

	Nombre del documento: Formato de Programa de Estudio de asignatura de Especialidad	Código: TecNM-AC-PO-007-02
		Revisión: 0
	Referencia a la Norma ISO 9001:2015 8.3, 8.3.1	Página 1 de 29

1. Datos Generales de la asignatura

Nombre de la asignatura:	Análisis y Clasificación de Datos con Python.
Clave de la asignatura:	CDD-2501
SATCA¹:	2 - 3 - 5
Carrera:	Ingeniería en Sistemas Computacionales

2. Presentación

Caracterización de la asignatura
La asignatura Programación con Python para Análisis y Clasificación de Datos* es una pieza clave dentro de la formación de un estudiante de Ingeniería en Sistemas Computacionales, ya que contribuye significativamente al desarrollo de habilidades esenciales para el manejo de grandes volúmenes de datos, la creación de algoritmos, y la capacidad para implementar soluciones que involucran análisis y toma de decisiones basada en datos. Esta asignatura contribuye directamente al perfil de egreso al desarrollar competencias en el área de Ciencia de Datos, Programación Avanzada y Aprendizaje Automático. El Ingeniero egresado estará capacitado para analizar, diseñar y optimizar soluciones tecnológicas, debido a que la ciencia de datos es un componente clave en este proceso. Los conocimientos adquiridos en esta asignatura permiten al egresado procesar, analizar y clasificar grandes conjuntos de datos, que son habilidades fundamentales en sectores como la inteligencia de negocios, el análisis predictivo y la automatización de procesos. Además, la capacidad de aplicar modelos de aprendizaje automático, fortalece la habilidad del egresado para diseñar sistemas que aprendan y mejoren de manera autónoma, un aspecto crucial en el desarrollo de software inteligente. La importancia de esta asignatura radica en la creciente demanda de científicos de datos y desarrolladores de inteligencia artificial en el ámbito profesional. Las empresas de todos los sectores están buscando Ingenieros que puedan manejar datos y transformar esa información en valor para la organización, ya sea mediante la optimización de procesos, la predicción de tendencias o la automatización de tareas. Python se ha convertido en uno de los lenguajes de programación más utilizados en este campo debido a su simplicidad y la potencia de sus bibliotecas para el análisis de datos y el machine learning.

¹ Sistema de Asignación y Transferencia de Créditos Académicos

	Nombre del documento: Formato de Programa de Estudio de asignatura de Especialidad	Código: TecNM-AC-PO-007-02
		Revisión: 0
	Referencia a la Norma ISO 9001:2015 8.3, 8.3.1	Página 2 de 29

La asignatura se centra en proporcionar las bases de programación en Python y su aplicación a problemas de análisis y clasificación de datos. A lo largo del curso, los estudiantes aprenden a manejar bibliotecas clave como ****Pandas****, ****NumPy****, ****Matplotlib**** y ****Seaborn**** para la manipulación y visualización de datos. Además, se les introduce a técnicas de machine learning mediante el uso de ****Scikit-Learn**** para implementar algoritmos de clasificación, tales como regresión logística, KNN y Random Forest, fundamentales para la predicción y toma de decisiones en diversas áreas de aplicación. También se profundiza en la importancia del preprocesamiento de datos y la evaluación de modelos, dotando a los estudiantes de un enfoque integral en la resolución de problemas complejos.

La asignatura se relaciona con diversas materias del plan de estudios como Estructuras de Datos, Bases de Datos, Inteligencia Artificial, Gestión de Proyectos de Software y Tópicos Avanzados de programación entre otras, generando así un marco que permite desarrollar proyectos integradores:

En conclusión, la asignatura no solo es fundamental para el desarrollo de competencias clave en la ciencia de datos y el machine learning, sino que también se integra con otras materias del plan de estudios, fomentando la creación de proyectos multidisciplinarios que preparan a los estudiantes para enfrentar los desafíos del mundo laboral actual.

Intención didáctica

La asignatura está diseñada para que los estudiantes desarrollen competencias en el uso de Python aplicado a la ciencia de datos y el aprendizaje automático, con un enfoque práctico en la manipulación, análisis y visualización de datos, así como en la implementación de modelos de clasificación. A lo largo del curso, los estudiantes se sumergirán en temas que van desde la instalación y configuración del entorno de trabajo hasta la realización de proyectos integradores que abarcan todas las fases del ciclo de vida de un proyecto de ciencia de datos.

En el tema 1, los estudiantes aprenderán a instalar Python y herramientas fundamentales como Anaconda y Jupyter Notebooks, y explorarán los conceptos básicos del lenguaje, como la sintaxis, tipos de datos, estructuras de control y el uso de funciones. Además, se introducirá el uso de bibliotecas especializadas en ciencia de datos, como NumPy, Pandas y Matplotlib, que son esenciales para la manipulación y visualización de datos. El tema 2, profundiza en el uso de Pandas para gestionar estructuras de datos como Series y DataFrames, así como en la lectura y escritura de distintos formatos de archivo (CSV, Excel, SQL). Los estudiantes aprenderán a realizar tareas críticas de limpieza y preprocesamiento de datos, como el manejo de valores faltantes y duplicados, así como la transformación y normalización de los datos para facilitar su análisis. También se abordará el análisis descriptivo mediante estadísticas básicas y visualización.

El tema 3, estudiarán los principios de la visualización efectiva. Con Matplotlib, los estudiantes crearán gráficos básicos como líneas, barras y gráficos de dispersión,

	Nombre del documento: Formato de		Código: TecNM-AC-PO-007-02
	Programa de Estudio de asignatura de Especialidad		Revisión: 0
	Referencia a la Norma ISO 9001:2015	8.3, 8.3.1	Página 3 de 29

mientras que con Seaborn, explorarán visualizaciones avanzadas, como heatmaps, pairplots y boxplots, que permiten obtener información más profunda de los datos.

Tema 4, introduce los conceptos fundamentales de machine learning, diferenciando entre aprendizaje supervisado y no supervisado. Se utilizará la biblioteca Scikit-Learn para la implementación de modelos, y los estudiantes aprenderán a preparar conjuntos de datos para su uso en estos modelos, incluyendo la división en conjuntos de entrenamiento y prueba.

Tema 5, se presentarán varios algoritmos de clasificación, como la regresión logística, k-vecinos más cercanos (KNN), árboles de decisión y Random Forest. Los estudiantes aprenderán a evaluar el rendimiento de estos modelos utilizando métricas como precisión, recall, F1-score, así como curvas ROC y AUC. Además, se enseñarán técnicas de optimización de modelos mediante el ajuste de hiperparámetros y la validación cruzada. Finalmente, la Unidad 6, la integración de todos los conceptos aprendidos en un proyecto final en el que los estudiantes desarrollarán una solución completa, desde la limpieza de los datos hasta la clasificación y presentación de resultados mediante reportes y visualizaciones interactivas. Se discutirá también la ética y responsabilidad en el manejo de datos y la aplicación de algoritmos de aprendizaje automático.

3. Participantes en el diseño y seguimiento curricular del programa

Lugar y fecha de elaboración o revisión	Participantes	Observaciones
Zihuatanejo, Gro. del 20 Mayo al 28 de Octubre de 2024	Instituto Tecnológico de la Costa Grande	Reunión de Academia semanal e intersemestral

4. Competencia(s) a desarrollar

Competencia(s) específica(s) de la asignatura
El estudiante instalara y configurara entornos de desarrollo para ciencia de datos utilizando Python con bibliotecas esenciales, implementando estructuras de control y módulos de programación en Python, utilizando herramientas como NumPy, Pandas y Matplotlib para la manipulación de datos. Manipulando grandes volúmenes de información mediante la lectura, limpieza, y preprocesamiento de datos, desarrollara visualizaciones efectivas con Matplotlib y Seaborn. Además, deberá aplicar modelos de aprendizaje automático supervisado e implementa algoritmos de clasificación como regresión logística, KNN y Random Forest. Finalmente, el estudiante debe evaluar y optimizar los modelos de clasificación usando métricas de rendimiento y validando sus resultados en proyectos integradores, considerando aspectos éticos en el manejo de datos.

