Código de la Asignatura: IO9F3

Nombre del programa académico	Maestría en Investigación Operativa y Ciencia de Datos
Nombre completo de la asignatura	Análisis de Datos no Estructurados
Área académica o categoría	Estadística Enfocada a la Ciencia de Datos
Semestre y año de actualización	Primer semestre de 2025
Tipo de asignatura	[x] Obligatoria [ ] Electiva
Número de créditos ECTS	5.22 ECTS
Intensidad horaria	10 horas Semanales – 40 Mensuales
Director o contacto del programa	Eliana Mirledy Toro Ocampo
Coordinador o contacto de la asignatura	Julián David Echeverry Correa - jde@utp.edu.co Google Scholar Julian David Echeverry

Descripción y Contenidos

1. Breve descripción

La asignatura proporciona una comprensión detallada de los datos que carecen de una forma organizada predefinida. Se exploran conceptos fundamentales como la definición de datos no estructurados, las diferencias con los datos estructurados y los desafíos asociados con su análisis. Además, se estudian diversas fuentes y tipos de datos no estructurados, como texto, imágenes, videos, audio y voz. A lo largo del curso, los estudiantes aprenden técnicas especializadas de procesamiento para cada tipo de dato, incluyendo tokenización y segmentación de texto, preprocesamiento de imágenes, y extracción de características de audio. Se introducen herramientas y tecnologías específicas para el análisis de datos no estructurados, así como el uso de bibliotecas y frameworks relevantes en este campo.

2. Objetivos del programa académico

OP1. Analizar y combinar conocimientos y técnicas de la estadística y la optimización matemática bajo enfoques sostenibles, con el objetivo de tomar decisiones informadas, esenciales para abordar los desafíos actuales y futuros en ambientes académicos y empresariales. OP2. Identificar herramientas y enfoques efectivos para el uso de los recursos de las organizaciones con el fin de optimizar sus procesos, combinando las prácticas de investigación de operaciones y estadística, que permitan mejorar la eficiencia, optimizar la toma de decisiones, prever tendencias y mantener su competitividad en un mercado cada vez más complejo y basado en datos. OP3. Analizar y extraer conocimientos significativos a partir de enfoques estadísticos multivariados, tanto cualitativos como cuantitativos, para examinar simultáneamente múltiples variables o factores en un conjunto de datos, con el objetivo de obtener información más completa y profunda sobre las relaciones y patrones entre ellas, aplicados en diferentes campos, como la investigación científica, la ingeniería, la economía, la medicina, entre otras.

Objetivos de la asignatura

OA1. Comprender los conceptos fundamentales de los datos no estructurados y sus diferencias con los datos estructurados. OA2. Identificar y analizar diferentes tipos de datos no estructurados, como texto, imágenes, videos, audio y voz. OA3. Desarrollar habilidades para el procesamiento de texto, imágenes, videos, audio y voz. OA4. Familiarizarse con herramientas y tecnologías utilizadas en el análisis de datos no estructurados. OA5. Aplicar técnicas de análisis de datos no estructurados en casos prácticos para resolver problemas reales.

3. Resultados de aprendizaje de la asignatura

RAA1. Distinguir entre datos estructurados y no estructurados, comprendiendo sus características y desafíos únicos. RAA2. Identificar y manejar eficientemente diversos tipos de datos no estructurados, como texto, imágenes, videos, audio y voz. RAA3. Aplicar técnicas avanzadas de procesamiento de texto, imágenes, videos, audio y voz para extraer información valiosa. RAA4. Utilizar herramientas y tecnologías especializadas en el análisis de datos no estructurados de manera efectiva. RAA5. Aplicar de manera práctica los conocimientos adquiridos en la resolución de problemas reales de análisis de datos no estructurados.

4. Contenido

T1. Introducción a los Datos No Estructurados (22 h: 6 h con docente y 16h independientes) Concepto de datos no estructurados, Diferencias entre datos estructurados y no estructurados, Desafíos en el análisis de datos no estructurados T2. Fuentes y Tipos de Datos No Estructurados (22 h: 6 h con docente y 16h independientes) Texto no estructurado Imágenes y videos, Audio y voz T3. Procesamiento de Texto (32 h: 9 h con docente y 23 h independientes) Tokenización y segmentación de texto, Extracción de características de texto, Análisis de sentimientos y opiniones T4. Procesamiento de Imágenes y Videos (32 h: 9 h con docente y 23 h independientes) Preprocesamiento de imágenes, Extracción de características de imágenes, Reconocimiento de objetos y clasificación de imágenes T5. Procesamiento de Audio y Voz (22 h: 6 horas con docente y 16h independientes) Preprocesamiento de audio, Extracción de características de audio, Reconocimiento de voz y transcripción de audio T6. Herramientas y Tecnologías para el Análisis de Datos No Estructurados (14 h: 4 h con docente y 10h independientes) Introducción a herramientas de análisis de datos no estructurados, Uso de bibliotecas y frameworks

5. Requisitos

- Programación en Python. Conocimientos básicos de estadística y manejo de datos.

6. Recursos

Libros de texto Jurafsky, D., & Martin, J. H. (2024). Speech and Language Processing (3rd Edition - draft). Pearson. ● Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. O'Reilly Media. ● Kipper, S., & Hobson Lane. (2022). Natural Language Processing in Action. Manning Publications. ● Eisenstein, J. (2019). Natural Language Processing. MIT Press. ● McLoughlin, I. (2009). Applied Speech and Audio Processing: With MATLAB Examples. Cambridge University Press. ●O'Shaughnessy, D. (2017). Speech and Audio Signal Processing (2nd Edition). IEEE Press. ● Yang, et al. (2025). Omni-AVSR: Towards Unified Multimodal Speech Recognition with Large Language Models. (arXiv). ● Rabiner, L., & Schafer, R. (2011). Theory and Applications of Digital Speech Processing. Pearson. ● Müller, M. (2015). Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications. Springer. ●Szeliski, R. (2022). Computer Vision: Algorithms and Applications (2nd Edition). Springer. ● Gonzalez, R.C., & Woods, R.E. (2018). Digital Image Processing (4th Edition). Pearson. ● Lee, D., et al. (2023). ATFusion: An Alternate Cross-Attention Transformer Network for Infrared and Visible Image Fusion. (arXiv).

7. Herramientas técnicas de soporte para la enseñanza

Clases Magistrales, Se disponen de videos explicativos sobre los contenidos del curso.

8. Trabajos en laboratorio y proyectos

Durante el semestre el profesor implementa tareas y talleres que fomentan el aprendizaje, y lo relacionan con la especialidad de cada estudiante. Talleres en Python por unidad (T3 a T6: 4 talleres de 10 horas, 40 horas). Proyecto de aplicación práctica (40 horas), que incluye el análisis de ética de datos: privacidad, consentimiento y sesgos en texto, imagen y voz (6 horas).

9. Métodos de aprendizaje

Metodologías activas, talleres y ejercicios de modelación

10. Métodos de evaluación

Talleres: 40% - se realizarán talleres de los temas vistos durante las clases. Estos talleres podrán ser resueltos en grupos de 2 personas. Proyecto: 60% - se deberá entregar un proyecto de aplicación práctica del uso de las técnicas vistas en clase. Este proyecto puede ser desarrollado en grupos de 2 personas.