



1. Datos Generales de la asignatura

Nombre de la asignatura:	Ciencia de Datos
Clave de la asignatura:	ING-2507
SATCA¹:	3-3-6
Carreras:	Ingeniería en Sistemas Computacionales, Ingeniería Industrial, Ingeniería en Gestión Empresarial, Licenciatura en Administración, Ingeniería en Administración.

2. Presentación

Caracterización de la asignatura
<p>Esta asignatura aporta al perfil del estudiante de nivel licenciatura las competencias profesionales:</p> <ul style="list-style-type: none">• Aplicar algoritmos de transformación de datos para preparar su ingreso a modelos de machine learning con un enfoque interdisciplinaria.• Reducir la dimensionalidad de los datos para disminuir el costo computacional, mejorando el aprovechamiento de los recursos disponibles.• Generar modelos de machine learning para clasificar, predecir y/o segmentar los datos ya transformados.• Evaluar los modelos de machine learning para elegir la opción optima en relación a los casos de estudio, implementado soluciones efectivas en el campo de la toma de decisiones basada en modelos matemáticos. <p>La importancia de esta asignatura, es que permite mejorar la toma de decisiones, aumentando el nivel de certidumbre en las organizaciones al implementar algoritmos y modelos que generan resultados más precisos en las métricas, haciendo uso de la capacidad de cómputo disponible.</p>

¹ Sistema de Asignación y Transferencia de Créditos Académicos



Intención didáctica

La asignatura de Ciencia de Datos está diseñada para proporcionar a los estudiantes las habilidades necesarias para manejar, transformar y analizar grandes volúmenes de datos de manera eficiente. A través de una combinación de teoría y práctica, los estudiantes aprenderán a aplicar algoritmos avanzados, a realizar procesos de reducción y transformación de datos, y a construir y evaluar modelos que les permitirán descubrir patrones y tendencias significativas. Los estudiantes conocerán las distintas métricas de evaluación, permitiéndoles comparar y validar la eficacia de los modelos generados.

Unidad 1 Procesamiento de Datos: En esta primera unidad, los estudiantes conocerán los fundamentos del procesamiento de datos y aprenderán a realizar transformaciones necesarias para la preparación de los datos. Se abordan métodos de escalamiento, estandarización, normalización y binarización, así como técnicas avanzadas como las transformaciones BOXCOX y YEOJOHNSON. Además, se introduce el concepto de validación cruzada y la división de datos, que son clave para garantizar que los modelos trabajen con datos de manera confiable. Esta unidad sienta las bases para que los estudiantes desarrollen un conocimiento profundo en la estructuración y preprocesamiento de datos, asegurando la calidad en las etapas posteriores del análisis.

Unidad 2 Tratamiento de Datos: En esta unidad, los estudiantes explorarán técnicas para seleccionar y reducir las dimensiones de los datos, facilitando un análisis más eficiente y centrado en los atributos más relevantes. Se introduce la importancia de las características y el proceso de selección, así como la reducción de dimensiones mediante el Análisis de Componentes Principales. Con estas habilidades, los estudiantes podrán simplificar conjuntos de datos complejos, identificar patrones significativos y optimizar los modelos, mejorando tanto su rendimiento como su interpretabilidad.

Unidad 3 Modelado: Los estudiantes aprenden a aplicar algoritmos de machine learning para clasificar, predecir y agrupar datos. Se introduce una variedad de algoritmos como la Regresión Logística, K-Nearest Neighbors, Regresión Lineal y Lasso, así como el algoritmo de clustering K-MEANS. A través de estos modelos, los estudiantes comprenderán cómo realizar predicciones y segmentaciones de datos basadas en patrones detectados, sentando las bases para el desarrollo de soluciones que aprovechen el aprendizaje automático en escenarios prácticos.



Unidad 4 Evaluación de Modelos: Para culminar el curso, esta unidad enseña a los estudiantes a evaluar los modelos generados con métricas de precisión y confiabilidad. A través de métricas de clasificación, regresión y clustering, los estudiantes aprenderán a interpretar los resultados de los modelos y a compararlos para determinar su efectividad. Esta unidad incluye el análisis de métricas como Matriz de Confusión, Accuracy, Precision, Recall, ROC, MAE, MSE, R2 y otras, asegurando que los estudiantes puedan validar sus modelos y tomar decisiones fundamentadas. Con esta base, los estudiantes estarán preparados para aplicar sus conocimientos en proyectos reales y optimizar sus modelos en base a evaluaciones detalladas.

