

Asignatura: Introducción a la investigación con estrategias computacionales

Modalidad: Seminario de investigación

Carga horaria total: 45 horas (15 clases semanales de 3 horas)

Profesor Titular: Lucas Rubinich

Equipo docente y funciones: Paula Miguel (Adjunta a cargo), Ezequiel Saferstein, Pedro Orden

Cuatrimestre y año de dictado: 1º o 2º cuatrimestre 2024

Fundamentos y objetivos de la asignatura:

Las ciencias sociales se encuentran ante una expansión sin precedentes de sus fronteras ante los procesos de digitalización de la vida cotidiana. Cada vez más prácticas y relaciones sociales comprenden un comportamiento digital que genera datos y deja huellas. En ese sentido, las tecnologías digitales han tenido y tienen un fuerte impacto cultural, social, epistémico e inauguran áreas de estudio interdisciplinario como las llamadas ciencias sociales computacionales y las humanidades digitales. El volumen, la complejidad y la disponibilidad de datos ha crecido como nunca antes. El mundo digital no solamente ofrece nuevas fuentes y formatos de datos, sino también nuevas formas de vínculo social, que merecen ser abordados con herramientas adecuadas y metodologías pertinentes a la investigación desde la teoría social. Así, crece la investigación que explora los impactos culturales y sociales de las tecnologías emergentes, especialmente en torno a datos y desarrollos relacionados con la inteligencia artificial; metodologías de investigación digital; investigación basada en archivos digitales, entre otras posibilidades.

Este seminario invita a que los estudiantes puedan introducirse en la ciencia de datos y valerse de métodos y estrategias computacionales para trabajar de manera dinámica en la recolección, procesamiento y análisis de datos complejos. Para eso, propone una introducción a la investigación aplicando esas herramientas y técnicas sobre diferentes fuentes, tipos y volúmenes de datos digitales. Esto implica una reflexión, no solamente en cuanto a los aportes de estas metodologías para el análisis de lo social sino también en relación con su lugar teórico-epistemológico, recuperando los alcances y límites de las técnicas específicas y evaluando de manera crítica sus fortalezas, sesgos y tensiones éticas.

El objetivo del seminario es que los estudiantes adquieran herramientas y estrategias que les permitan profundizar en el análisis de diferentes fuentes y volúmenes de datos complejos. Esto implica incorporar el flujo de trabajo con datos: comprender y plantear un problema, obtener y entender la información, minar, refinar y construir un modelo interpretativo sobre los datos. Para eso la modalidad de trabajo prevé el desarrollo de ejercicios y aplicaciones prácticas abordando casos específicos, utilizando bases de datos públicos y/o datos recolectados a tal efecto. Esto permitirá avanzar sobre aspectos epistemológicos y metodológicos, a la vez que el acercamiento a las estrategias computacionales desde la práctica permitirá sopesar su aporte para las ciencias sociales, observando los contextos de uso y aplicación, sus alcances y sus límites.

Eso implica:

- Conocer e incorporar distintas herramientas disponibles y su marco de referencia teórico-conceptual.
- Incorporar fundamentos básicos de programación para el análisis de datos complejos y el manejo de bases de datos.
- Incorporar herramientas y habilidades para la obtención, análisis y visualización de datos.
- Familiarizarse con la aplicación e implementación de algoritmos y modelización de datos.
- Evaluar fortalezas y límites de las diferentes metodologías abordadas para poder tomar decisiones metodológicas justificadas en función de sus intereses teórico-conceptuales.
- Integrar los conocimientos específicos en diseños de investigación social coherentes, consistentes y viables.

Contenidos desglosados por unidades:

Unidad 1. Las estrategias computacionales como caja de herramientas

Introducción a la ciencia de datos para ciencias sociales. Cuestiones éticas. Introducción a la programación, herramientas para la materia. Distintos tipos de datos. Flujo de trabajo con datos. Errores comunes.

Unidad 2. Captura y limpieza de datos

Posibles fuentes de datos (APIs, scrapping, OCR). Integrar datos de múltiples fuentes. Bases de datos. Preparación de datos numéricos: estandarización, normalización, trabajo con *outliers*, datos faltantes y nulos. Preprocesado y limpieza de datos textuales.

