Programa de actualización en**Estrategias computacionales para la investigación social****Año 2025****Coordinadora: Paula Miguel****Fundamentación**

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. El volumen, la complejidad y la disponibilidad de datos ha crecido como nunca antes. 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 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 grandes volúmenes de datos y desarrollos relacionados con la inteligencia artificial; metodologías de investigación digital; investigación basada en archivos digitales, entre otras posibilidades. Esos aportes pueden ser aplicados a la toma de decisiones y el desarrollo de nuevos procesos en la gestión pública y productiva, en la elaboración de estudios de mercado y de opinión pública, en la gestión y análisis de la comunicación digital, la sistematización de información y la investigación académica, entre otras. Al mismo tiempo, se requiere de análisis apropiados desde lo técnico que puedan valerse de los problemas y preguntas más clásicas de la teoría social para generar respuestas y soluciones relevantes.

Este programa invita a que las y los estudiantes puedan valerse de los aportes de la ciencia de datos y acercarse a las herramientas analíticas, métodos y estrategias computacionales para trabajar de manera dinámica en la recolección, procesamiento y análisis de datos complejos. Para eso, propone aplicar 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 programa 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, construir un modelo

interpretativo sobre los datos y poder presentar los resultados. 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.

Objetivos

Objetivo general:

El objetivo del programa es que los participantes adquieran herramientas computacionales que les permitan profundizar en el análisis de diferentes fuentes y volúmenes de datos complejos y avanzar sobre los aspectos epistemológicos y metodológicos de las estrategias computacionales para las ciencias sociales. Esto implica evaluar de manera crítica el contexto de uso y aplicación, alcances y límites de las técnicas específicas para poder tomar decisiones metodológicas justificadas en función de los intereses teórico-conceptuales.

Objetivos específicos:

- Conocer e incorporar distintas herramientas disponibles y su marco de referencia teórico-conceptual, evaluando sus fortalezas y límites.
- Incorporar fundamentos básicos de programación y manejo de bases de datos para la extracción, análisis y visualización de análisis de datos complejos.
- Familiarizarse con la aplicación e implementación de algoritmos y modelización de datos.
- Integrar los conocimientos específicos en diseños de investigación coherentes, consistentes y factibles

Destinatarios

Los contenidos de este programa de actualización están orientados especialmente a profesionales de las ciencias sociales que trabajan en sectores que requieren del análisis de datos y la sistematización de información para la administración pública; la elaboración de estudios de mercado y de opinión pública; la gestión productiva; la gestión y análisis de la comunicación digital; la investigación académica, entre otras posibilidades. También está abierto a profesionales de otras áreas y al público en general interesado en la temática.

Se esperan postulantes que cuenten con título de carreras de al menos 4 años, provenientes de universidades nacionales o extranjeras o formación profesional equivalente y cumplan con los requisitos que determine el programa.

Modalidad de cursada, estrategia didáctica y evaluación:*Modalidad de cursada*

El programa se dicta en modalidad a distancia. Se prevé que la duración de la cursada abarque un cuatrimestre (32 clases), con una carga horaria total de 128 horas. Cada clase consiste en un encuentro sincrónico de dos horas, que se complementa con dos horas de trabajo asincrónico.

Estrategia didáctica

En los encuentros sincrónicos se dictarán clases teórico-prácticas, modalidad taller, para abordar los contenidos del programa. Se espera una comprometida dedicación y activa participación de los estudiantes que desarrollarán prácticas, ejercicios y análisis sobre los temas propuestos. Se apuntará a que se construyan preguntas y planteos de problemas desde los distintos ejes de contenidos, pensando en su aplicación a casos concretos de investigación. Las actividades indicadas en el programa serán de carácter obligatorio.

Recursos y fuentes

Los módulos temáticos del programa se complementan con recursos para obtener bases de datos públicas, fuentes documentales, etc. que presenten casos de interés para realizar prácticas y análisis a lo largo de la cursada (entorno Python, Google Colab, Campus, etc.).

Evaluación

Se tendrá en cuenta la asistencia a los encuentros sincrónicos y la resolución de las actividades y ejercicios propuestos a lo largo de la cursada de los módulos. El programa prevé la elaboración de un Trabajo Integrador Final (TIF) que podrá ser diseñado y resuelto en pequeños grupos. Se espera que en este trabajo se incorporen los temas, estrategias metodológicas y analíticas en un caso concreto de investigación e implementación.

Seminarios y contenidos mínimos

Módulo	Carga Horaria	Créditos
Módulo I. Introducción a la programación en Python	32	2
Módulo II. Analítica de datos	24	1.5
Módulo III. Fundamentos del aprendizaje automático	24	1.5
Módulo IV. Introducción al procesamiento del lenguaje natural (NLP)	24	1.5
Módulo V. Implementación y desarrollo de TIF	24	1.5
Carga Horaria Total	128 Hs	8

Contenidos mínimos

Módulo I. Introducción a las estrategias computacionales y programación en Python

Ciencia de datos y computación aplicada a las ciencias sociales. Actualización, debates, aportes, mirada crítica. Las posibilidades, límites, alcances de las estrategias de abordaje. Distintos tipos de datos. Principios de programación orientada a objetos (POO). Lógica y estructuración de código (Python). Algoritmos y secuencialidad. Sintaxis de un programa. Fuentes de información estadística. Acceso, recolección y extracción de datos (APIs, scrapping, pedidos). Buenas prácticas. Cuestiones éticas.

Módulo II. Analítica de datos

Datos estructurados y no estructurados. Estructura de los datos. Manejo de base de datos. Consultas. Preparación de datos numéricos. Estandarización, normalización, trabajo con outliers y datos faltantes. Fundamentos del análisis estadístico descriptivo. Medidas de tendencia central y de dispersión. Básicos de probabilidad. Estimación, probabilidad y distribución. Outputs y visualización de datos (tipos de información, tablas y gráficos básicos).