3. Participantes en el diseño y seguimiento curricular del programa

Lugar y fecha de elaboración o revisión	Participantes	Observaciones
Agosto-octubre 2024	Instituto Tecnológico de: Zitácuaro Huetamo Morelia Jiquilpan Lázaro Cárdenas La Piedad Uruapan	



4. Competencia(s) a desarrollar

Competencia(s) específica(s) de la asignatura

Las competencias que los estudiantes desarrollarán en la asignatura de Ciencia de Datos incluyen el manejo y procesamiento de datos, fortaleciendo su capacidad para estructurar y transformar información mediante técnicas de preprocesamiento que optimizan los resultados en las etapas de análisis. Adquirirán habilidades de análisis crítico para identificar y seleccionar las características más relevantes en un conjunto de datos. Además, aprenderán a implementar algoritmos de Machine Learning para construir modelos predictivos.

La evaluación de modelos predictivos y analíticos, donde los estudiantes aplicarán y compararán métricas de evaluación para clasificar, predecir y agrupar datos mediante indicadores, que les permitirán validar y mejorar la efectividad de los modelos en escenarios prácticos.

Se desarrollará la capacidad de interpretar resultados de modelos de ciencia de datos para generar conclusiones fundamentadas que guíen la toma de decisiones, garantizando que las soluciones propuestas se basen en análisis rigurosos.

Se fomentará una mentalidad innovadora y adaptable, permitiendo a los estudiantes aplicar técnicas de procesamiento, análisis y modelado en distintos contextos y tipos de datos, fortaleciendo su capacidad para enfrentar desafíos en un entorno dinámico.

5. Competencias previas

- Para el alumno de la especialidad Inteligencia de Negocios y Ciencia de Datos, es fundamental que el estudiante cuente con competencias previas en el análisis exploratorio de datos. Es necesario que el estudiante tenga conocimientos sólidos en el análisis y comprensión de datos, así como en la identificación y manejo de diversas fuentes de datos. Además, debe contar con habilidades en estadística descriptiva, lo que incluye calcular y entender medidas de tendencia central, dispersión y otras estadísticas básicas. La capacidad de visualización de datos es esencial, ya que se espera que el estudiante sepa interpretar y crear gráficos como histogramas, diagramas de densidad, boxplots, así como matrices de correlación y dispersión. Estos conocimientos previos son relevantes para que el estudiante pueda aprovechar al máximo los contenidos de Ciencia de Datos y avanzar hacia técnicas y modelos de mayor complejidad.



6. Temario

No.	Temas	Subtemas
1	Procesamiento de Datos.	1.1. Procesamiento de Datos. 1.2. Métodos de Transformación de Datos. 1.3. Métodos de Re muestreo. 1.4. Escalamiento y estandarización 1.5. Normalización y Binarización. 1.6. BOXCOS y YEOJOHNSON 1.7. Validación Cruzada 1.8. División por porcentaje
2	Tratamiento de Datos	2.1. Feature Selection. 2.2. Feature Importance. 2.3. Reducción de dimensiones (PCA).
3	Modelado	3.1. Introducción a los Algoritmos de Machine Learning. 3.2. Algoritmos de Clasificación. 3.2.1. Regresión Logística 3.2.2. KNN 3.3. Algoritmos de Regresión. 3.3.1. Lineal. 3.3.2. Lasso. 3.4. Algoritmos de Clustering. 3.4.1. KMEANS
4	Evaluación de Modelos	4.1. Introducción a la Evaluación de Modelos. 4.2. Métricas de Clasificación. 4.2.1. Matriz de Confusión 4.2.2. Accuracy, Precision y Recall 4.2.3. Kappa 4.2.4. ROC 4.2.5. Reporte de Clasificación 4.3. Métricas de Regresión. 4.3.1. Métrica MAE 4.3.2. Métrica MSE, R2 y R2 Ajustado 4.3.3. Correlación Pearson 4.4. Métricas de Clustering. 4.5. Comparación de algoritmos.