Unidad 3. Modelizar, analizar y visualizar datos

Básicos de estadística descriptiva y probabilidad. Métodos de aprendizaje supervisado y no supervisado. Modelos de regresión, modelos de clasificación y clusterización. Entrenamiento y validación. Reducción de la dimensionalidad. Métricas de performance. Árboles de decisión. Ensamblados aleatorios. Básicos del análisis de redes. Visualización de datos. Tipos de diagramas básicos.

Unidad 4. El análisis de datos textuales

Cómo analizar textos. Introducción al procesamiento del lenguaje natural (NLP). Matriz de frecuencias, TF-IDF, nubes de palabras, análisis de sentimiento. Clasificación de documentos. Detección y modelado de tópicos. Visualización de datos. Un acercamiento a la etnografía digital.

Unidad 5. La aplicación en la investigación (transversal a la cursada)

La construcción de una pregunta de investigación. Diseños de investigación integrales. El trabajo con conceptos y categorías analíticas. Los debates, aportes, lecturas críticas y el estado de la cuestión. Las posibilidades, límites, alcances de los datos y las estrategias de abordaje. La presentación de resultados.

Recursos y fuentes

Las unidades temáticas del programa se complementan con recursos para obtener bases de datos públicos, fuentes documentales y otros datos que presenten casos para prácticas y análisis a lo largo de la cursada.

- Datos Argentina (datos.gob.ar): 1176 datasets disponibles.
- Buenos Aires Data (data.buenosaires.gob.ar): 548 datasets disponibles.
- Otras fuentes para datasets, modelos, código: GitHub, Huggingface, Kaggle, entre otras.

Bibliografía General dentro de cada unidad (obligatoria)

Unidad 1.

Cady, Field (2017). *The data science handbook*. Hoboken, NJ: John Wiley & Sons.

Downey, A., Elkner, J., Meyers, C. (2002). *Aprenda a pensar como un programador con Python*. Wellesley, MA: Green Tea Press.

Kelleher, J. D., & Tierney, B. (2018). *Data science*. MIT Press.

VanderPlas, Jake. (2017). *Python data science handbook. Essential tools for working with data*. Sebastopol, CA: O'Reilly.

Salganik, M. (2018). *Bit by Bit. Social research in the digital age*. Princeton: Princeton University Press.

Unidad 2.

Foster, I., Ghani, R., Jarmin, R. et al. (eds.) (2017). *Big Data and Social Science. A Practical Guide to Methods and Tools*. Boca Raton, FL: CRC Press.

Kelleher, J. D. and Tierney, B. (2018). *Data science*. Cambridge, MA: The MIT Press.

vanden Broucke, S. and Baesens, B. (2018). *Practical Web Scraping for Data Science. Best Practices and Examples with Python*. Nueva York, NY: Apress.

Unidad 3.

Cramer, D. and Howitt, D. (2004). *The SAGE Dictionary of Statistics. A practical resource for students in the social sciences*. Londres: SAGE.

Foster, I., Ghani, R., Jarmin, R. et al. (eds.) (2017). *Big Data and Social Science. A Practical Guide to Methods and Tools*. Boca Raton, FL: CRC Press.

Lynch, Scott M. (2007). *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. Nueva York, NY: Springer.

Illiinsky, N. and Steele, J. (2011). *Designing Data Visualizations*. Wellesley, MA: Green Tea Press.

Unidad 4.

Bengfort, Benjamin; Bilbro, Rebecca y Ojeda, Tony (2018). *Applied Text Analysis with Python*. Sebastopol, CA: O'Reilly Media.

Dipanjan, Sarkar (2016). *Text Analytics with Python. A Practical Real-World Approach to Gaining Actionable Insights from your Data*. Apress.

Pink, Sarah. et al. (2019). *Etnografía digital. Principios y prácticas*. Madrid: Morata.

Unidad 5.

Casos de estrategias aplicadas

Chan, F. y L. Matyas (eds.) *Econometrics with Machine Learning*, Advanced Studies in Theoretical and Applied Econometrics 53.

Cotik, Viviana; Debandi, Natalia; Luque, Franco; Miguel, Paula; Moro, Agustín; Pérez, Juan Manuel; Serrati, Pablo; Zajac, Joaquín; Zayat, Demián (2020). *A study of Hate Speech in Social Media during the COVID-19 outbreak*. En: Open Review.