	Nombre del documento: Formato de		Código: TecNM-AC-PO-007-02
	Programa de Estudio de asignatura de Especialidad		Revisión: 0
	Referencia a la Norma ISO 9001:2015	8.3, 8.3.1	Página 4 de 29

5. Competencias previas

El estudiante deberá conocer los fundamentos de programación básica, incluyendo sintaxis, tipos de datos, y estructuras de control, bajo la condición de haber cursado asignaturas previas de introducción a la programación. Manejo de conceptos de álgebra básica y probabilidad para poder entender análisis de datos y algoritmos de clasificación. Comprender estructuras de datos y algoritmos simples para procesar información de manera eficiente. Además, utilizar herramientas de desarrollo de software y bases de datos (lectura y escritura de archivos CSV, Excel o SQL) en un entorno previo de trabajo, lo que le permitirá abordar con fluidez la manipulación de datos en Python.

6. Temario

No.	Temas	Subtemas
1	Introducción a Python y herramientas de Ciencia de Datos	1.1) Instalación y configuración del entorno 1.1.1) Instalación de Python y bibliotecas esenciales (Anaconda, Jupyter Notebooks). 1.2) Fundamentos de Python 1.2.1) Sintaxis básica, tipos de datos, estructuras de control. 1.2.2) Funciones y módulos en Python. 1.3) Bibliotecas para Ciencia de Datos 1.3.1) Introducción a NumPy, Pandas y Matplotlib.
2	Manipulación y análisis de Datos	2.1) Introducción a Pandas 2.1.1) Estructuras de datos: Series y DataFrames. 2.1.2) Lectura y escritura de datos (CSV, Excel, SQL). 2.2) Limpieza y preprocesamiento de datos 2.2.1) Manejo de datos faltantes y duplicados. 2.2.2) Transformación y normalización de datos. 2.3) Exploración de datos 2.3.1) Análisis descriptivo: estadísticas básicas y visualización.
3	Visualización de datos	3.1) Introducción a la Visualización 3.1.1) Principios de visualización efectiva. 3.2) Visualización con Matplotlib 3.2.1) Gráficas básicas: líneas, barras, dispersión. 3.3) Visualización avanzada con Seaborn 3.3.1) Creación de gráficos complejos: heatmaps, pairplots, boxplots.

	Nombre del documento: Formato de		Código: TecNM-AC-PO-007-02
	Programa de Estudio de asignatura de Especialidad		Revisión: 0
	Referencia a la Norma ISO 9001:2015	8.3, 8.3.1	Página 5 de 29

4	Introducción a aprendizaje automático	4.1) Conceptos Básicos de Aprendizaje Automático 4.1.1) Tipos de aprendizaje: supervisado vs. no supervisado. 4.2) Bibliotecas de Aprendizaje Automático 4.2.1) Introducción a Scikit-Learn. 4.3) Preparación de Datos para Modelos de Aprendizaje Automático. 4.3.1) División de conjuntos de datos: entrenamiento y prueba.
5	Clasificación de datos	5.1) Algoritmos de clasificación 5.1.1) Regresión logística, k-vecinos más cercanos (KNN). 5.1.2) Árboles de decisión y Random Forest. 5.2) Evaluación de modelos de clasificación. 5.2.1) Métricas de rendimiento: precisión, recall, F1-score. 5.2.2) Curvas ROC y AUC. 5.3) Optimización de Modelos. 5.3.1) Ajuste de hiperparámetros y validación cruzada.
6	Proyectos prácticos y aplicaciones	6.1) Proyecto integrador. 6.1.1) Desarrollo de un proyecto que integre todos los conceptos aprendidos. 6.1.2) Aplicación de técnicas de limpieza, análisis, visualización y clasificación de datos. 6.2) Presentación de resultados. 6.2.1) Creación de reportes y visualizaciones interactivas. 6.3) Discusión sobre Ética y responsabilidad en Ciencia de Datos. 6.3.1) Consideraciones éticas en el manejo de datos y algoritmos de aprendizaje automático.

	Nombre del documento: Formato de Programa de Estudio de asignatura de Especialidad	Código: TecNM-AC-PO-007-02
		Revisión: 0
	Referencia a la Norma ISO 9001:2015 8.3, 8.3.1	Página 6 de 29

7. Actividades de aprendizaje de los temas

Unidad I Introducción a Python y herramientas de Ciencia de Datos	
Competencias	Actividades de aprendizaje
Específica(s): Conoce y aplica la Instalación y configuración de entorno de desarrollo para ciencia de datos, incluyendo Python y bibliotecas esenciales para una programación eficiente. Genéricas: • Habilidades de gestión de información (habilidad para buscar y analizar información proveniente de fuentes diversas). • Capacidad de análisis y síntesis. • Capacidad de comunicación oral y escrita. • Capacidad de aplicar los conocimientos en la práctica • Habilidades en el uso de las tecnologías de la información y de la comunicación. • Capacidad para analizar y organizar datos de forma estructurada.	• El docente proporciona materiales y recursos en línea sobre la instalación y configuración de Python y bibliotecas esenciales, **mientras que el estudiante realiza la instalación de Anaconda y Jupyter Notebooks, creando su propio entorno de trabajo. • El docente expone y explica ejemplos sobre los fundamentos de Python, propone ejercicios para desarrollarse en clase con el objetivo que el estudiante complemente actividades mediante el desarrollo de escribir scripts utilizando sintaxis básica, estructuras de control, y funciones, reforzando su comprensión a través de las prácticas propuestas por el docente de forma extraclase. • El docente organiza sesiones de laboratorio donde se introducen bibliotecas como NumPy, Pandas y Matplotlib, el estudiante participa en la manipulación y visualización de conjuntos de datos, aplicando los conceptos aprendidos en casos prácticos. • El docente promueve discusiones grupales sobre las mejores prácticas en programación y análisis de datos, mientras que el estudiante colabora en grupos para intercambiar ideas y resolver problemas, fomentando la comunicación y el aprendizaje colaborativo. • El docente plantea proyectos que fomente la búsqueda y selección de información de diversas fuentes sobre aplicaciones de ciencia de datos, el estudiante investiga y presenta sus hallazgos, desarrollando habilidades de inducción-deducción y análisis-síntesis.

