

PROGRAMA DE CURSO

Código (a completar por la Escuela)	Nombre (<i>Nombre oficial del curso según la normativa del plan de formación vigente o del organismo académico que lo desarrolla. No debe incluir espacios ni caracteres especiales antes del comienzo del nombre.</i>)	
MG_CIE_DATOS	Ciencia de Datos para Políticas Públicas en Educación	
Nombre en inglés		
Data Science for public policy and decision making on education		
Unidad Académica u organismo de la unidad académica que lo desarrolla		
Instituto de Estudios Avanzados en Educación		
Docente responsable del curso		Semestre
Patricio Rodríguez Valdés		1 ^{er} Semestre 2023
SCT (Cantidad de horas de trabajo que el estudiante debe dedicar al curso para el logro de los resultados de aprendizajes expresados acorde con el Sistema de Créditos Transferibles. Un SCT corresponde a 30 horas cronológicas)	Horas de trabajo presencial	Horas de Trabajo no presencial de la/el estudiante
3	24 (Semestral)	48 (Semestral)
Horario Semanal (<i>Día y Hora</i>)		Modalidad (<i>Presencial, Semi Presencial, A Distancia, Online, Clases Sincrónicas, Clases No Sincrónicas</i>)
Jueves 18:00 - 19:20, 19:40 - 21:00, Primer Bimestre		Presencial
Requisitos (<i>Indique los requisitos de la asignatura de acuerdo con lo establecido en el plan de estudios y/o reglamento de carrera</i>)		Carácter del curso (<i>Señale si el curso es obligatorio, electivo o libre</i>)
Metodología Cuantitativa I		Electivo
Propósito general del curso (<i>Indique el propósito del curso consignado en el documento ficha de curso</i>)		
La Ciencia de Datos es una metodología que 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 se introducen 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 en educación		
Competencias a las que contribuye el curso (<i>Indique la o las competencias a las que tributa el curso, consignada(s) en el documento ficha de curso</i>)		
<ul style="list-style-type: none"> Reconoce el potencial de la captura, curación y almacenamiento de datos para la toma de decisiones y para el diseño, evaluación y seguimiento de políticas públicas con énfasis en educación. Distingue a partir de distintas métricas de ajuste, cuál es el modelo que obtiene mejores resultados. Reconoce distintos tipos de posibilidades de modelos de análisis de datos y selecciona algoritmos apropiados para generarlos. Crea e interpreta modelos predictivos o descriptivos utilizando el lenguaje de programación Python. Identifica la posibilidad de sobreajuste de modelo y aplica métodos para corregirlos. Conoce la importancia de reducir el número de variables para la captura e interpretabilidad de los modelos y aplica 		