Kessler, G., Vommaro, G., & Paladino, M. (2022). Antipopulistas reaccionarios en el espacio público digital. *Estudios Sociológicos de El Colegio de México*, 40(120), 651–692.

Pérez, Juan Manuel; Luque, Franco M.; Zayat, Demian; Kondratzky, Martín; Moro, Agustín; Serrati, Pablo Santiago; Zajac, Joaquín; Miguel, Paula; Debandi, Natalia; Gravano, Agustín; Cotik, Viviana y Cotik, V. (2023). Assessing the Impact of Contextual Information in Hate Speech Detection. *IEEE Access*, 11:30575-30590. doi: 10.1109/ACCESS.2023.3258973.

Perry, Barbara y Olsson, Patrik (2009) Cyberhate: the globalization of hate, *Information & Communications Technology Law*, 18:2, 185-199, DOI: 10.1080/13600830902814984

Rabbia, Hugo (2021) Religión, creencias conspirativas e ideología política como predictores del apoyo a la gestión gubernamental de la pandemia de COVID-19 en Argentina. *Psicología*, 40(1-2): 29-58.

Sumpter, David. (2018). *Outnumbered: From Facebook and Google to Fake News and Filter-bubbles*. Londres: Bloomsbury Sigma.

Bibliografía general (complementaria)

Metodología e investigación en Ciencias Sociales

Archenti, N., Marradi, A., & Piovani, J. I. (2018). Manual de metodología de las ciencias sociales. Buenos Aires: Siglo Veintiuno Editores.

Becker, Howard (2009) *Trucos del oficio. Cómo conducir su investigación en ciencias sociales*. Buenos Aires: Siglo Veintiuno Editores.

Becker, Howard (2012) *Manual de escritura para científicos sociales. Cómo empezar y terminar una tesis, un libro o un artículo*. Buenos Aires: Siglo Veintiuno Editores.

Bourdieu, Pierre; Chamboredon, J. C. y Passeron, J.C. (1987) *El oficio del sociólogo*. México: Siglo Veintiuno Editores.

Bourdieu, P. y Wacquant, L. (2005) *Una invitación a la sociología reflexiva*. Buenos Aires: Siglo Veintiuno Editores.

Champagne, P. et al. (1993) *Iniciación a la práctica sociológica*. México: Siglo Veintiuno Editores.

Sautu, R.; Boniolo, P.; Dalle, P.; Elbert, R. (2005). *Manual de metodología. Construcción del marco teórico, formulación de los objetivos y elección de la metodología*. Buenos Aires: CLACSO.

Seawright, J. (2016). *Multi-Method Social Science: Combining Qualitative and Quantitative Tools*. Cambridge: Cambridge University Press.

Vasilachis de Gialdino, Irene (coord.) (2006) *Estrategias de investigación cualitativa*. Buenos Aires: Gedisa.

Wainerman, C. y R. Sautu (2001). *La trastienda de la investigación*. Buenos Aires: Lumiere.

Metodología de la enseñanza

En el desarrollo del seminario se dictarán clases teórico-prácticas para abordar los contenidos del programa. Se espera una comprometida dedicación y activa participación de los estudiantes que, en las reuniones, desarrollarán lecturas, prácticas, ejercicios y análisis sobre los temas y problemas, pensando en la aplicación de los contenidos en casos concretos de investigación. Se apuntará a que se construyan preguntas y planteos de problemas desde los distintos ejes de contenidos con anclaje en las prácticas de investigación. Las lecturas indicadas en el programa serán de carácter obligatorio.

Régimen de promoción y evaluación

El seminario asigna 50 horas de investigación. La evaluación final consistirá en la elaboración de un trabajo final que integre temas y bibliografía del curso e incorpore de estrategias metodológicas y analíticas en un caso concreto de investigación.

Para obtener la regularidad en la materia y quedar habilitados para la acreditación final de horas de investigación los estudiantes deben cumplir y aprobar las actividades propuestas en la cursada y contar con el 75% de asistencia a clase (hasta 3 faltas). Una vez concluida la cursada, y habiendo cumplido con esos requisitos, los estudiantes quedan habilitados para presentar el trabajo final.

Los trabajos finales se presentan en fechas de exámenes, según calendario académico. Para ello los estudiantes deben inscribirse a través del SIU Guaraní. Las entregas finales deben ser enviadas al correo del seminario, indicando nombre completo, DNI y cuatrimestre de cursada.