	Nombre del documento: Formato de Programa de Estudio de asignatura de Especialidad	Código: TecNM-AC-PO-007-02
		Revisión: 0
	Referencia a la Norma ISO 9001:2015 8.3, 8.3.1	Página 7 de 29

	El docente guía en la organización de los trabajos individuales y grupales, el estudiante elabora planes de trabajo que contemplen la ejecución de actividades prácticas y la evaluación de resultados, fortaleciendo su capacidad de planeación y organización.
Unidad II Manipulación y Análisis de Datos	
Competencias	Actividades de aprendizaje
Específica(s): Utilizar la biblioteca Pandas, realizando la lectura y escritura de datos desde archivos CSV, Excel y bases de datos SQL, aplicando técnicas de limpieza y preprocesamiento, identificación y manejo datos. Genéricas: - Habilidades de gestión de información (habilidad para buscar y analizar información proveniente de fuentes diversas). - Capacidad de análisis y síntesis. - Capacidad de comunicación oral y escrita. - Capacidad de aplicar los conocimientos en la práctica - Habilidades en el uso de las tecnologías de la información y de la comunicación. - Capacidad para analizar y organizar datos de forma estructurada.	- El docente proporciona recursos en línea sobre Pandas y su aplicación en el manejo de datos, el estudiante investiga ejemplos prácticos sobre Series y DataFrames, promoviendo la búsqueda y análisis de información en diversas fuentes. - El docente organiza talleres con ejemplos para la lectura y escritura de datos desde diferentes formatos (CSV, Excel, SQL), el estudiante desarrolla ejercicios donde aplica estos métodos, fomentando el uso de nuevas tecnologías en el proceso de aprendizaje. - El docente guía a los estudiantes en la planificación de proyectos de análisis de datos, mientras que el estudiante desarrolla un cronograma para la limpieza y preprocesamiento de datos, favoreciendo la organización y estructuración de su trabajo. - El docente plantea ejercicios de casos de estudio donde los estudiantes deben realizar análisis descriptivos, aplicando inducción-deducción para identificar tendencias, patrones, desarrollando habilidades de análisis y síntesis. - El docente propone practicas mediante problemas que integran conceptos de manipulación de datos y análisis estadístico, el estudiante resuelve estos problemas, aplicando los conocimientos, habilidades adquiridas y relacionando contenidos con otras asignaturas.

	Nombre del documento: Formato de	Código: TecNM-AC-PO-007-02
	Programa de Estudio de asignatura de Especialidad	Revisión: 0
	Referencia a la Norma ISO 9001:2015 8.3, 8.3.1	Página 8 de 29

Unidad III Visualización de datos	
Competencias	Actividades de aprendizaje
Específica(s): Comprender y aplicar los principios fundamentales de la visualización efectiva, seleccionando y utilizando de manera adecuada las herramientas de visualización de datos en Python, específicamente con las bibliotecas Matplotlib y Seaborn. Genéricas: • Habilidades de gestión de información (habilidad para buscar y analizar información proveniente de fuentes diversas). • Capacidad de análisis y síntesis. • Capacidad de comunicación oral y escrita. • Capacidad de aplicar los conocimientos en la práctica • Habilidades en el uso de las tecnologías de la información y de la comunicación. • Capacidad para analizar y organizar datos de forma estructurada.	• Fomentar la investigación y selección, donde el estudiante realiza búsquedas en distintas fuentes sobre los principios de visualización efectiva y el uso de herramientas como Matplotlib y Seaborn, permitiéndole desarrollar competencias de análisis crítico de información técnica. • Impulsar el uso de las tecnologías de la información y comunicaciones, el docente organiza sesiones con ejemplos, donde los estudiantes trabajan directamente con Python, utilizando Matplotlib y Seaborn para crear visualizaciones básicas y avanzadas. • Los estudiantes desarrollan proyectos que incluyen la planeación de gráficos para la presentación de datos. Esto requiere organizar el flujo de trabajo y planificar las visualizaciones necesarias para comunicar hallazgos de manera efectiva. • Se realizan actividades grupales donde los estudiantes deben interpretar datos y discutir la mejor manera de visualizarlos. Estos ejercicios permiten la argumentación y el intercambio de ideas, fortaleciendo habilidades de comunicación y colaboración. • Se propone a los estudiantes el análisis de datos reales y su síntesis en visualizaciones efectivas. Esta actividad impulsa la capacidad de análisis-deducción y fomenta el pensamiento crítico y la capacidad de abstracción para comunicar información visualmente. • El docente plantea prácticas donde se exploran variables de un conjunto de datos, los estudiantes experimentan con distintas opciones de visualización y seleccionan la más adecuada según los objetivos de análisis, promoviendo la

	Nombre del documento: Formato de Programa de Estudio de asignatura de Especialidad	Código: TecNM-AC-PO-007-02
		Revisión: 0
	Referencia a la Norma ISO 9001:2015 8.3, 8.3.1	Página 9 de 29

	observación, identificación de variables y planteamiento de hipótesis.
Unidad IV Introducción a aprendizaje automático	
Competencias	Actividades de aprendizaje
Específica(s): Comprender los conceptos fundamentales del aprendizaje automático y diferenciar entre los tipos de aprendizaje supervisado y no supervisado. Además, deberá adquirir habilidades en el uso de Scikit-Learn, una biblioteca clave de Python para implementar modelos básicos de aprendizaje automático. Genéricas: - Habilidades de gestión de información (habilidad para buscar y analizar información proveniente de fuentes diversas). - Capacidad de análisis y síntesis. - Capacidad de comunicación oral y escrita. - Capacidad de aplicar los conocimientos en la práctica - Habilidades en el uso de las tecnologías de la información y de la comunicación. - Capacidad para analizar y organizar datos de forma estructurada.	- El docente propone actividades de búsqueda de información sobre tipos de aprendizaje automático (supervisado vs. no supervisado). Esto permite al estudiante desarrollar competencias en la selección y análisis de información en fuentes confiables, fomentando el análisis crítico y el uso adecuado de conceptos técnico-científicos. - En sesiones de ejemplos prácticos proporcionado por el docente, los estudiantes trabajan con Scikit-Learn, explorando sus funcionalidades y estructuras para la implementación de modelos. Esto fortalece su competencia en el uso de nuevas tecnologías y en la resolución de problemas mediante herramientas específicas de aprendizaje automático. - Planeación de proyectos de aprendizaje automático planteado por el docente, donde se pide a los estudiantes planificar y estructurar un proyecto sencillo de modelado en el que definan los pasos para preparar datos y construir un modelo. Esta actividad promueve la organización y planificación, habilidades cruciales en la gestión de proyectos de análisis de datos en contextos profesionales. - El docente impulsa el trabajo en equipo y discusión de enfoques, en el cual los estudiantes forman grupos para discutir y comparar enfoques de aprendizaje supervisado y no supervisado en casos específicos, lo que fomenta el intercambio de ideas y el desarrollo de habilidades comunicativas y colaborativas.

	Nombre del documento: Formato de Programa de Estudio de asignatura de Especialidad	Código: TecNM-AC-PO-007-02
		Revisión: 0
	Referencia a la Norma ISO 9001:2015 8.3, 8.3.1	Página 10 de 29

	• Análisis y síntesis de modelos, los estudiantes aplican el análisis-síntesis para seleccionar y justificar el uso de un tipo de aprendizaje automático para cada problema específico, lo cual incentiva el desarrollo de habilidades de razonamiento crítico y la capacidad de formular hipótesis para resolución de problemas. • El docente propone prácticas de preparación de datos, a través de ejercicios prácticos, los estudiantes dividen conjuntos de datos en entrenamiento y prueba, experimentando con distintas configuraciones y observando cómo afectan los resultados del modelo. Esta práctica fortalece su comprensión del manejo de variables y la validación de resultados en aprendizaje automático. • Integración y solución de problemas interdisciplinares**: Se proponen problemas donde los estudiantes deben integrar conceptos de otras asignaturas, como estadística y programación, para optimizar sus modelos, mejorando así su capacidad de análisis y la integración de conocimientos.
Unidad V Clasificación de datos	
Específica(s): Aplicar diferentes algoritmos de clasificación, como regresión logística, k-vecinos más cercanos (KNN), árboles de decisión y Random Forest, para resolver problemas de clasificación de datos y ser competente para evaluar el rendimiento de estos modelos utilizando métricas como precisión, recall y F1-score, así como en interpretar curvas (ROC) y área bajo la curva (AUC) para análisis más detallados. Genéricas:	• Investigación y análisis de algoritmos, los estudiantes investigarán diferentes algoritmos de clasificación, como regresión logística, KNN, árboles de decisión y Random Forest, comparando sus características y aplicaciones en el análisis de datos. Esto fomenta la búsqueda y análisis de información y el uso de fuentes confiables. • Implementación de ejemplos en Python, utilizando bibliotecas como Scikit-Learn, los estudiantes crearán modelos de clasificación, evaluarán su rendimiento con métricas como precisión, recall y F1-score, y analizarán curvas ROC y AUC. Esto propicia el uso de tecnologías

