Módulo 2. Análisis automático de datos

Objetivos

Los objetivos de este módulo son:

- Brindar a los estudiantes una introducción práctica al proceso de análisis de datos, desde la exploración y preprocesamiento hasta la generación y evaluación de modelos.
- Incorporar conceptos clave de la Ciencia de Datos y de la Minería de Datos mediante el uso de una herramienta visual y accesible.
- Introducir técnicas de aprendizaje supervisado y no supervisado, evaluando su aplicación en contextos reales.
- Fortalecer la capacidad de comunicar hallazgos relevantes mediante visualizaciones claras y narrativas comprensibles, orientadas a públicos no técnicos.

Contenidos mínimos

- Introducción a la Ciencia de Datos. Su relación con la Minería de Datos y el Aprendizaje Automático.
- Proceso de Descubrimiento de Conocimiento en Bases de Datos (KDD): selección, preprocesamiento, transformación, minería de datos, evaluación e interpretación.
- Análisis exploratorio de datos: tipos de variables, análisis de distribuciones, valores
- faltantes y atípicos. Visualización con herramientas accesibles.
- Preprocesamiento de datos: limpieza, normalización, discretización, codificación de atributos categóricos.
- Aprendizaje supervisado. Modelos de clasificación y regresión. División de datos en entrenamiento, validación y prueba. Evaluación y validación de modelos predictivos.
- Aprendizaje no supervisado: agrupamiento de datos con características similares, detección de datos anómalos, identificación de patrones frecuentes.
- Comunicación de resultados: principios básicos de visualización con datos. Herramientas accesibles para la presentación de hallazgos.

Motivación

Este módulo busca introducir a los estudiantes en el proceso de descubrimiento de conocimiento a partir de datos, utilizando herramientas accesibles para usuarios sin conocimientos de programación. A través del uso de una plataforma visual y de código abierto, se guiará al estudiante en el análisis exploratorio, la limpieza de datos, la aplicación de modelos y la interpretación de resultados. Además, se fomentará la comunicación efectiva mediante visualizaciones y narrativas

Calle 120 y 50 – 2do piso (1900) La Plata
TEL-FAX: (54) 221-4273235
http://postgrado.info.unlp.edu.ar
E-Mail: postgrado@lidi.info.unlp.edu.ar

simples, esenciales para la toma de decisiones basada en datos.

Modalidad de evaluación

La evaluación se realizará mediante cuestionarios breves que abordarán tanto los contenidos teóricos vistos en clase como los resultados de los ejercicios prácticos desarrollados durante el curso.

Bibliografía básica

- Hernández Orallo, Ramírez Quintana, Ferri Ramírez. Introducción a la Minería de Datos.
- Prentice Hall. 2004. ISBN 84-205-4091-9.
- Ian H. Witten, Eibe Frank, Mark A. Hall. Data Mining: Practical Machine Learning Tools and Techniques, (Fourth Edition). Morgan Kaufmann. 2017. ISBN 978-0-12-804291-5.

Bibliografía complementaria

- Michael A. Nielsen. “Neural Networks and Deep Learning”, Determination Press, 2015.
- Ian Goodfellow. Yoshua Bengio y Aaron Courville, “Deep Learning”, MIT Press, 2016.
- Provost, F., & Fawcett, T. (2013). Data Science for Business. O'Reilly.
- Han, J., Kamber, M., Pei, J. (2011). Data Mining: Concepts and Techniques. Morgan Kaufmann.
- Sarah Guido, Andreas Müller. Introduction to Machine Learning with Python. O'Reilly Media. 2016.
- Jason Bell. Machine Learning: Hands-On for Developers and Technical Professionals. Wiley. 2015.

Calle 120 y 50 – 2do piso (1900) La Plata
TEL-FAX: (54) 221-4273235
http://postgrado.info.unlp.edu.ar
E-Mail: postgrado@lidi.info.unlp.edu.ar