básico
<b>1: Conceptos básicos</b>	Fundamentos de programación en R	24 de mayo	García, C. (2022). <a href="#">Curso: R para análisis de datos.</a>
	<i>Machine learning</i> : conceptos básicos de aprendizaje supervisado	31 de mayo	Sandoval, L (2018). <a href="#">Algoritmos de aprendizaje automático para la predicción de datos.</a>
<b>2: Desarrollo de modelos</b>	Regresión logística	7 de junio	Cruz, A, Urdinez, F. (2021). <a href="#">AnalizaR Datos Políticos.</a> Sección 8.2 “Modelos logísticos”.
	FICHA LECTURA 1		
	Árboles de clasificación	14 de junio	Orellana, J. (2018). <a href="#">Arboles de decisión y random forest.</a> Arboles de decisión parte I y II
	FICHA LECTURA 2		
	Bosques aleatorios	21 de junio	Orellana, J. (2018). <a href="#">Arboles de decisión y random forest.</a> Random Forest parte I y II
	FICHA LECTURA 3		
	Máquinas de soporte vectorial	28 de junio	Hernández, F. (2021). <a href="#">Modelos predictivos.</a> Capítulos 4 y 5.
FICHA LECTURA 4			
Ada-boost	5 de julio	Hernández, F. (2021). <a href="#">Modelos predictivos.</a> Capítulos 6.	
FICHA LECTURA 5			

### Materiales a utilizar

Betancourt, G. (2005). *Las máquinas de soporte vectorial (SVMs)*. Scientia et Technica, 27, 67-72.  
<https://revistas.utp.edu.co/index.php/revistaciencia/article/view/6895/4139>

- Breiman, L. et. al. (2021). *Classification and regression trees*. Routledge. <https://doi.org/10.1201/9781315139470>
- Casal, R. et. al. (2021). *Aprendizaje estadístico*. [https://rubenfcasal.github.io/aprendizaje\\_estadistico/bosques-aleatorios.html](https://rubenfcasal.github.io/aprendizaje_estadistico/bosques-aleatorios.html)
- Cruz, A, Urdinez, F. (2021). *AnalizaR Datos Políticos*. <https://arcruz0.github.io/libroadp/index.html>
- Freund, Y, Schapire, R. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of computer and system sciences*, 55 (1), 119-139. <https://doi.org/10.1006/jcss.1997.1504>
- Gil, C. (2018). *Máquinas de vector soporte*. [https://rpubs.com/Cristina\\_Gil/SVM](https://rpubs.com/Cristina_Gil/SVM)
- Hernández, F, Mazo, M. (2018). *Modelos de regresión con R*. [https://fhernanb.github.io/libro\\_regresion/](https://fhernanb.github.io/libro_regresion/)
- Hernández, F. (2021). *Modelos predictivos*. [https://fhernanb.github.io/libro\\_mod\\_pred/index.html](https://fhernanb.github.io/libro_mod_pred/index.html)
- León. D, Cuenca, D. (s.f). *Support Vector Machine*. [https://rstudio-pubs-static.s3.amazonaws.com/570352\\_e34015b16f1a47e883e04c6195d4711f.html](https://rstudio-pubs-static.s3.amazonaws.com/570352_e34015b16f1a47e883e04c6195d4711f.html).
- Medina-Merino, R. F., & Ñique-Chacón, C. I. (2017). Bosques aleatorios como extensión de los árboles de clasificación con los programas R y Python. *Interfases*, (010), 165-189. <https://doi.org/10.26439/interfases2017.n10.1775>
- Meyer, D. (2021). *Support Vector Machines. The interface to libsvm in package e1071*.
- Moral, I. (2006). Modelos de regresión: lineal simple y regresión logística en A. Guillén, R. Crespo (Ed.), *Métodos Estadísticos para Enfermería Nefrológica*. (pp. 195- 214). Sociedad Española de Enfermería Nefrológica.
- Orellana, J. (2018). *Arboles de decisión y random forest*. <https://bookdown.org/content/2031/>
- Ramírez, J, Ramos, M. (2015). Clasificación mediante máquinas de soporte vectorial. *Matemática*, 13 (2), 29-40. <http://www.revistas.espol.edu.ec/index.php/matematica/article/view/488/366>