	Nombre del documento: Formato de		Código: TecNM-AC-PO-007-02
	Programa de Estudio de asignatura de Especialidad		Revisión: 0
	Referencia a la Norma ISO 9001:2015	8.3, 8.3.1	Página 11 de 29

• Habilidades de gestión de información (habilidad para buscar y analizar información proveniente de fuentes diversas). • Capacidad de análisis y síntesis. • Capacidad de comunicación oral y escrita. • Capacidad de aplicar los conocimientos en la práctica • Habilidades en el uso de las tecnologías de la información y de la comunicación. • Capacidad para analizar y organizar datos de forma estructurada.	actuales y permite al estudiante experimentar con conceptos y metodologías de la asignatura. • Optimización y validación de modelos, mediante ejercicios con problemas, en equipos, los estudiantes aplicarán técnicas de ajuste de hiperparámetros y validación cruzada para mejorar la precisión de los modelos, desarrollando habilidades de planeación, análisis y trabajo en equipo. • Practica aplicando la clasificación: el docente propondrá un proyecto de un caso práctico, en el que los estudiantes integren los conocimientos y practiquen el análisis, la limpieza y el preprocesamiento de datos reales, eligiendo el algoritmo más adecuado para su clasificación y evaluando el modelo creado. Esto fomenta la resolución de problemas complejos, el uso de terminología científico-tecnológica y la colaboración en equipo, además de conectar sus competencias con problemáticas profesionales en el campo de la ciencia de datos.
--	--

Unidad VI Proyectos prácticos y aplicaciones

Específica(s): Capaz de resolver problemas donde aplique de manera efectiva todos los conocimientos adquiridos sobre la manipulación, análisis, visualización y clasificación de datos, integrando técnicas de limpieza de datos, análisis descriptivo y la utilización de algoritmos de aprendizaje automático para resolver problemas prácticos y presentar los resultados de forma clara, utilizando herramientas de visualización interactivas y reportes bien estructurados.	• Investigación y Análisis de Información: Los estudiantes deberán buscar y seleccionar información sobre problemas reales en ciencia de datos, revisando fuentes confiables y relevantes. Esto los conducirá a identificar datos útiles y a plantear hipótesis que les permitan aplicar las técnicas aprendidas durante el curso. • Aplicación de Herramientas Tecnológicas: El estudiante usará tecnologías y herramientas modernas como Python, Pandas, Matplotlib, Scikit-Learn, y otras bibliotecas de visualización y aprendizaje automático para procesar y analizar datos, generando insights a partir de conjuntos de datos reales.
---	---

	Nombre del documento: Formato de Programa de Estudio de asignatura de Especialidad	Código: TecNM-AC-PO-007-02
		Revisión: 0
	Referencia a la Norma ISO 9001:2015 8.3, 8.3.1	Página 12 de 29

Genéricas:

- Habilidades de gestión de información (habilidad para buscar y analizar información proveniente de fuentes diversas).
- Capacidad de análisis y síntesis.
- Capacidad de comunicación oral y escrita.
- Capacidad de aplicar los conocimientos en la práctica
- Habilidades en el uso de las tecnologías de la información y de la comunicación.
- Capacidad para analizar y organizar datos de forma estructurada.

- Planeación y Organización del Proyecto: Los estudiantes desarrollarán un proyecto integrador que involucra planificación y organización, gestionando el flujo del trabajo, la limpieza de datos, su análisis y clasificación, así como la construcción de modelos predictivos.
- Trabajo en Equipo y Comunicación: A través de actividades grupales, los estudiantes compartirán ideas, discutirán enfoques y colaborarán en el desarrollo del proyecto, fortaleciendo habilidades de comunicación y trabajo colaborativo.
- Desarrollo de Pensamiento Crítico y Solución de Problemas: Mediante la aplicación práctica de los modelos y conceptos aprendidos, los estudiantes desarrollarán habilidades de análisis y síntesis para resolver problemas complejos. Esto incluirá la toma de decisiones basada en resultados obtenidos de los modelos de clasificación y su validación.
- Presentación de Resultados: Los estudiantes presentarán sus proyectos mediante reportes y visualizaciones interactivas, desarrollando habilidades de comunicación profesional y el uso adecuado de las herramientas tecnológicas para la creación de informes comprensibles y visualmente atractivos.
- Ética y Responsabilidad en Ciencia de Datos: Finalmente, se fomentará la reflexión sobre los aspectos éticos relacionados con el manejo y análisis de datos, promoviendo en los estudiantes un enfoque responsable en el uso de datos y algoritmos de aprendizaje automático, respetando la privacidad y la justicia en la toma de decisiones automatizada.

	Nombre del documento: Formato de Programa de Estudio de asignatura de Especialidad	Código: TecNM-AC-PO-007-02
		Revisión: 0
	Referencia a la Norma ISO 9001:2015 8.3, 8.3.1	Página 13 de 29

8. Práctica(s)

Unidad I

Práctica 1: Introducción a NumPy - Operaciones y Manipulación de Arrays

Objetivo: Familiarizarse con la biblioteca NumPy, trabajar con arrays y realizar operaciones matemáticas básicas.

Instrucciones:

1. Crear arrays unidimensionales y multidimensionales con datos generados aleatoriamente.
2. Realizar operaciones básicas de suma, resta, multiplicación y división entre arrays.
3. Calcular estadísticos como media, desviación estándar y suma acumulada en los arrays.
4. Filtrar elementos de un array utilizando condiciones lógicas (por ejemplo, seleccionar solo los valores mayores a 10).
5. Realizar operaciones avanzadas, como transponer un array y realizar productos matriciales entre dos matrices.

Práctica 2: Análisis de Datos con Pandas - Exploración y Limpieza de Datos

Objetivo: Aprender a utilizar Pandas para leer, explorar y limpiar un conjunto de datos.

Instrucciones:

1. Leer un archivo CSV con Pandas (puede ser un dataset de ejemplo, como datos de ventas o clima).
2. Explorar el dataset mostrando las primeras y últimas filas, el tamaño del dataset y sus estadísticas descriptivas.
3. Identificar y tratar valores faltantes (por ejemplo, eliminarlos o sustituirlos con la media o la mediana).
4. Crear nuevas columnas derivadas, como convertir datos de fechas en día, mes y año.
5. Realizar filtrados y agrupamientos para obtener estadísticas específicas, como el promedio de ventas por mes o el total por categoría.

Práctica 3: Visualización Básica con Matplotlib - Gráficos de Datos

Objetivo: Usar Matplotlib para visualizar datos, crear gráficos y personalizar su apariencia.

Instrucciones:

1. Generar un gráfico de líneas que represente el cambio de una variable en el tiempo (por ejemplo, una serie de valores de temperatura).
2. Crear un gráfico de barras que muestre comparaciones entre categorías (por ejemplo, ventas por región).
3. Realizar un gráfico de dispersión para observar la relación entre dos variables (como edad vs. ingresos).
4. Personalizar los gráficos agregando títulos, etiquetas en los ejes, colores y leyendas.
5. Guardar los gráficos generados en formato PNG o JPG.