7. Actividades de aprendizaje de los temas

1. Tratamientos de Datos	
Competencias	Actividades de aprendizaje
<p><i>Específica(s):</i></p> <ul style="list-style-type: none">• Desarrollar la capacidad de manipular hojas de cálculo con grandes volúmenes de datos, aplicando principios de normalización.• Identificar y discriminar patrones de datos incorrectos y no estandarizados.• Dominar el uso de herramientas tecnológicas para la manipulación de datos, como hojas de cálculo y lenguajes de programación.	<ul style="list-style-type: none">• Presentar una visualización de datos que hayan sido normalizados y escalados.• Entregar un conjunto de datos transformados, listos para ser utilizados en modelos de análisis.
2. Tratamiento de Datos	
Competencias	Actividades de aprendizaje
<ul style="list-style-type: none">• Comprender la importancia de las variables en el contexto del problema a resolver.• Utilizar herramientas para transformar y representar adecuadamente los datos.	<ul style="list-style-type: none">• Presentar gráficos y tablas que muestren la reducción de los datos y la selección de variables más relevantes.
3. Modelado	
Competencias	Actividades de aprendizaje
<ul style="list-style-type: none">• Desarrollar la habilidad para identificar y emplear diferentes modelos de regresión, clasificación y clustering.• Seleccionar y aplicar el modelo más adecuado para el proyecto en desarrollo.• Utilizar herramientas tecnológicas para la implementación de los modelos.	<ul style="list-style-type: none">• Entregar un modelo de Regresión, Clasificación o Clustering, dependiendo de la naturaleza del proyecto en desarrollo.



4. Evaluación de Modelos	
Competencias	Actividades de aprendizaje
<ul style="list-style-type: none">Reconocer y aplicar diversas métricas para medir la calidad de los modelos analizados.Utilizar herramientas para presentar las métricas obtenidas mediante el entrenamiento y prueba de los modelos.Desarrollar habilidades de redacción y persuasión para comunicar los resultados del proyecto a las partes interesadas.	<ul style="list-style-type: none">Elaborar y entregar un reporte detallado del análisis de datos realizado, incluyendo la evaluación de modelos y recomendaciones basadas en los resultados obtenidos.

8. Práctica(s)

Práctica 1: Lectura y transformación de un conjunto de datos

Objetivo: Familiarizar a los estudiantes con la manipulación de datos básicos, integrando conjuntos obtenidos en materias previas y realizando transformaciones iniciales.

Actividades sugeridas:

- Importación de datos y limpieza inicial: Utilizar herramientas como Excel o software de análisis de datos (Python, R) para importar el conjunto de datos. Identificar valores nulos o inconsistentes y realizar una limpieza básica.
- Exploración de la estructura de los datos: Analizar las variables y sus tipos, realizando un resumen descriptivo de los datos. Esto puede incluir calcular medidas de tendencia central y dispersión.
- Transformación de datos: Aplicar transformaciones básicas, como normalización o estandarización, y registrar los cambios realizados. Discutir la importancia de estos pasos para la preparación de los datos en proyectos de análisis.

Práctica 2: Selección de características importantes mediante PCA y técnicas de selección de características

Objetivo: Guiar a los estudiantes en la identificación de las variables más relevantes, utilizando métodos de reducción de dimensionalidad y selección de características.

Actividades sugeridas:



- Introducción a PCA y selección de características: Revisar los conceptos básicos de PCA (Análisis de Componentes Principales) y otras técnicas de selección de características de manera teórica. Utilizar ejemplos prácticos para explicar su aplicación.
- Aplicación práctica de PCA y selección de características: Guiar a los estudiantes en la aplicación de PCA y otras técnicas de selección en sus datos, utilizando herramientas como Excel (si tienen complementos) o software de análisis (Python o R). Interpretar los resultados y seleccionar las variables más relevantes.
- Comparación y justificación de las variables seleccionadas: Comparar las variables seleccionadas mediante distintas técnicas y justificar su relevancia para el proyecto en desarrollo. Documentar el proceso en un informe breve.

Práctica 3: Construcción de modelos de clasificación, regresión y/o clustering

Objetivo: Introducir a los estudiantes en la implementación de modelos básicos de machine learning, enfocados en clasificación, regresión o clustering según los datos disponibles.

Actividades sugeridas:

- Selección del modelo adecuado: Discutir los tipos de problemas que pueden abordarse con clasificación, regresión o clustering y seleccionar el modelo más adecuado para los datos de cada estudiante.
- Implementación del modelo: Utilizar una herramienta accesible, como una hoja de cálculo o plataformas de machine learning de bajo código (como Google Colab, Weka o DataRobot), para crear el modelo. Si los estudiantes tienen conocimientos básicos de programación, pueden realizarlo en Python o R, con asistencia guiada.
- Evaluación inicial del modelo: Realizar una evaluación preliminar del modelo utilizando métricas básicas, y documentar los resultados obtenidos en términos de precisión, error o agrupación.