Módulo III. Fundamentos del aprendizaje automático

Modelos de aprendizaje supervisado y no supervisado. Regresión lineal, cuadrados mínimos, regresión logística. Interpretación y validación. Problemas y errores típicos (overfitting, underfitting). Regularización. Entrenamiento y validación. Métricas para la medición de performance en regresión y clasificación. Modelo de árboles de decisiones y clusterización. Árboles de decisión. Ensamblados aleatorios (random forest). Clustering de datos y reducción de la dimensionalidad.

Módulo IV. Análisis de textos y procesamiento del lenguaje natural (NLP)

Extracción de datos textuales (APIs, OCR, scrapping). Preprocesado y limpieza de datos textuales. Expresiones regulares. Categorías gramaticales (POS). Reconocimiento de entidades (NER). Matriz de frecuencias, TF-IDF. Aprendizaje supervisado y no-supervisado. Clasificación de documentos. Detección de tópicos. Grandes modelos de lenguaje (LLMs).

Módulo V. Implementación y desarrollo de TIF

(Módulo transversal a la cursada. Acompaña la preparación del TIF)

Diseños de investigación integrales. Modularización de código. Python autoejecutable. Entornos: virtual-environment y conda. Buenas prácticas de desarrollo. Estandarización, pruebas y depuración de código. Documentación del desarrollo. Presentación de resultados, repositorios (Git). Parámetros de calidad.

Bibliografía general:

- Bengfort, Benjamin; Bilbro, Rebecca y Ojeda, Tony (2018). *Applied Text Analysis with Python*. Sebastopol, CA: O'Reilly Media.
- Bhimavarapu, U., & Hemanth, J. D. (2023). *Learning Professional Python: Volume 1: The Basics*. CRC Press.
- Cady, Field (2017). *The data science handbook*. Hoboken, NJ: John Wiley & Sons.
- Chen, D. Y. (2017). *Pandas for everyone: Python data analysis*. Addison-Wesley Professional.
- Dipanjan, Sarkar (2016). *Text Analytics with Python. A Practical Real-World Approach to Gaining Actionable Insights from your Data*. Apress.
- Downey, A., Elkner, J., Meyers, C. (2002). *Aprenda a pensar como un programador con Python*. Wellesley, MA: Green Tea Press.
- 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.
- Illiinsky, N. and Steele, J. (2011). *Designing Data Visualizations*. Wellesley, MA: Green Tea Press.
- James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An introduction to statistical learning: With applications in python*. Springer Nature.
- Lynch, Scott M. (2007). *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. Nueva York, NY: Springer.
- Kelleher, J. D., & Tierney, B. (2018). *Data science*. MIT Press.
- Salganik, M. (2018). *Bit by Bit. Social research in the digital age*. Princeton: Princeton University Press.
- vanden Broucke, S. and Baesens, B. (2018). *Practical Web Scraping for Data Science. Best Practices and Examples with Python*. Nueva York, NY: Apress.
- VanderPlas, Jake. (2017). *Python data science handbook. Essential tools for working with data*. Sebastopol, CA: O'Reilly.
- Wilke, C. O. (2019). *Fundamentals of data visualization: a primer on making informative and compelling figures*. O'Reilly Media.

CV Resumido Plantel Docente

Los contenidos de cada módulo se encuentran a cargo de profesionales de las ciencias sociales con experiencia en programación y en la aplicación de estas estrategias en el campo profesional y académico. También se contará con la participación de especialistas en temáticas específicas.

Paula Miguel (Coordinadora)

Socióloga y Doctora en Ciencias Sociales por la Universidad de Buenos Aires, especializada en análisis cultural, análisis cualitativo y ciencia de datos. Es Profesora Titular Regular de Sociología e Investigadora Adjunta CONICET en la Universidad de Buenos Aires. Desde 2012 dirige proyectos de investigación acreditados y financiados

por la Universidad de Buenos Aires y la Agencia I+D+i. Entre ellos, coordina una línea de investigación interdisciplinaria enfocada en el procesamiento del lenguaje natural y detección de discursos discriminatorios en redes sociales.

Zacarías Abuchanab

Es Licenciado en Sociología (UBA). Maestrando en Metodología de la investigación social (UNTREF). Posgrado en Estadística Matemática - sin finalizar - (UBA). Consultor permanente de arquitectura de datos para la División de Estadísticas de la CEPAL (UN). Profesor de la materia "Fundamentos de Programación para Ciencias Sociales" (Universidad Austral) y del seminario de "Procesamiento de Lenguaje Natural" en la Diplomatura en Ciencias Sociales Computacionales (EIDAES-UNSAM).

Ariana Bardauil

Es Licenciada en ciencia política por la Universidad Nacional de La Matanza (UNLaM) y candidata a magíster en Generación y Análisis de la Información Estadística en la Universidad Nacional de Tres de Febrero (UNTREF). Docente de minería de texto en la Universidad de Flores (UFLO) y de demografía social en la UNLaM. Se desempeña como Data scientist en el Ministerio de Salud (GCBA), Co-organizadora de R en Buenos Aires y coordinadora del Núcleo de Innovación Social.

Ramiro Fernández

Es Licenciado en Sociología (UBA). Se encuentra finalizando su tesis de Maestría en Explotación de Datos y Descubrimiento del Conocimiento (UBA). Cuenta con una Diplomatura en Cloud Data Engineer (ITBA). Ha trabajado como cientista de datos en organismos públicos y privados. Actualmente, se desempeña como Data Engineer Manager en ADDI. Coordina la Diplomatura en Ciencias Sociales Computaciones en la UNAB.