

## **Especialización en Economía- Opción Ciencia de Datos**

### **Edición 2024**

#### **Descripción sintética de asignaturas**

##### **Módulo 1: Introducción a la ciencia de datos**

###### **Introducción al trabajo con datos (6 créditos)**

Leonardo Barceló

El objetivo de la materia es introducir a los estudiantes al mundo de la ciencia de datos, brindarles los conceptos básicos sobre datos, de qué forma se almacenan, en qué formatos vienen y qué herramientas hay disponibles para procesarlos. Al final el curso los/las estudiantes podrán distinguir entre datos e información, conocerán las distintas categorizaciones por tipo de dato; trabajarán con las principales herramientas para procesamiento de datos: R y RStudio, sus principales librerías (dplyr, tidy), Python y sus principales librerías (pandas, numpy), MySQL y principales conceptos de bases relacionales y joins, Google colab, Jupyter Notebooks y aprenderán sobre control de versiones en github. Finalmente se dará una introducción a datos no estructurados y APIs.

###### **Proceso de desarrollo de programación (3 créditos)**

German Tejería

Esta materia tiene como objetivo introducir a los/las estudiantes al desarrollo de proyectos de programación. El curso cubre el paradigma de la programación funcional y la programación basada en objetos, ejemplos con creación de algoritmos básicos y funciones, detección de bugs, testeos e iteraciones, buenas prácticas en programación y desarrollo de código, y una introducción a las principales metodologías en programación y desarrollo de código.

###### **Introducción al análisis de datos con Python (3 créditos)**

Daniel Miles

El objetivo de esta materia es acercar a los/las estudiantes a la programación en Python orientada a analizar datos para resolver problemas. Aprender a programar para abordar los típicos problemas de las ciencias sociales desde otra perspectiva.

###### **Datos no estructurados y semiestructurados (4 créditos)**

Elina Gomez, Nicolas Schmidt

El objetivo de esta materia es conocer las principales herramientas y algoritmos específicos que permiten analizar datos no estructurados o semiestructurados. Se utilizarán técnicas de extracción de datos de la web: scraping con, se verán herramientas para lectura de texto y documentos en formato .docx y .pdf, herramientas para la lectura de imágenes o .pdf guardados como imágenes.

## Módulo 2. Procesamiento de datos

### Análisis exploratorio (4 créditos)

Guillermo Lezama

Esta materia tiene como objetivo introducir a los/las estudiantes a las principales técnicas de análisis exploratorio de datos a partir de las cuales obtener importantes insights y detectar patrones. La primera parte del curso estará orientada a las estadísticas descriptivas, filtrado y armado de bases auxiliares. La segunda parte estará orientada al manejo de variables: creación, modificación y eliminación. La tercera parte estará dedicada a gráficos exploratorios, correlación de variables y detección de missings (diferencia entre NAN y NA).

### Limpieza y transformación de datos (4 créditos)

Gladys Cardozo

Esta materia tiene como objetivo que los/las estudiantes aprendan técnicas de limpieza de datos y transformación que les permitan definir el modelo de datos que mejor se adapte a sus fines de análisis. A lo largo del curso se verán funciones condicionales, expresiones regulares, librerías para el manejo de datos de texto y fecha, automatización del armado de bases de datos, tratamiento de missings, métodos de imputación, análisis de componentes principales, normalización y estandarización de variables.

### Visualización (4 créditos)

Iael Klaczko

Esta materia tiene como objetivo el dominio de los conceptos de visualización para lograr una presentación atractiva y eficiente de los datos. Al finalizar los/las estudiantes podrán definir la visualización que mejor se adapte a cada proyecto de análisis de datos particular, podrán desarrollar presentaciones de proyectos de ciencias de datos atractivas y autocontenidas. El curso cubre los siguientes temas: grammar of graphics, ggplot2 y plotly, concepto de dashboards (dashboards en Python y en R con flexdashboard), introducción a softwares de visualización (Tableau y PowerBI), mejores prácticas para la visualización de datos (elección de colores, paletas y formas), ventajas y desventajas de las visualizaciones más utilizadas.

### Visualizaciones interactivas (2 créditos)

Iael Klaczko

Esta materia tiene como objetivo introducir a los/las estudiantes a las principales herramientas de presentación de hallazgos y resultados obtenidos de la etapa de visualización. El curso comprenderá la construcción de aplicaciones web con shiny que permiten la interacción con visualizaciones y datos, así como el armado de reportes con R Markdown y la integración de ambas.

## Módulo 3: Técnicas de análisis de datos

### Machine Learning I (4 créditos)

El objetivo de esta materia es presentar a los estudiantes los aspectos generales de las distintas técnicas de Machine learning e introducir a los estudiantes a los principales conceptos de esta área de conocimiento. Se presentarán las diferencias entre las técnicas de machine learning y la econometría, la importancia del tamaño de la base de datos, las técnicas de segmentación de la base de datos y los métodos que existen para evaluar las predicciones fuera de la muestra cross-validation y k-folds.

Adicionalmente se presentan las diferencias entre los problemas de clasificación y regresión y entre el aprendizaje supervisado y no supervisado. Se discuten los aspectos a considerar a la hora de elegir los algoritmos o métodos a aplicar en base a la pregunta y los datos disponibles. Finalmente se presentan las formas de evaluar la performance predictiva del modelo y el concepto de tuning de parámetros.

## Machine Learning II (4 créditos)

El objetivo de esta materia es introducir a los estudiantes los distintos algoritmos de machine learning en base a las técnicas comúnmente utilizadas en aprendizaje supervisado y no supervisado. El enfoque adoptado es práctico. Cada uno de los algoritmos presentados será implementado por parte de los estudiantes a un conjunto de datos reales para responder a un problema de clasificación o regresión utilizando python (y alternativamente se brindarán los códigos en R). Se proponen los siguientes algoritmos:

1. Perceptron algorithm, SVM, Regresión lineal, Regresión logística, Lasso, Ridge, ElasticNet, Árboles de decisión, Random Forest, Modelos de promedios bayesianos, KNN.
2. Natural Language processing techniques, generative models, Bag of words.
3. Aprendizaje no supervisado: K-means, K-medoids, Distancia y variables cualitativas, DBSCAN. Hierarchical clustering, Birch.

## Deep Learning (4 créditos)

Johann Perez

El objetivo de esta materia es introducir a los estudiantes a las redes neuronales y presentar los principales tipos de redes y su implementación con python. El curso será aplicado y cada una de las redes neuronales presentadas será implementada por parte de los estudiantes a un conjunto de datos reales.

## Análisis de redes (3 créditos)

Victoria Landaberry, Andrea Baron

El objetivo de esta materia es introducir a los estudiantes al análisis de redes. En particular se presentarán los principales tipos de redes, los gráficos de redes, las medidas y topologías de redes y los métodos de imputación en la construcción de redes con datos faltantes. Se realizará una aplicación práctica con redes sociales.

## Proyectos y conceptos tecnológicos (3 créditos)

Nicolas Serrano y Johann Perez

El objetivo de esta materia es introducir a los estudiantes a un conjunto de conceptos y prácticas relacionadas con la tecnología y los procesos de innovación tecnológica. Busca reducir la brecha de conocimientos entre el área social y el campo tecnológico. Esta materia se realizará en forma de talleres. En cada uno de ellos se abordará un concepto o práctica particular (Gestión de proyectos tecnológicos e introducción a las metodologías ágiles (Nicolás Serrano), Seguridad en big data y Machine learning (Nicolás Serrano), Control de versiones en el desarrollo de productos basados en datos (Johann Perez), Cloud Computing, Internet of Things.