

Programa de asignatura Énfasis II Aprendizaje No Supervisado para el Análisis de Datos

1. Identificación

Nombre Escuela:	Escuela de Administración
Nombre Departamento:	Área de gestión de la información y riesgos
Nombre Programa Académico:	Administración de Negocios
Nombre Programa Académico (En inglés):	Business Management
Asignatura Prerrequisito	N/A
Semestre de Ubicación	6 semestre
Intensidad Horaria Semanal	3 horas
Intensidad Horaria Semestral	48 Horas
Créditos	3
Características	No suficienteable

2. Justificación

Si bien las diferentes técnicas estadísticas y de análisis de datos permiten dar cuenta del comportamiento de los datos y hacer predicciones con base en ellos, hay cierto tipo de información que no cumple los supuestos necesarios para aplicar este tipo de técnicas. En estos casos se hace necesaria la aplicación de algoritmos basados en datos o algoritmos de *aprendizaje de máquina*. En este caso, es la misma estructura de los datos y las relaciones que tienen entre sí, las que dan forma a las funciones que definirán la forma en la que se analizará la información.

Después de cursar esta asignatura, se espera que el estudiante esté en la capacidad de desarrollar, de forma correcta, un *pipeline* para proyectos de aprendizaje de máquina no supervisado, con el fin de analizar información relevante para la toma de decisiones dentro de la organización.

3. Competencias y resultados de aprendizaje

Al culminar esta materia, el estudiante habrá fortalecido las siguientes competencias genéricas y habrá avanzado en los siguientes resultados de aprendizaje:

3.1. Competencias Genéricas:

Áreas de Desempeño	Competencias genéricas
Integridad	Reconoce dilemas éticos en el desempeño de su profesión
Pensamiento crítico	Escoge cursos de acción apropiados con base en argumentos y evidencias en contextos económicos y de negocios
Gestión estratégica	Integra saberes para la solución de problemas en las organizaciones y administrar estratégicamente negocios sostenibles, en las dimensiones social, ambiental y económica

3.2. Resultados de Aprendizaje:

Áreas de Desempeño	Competencias	Descriptor de la competencia	Resultados de Aprendizaje
Análisis de información y toma de decisiones en organizaciones	Emplea la analítica de datos y las diferentes tecnologías de la información para el ejercicio de su profesión	Selecciona los métodos de analítica de datos más adecuados, de acuerdo con la información disponible y el tipo de decisión a tomar en las organizaciones	Comprende la taxonomía del <i>Machine Learning</i> y las áreas de aplicación de la teoría del aprendizaje supervisado y no supervisado
			Implementa modelos por aprendizaje no supervisado para el análisis de datos de tipo estructurado y no estructurado
			Crea soluciones en función de los datos en donde la organización requiere escalamiento, alcance e innovación en sus procesos

4. Contenidos

Unidad 1: Introducción al Aprendizaje de Máquina (*Business Intelligence BI*)

Temas:

- Evolución del Aprendizaje de Máquina y aplicaciones
- Teoría del aprendizaje supervisado
- Teoría del aprendizaje no supervisado
- Teoría del aprendizaje por refuerzo

Unidad 2: Preprocesamiento e Identificación de la Estructura de los datos

Temas:

- Datos numéricos, nominales, variables indicadoras
- Etiquetamiento, problemas multiclase, problemas multietiqueta, *onehot encoding*
- Regularización L1 Lasso, L2 Ridge

Unidad 3: Métodos y modelos por aprendizaje supervisado

Temas:

- Árboles de decisión – clasificación
- Métodos Bayesianos (*Naïve Bayes* – Redes Bayesianas)
- Métodos de Ensamble (*Boosting*, Bosques Aleatorios)
- Validación de modelos (Curvas ROC-Lift)

Unidad 4: Métodos y modelos por aprendizaje no supervisado

Temas:

- Métodos de agrupamiento o *clustering*: EM *Expectation Maximization*
- Método k-Means, Fuzzy C-Means
- Mapas Auto-organizados (SOM)
- Validación de modelos de clasificación y agrupamiento (Matriz de Confusión, Índice de Entropía)

5. Estrategias metodológicas y cronograma

5.1. Metodología

La metodología tendrá una combinación de componente magistral, donde el docente explicará los conceptos teóricos, y de componente práctico en R, SQL-Server, Orange Data Mining o Python en donde se implementarán los diferentes algoritmos para el proceso de creación de modelos machine learning por aprendizaje no supervisado, aplicados a casos reales e hipotéticos.