Práctica 4: Integración de NumPy, Pandas y Matplotlib - Análisis Exploratorio de Datos

Objetivo: Realizar un análisis exploratorio de datos utilizando las tres bibliotecas, combinando manipulación, análisis y visualización de datos.

Instrucciones:

	Nombre del documento: Formato de Programa de Estudio de asignatura de Especialidad	Código: TecNM-AC-PO-007-02
		Revisión: 0
	Referencia a la Norma ISO 9001:2015 8.3, 8.3.1	Página 14 de 29

1. Cargar un dataset con Pandas y explorar sus estadísticas generales.
2. Limpiar el dataset eliminando valores faltantes y aplicando transformaciones (por ejemplo, normalización usando NumPy).
3. Calcular estadísticas específicas usando Pandas y NumPy (como media y desviación de columnas numéricas).
4. Visualizar diferentes aspectos del dataset usando Matplotlib: por ejemplo, crear un histograma para observar la distribución de una variable y un gráfico de caja (boxplot) para ver la dispersión de datos.
5. Concluir el análisis interpretando los gráficos y estadísticas obtenidas.

Unidad II

Práctica 1: Estructuras de Datos en Pandas - Series y DataFrames

Objetivo: Familiarizarse con las estructuras de datos Series y DataFrames en Pandas, así como aprender a crear y manipular datos.

Instrucciones:

1. Crear una Serie con datos de ejemplo (por ejemplo, calificaciones de estudiantes o precios de productos) y practicar operaciones como obtener la media, máximo y mínimo.
2. Crear un DataFrame a partir de un diccionario de listas (por ejemplo, datos de estudiantes con columnas como nombre, edad, calificación).
3. Acceder y seleccionar datos específicos en el DataFrame usando índices y etiquetas.
4. Añadir, modificar y eliminar columnas y filas del DataFrame.
5. Realizar operaciones matemáticas en columnas numéricas y aplicar funciones en columnas específicas.

Práctica 2: Lectura y Escritura de Datos - Archivos CSV, Excel y SQL

Objetivo: Practicar la lectura y escritura de datos en diferentes formatos utilizando Pandas.

Instrucciones:

1. Cargar un archivo CSV en un DataFrame (por ejemplo, datos de ventas o inventario) y visualizar las primeras filas.
2. Cargar un archivo Excel en un DataFrame, especificando qué hoja cargar.
3. Conectar a una base de datos SQL (utilizando SQLite, por ejemplo) y cargar una tabla en un DataFrame.
4. Realizar algunas operaciones simples de exploración, como contar filas, obtener estadísticas básicas y filtrar datos.
5. Guardar el DataFrame en un archivo CSV o Excel después de realizar cambios en los datos.

Práctica 3: Limpieza y Preprocesamiento de Datos - Datos Faltantes y Duplicados

Objetivo: Aprender a identificar y manejar datos faltantes y duplicados en un DataFrame.

Instrucciones:

1. Cargar un dataset con valores faltantes (puedes generarlo manualmente o descargar uno de sitios como Kaggle).
2. Identificar y contar los valores faltantes en cada columna.

	Nombre del documento: Formato de Programa de Estudio de asignatura de Especialidad	Código: TecNM-AC-PO-007-02
		Revisión: 0
	Referencia a la Norma ISO 9001:2015 8.3, 8.3.1	Página 15 de 29

3. Implementar diferentes métodos para tratar los datos faltantes, como eliminación de filas con valores nulos, sustitución con la media o la mediana.
4. Detectar y eliminar filas duplicadas en el dataset.
5. Normalizar o escalar una columna numérica, por ejemplo, convirtiendo valores a una escala de 0 a 1.

Práctica 4: Exploración de Datos - Análisis Descriptivo y Visualización

Objetivo: Realizar un análisis exploratorio de datos, utilizando estadísticas descriptivas y visualizaciones.

Instrucciones:

1. Cargar un dataset y explorar su estructura y tipos de datos.
2. Generar estadísticas descriptivas como la media, mediana, desviación estándar y percentiles de columnas numéricas.
3. Realizar un análisis visual creando gráficos de resumen de los datos usando Matplotlib o Seaborn, por ejemplo:
 - Gráfico de barras para comparar categorías (como productos más vendidos).
 - Histograma para observar la distribución de una variable numérica.
 - Boxplot para identificar valores atípicos en una columna.
4. Crear un reporte de resumen de los datos principales y visualizar los hallazgos en los gráficos generados.
5. Concluir interpretando las estadísticas y gráficos para entender mejor los datos.

Unidad III

Práctica 1: Principios de Visualización Efectiva - Gráficos Claros y Precisos

Objetivo: Aprender los principios básicos de visualización efectiva, aplicando buenas prácticas para crear gráficos claros y bien diseñados.

Instrucciones:

1. Crear un conjunto de datos sencillo (por ejemplo, ventas mensuales de productos) e implementarlo en un gráfico de líneas.
2. Añadir etiquetas claras en los ejes y un título descriptivo al gráfico.
3. Aplicar personalización de colores, estilos de líneas y marcadores para hacer el gráfico visualmente atractivo y legible.
4. Crear un gráfico de barras comparativo y ajustar el tamaño y la rotación de las etiquetas en el eje x para mejorar la legibilidad.
5. Practicar la adición de anotaciones en puntos específicos para resaltar valores importantes (por ejemplo, el valor de ventas más alto).

Práctica 2: Visualización Básica con Matplotlib - Gráficos de Líneas, Barras y Dispersión

Objetivo: Dominar la creación de gráficos básicos en Matplotlib, explorando sus aplicaciones y personalización.

Instrucciones:

1. Generar datos de ejemplo, como temperatura por horas o ingresos por mes, y representar estos datos en un gráfico de líneas.
2. Crear un gráfico de barras para comparar diferentes categorías (por ejemplo, ventas por producto o calificaciones promedio por grupo).

	Nombre del documento: Formato de Programa de Estudio de asignatura de Especialidad	Código: TecNM-AC-PO-007-02
		Revisión: 0
	Referencia a la Norma ISO 9001:2015 8.3, 8.3.1	Página 16 de 29

3. Usar un gráfico de dispersión para observar la relación entre dos variables (por ejemplo, edad vs. ingresos, altura vs. peso).
4. Personalizar los gráficos cambiando colores, estilos de líneas, tamaños y añadir leyendas.
5. Exportar los gráficos en formato de imagen (PNG o JPG) para su uso posterior.

Práctica 3: Gráficos Avanzados con Seaborn - Heatmaps, Pairplots y Boxplots

Objetivo: Crear visualizaciones avanzadas con Seaborn para el análisis de relaciones y patrones en los datos.

Instrucciones:

1. Cargar un dataset en Pandas (por ejemplo, un conjunto de datos de Iris o titánico) y analizar sus estadísticas descriptivas.
2. Usar Seaborn para crear un heatmap que muestre la correlación entre variables numéricas en el dataset.
3. Generar un pairplot para explorar visualmente la relación entre múltiples variables (útil para datasets con varias características).
4. Crear un boxplot para observar la distribución de valores y detectar posibles valores atípicos en una variable numérica (por ejemplo, salario por grupo de edad).
5. Interpretar los gráficos generados y hacer anotaciones sobre las relaciones o patrones observados.

Práctica 4: Proyecto Integrado de Visualización - Análisis de un Dataset Complejo

Objetivo: Aplicar los conocimientos de Matplotlib y Seaborn en un análisis visual completo de un conjunto de datos más complejo.