métodos para reducirlas.		
Resultados de Aprendizaje (<i>Enunciados que establecen lo que el o la estudiante debe saber hacer en términos de actuaciones complejas al finalizar el curso. El conjunto de los resultados de aprendizaje debe evidenciar el logro del propósito del curso</i>)		
Al término del curso el/la estudiante demuestra que: <ul style="list-style-type: none"> – 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. 		
Metodología (<i>Descripción sucinta de las principales estrategias metodológicas que se desplegarán en el curso, coherente con un enfoque por competencias</i>)		Evaluación (<i>Descripción sucinta de las principales herramientas y situaciones de evaluación que permiten constatar el logro de los resultados de aprendizajes</i>)
Exposición en clases, lecturas complementarias y aprendizaje basado en proyectos.		<ul style="list-style-type: none"> • Controles de Lectura • Proyecto durante el curso.
Requisitos de aprobación (<i>requerimientos necesarios para la aprobación del curso, acordes con su propósito y normativa general que lo regula</i>). Incluir requisitos de asistencia en caso de existir.		
<ul style="list-style-type: none"> • Tener promedio de controles mayor e igual que 4.0 • Nota individual de cada tarea mayor e igual que 4.0 		
Unidades Temáticas		
Unidad	Nombre de la Unidad	Duración en semanas
1	Introducción a la Ciencia de Datos	1
Contenidos	Resultados de Aprendizaje de la Unidad	
En este primer módulo se desarrollarán los principales conceptos y definiciones de Ciencia de Datos, permitiendo entender su aplicación en gestión y políticas públicas. Se verán los objetivos de diseño de la ciencia de datos, otorgando particular importancia al diseño desde su origen. Se introducirá la principal herramienta a utilizar: Python y Jupyter notebooks.	<i>El/la estudiante:</i> <ul style="list-style-type: none"> - Conoce conceptos y definiciones de Ciencia de Datos, objetivos de diseño y una introducción a Python. - Declara los conceptos y definiciones a utilizar a lo largo del curso, dando particular importancia al diseño desde su origen. - Conoce la principal herramienta a utilizar: Python. 	
Unidad	Nombre de la Unidad	Duración en semanas
2	Adquisición, exploración, preparación y minería de datos Python	2
Contenidos	Resultados de Aprendizaje de la Unidad	
En este segundo módulo se desarrollarán los contenidos para comprender la correcta elección de entidades, entendimiento de sus atributos y el diseño de tratamientos para los datos, previo a cualquier análisis de la herramienta Python. Se establecerán contenidos que permitan manejar la limpieza y preparación de la información en su formato y forma según objetivos del diseño, asegurando la relevancia del insumo previo a análisis y procesos. Se aplicarán métodos estadísticos u otros	<i>El/la estudiante:</i> <ul style="list-style-type: none"> - Comprende la correcta elección de entidades, entendimiento de sus atributos y el diseño de tratamientos para los datos, previo a cualquier análisis de la herramienta Python. - Maneja la limpieza y preparación de la información en su formato y forma según objetivos del diseño, asegurando la relevancia del insumo previo a análisis y procesos. - Aplica métodos estadísticos u otros métodos avanzados para identificar patrones, utilizando Python. 	

<p>métodos avanzados para identificar patrones, utilizando Python. De esta manera, se entregarán conocimientos para la adquisición y pre-procesamiento de datos asegurando una mínima calidad de la información a trabajar, utilizando Python. Se entregarán contenidos para la preparación de datos con Python: estructura y filtros y minería de datos con Python.</p>		
Unidad	Nombre de la Unidad	Duración en semanas
3	Modelación de datos para Políticas Públicas	4
<p>En el tercer módulo se presentarán algunos algoritmos que permitan producir modelos descriptivos o predictivos según sea el caso, para llevar a cabo los análisis que generarán la evidencia para la toma de decisiones.</p> <p>En particular se revisarán los aspectos teóricos de los algoritmos de regresión (lineal y logística), clasificación (árboles de decisión puros y mezcla de árboles) además de agrupamiento (k-means) y como se pueden implementar usando las herramientas del curso.</p> <p>Se revisará un ejemplo práctico de un proyecto de Machine Learning en Python desde la limpieza de datos hasta la generación de un modelo, y se discutirán las implicancias para generar evidencia para focalizar los desafíos post crisis utilizando datos de fuentes abiertas de Chile u otros países.</p> <p>Se incorpora el uso de técnicas de explicación de modelos, que permiten abrir las cajas negras de la predicción descomponiendo el aporte de cada variable utilizada en la predicción final del resultado, utilizando SHAP (SHapley Additive exPlanations).</p>	<p>El/la estudiante:</p> <ul style="list-style-type: none"> - Es capaz de definir el diseño de modelos de acuerdo con el objetivo de diseño declarado y consolidado, involucrando una vista global del ciclo de Ciencia de Datos, y manteniendo siempre en el radar cuál es el objetivo. - Desarrolla modelos, de acuerdo con el objetivo, aplicando ML y/o HPC utilizando Python. - Maneja herramientas de validación de modelos en sus puntos críticos. 	
Unidad	Nombre de la Unidad	Duración en semanas
4	Evaluación, rediseño, interpretación y comunicación de resultados	1
<p>Los/las estudiantes aprenderán cómo comunicar efectivamente los resultados de los análisis efectuados a través del proceso de la Ciencia de Datos y cómo distintos tipos de visualizaciones pueden comunicar mejor el mensaje que se quiere entregar.</p> <p>Se discutirán ejemplos específicos de visualizaciones históricas que ayudaron a tomar decisiones que incluso salvaron vidas.</p>	<p>El/la estudiante:</p> <ul style="list-style-type: none"> - Analiza la calidad y detecta falencias en modelos desarrollados. - Es capaz de generar adecuaciones al proceso de análisis del modelo. - Interpreta correctamente los resultados y/o hallazgos. - Es capaz de comunicar los resultados a través de herramientas y métodos de visualización de datos interactiva. 	
<p>Bibliografía (Textos de referencia (obligatorios y sugeridos) a ser consultados por los y las estudiantes para favorecer los aprendizajes). Ordenar por unidad</p>		