Los estudiantes deberán solucionar, como parte del trabajo independiente, diferentes problemáticas planteadas por el docente relacionadas con las temáticas de la asignatura. Finalmente, los estudiantes presentarán un proyecto final a manera de idea de negocio en función de los datos, tomando como base diferentes tipologías de proyectos, y en donde la organización requiera alcance, escalamiento e innovación.

ADMINISTRACIÓN DE NEGOCIOS	CONTADURÍA	MERCADEO	NEGOCIOS INTERNACIONALES
Facts-Driven Decision Making (DDDM) • Oportunidades Negocio ✓ SmartContracts ✓ SmartPricing ✓ BlockChain ✓ E-Commerce • Escenarios TD utilizando realidades mixtas - Metaverso	Robot Process Automation (RPAs) ✓ Detección de Fraudes y Lavado de Activos ✓ Insurance Pricing Strategies-AgroBusiness ✓ Credit & Operational Risk Modelling (NuBank-Auditoria 4.0)	Marketing Intelligence (IOT & IOB) ✓ Neuromarketing - BCI (Brain Computer Interface) ✓ Patrones E-Tracker y comportamiento. ✓ Marketing Digital y ChatBots. ✓ Segmentación Mercados	Big Data Market Intelligence (BDMI) ✓ Expansión a nuevos mercados. ✓ Optimización de rutas logísticas (Sustainable). ✓ Bots - trazabilidad en la cadena de suministro

5.2. Cronograma

Unidad	Duración en semanas	Duración en horas
Unidad 1	2 semanas	6 horas
Unidad 2	2 semanas	6 horas
Unidad 3	6 semanas	18 horas
Unidad 4	5 semanas	15 horas
Trabajo final	1 semanas	3 horas

6. Recursos

6.1 Locativos:

- Sala de cómputo
- Aula invertida
- Salón de clase

6.2 Tecnológicos:

- Computadores
- Videobeam

6.3 Didácticos:

- Bases de datos especializadas
- Casos de estudio
- Material audiovisual

7. Bibliografía

Libros

- Gao, J., & Lu, J. (2019). Dimensionality Reduction in Machine Learning. In Proceedings of the International Conference on Computational Science (ICCS 2019) (pp. 413-426). Springer.
- Aggarwal, C. C. (2015). Data mining: the textbook. Springer.
- Kuhn, M., & Johnson, K. (2013). Applied Predictive Modelling: Springer.
- Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (2013). Machine learning: An artificial intelligence approach: Springer Science & Business Media.
- Vercellis, Carlo (2013) Business Intelligence: Data Mining and Optimization for Decision Making, John Wiley and Sons, ISBN-978-0-470-51138-1,
- Witten, I., & Eibe, F. (2005). Data Mining, Practical Machine Learning Tools and Techniques. San Francisco: Elsevier.
- Bishop, C. (2006). Pattern Recognition and Machine Learning: Springer.
- Dietterich, T. G. (1989). MACHINE LEARNING. Annual Review of Computer Science, 4, 255-306.
- Dietterich, T. G. (1997). Machine-learning research - Four current directions. Ai Magazine, 18(4), 97-136.
- Mitchell, T. M. (1997). Machine Learning: McGraw-Hill.

Bases de datos

- Bases de Datos de Riesgo Crédito.
- Base de Datos Adventure Works (<https://docs.microsoft.com/en-us/sql/samples/adventureworks-install-configure?view=sql-server-ver15&tabs=ssms>)
- Bases de Datos Científicas (<https://www.sciencedirect.com/journal/data-in-brief>)

8. Requisitos del proceso de aseguramiento de la calidad

Versión número:	2
Fecha elaboración:	09/05/2023
Responsable:	Alejandro Peña P., Lina María Sepulveda C.