Instrucciones:

1. Elegir un dataset (por ejemplo, datos de salud, mercado o financiero) y cargarlo en un DataFrame de Pandas.
2. Explorar y limpiar los datos si es necesario, eliminando valores nulos o transformando variables.
3. Crear una serie de visualizaciones:
 - Gráfico de líneas para observar cambios en una variable en el tiempo (por ejemplo, precios de acciones).
 - Gráfico de dispersión para analizar la relación entre dos variables clave.
 - Heatmap para visualizar la correlación entre varias variables.
 - Boxplot para observar la distribución de una variable específica en función de diferentes categorías.
4. Crear un reporte visual que incluya cada gráfico, acompañado de una breve interpretación de lo que muestra cada visualización.
5. Concluir con un resumen de los hallazgos principales a partir de los gráficos generados.

Unidad IV

Práctica 1: Conceptos Básicos de Aprendizaje Automático - Tipos de Aprendizaje

Objetivo: Entender la diferencia entre aprendizaje supervisado y no supervisado.

Instrucciones:

	Nombre del documento: Formato de		Código: TecNM-AC-PO-007-02
	Programa de Estudio de asignatura de Especialidad		Revisión: 0
	Referencia a la Norma ISO 9001:2015	8.3, 8.3.1	Página 17 de 29

1. Investigar y anotar ejemplos de problemas que pueden resolverse mediante aprendizaje supervisado (por ejemplo, clasificación de correos electrónicos como spam o no spam) y no supervisado (por ejemplo, agrupamiento de clientes).
2. Crear un conjunto de datos sencillo con características y etiquetas (por ejemplo, características de frutas como color, peso, tipo) para simular un problema de clasificación supervisada.
3. Generar un conjunto de datos sin etiquetas y realizar un agrupamiento simple para explorar el aprendizaje no supervisado.
4. Escribir una reflexión breve sobre las diferencias en cómo se entrenan los modelos supervisados y no supervisados.

Práctica 2: Introducción a Scikit-Learn - Clasificación y Regresión

Objetivo: Familiarizarse con Scikit-Learn y aplicar algoritmos de clasificación y regresión básicos.

Instrucciones:

1. Cargar un conjunto de datos de Scikit-Learn, como iris (clasificación) o boston (regresión).
2. Explorar el dataset e identificar características y etiquetas.
3. Utilizar Scikit-Learn para dividir el conjunto de datos en entrenamiento y prueba.
4. Entrenar un modelo de clasificación (por ejemplo, KNeighborsClassifier para el iris) y un modelo de regresión (por ejemplo, LinearRegression para el boston).
5. Evaluar el rendimiento de cada modelo calculando la precisión (clasificación) o el error cuadrático medio (regresión) y discutir los resultados obtenidos.

Práctica 3: Preparación de Datos para Modelos - Manejo y Escalado de Datos

Objetivo: Practicar la preparación de datos para modelos de aprendizaje automático.

Instrucciones:

1. Cargar un conjunto de datos con características numéricas y categóricas (por ejemplo, el dataset titanic de Kaggle).
2. Identificar y manejar los datos faltantes, ya sea eliminando filas o imputando valores.
3. Convertir variables categóricas en variables numéricas utilizando métodos como OneHotEncoding.
4. Escalar las características numéricas usando StandardScaler o MinMaxScaler para estandarizar los datos.
5. Dividir el conjunto de datos preprocesado en conjuntos de entrenamiento y prueba, listos para el entrenamiento de modelos.

Práctica 4: Proyecto Completo de Aprendizaje Automático - Implementación de un Modelo Predictivo

Objetivo: Aplicar todo el flujo de trabajo de aprendizaje automático, desde la carga y preparación de datos hasta la construcción y evaluación de un modelo.

Instrucciones:

1. Seleccionar o cargar un conjunto de datos adecuado para un problema de clasificación o regresión (por ejemplo, predicción de ingresos con el dataset adult o predicción de supervivencia con el titanic).
2. Realizar una exploración de los datos para conocer las características y su tipo, y preparar los datos (manejo de datos faltantes, escalado, codificación de variables categóricas).

	Nombre del documento: Formato de		Código: TecNM-AC-PO-007-02
	Programa de Estudio de asignatura de Especialidad		Revisión: 0
	Referencia a la Norma ISO 9001:2015	8.3, 8.3.1	Página 18 de 29

3. Dividir los datos en conjuntos de entrenamiento y prueba.
4. Entrenar un modelo usando un algoritmo de Scikit-Learn adecuado (por ejemplo, RandomForestClassifier o LinearRegression).
5. Evaluar el rendimiento del modelo en el conjunto de prueba usando métricas como precisión, recall o MAE, y generar un reporte de los resultados obtenidos.
6. Presentar un análisis final sobre el rendimiento del modelo y las mejoras que podrían hacerse.

Unidad V

Práctica 1: Implementación de Algoritmos de Clasificación Básicos

Objetivo: Implementar y comprender el funcionamiento de algoritmos de clasificación como regresión logística y k-vecinos más cercanos (KNN).

Instrucciones:

1. Cargar el conjunto de datos Iris (o cualquier otro conjunto de datos de clasificación, como Titanic).
2. Dividir los datos en conjuntos de entrenamiento y prueba.
3. Implementar un modelo de regresión logística para clasificar los datos y entrenarlo con el conjunto de entrenamiento.
4. Implementar un modelo de k-vecinos más cercanos (KNN) y entrenarlo con el mismo conjunto de entrenamiento.
5. Evaluar ambos modelos utilizando el conjunto de prueba, calculando métricas de rendimiento como precisión y F1-score.
6. Comparar el rendimiento de ambos modelos y discutir en qué casos cada uno sería más adecuado.

Práctica 2: Construcción y Evaluación de Árboles de Decisión y Random Forest

Objetivo: Desarrollar un árbol de decisión y un modelo de bosque aleatorio (Random Forest) y evaluar sus métricas de rendimiento.

Instrucciones:

1. Cargar un conjunto de datos (como Breast Cancer de Scikit-Learn).
2. Dividir el conjunto de datos en conjuntos de entrenamiento y prueba.
3. Entrenar un modelo de árbol de decisión con el conjunto de entrenamiento.
4. Entrenar un modelo de bosque aleatorio (Random Forest) con el mismo conjunto de entrenamiento.
5. Evaluar ambos modelos utilizando el conjunto de prueba y calcular las métricas de precisión, recall y F1-score.
6. Comparar los resultados de ambos modelos y analizar las fortalezas y limitaciones de cada enfoque.

Práctica 3: Análisis de Curvas ROC y AUC en Modelos de Clasificación

Objetivo: Interpretar y analizar el rendimiento de modelos utilizando la curva ROC y el área bajo la curva (AUC).

Instrucciones:

1. Cargar un conjunto de datos binario, como Breast Cancer de Scikit-Learn.
2. Dividir los datos en conjuntos de entrenamiento y prueba.

	Nombre del documento: Formato de		Código: TecNM-AC-PO-007-02
	Programa de Estudio de asignatura de Especialidad		Revisión: 0
	Referencia a la Norma ISO 9001:2015	8.3, 8.3.1	Página 19 de 29

3. Entrenar un modelo de regresión logística con el conjunto de entrenamiento.
4. Usar las funciones de Scikit-Learn para trazar la curva ROC del modelo entrenado.
5. Calcular el área bajo la curva (AUC) para el modelo de regresión logística.
6. Entrenar un modelo de KNN y repetir los pasos 4 y 5.
7. Comparar las curvas ROC de ambos modelos y analizar qué modelo es más efectivo a diferentes umbrales de decisión.

Práctica 4: Optimización de Modelos mediante Ajuste de Hiperparámetros y Validación Cruzada

Objetivo: Optimizar un modelo de clasificación mediante ajuste de hiperparámetros y validación cruzada.

