**FACULTAD DE CIENCIAS SOCIALES**

**CARRERA SOCIOLOGÍA**

**PROGRAMA DE CURSOS ELECTIVOS**

PROFESOR (ES / AS) : Patricio Rodríguez Valdés

E-MAIL : prodriguez@uchile.cl

CURSO ELECTIVO CORRESPONDIENTE AL ÁREA DE

*(Marque con una X la casilla a la que corresponde este curso electivo):*

|  |  |
| --- | --- |
| Profundización metodológica | X |
| Profundización Teórica |  |
| Sociologías de Especialidad |  |
| Transformaciones de la Sociedad Chilena |  |

|  |
| --- |
| **BREVE RESUMEN DEL CURSO ELECTIVO** La Ciencia de Datos es una metodología usa un amplio conjunto de técnicas provenientes de múltiples disciplinas como Ciencias de la Computación, Matemáticas, Estadísticas, Econometría e Investigación Operativa para extraer información útil del ‘Big Data’ y apoyar la toma de decisiones. Este curso es corresponde a una introducción a las técnicas que utiliza la ciencia de datos, además de cómo se puede seguir sistemáticamente sus pasos para transformar datos en evidencia que permita apoyar la toma de decisiones y la elaboración de políticas públicas. |

|  |
| --- |
| **PROGRAMA** |
| 1. **Nombre de la actividad curricular electiva**

Introducción a la Ciencia de Datos. |
| 1. **Nombre de la actividad curricular electiva en inglés**

Introduction to Data Science |
| **3. Unidad Académica / organismo de la unidad académica que lo desarrolla**Departamento de Sociología  |
| **4. Ámbito:** Investigación, intervención social y política. |
| **5. Horas de trabajo**  | presencial (del estudiante) 3 | no presencial (del estudiante) 3 |
| **6. Tipo de créditos***SCT**(Corresponde al Sistema de Creditaje de diseño de la asignatura, de acuerdo a lo expuesto en la normativa de los planes de estudio en que esta se desarrolla.)* | *((indique la distribución de horas definida en el plan de formación. Corresponde a la traducción en carga horaria de los sct)* | *((indique la distribución de horas definida en el plan de formación. Corresponde a la traducción en carga horaria de los sct)* |
| **7. Número de créditos SCT – Chile****4** |
| **8. Horarios** Viernes 14:30 -16:00 hrs., 16:15 – 17:45 |
| **9. Salas** |
| **10. Requisitos** | CURSO ELECTIVO, NO REQUIERE REQUISITOS |
| **11. Propósito general del curso** | Conocer, comprender y aplicar los conceptos y nociones necesarias para diseñar y llevar a cabo un proceso de Ciencia de Datos, evaluando críticamente los procesos y necesidades comprehendidas en cada etapa  |
| **12. Resultados de Aprendizaje** * Comprender y aplicar nociones de matemática, estadística y programación, al servicio de la Ciencia de Datos.
* Comprender y aplicar buenas prácticas en el diseño de modelos según los objetivos y preguntas planteadas.
* Comprender y aplicar críticamente diversas técnicas de análisis de datos, reconociendo su pertinencia en función de los objetivos de análisis y datos utilizados.
* Comprender y aplicar análisis geoespaciales básicos, al servicio de la construcción de mejores modelos.
 |
| **13. Saberes / contenidos**Se espera que, al terminar el curso de Introducción a la Ciencia de Datos, los estudiantes dominen los contenidos necesarios para realizar el proceso completo del ciclo de análisis de datos. Específicamente, se espera que los participantes puedan explorar, preparar datos, planificar modelos, elaborar modelos, comunicar resultados, y tomar decisiones en función de dichos resultados. Lo anterior, con un énfasis en las habilidades de programación necesarias para llevar a cabo el proceso.  |
| **14. Metodología**Trabajo individual y proyecto de curso grupal. |
| **15. Evaluación**Se efectuarán 5 controles de lectura individuales y el desarrollo de un proyecto de Ciencia de Datos se efectuará en forma **grupal**, y se entregarán a través de la plataforma U-cursos. Las fechas de los controles de lectura serán calendarizadas la primera clase.La programación de las entregas de los trabajos es la siguiente:

|  |  |  |
| --- | --- | --- |
| **Tarea** | **Tema y ponderación** | **Fecha** |
| $T1$  | Limpieza y descriptiva de datos (20%) |  |
| $T2$  | Modelo de datos (modelo *machine learning*) (30%). |  |
| $T3$  | Visualización de datos (presentación del análisis) (50%) |  |

