

PROGRAMA DE LA ASIGNATURA
FUNDAMENTOS DE CIENCIA DE DATOS

CARRERA: LICENCIATURA EN ASTRONOMÍA – GEOFÍSICA – LICENCIATURA EN METEOROLOGÍA Y CIENCIAS DE LA ATMÓSFERA

CARGA HORARIA SEMANAL: 4 HORAS DE TEÓRICO-PRÁCTICO

CARÁCTER: SEMESTRAL

PROFESOR A CARGO: Dra. ERIKA GULARTE y Dr. GUSTAVO BAUME

OBJETIVOS

La ciencia de datos es un campo interdisciplinario que utiliza técnicas estadísticas, computacionales y matemáticas para extraer información y conocimientos a partir de los datos. El objetivo general del curso consiste en adquirir los conceptos, las técnicas y las herramientas fundamentales de la ciencia de datos.

Por su parte, los objetivos específicos delineados para este curso son los siguientes:

- Desarrollar habilidades para recopilar, limpiar y manipular datos para su análisis.
- Aplicar distintas técnicas de modelado, mediante aprendizaje automático, a fin de identificar patrones y tendencias en los datos, predecir resultados y tomar decisiones informadas.
- Adquirir habilidades de visualización, tanto sea para analizar los datos como para comunicar los resultados de manera clara y efectiva.

Así mismo, las aplicaciones prácticas estarán orientadas a los temas de interés profesional de la Astronomía, Geofísica y/o Meteorología.

CONTENIDO TEMÁTICO

El curso se ha distribuido en ocho capítulos, contemplando los tópicos de ciencia de datos de mayor interés para los estudiantes de la FCAG. En el Capítulo I se presentan los conceptos y herramientas básicas. El Capítulo II aborda los distintos tipos de datos y su adquisición/recopilación a través de diversas fuentes. Tanto el análisis inicial como el análisis exploratorio de los datos se presentan en el Capítulo III. Este capítulo se complementa con el proceso de homogeneización de los datos y con la selección de los atributos a ser considerados para el análisis. Luego, en los Capítulos IV y V, se plantea el aprendizaje automático mediante las técnicas de aprendizaje supervisado y no supervisado. En los dos capítulos siguientes, se abordan dos tópicos específicos de interés. Por un lado, en el Capítulo VI se detalla el aprendizaje profundo a través de las redes neuronales. Por otro lado, en el Capítulo VII se analiza el caso particular de las series temporales en el contexto de la ciencia de datos. Finalmente, debido a la dinámica que representan los contenidos de la asignatura, el Capítulo VIII se deja sin especificar. El mismo abarcará diferentes tópicos relevantes y de actualidad.

A continuación, se detalla el contenido de cada uno de los capítulos mencionados:

Capítulo I: INTRODUCCIÓN.

Evolución histórica. Conceptos generales. El proceso KDD y la minería de datos. Aprendizaje automático. Lenguajes de programación. Software.

Capítulo II: EL DATO.

Tipos de datos. Visualización de los datos. Formatos de archivos. Bases de datos. Otras fuentes de datos.

Capítulo III: PREPROCESAMIENTO.

Análisis inicial de los datos. Análisis exploratorio de los datos. Limpieza e integración de los datos. Selección y transformación de atributos.

Capítulo IV: APRENDIZAJE NO SUPERVISADO.

Concepto básico. Métricas. Métodos de agrupamiento. Reglas de asociación. Comparación de métodos. Evaluación.

Capítulo V: APRENDIZAJE SUPERVISADO.

Concepto básico. Partición de datos. Métodos de clasificación. Métodos de regresión. Comparación de métodos. Evaluación.

Capítulo VI: APRENDIZAJE PROFUNDO.

Concepto básico. Redes Neuronales. Tipos de redes neuronales. Funcionamiento de la red. Visualización de la red. Aplicaciones.

Capítulo VII: SERIES DE TIEMPO.

Concepto básico. Tipos de datos temporales. Estructura interna de la serie de tiempo. Modelos lineales y no lineales. Selección y validación de modelos. Modelos con memoria de largo y corto plazo.

BIBLIOGRAFÍA

- Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems (2022). O'Reilly Media.
- Gray A., J.T. VanderPlas, A.J. Connolly, Ž. Ivezić. Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data (2020). Princeton University Press.
- Han J., M. Kamber, J. Pei. Data Mining: Concepts and Techniques 3rd Edition (2011). Elsevier, Morgan Kaufmann.
- Hastie T., R. Tibshirani, J. Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2013). Springer.
- Müller A.C., S. Guido. Introduction to Machine Learning with Python: A Guide for Data Scientists (2016). O'Reilly Media.
- Orallo J.H., M.J. Ramírez Quintana, C. Ferri Ramírez. Introducción a la Minería de Datos (2004). Pearson Prentice Hall.
- Press William H., S.A. Teukolsky, W.T. Vetterling, B.P. Flannery. Numerical Recipes: The Art of Scientific Computing 3rd Edition (2007). Cambridge University Press.
- Witten I.H., E. Frank, M.A. Hall, C.J. Pal. Data Mining Practical Machine Learning Tools and Techniques (2007). Morgan Kaufmann.

La bibliografía presentada se complementará con información disponible en diferentes páginas web de sitios especializados. Adicionalmente, la naturaleza de la asignatura hace necesaria una actualización y mejora regular de la bibliografía involucrada que será brindada oportunamente a los alumnos.