Instrucciones:

1. Cargar el conjunto de datos Iris o Breast Cancer.
2. Dividir los datos en conjuntos de entrenamiento y prueba.
3. Entrenar un modelo de Random Forest con hiperparámetros predeterminados y evaluarlo.
4. Implementar GridSearchCV o RandomizedSearchCV para realizar una búsqueda de hiperparámetros óptimos para el modelo de Random Forest.
5. Utilizar validación cruzada para evaluar el modelo con los mejores hiperparámetros obtenidos.
6. Comparar el rendimiento del modelo antes y después de la optimización y analizar los resultados.

Unidad VI

Objetivo

Estos proyectos permiten a los estudiantes aplicar todos los conocimientos adquiridos durante el curso, y al mismo tiempo los preparan para enfrentar problemas reales en el campo de la ciencia de datos. Además, desarrollan competencias críticas como la comunicación efectiva de resultados, la colaboración en equipo y el análisis de datos de manera ética.

Proyecto 1: Análisis y Predicción de Precios de Viviendas

- **Descripción:** Los estudiantes deberán desarrollar un modelo de predicción de precios de viviendas utilizando un conjunto de datos real (por ejemplo, el dataset de precios de viviendas de Boston o cualquier otro conjunto de datos relacionado). El proyecto integrará:
 - **Limpieza de Datos:** Manejo de valores faltantes y eliminación de duplicados.
 - **Análisis Descriptivo y Exploratorio:** Estadísticas básicas, visualización de la distribución de precios, correlaciones entre las variables.
 - **Modelado Predictivo:** Aplicación de regresión lineal, árboles de decisión o Random Forest para predecir los precios de las viviendas.
 - **Evaluación de Modelos:** Uso de métricas como MSE, R^2 y evaluación con curvas ROC si es necesario.

	Nombre del documento: Formato de		Código: TecNM-AC-PO-007-02
	Programa de Estudio de asignatura de Especialidad		Revisión: 0
	Referencia a la Norma ISO 9001:2015	8.3, 8.3.1	Página 20 de 29

- **Habilidades:** Análisis de datos, limpieza de datos, modelado de regresión, visualización, evaluación de modelos.

Proyecto 2: Análisis de Datos de Tráfico Vehicular y Predicción de Congestión

- **Descripción:** Los estudiantes trabajarán con datos de tráfico vehicular (pueden utilizar un dataset público relacionado con la congestión en ciudades) para predecir la congestión de tráfico en diferentes momentos del día. El proyecto integrará:
 - **Preprocesamiento de Datos:** Limpieza de datos faltantes, normalización y transformación de características (por ejemplo, hora, día de la semana).
 - **Visualización de Datos:** Uso de gráficos de dispersión y líneas para mostrar la relación entre variables como la hora y la cantidad de vehículos.
 - **Modelado Predictivo:** Uso de KNN y Árboles de Decisión para predecir la congestión en diferentes intervalos de tiempo.
 - **Evaluación de Modelos:** Implementación de métricas de precisión y recall, además de la curva ROC.
- **Habilidades:** Limpieza de datos, análisis de tráfico, clasificación y regresión, visualización, evaluación de modelos.

Proyecto 3: Predicción de Enfermedades Cardíacas a partir de Datos Médicos

- **Descripción:** Los estudiantes deberán desarrollar un modelo de predicción para determinar la probabilidad de que un paciente padezca una enfermedad cardíaca en función de diversas características médicas (como la edad, el colesterol, la presión arterial, el nivel de azúcar en sangre, etc.). Este proyecto integrará:
 - **Limpieza y Preprocesamiento de Datos:** Manejo de valores faltantes, tratamiento de outliers y normalización de los datos.
 - **Exploración y Análisis:** Análisis descriptivo de las variables, correlaciones y distribuciones de los datos.
 - **Modelado de Clasificación:** Implementación de algoritmos de clasificación como regresión logística, k-vecinos más cercanos (KNN), y Random Forest.
 - **Evaluación del Modelo:** Utilización de métricas como precisión, recall, F1-score y curva ROC para evaluar el rendimiento.
- **Habilidades:** Preprocesamiento de datos, análisis exploratorio, clasificación supervisada, evaluación de modelos, análisis de salud.

Proyecto 4: Sistema de Recomendación de Películas

- **Descripción:** Este proyecto tiene como objetivo crear un sistema de recomendación de películas utilizando un dataset que contenga información sobre las preferencias de los usuarios y las características de las películas (por ejemplo, el género, las valoraciones, etc.). Los estudiantes aplicarán:
 - **Preprocesamiento de Datos:** Limpieza de datos, eliminación de valores nulos y creación de matrices de usuario-item.
 - **Análisis y Visualización:** Exploración de las tendencias en las preferencias de los usuarios y visualización de las relaciones entre películas y géneros.

	Nombre del documento: Formato de		Código: TecNM-AC-PO-007-02
	Programa de Estudio de asignatura de Especialidad		Revisión: 0
	Referencia a la Norma ISO 9001:2015	8.3, 8.3.1	Página 21 de 29

- **Modelado de Recomendación:** Implementación de un sistema de recomendación basado en filtrado colaborativo o basado en contenido.
- **Evaluación:** Validación del sistema mediante la precisión de las recomendaciones y el análisis de la matriz de confusión.
- **Habilidades:** Sistemas de recomendación, análisis de preferencias de usuarios, análisis colaborativo, visualización de datos.

Proyecto 5: Análisis de Comportamiento de Clientes para Predecir la Churn (Cancelación de Servicio)

- **Descripción:** En este proyecto, los estudiantes desarrollarán un modelo predictivo para predecir si un cliente abandonará un servicio (churn) basándose en datos históricos de comportamiento del cliente. Este proyecto integrará:
 - **Limpieza y Preprocesamiento de Datos:** Manejo de valores faltantes, normalización de datos, y transformación de variables categóricas.
 - **Análisis y Visualización:** Exploración de las características que afectan el churn, visualización de las tasas de cancelación por diferentes segmentos de clientes.
 - **Modelado de Clasificación:** Uso de algoritmos como regresión logística, SVM, y árboles de decisión para clasificar a los clientes en "churn" o "no churn".
 - **Evaluación del Modelo:** Uso de precisión, recall, F1-score y curva ROC para medir la efectividad del modelo.
- **Habilidades:** Análisis de clientes, predicción de churn, clasificación supervisada, visualización de datos.

Proyecto 6: Predicción de la Demanda de Productos en una Tienda en Línea

- **Descripción:** Este proyecto tiene como objetivo predecir la demanda futura de productos en una tienda en línea usando datos históricos de ventas. Los estudiantes deberán:
 - **Limpieza y Preprocesamiento de Datos:** Tratar valores faltantes, transformar variables de fechas, y normalizar los datos de ventas.
 - **Análisis y Visualización:** Analizar las tendencias de ventas, los productos más vendidos y las estaciones del año que afectan la demanda.
 - **Modelado Predictivo:** Implementación de modelos de series temporales como ARIMA, o modelos de regresión para predecir la demanda futura de productos.
 - **Evaluación de Modelos:** Utilización de métricas como MAPE (Mean Absolute Percentage Error) y RMSE para evaluar las predicciones de la demanda.
- **Habilidades:** Análisis de series temporales, predicción de demanda, modelado predictivo, visualización de tendencias.

