

Diploma Universitario en Ciencias Sociales Computacionales y Humanidades Digitales

Coordinador: Dr. Germán Rosati

Fundamentación

Muchas áreas del conocimiento se han beneficiado de la aplicación creativa de las técnicas englobadas bajo el nombre de “Machine Learning” o “Aprendizaje Automático”. Sin embargo, las ciencias sociales argentinas no parecen aún haber hecho uso extensivo de las mismas. Junto con la aparición de nuevas fuentes y formatos de datos (vinculados a las tecnologías mobile, a las redes sociales, a la Internet of Things, etc.) se comienza a redefinir un área disciplinar que intersecta problemas clásicos de las ciencias sociales, técnicas avanzadas de análisis de datos y nuevos datos: las Ciencias Sociales Computacionales -CSS-.

A su vez, la administración pública y los sectores productivos requieren analistas calificados no sólo en aspectos técnicos, sino también en la capacidad de generar problemas, preguntas y análisis relevantes. Las CSS cumplen un rol fundamental en este aspecto.

La presente diplomatura se propone realizar una primera aproximación al rol de las CSS y a algunas de sus técnicas principales. Los resultados se observan cotidianamente: desde los servicios provistos por las compañías más importantes, pasando por la implementación y evaluación de políticas públicas, llegando hasta el abordaje de problemas clásicos de las ciencias sociales como la desigualdad, la estructura social, el conflicto, las construcciones culturales, etc. En todos estos casos, se estudian comportamientos humanos (ciencias sociales) por medios computacionales

Se buscará un acercamiento desde un enfoque conceptual (fundamentos teórico-metodológicos, casos y problemas de aplicación, etc.) como técnico (análisis de

algoritmos, herramientas con interfaces gráficas, etc.) de las mismas. Se realizarán lecturas y exposiciones teóricas sobre las temáticas, pero también se hará especial énfasis en su aplicación práctica, particularmente centrada en problemas relevantes de las Ciencias Sociales Computacionales.

Objetivos

La diplomatura se propone que los asistentes:

- realicen una primera aproximación a las discusiones epistemológicas y metodológicas acerca de la aplicación de las CSS en la actividad científica y las diferencias con el modelado estadístico tradicional.
- incorporen algunos fundamentos de programación para avanzar en análisis de datos.
- logren construir visualizaciones efectivas de los datos que analizan.
- incorporen fundamentos conceptuales del flujo de trabajo y de la metodología asociadas a estas técnicas.
- conozcan y sepan implementar diversos algoritmos y métodos de modelización.
- identifiquen situaciones de aplicación de este tipo de modelos a problemas de investigación básica y aplicada.

Requisitos de ingreso

El postulante deberá contar con título superior no universitario, título universitario o formación profesional equivalente y reunir los prerrequisitos que determine el Comité Académico de la carrera, a fin de comprobar que su formación resulte compatible con las exigencias del Diploma. En casos excepcionales de postulantes que se encuentren fuera de los términos precedentes, podrán ser admitidos siempre que demuestren, a través de las evaluaciones y los requisitos que la Universidad establezca, poseer preparación y experiencia laboral acorde con los estudios de posgrado que se proponen iniciar así como aptitudes y conocimientos suficientes para cursarlos satisfactoriamente.

Asignaturas o módulos

Los asistentes deberán cursar los siguientes módulos para aprobar la diplomatura:

- Módulo 1. Introducción a la Ciencia de Datos para Científicos Sociales.
- Módulo 2. Visualización de datos.
- Módulo 3. Introducción al modelado de datos.
- Módulo 4. Fundamentos de Machine Learning.
- Módulo 5. Minería de Texto + webscraping.

Carga horaria y duración

El Diploma en Ciencias Sociales Computacionales y Humanidades Digitales está conformado por cinco materias divididas en dos cuatrimestres, con un total de 120 horas (presenciales y de trabajo domiciliario).