Unidad 1.

- 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>.
- 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>.
- “Python for Developers”. <https://ricardoduarte.github.io/python-for-developers>
- 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>.

Unidad 2.

- 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>
- 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>
- “Jupyter Notebook Viewer”. [Online]. Disponible en: <https://bit.ly/2CqpH9h>.
- R. Kumar, IPython Notebooks to learn Python. Contribute to rajathkmp/Python-Lectures development by creating an account on GitHub. 2018.

Unidad 3.

- 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/>.
- Buenadicha Sánchez, C., Galdon, G., Hermosilla, M., Loewe, D., & Pombo, C. (2019). La gestión ética de los datos. Inter-American Development Bank. <https://doi.org/10.18235/0001623>.
- Butler, D. (2013). When Google got flu wrong. *Nature News*, 494(7436), 155. <https://doi.org/10.1038/494155a>
- 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>
- 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>
- 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>.

Unidad 4:

- 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/>.
- Doumont, J. (Eds) (2010). *English Communication for Scientists*. Cambridge, MA: NPG Education, 2010. Disponible en <https://www.nature.com/scitable/ebooks/english-communication-for-scientists-14053993/>.
- Dumont (2009). *Trees, maps, and Theorems: Effective Communications for Rational Minds*. Principae.
- The Python graph gallery (2017). The Python graph gallery. Disponible en <https://python-graph-gallery.com/>.

Recursos web (Recursos de referencia disponibles en plataformas digitales para el apoyo del proceso formativo del estudiante)

Python y Jupyter Notebooks:

- J. Vanderplas, The Jupyter Notebooks behind my O'Reilly report, "A Whirlwind Tour of Python": [jakevdp/WhirlwindTourOfPython](https://github.com/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/2oNFsOG> .
- "Tutorial de Python 3.6.3". [Online]. Disponible en: <https://bit.ly/1rfUbn>.
- "Azure Notebooks": <https://docs.microsoft.com/en-us/azure/notebooks/>
- Crear y ejecutar un Jupyter Notebook de Python en Azure notebooks: <https://docs.microsoft.com/en-us/azure/notebooks/tutorial-create-run-jupyter-notebook>
- Crear y compartir un notebook en Azure: <https://docs.microsoft.com/en-us/azure/notebooks/quickstart-create-share-jupyter-notebook>

Machine learning:

- Intel AI developer program (s.f.). Machine Learning. <https://software.intel.com/en-us/ai/courses/machine-learning>
- Microsoft Azure (2019). ML Notebooks for Azure Machine Learning. Disponible en <https://github.com/Azure/MachineLearningNotebooks>
- Müller (2020). Introduction to Machine learning with scikit-learn. Disponible en <https://github.com/amueller/ml-workshop-short>

Visualización y presentación de resultados:

- Doumont, J. (Eds) (2010). *English Communication for Scientists*. Cambridge, MA: NPG Education, 2010. Disponible en <https://www.nature.com/scitable/ebooks/english-communication-for-scientists-14053993/>
- Doumont, J. (2013). Creating effective slides: Design, Construction, and Use in Science. Disponible en <https://www.youtube.com/watch?v=meBXuTIPJQk>.
- The Python graph gallery (2017). The Python graph gallery. Disponible en <https://python-graph-gallery.com/>