

UNIVERSIDAD CENTRAL DE VENEZUELA
FACULTAD DE CIENCIAS
ESCUELA DE COMPUTACIÓN
MINERÍA DE DATOS
(6213)

I-2024

Instructor:	Wilmer Gonzalez	Horario:	Mar, Jue @ 9am-11am
Email:	will.all.gs@gmail.com	Lugar:	Online.

Objetivos

- Comprender los fundamentos teóricos y prácticos asociados al proceso de minería de datos.
- Comprender y aplicar técnicas de minería de datos para descubrir conocimiento en problemas del mundo real.

Contenido temático

1. **Introducción al descubrimiento de conocimiento a partir de datos:** Objetivo de la Minería de datos. Definición del problema. Taxonomía de las técnicas de Minería de datos. [6] [4]
2. **Proceso de Minería de datos:** Metodología CRISP-DM. [12] [3] [1]
3. **Análisis exploratorio de datos (EDA):** Objetivos del EDA. Visualización de datos. [9] [11] [7]
4. **Pre-procesamiento de los datos:** Técnicas de normalización de datos. Tratamiento de datos ausentes. Reducción de dimensionalidad. [6]
5. **Algoritmos de Minería de datos:** Algoritmos predictivos. Algoritmos descriptivos. Medidas de similitud. [5] [6]
6. **Aplicaciones de Minería de datos:** Discusiones de aplicaciones de algoritmos de Minería de datos a problemas.

Evaluaciones

Tema (t)	Evaluación	Ponderación	Fecha estimada
1 y 2	Parcial	30%	07-05-2024
3, 4 y 5	Quiz (<i>t</i>)	30% (10% c/u)	7-8 días después de (<i>t</i>)
3, 4 y 5	Asignación (<i>t</i>)	30% (10% c/u)	7-8 días después de (<i>t</i>)
6	Presentación	10%	Antes de 05-08-2024

Logística

- Usaremos [Slack](#) como principal medio de comunicación.
- Usaremos dm.ucv.ai como página principal del curso.

Referencias bibliográficas

- [1] Ahmed Banimustafa and Nigel W. Hardy. “A Scientific Knowledge Discovery and Data Mining Process Model for Metabolomics”. In: *IEEE Access* 8 (2020), pp. 209964–210005. URL: <https://api.semanticscholar.org/CorpusID:227277430>.
- [2] Katy Börner, Andreas Bueckle, and Michael Ginda. “Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments”. In: *Proceedings of the National Academy of Sciences* 116 (2019), pp. 1857–1864. URL: <https://api.semanticscholar.org/CorpusID:73417466>.
- [3] Peter Chapman. “CRISP-DM 1.0: Step-by-step data mining guide”. In: 2000. URL: <https://api.semanticscholar.org/CorpusID:59777418>.
- [4] Usama M. Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. “The KDD process for extracting useful knowledge from volumes of data”. In: *Commun. ACM* 39 (1996), pp. 27–34. URL: <https://api.semanticscholar.org/CorpusID:56635038>.
- [5] Trevor J. Hastie, Robert Tibshirani, and Jerome H. Friedman. “The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd Edition”. In: *Springer Series in Statistics*. 2001. URL: <https://api.semanticscholar.org/CorpusID:40654213>.
- [6] Jure Leskovec, Anand Rajaraman, and Jeffrey D. Ullman. “Mining of Massive Datasets, 3rd Ed”. In: 2014. URL: <http://www.mmds.org/>.
- [7] Paul C. Parsons. “Understanding Data Visualization Design Practice”. In: *IEEE Transactions on Visualization and Computer Graphics* PP (2021), pp. 1–1. URL: <https://api.semanticscholar.org/CorpusID:237194764>.
- [8] Xuedi Qin et al. “Making data visualization more efficient and effective: a survey”. In: *The VLDB Journal* 29 (2019), pp. 93–117. URL: <https://api.semanticscholar.org/CorpusID:208171945>.
- [9] Pedro M. Valero-Mora. “ggplot2: Elegant Graphics for Data Analysis”. In: *Journal of Statistical Software, Book Reviews* 35.1 (2010), pp. 1–3. DOI: [10.18637/jss.v035.b01](https://doi.org/10.18637/jss.v035.b01). URL: <https://www.jstatsoft.org/index.php/jss/article/view/v035b01>.
- [10] Zezhong Wang et al. “Cheat Sheets for Data Visualization Techniques”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020). URL: <https://api.semanticscholar.org/CorpusID:218482669>.
- [11] Hadley Wickham. “Tidy Data”. In: *Journal of Statistical Software* 59.10 (2014), pp. 1–23. DOI: [10.18637/jss.v059.i10](https://doi.org/10.18637/jss.v059.i10). URL: <https://www.jstatsoft.org/index.php/jss/article/view/v059i10>.
- [12] Rüdiger Wirth and Jochen Hipp. “CRISP-DM: Towards a Standard Process Model for Data Mining”. In: 2000. URL: <https://api.semanticscholar.org/CorpusID:1211505>.
- [13] Aoyu Wu et al. “AI4VIS: Survey on Artificial Intelligence Approaches for Data Visualization”. In: *IEEE Transactions on Visualization and Computer Graphics* 28 (2021), pp. 5049–5070. URL: <https://api.semanticscholar.org/CorpusID:236090307>.

Otros recursos:

- kdd.org - The community for data mining, data science and analytics - ACM.
- datavizcatalogue.com, datavizproject.com, [visualizationcheatsheets](http://visualizationcheatsheets.com) - Comunidades de visualización de datos y herramientas.
- [Design, Visualization, & Cognition Lab at Purdue University](http://design.viz.cognition.purdue.edu).
- Artículos útiles sobre visualización de datos [13] [2] [8] [10]