Código de la Asignatura: IO6E3

Nombre del programa académico	Maestría en Investigación Operativa y Ciencia de Datos
Nombre completo de la asignatura	Técnicas de Exploración de Datos
Área académica o categoría	Estadística Enfocada a la Ciencia de Datos
Semestre y año de actualización	Segundo semestre de 2023
Tipo de asignatura	[ ] Obligatoria [x] 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

El curso se centrará en el desarrollo de un proyecto de analítica de datos, abordando preprocesamiento, modelos supervisados y no supervisados, evaluación y despliegue. Se trabajará con la metodología CRISP-DM, que permite definir el problema, seleccionar técnicas, evaluar métricas y presentar resultados.

2. Objetivos del programa académico

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. OP5. Fomentar dimensiones humanas que permitan generar espacios participativos de diálogo y reflexión con un enfoque sostenible, que contribuyan a la formación de profesionales capaces de abordar desafíos de manera integral y responsable.

Objetivos de la asignatura

OA1: Presentar y estructurar metodologías de analítica avanzada (como CRISP-DM o SEMMA) para resolver problemas complejos de ingeniería, empleando el pensamiento crítico para diagnosticar necesidades reales del entorno. OA2: Ilustrar e implementar técnicas de minería de datos (aprendizaje supervisado y no supervisado) bajo principios de rigor científico, garantizando que el descubrimiento de patrones se fundamente en la ética de la formación humana

OA3: Desarrollar y evaluar modelos de analítica orientados a la sostenibilidad y la eficiencia de recursos, combinando herramientas para la toma de decisiones que fortalezcan la transparencia y la ciudadanía.

3. Resultados de aprendizaje de la asignatura

RAA1: Aplica las etapas de la metodología CRISP-DM con pensamiento crítico (60%), desde el entendimiento del negocio hasta el despliegue, asegurando que el flujo de datos sea coherente y apunte a la sostenibilidad institucional. RAA2: Discrimina entre problemas de aprendizaje supervisado (predicción) y no supervisado (agrupamiento) según las necesidades estratégicas, validando la idoneidad de la técnica con el rigor técnico propio de la formación humana. RAA3: Utiliza técnicas estadísticas y de investigación operativa para la limpieza, transformación y modelado de grandes volúmenes de datos, optimizando el procesamiento para garantizar un enfoque sostenible en el uso de recursos computacionales. RAA4:Evalúa indicadores de desempeño y combina modelos en herramientas especializadas (KNIME, R, Python), sustentando resultados transparentes que faciliten la comunicación de hallazgos complejos a audiencias multidisciplinarias. RAA5: Se integra en equipos multidisciplinarios para resolver desafíos de analítica, ejerciendo su rol con la responsabilidad ética y el compromiso social que exige la formación humana.

4. Contenido

T1: Conceptos básicos de análisis de datos, tipos de variables y preprocesamiento (16 h: 4 h con docente y 12h independientes). ● T2: Metodología CRISP-DM y selección de características (16 h: 4 h con docente y 12 h independientes). ● T3: Medidas de distancia y análisis no supervisado (32 h: 10 docente y 22 h independientes). ● T4: Enfoque supervisado y métodos de clasificación (32 h: 10 con docente y 22 h independientes) ● T5: Métricas de evaluación (16 h: 4 con docente y 12 independientes). ●T6: Análisis de canasta de mercado (16 h: 4 con docente y 12 independientes). ●T7: Minería de texto y fundamentos de PLN (16 h: 4 con docente y 12 independientes)

5. Requisitos:Álgebra Lineal, Teoría de la Probabilidad, Estadística

6. Recursos

Libros de texto: Witten et al. (2011); Bell (2015); Murphy (2012); Duda, Hart & Stork (2001); Hastie, Tibshirani & Friedman (2009); Schölkopf & Smola (2001); Schutt & O'Neil (2013); Vercellis (2009), sobre minería de datos, aprendizaje automático y business intelligence. Artículos académicos y lecturas complementarias proporcionadas durante el curso. **Software:** KNIME (con complementos), Orange, Python, RapidMiner, Weka y Matlab. **Bases de datos:** Suscritas por la universidad.

7. Herramientas técnicas de soporte para la enseñanza

Clases magistrales con apoyo de diapositivas y lecturas previas. Lecturas autónomas sobre tidy data, big data, clustering, árboles de decisión y SVM. Talleres en clase con KNIME y datos reales (UCI), y trabajos grupales fuera del aula en analítica de datos. Videos sobre visión por computador y PLN (Fei-Fei Li) y sobre ética en minería de datos ("Weapons of Math Destruction").

8. Trabajos en laboratorio y proyectos

Taller de temas T1 y T2 (8 horas) ● Taller de temas T3, T4 y T5 (12 horas) ● Taller de temas T6 y T7 (8 horas) ● Proyecto final en grupos de 2 estudiantes con una dedicación de 20 horas por estudiante.

9. Métodos de aprendizaje

Clases magistrales apoyadas en TIC, lecturas especializadas como trabajo autónomo semanal, trabajos

grupales, talleres dentro y fuera del aula, y tutorías.

10. Métodos de evaluación

Entrega 1 (20%): Taller individual sobre conceptos básicos de análisis de datos y metodología CRISP-DM. ● **Entrega 2 (20%)**: Taller grupal sobre medidas de distancia, enfoques supervisado y no supervisado, clasificación y métricas de evaluación ● **Entrega 3 (20%)**: Taller grupal sobre análisis de canasta de mercado y minería de texto/PLN ● **Proyecto final (40%)**: Desarrollo grupal de un sistema de procesamiento y análisis de datos, evaluando todos los temas