Duración:

Modalidad de evaluación

Para obtener el Diploma debe contarse con un 75% de asistencia a clases y con la aprobación de un trabajo final escrito en cada uno de los módulos, acordado con el profesor respectivo y monitoreado por el mismo.

Contenidos mínimos

Módulo 1. Introducción a la Ciencia de Datos para Científicos Sociales

- ¿Qué es la ciencia de datos? Su relación y aplicaciones en las ciencias sociales (construcción de nuevas métricas de desigualdad, análisis de determinantes sociales de la salud, estudio de tópicos en textos legislativos).
- Elementos básicos de programación estadística en R. Objetos (vectores, listas, matrices, data frames, tibbles).
- Procesamiento de datos con tidyverse. Data wrangling (limpieza de datos). Análisis estadístico descriptivo con R.

Módulo 2. Visualización de datos

- Introducción conceptual (Infografías vs. gráficos; ejemplos históricos. Joseph Minard, William Playfair, John Snow, Florence Nightingale, Harry Beck)
- Replicabilidad
- Gramática de gráficos
- Visualización y generación de gráficos en R: Nociones de graficación (forma, color, tamaño, color).
- Introducción a ggplot2.

Módulo 3. Introducción al modelado de datos

- Modelos básicos: regresión lineal y logística. Estimación, mínimos cuadrados, máxima verosimilitud. Interpretación y validación: medidas de ajuste, pruebas de hipótesis, diagnósticos.
- Nociones de aprendizaje automático. Tipos de problemas en aprendizaje supervisado: clasificación y regresión. Error de entrenamiento, error de testing. Sobre-ajuste.
- Regularización: LASSO y Ridge (?).
- Introducción a tidymodels.

Módulo 4. Fundamentos de Machine Learning

- Clasificadores basados en árboles: CART.
- Algoritmos de Ensamble: bagging, random forest, boosting, Gradient Boosting. Interpretable Machine Learning: Herramientas para la interpretación de modelos de caja negra (Feature Importance, Partial Dependence Plots, Individual Conditional Expectancy Plots).

Módulo 5. Minería de Texto + webscraping

- Webscraping: construyendo datasets desde la web. HTML, tags y atributos. Introducción a rvest.
- APIs
- Etapas del preprocesamiento: tokenización, normalización (lemas y stemming), stopwords.

- Vectorización de texto: modelo Bag of Words, Term-frequency Matrix (TFM), diferentes formas de ponderación de la TFM. Distancias entre palabras (caracteres y conceptual).
- Topic modelling: Latent Dirichlet Allocation, Non-Negative Matrix Factorization.
- Word Vectors

Certificación

La certificación de la Diplomatura será emitida por el Instituto de Altos Estudios Sociales, de enseñanza universitaria de carreras de grado y posgrado, y de investigación en ciencias sociales de la Universidad Nacional de San Martín.

Las certificaciones se confeccionarán de acuerdo con las especificaciones técnicas que se detallan y presentan en el siguiente modelo:

Formato:

- Orientación de página "Vertical"
- Tamaño: A4
- Fuente: Times New Roman 18

Contenido:

Datos institucionales y académicos:

- Unidad Académica.
- Denominación completa del Curso.
- Duración del curso expresada en horas.
- Nombre y Apellido del cursante.
- Firma del Decano y Secretario Académico de la Unidad Académica respectiva.

Plantel Docente

El plantel docente está compuesto por profesionales e investigadores en ciencias sociales con amplia experiencia vinculada al análisis cuantitativo de información tanto en el mundo académico como en el ámbito profesional.

- Pablo Tiscornia (DNME-MINTUR, FLACSO)
- Guido Weksler (CONICET, FCE-UBA)

- Antonio Vázquez Brust (UTDT, FADU-UBA)
- Martín Iván Schuster (GCBA, UCL, IDAES-UNSAM)
- Victoria O'Donnell (CEPAL, UBA)
- Germán Rosati (CONICET, IDAES-UNSAM)

Ayudantes

- Laia Domenech Burin (IDAES-UNSAM)
- Tomás Maguire (IDAES-UNSAM)