Práctica 4: Evaluación y selección del modelo más adecuado

Objetivo: Evaluar los modelos construidos en la práctica anterior y seleccionar el que ofrezca el mejor rendimiento para el problema en cuestión.

Actividades sugeridas:

- Aplicación de métricas de evaluación: Guiar a los estudiantes en la aplicación de métricas como precisión, recall, MSE o silhouette score, según el tipo de modelo (clasificación, regresión o clustering).
- Análisis comparativo de modelos: Comparar el desempeño de diferentes modelos en base a las métricas obtenidas y discutir cuál es el más adecuado para el proyecto, justificando la elección.



- Ajuste y refinamiento del modelo seleccionado: Realizar pequeños ajustes al modelo elegido (como la optimización de hiperparámetros en modelos con múltiples configuraciones) para mejorar su rendimiento. Documentar el proceso y los cambios obtenidos.

Práctica 5: Comparación de algoritmos y evaluación de resultados

Objetivo: Analizar y comparar los algoritmos utilizados durante la asignatura, evaluando los resultados de cada uno para identificar sus ventajas y limitaciones.

Actividades sugeridas:

- Revisión teórica de los algoritmos empleados: Repasar los algoritmos aplicados durante la materia, destacando sus principales ventajas y desventajas en distintos contextos.
- Comparación práctica de algoritmos en un mismo conjunto de datos: Aplicar dos o tres algoritmos diferentes al mismo conjunto de datos para observar y comparar sus resultados, utilizando las métricas de evaluación ya conocidas.

Análisis de resultados y conclusiones: Redactar un informe comparativo que incluya los resultados de cada algoritmo y una reflexión sobre cuál sería más adecuado para distintas problemáticas. Este informe debe incluir una justificación técnica y práctica basada en el desempeño y aplicabilidad de cada modelo.

9. Proyecto de asignatura

Desarrollo de un proyecto de ciencia de datos aplicado a un problema real, en el cual los estudiantes analicen y visualicen los datos desde una perspectiva interdisciplinaria, integrando conocimientos y habilidades de distintas carreras. El proyecto fomentará la colaboración entre disciplinas, permitiendo a los estudiantes abordar el problema desde diversos enfoques y enriquecer el análisis y la interpretación de los resultados.



10. Evaluación por competencias

La evaluación debe ser continua y cotidiana, por lo que se debe considerar el desempeño en cada una de las actividades de aprendizaje, haciendo especial énfasis en:

- Rúbrica o productos, señalados en cada unidad académica dentro de las actividades de aprendizaje.
- Prácticas propuestas y su presentación y exposición en plenaria. Algunas evaluarán por equipo.
- Portafolio de evidencias de información recabada durante las consultas e investigaciones solicitadas, plasmadas en documentos escritos.
- Descripción de otras experiencias concretas que se obtendrán al participar en discusiones, exposiciones o cualquier otro medio didáctico profesional que trate sobre la materia y que deberán realizarse durante el curso académico.
- Exámenes teóricos prácticos para comprobar la efectividad del estudiante en la resolución de casos prácticos.

11. Fuentes de información

1. Bobadilla, J. (2020). Machine Learning y Deep Learning: Usando Python, Scikit y Keras: (1 ed.). RA-MA Editorial. (Disponible en E-Libro)
2. Ortega Candell, J. M. (2022). Big data, machine learning y data science en Python: (1 ed.). RA-MA Editorial. (Disponible en E-Libro)
3. Velasco Rebolledo, J. (2024). Machine Learning: fundamentos, algoritmos y aplicaciones para los negocios, industria y finanzas: (1 ed.). Ediciones Díaz de Santos.
<https://elibro.net/es/lc/huetamo/titulos/275096>
4. Alpaydin, E. (2021). Machine learning. MIT press.
5. Lee, W. M. (2019). Python machine learning. John Wiley & Sons
6. Martino, J. C. R. (2019). Hands-On Machine Learning with Microsoft Excel 2019: Build complete data analysis flows, from data collection to visualization. Packt Publishing Ltd.