 |
| **16. Requisitos de aprobación*** $Nota Final : NF = 0,2 T1+0,3 T2+0,5 T3$

Para aprobar, todas las notas tienen que tener nota superior a 4.0 |
| **17. Palabras Clave**Big Data, Ciencia de Datos, aprendizaje automático (machine Learning), lenguajes de programación, toma de decisiones, visualización. |
| **18. Bibliografía Obligatoria** Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information, communication & society, 15(5), 662–679. <https://doi.org/10.1080/1369118X.2012.678878>Edelman, B. G., & Luca, M. (2014). Digital Discrimination: The Case of Airbnb.com (SSRN Scholarly Paper No. ID 2377353). Rochester, NY: Social Science Research Network. Disponible en [https://papers.ssrn.com/abstract=2377353](https://papers.ssrn.com/abstract%3D2377353)Goldsmith, S., Crawford, S., & Weinryb Grohsgal, B. (2016). Innovaciones en la prestación de servicios públicos: Número 4: El análisis predictivo: Impulsando la mejora a partir de los datos (Discussion Papers & Presentations). Inter-American Development Bank. <https://doi.org/10.18235/0000421>Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015). Prediction policy problems. The American Economic Review, 105(5), 491–495. <https://doi.org/10.1257/aer.p20151023>Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big Data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. Disponible en <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>Rodríguez, P., Palomino, N., & Mondaca, J. (2017). El uso de datos masivos y sus técnicas analíticas para el diseño e implementación de políticas públicas en Latinoamérica y el Caribe (Discussion Papers & Presentations). Inter-American Development Bank. <https://doi.org/10.18235/0000694>Tomar, L., Guicheney, W., Kyarisiima, H., & Zimani, T. (2016). Big Data in the public sector: Selected applications and lessons learned. Inter-American Development Bank. Disponible en <https://publications.iadb.org/handle/11319/7884> |
| **19. Bibliografía Complementaria**Anderson, C. (2008, junio 23). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. Disponible en <https://www.wired.com/2008/06/pb-theory/>Bahamonde, J., Hevia, A., Font, G., Bustos-Jiménez, J., & Montero, C. (2014). Mining Private Information from Public Data: The Transantiago Case. IEEE Pervasive Computing, 13(2), 37–43. <https://doi.org/10.1109/MPRV.2014.30> Butler, D. (2013). When Google got flu wrong. Nature News, 494(7436), 155. <https://doi.org/10.1038/494155a> Friedman, V. (2008, enero 14). Data Visualization and Infographics – Smashing Magazine. Disponible en <https://www.smashingmagazine.com/2008/01/monday-inspiration-data-visualization-and-infographics/> Harford, T. (2014, marzo 28). Big data: are we making a big mistake?. Disponible en <https://www.ft.com/content/21a6e7d8-b479-11e3-a09a-00144feabdc0> Hilbert, M., & López, P. (2011). The world’s technological capacity to store, communicate, and compute information. Science, 332(6025), 60–65. <https://doi.org/10.1126/science.1200970>The Economist. (2010, febrero). Data, data everywhere | The Economist. Recuperado de <http://www.economist.com/node/15557443>Miller, C. (2015, julio 9). When Algorithms Discriminate. The New York Times. Disponible en <https://www.nytimes.com/2015/07/10/upshot/when-algorithms-discriminate.html>  |
| **20. Recursos web** J. Vanderplas, The Jupyter Notebooks behind my O’Reilly report, “A Whirlwind Tour of Python”: jakevdp/WhirlwindTourOfPython. 2018.R. Kumar, IPython Notebooks to learn Python. Contribute to rajathkmp/Python-Lectures development by creating an account on GitHub. 2018.“Jupyter Notebook Viewer”. [Online]. Disponible en: <https://bit.ly/2CqpH9h> .“Python for Developers”. [Online]. Disponible en: <https://bit.ly/2oNFs0G> .“Tutorial de Python 3.6.3”. [Online]. Disponible en: <https://bit.ly/1IrfUbn>.Aprendizaje de lenguajes de programación gratuita: <https://www.sololearn.com/> (también disponible como app de celular) |
| **21. Programación por sesiones****Sesión 1 - Introducción a Ciencia de Datos, *Métodos de definición de objetivos* y Python.** **Sesión 2 – Adquisición y exploración de datos con Python.****Sesión 3 - Preparación de datos con Python: estructura y filtros.****Sesión 6 - Minería de datos con Python.****Sesión 7 - Planificación y diseño de modelos.****Sesión 8 – Desarrollo de modelos.****Sesión 9 – Desarrollo y validación de modelos.****Sesión 10 – Evaluación y rediseño.****Sesión 11 – Interpretación y comunicación de resultados a través de visualizaciones.****Sesión 12 – Presentación de proyectos de curso** |