Requisitos Comunes de los Proyectos:

- **Uso de Bibliotecas de Python:** Los estudiantes utilizarán bibliotecas como Pandas para manipulación de datos, Matplotlib y Seaborn para visualización,

	Nombre del documento: Formato de		Código: TecNM-AC-PO-007-02
	Programa de Estudio de asignatura de Especialidad		Revisión: 0
	Referencia a la Norma ISO 9001:2015	8.3, 8.3.1	Página 22 de 29

Scikit-learn para modelado de aprendizaje automático, y Statsmodels para análisis de series temporales.

- **Documentación y Presentación:** Los estudiantes deberán documentar el proceso de cada proyecto, incluyendo la preparación de datos, análisis, modelado y evaluación de resultados. Se requiere una presentación final de los resultados, con reportes claros y visualizaciones interactivas.
- **Trabajo en Equipo:** Los proyectos pueden desarrollarse en equipos para fomentar la colaboración y la discusión técnica entre los estudiantes.

	Nombre del documento: Formato de		Código: TecNM-AC-PO-007-02
	Programa de Estudio de asignatura de Especialidad		Revisión: 0
	Referencia a la Norma ISO 9001:2015	8.3, 8.3.1	Página 23 de 29

9. Proyecto de asignatura

El objetivo del proyecto que planteé el docente que imparta esta asignatura, es demostrar el desarrollo y alcance de la(s) competencia(s) de la asignatura, considerando las siguientes fases:

- **Fundamentación:** marco referencial (teórico, conceptual, contextual, legal) en el cual se fundamenta el proyecto de acuerdo con un diagnóstico realizado, mismo que permite a los estudiantes lograr la comprensión de la realidad o situación objeto de estudio para definir un proceso de intervención o hacer el diseño de un modelo.
- **Planeación:** con base en el diagnóstico en esta fase se realiza el diseño del proyecto por parte de los estudiantes con asesoría del docente; implica planificar un proceso: de intervención empresarial, social o comunitaria, el diseño de un modelo, entre otros, según el tipo de proyecto, las actividades a realizar los recursos requeridos y el cronograma de trabajo.
- **Ejecución:** consiste en el desarrollo de la planeación del proyecto realizada por parte de los estudiantes con asesoría del docente, es decir en la intervención (social, empresarial), o construcción del modelo propuesto según el tipo de proyecto, es la fase de mayor duración que implica el desempeño de las competencias genéricas y específicas a desarrollar.
- **Evaluación:** es la fase final que aplica un juicio de valor en el contexto laboral-profesión, social e investigativo, ésta se debe realizar a través del reconocimiento de logros y aspectos a mejorar se estará promoviendo el concepto de “evaluación para la mejora continua”, la metacognición, el desarrollo del pensamiento crítico y reflexivo en los estudiantes.

Un aspecto innovador e importante en el proceso de formación de los estudiantes es el proyecto de asignatura.

Se genera a partir de la definición de un problema del contexto a resolver y que esté directamente relacionado con la(s) competencia(s) a desarrollar en la asignatura.

- Fundamentación.
- Planeación.
- Ejecución.
- Evaluación.

	Nombre del documento: Formato de		Código: TecNM-AC-PO-007-02
	Programa de Estudio de asignatura de Especialidad		Revisión: O
	Referencia a la Norma ISO 9001:2015	8.3, 8.3.1	Página 24 de 29

10. Evaluación por competencias

Son las técnicas, instrumentos y herramientas sugeridas para constatar los desempeños académicos de las actividades de aprendizaje.

Desempeño	Nivel de Desempeño	Indicadores del alcance	Valoración numérica
COMPETENCIA ALCANZADA	Excelente	Cumple al menos cinco de los siguientes indicadores a) Se adapta a situaciones y contextos complejos. Puede trabajar en equipo, reflejar sus conocimientos en la interpretación de la realidad. Inferir comportamientos o consecuencias de los fenómenos o problemas en estudio. Incluir más variables en dichos casos de estudio. b) Hace aportaciones a las actividades académicas desarrolladas. Pregunta integrando conocimientos de otras asignaturas o de casos anteriores de la misma asignatura. Presenta otros puntos de vista que complementan al presentado en la clase. Presenta fuentes de información adicionales (Internet, documentales), usa más bibliografía, consulta fuentes en un segundo idioma, etc. c) Propone y/o explica soluciones o procedimientos no vistos en clase (creatividad). Ante problemas o casos de estudio propone perspectivas diferentes, para abordarlos y sustentarlos correctamente. Aplica procedimientos aprendidos en otra asignatura o contexto para el problema que se está resolviendo. d) Introduce recursos y experiencias que promueven un pensamiento crítico; (por ejemplo el uso de las tecnologías de la información estableciendo previamente un criterio). Ante temas de una asignatura, introduce cuestionamientos de tipo ético, ecológico, histórico, político, económico, etc.; que deben tomarse en cuenta para comprender mejor, o a futuro dicho tema. Se apoya en foros, autores, bibliografía, documentales, etc. para sustentar su punto de vista. e) Incorpora conocimientos y actividades interdisciplinarias en su aprendizaje. En el desarrollo de los temas de la asignatura, incorpora conocimientos y actividades desarrollados en otras asignaturas para lograr la competencia. f) Realiza su trabajo de manera autónoma y autorregulada. Es capaz de organizar su tiempo y trabajar sin necesidad de una supervisión estrecha y/o coercitiva. Aprovecha la planeación de la asignatura presentada por el (la) profesor(a) (instrumentación didáctica) para presentar propuestas de mejora de la temática vista durante el curso. Realiza actividades de investigación para participar activamente durante el curso.	95-100
	Notable	Cumple cuatro de los indicadores definidos en desempeño excelente.	85-94
	Bueno	Cumple tres de los indicadores definidos en el desempeño excelente.	75-84
	Suficiente	Cumple dos de los indicadores definidos en el desempeño excelente.	70-74
COMPETENCIA NO ALCANZADA	Desempeño insuficiente	No se cumple con el 100% de evidencias conceptuales, procedimentales y actitudinales de los indicadores definidos en el desempeño excelente.	NA (no alcanzada)

	Nombre del documento: Formato de		Código: TecNM-AC-PO-007-02
	Programa de Estudio de asignatura de Especialidad		Revisión: 0
	Referencia a la Norma ISO 9001:2015	8.3, 8.3.1	Página 25 de 29

11. Fuentes de información

Referencias bibliográficas

- Grus, J. (2023). Data science from scratch: First principles with Python (3rd ed.). O'Reilly Media.
- McKinney, W. (2022). Python for data analysis: Data wrangling with pandas, NumPy, and Jupyter (3rd ed.). O'Reilly Media.
- Mueller, J. P., & Massaron, L. (2023). Python for data science for dummies (3rd ed.). John Wiley & Sons.
- Géron, A. (2022). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems (3rd ed.). O'Reilly Media.
- VanderPlas, J. (2023). Python data science handbook: Essential tools for working with data (2nd ed.). O'Reilly Media.
- Cohn, T. (2023). Data visualization with Python: A hands-on approach using Matplotlib and Seaborn. Wiley.
- Graves, T. (2022). Effective data visualization for data science: Principles, techniques, and applications with Python. Springer.
- McKinney, W. (2022). Python for data analysis: Data wrangling with Pandas, NumPy, and Jupyter (3rd ed.). O'Reilly Media.
- Nargundkar, A., & Burkholder, E. (2023). Seaborn essentials: Practical visualizations for data science. Packt Publishing.
- VanderPlas, J. (2021). Python data science handbook: Essential tools for working with data. O'Reilly Media.
- Waskom, M. L. (2021). Seaborn: Statistical data visualization. Journal of Open Source Software, 6(60), 3021. <https://doi.org/10.21105/joss.03021>
- Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (3rd ed.). O'Reilly Media.
- Raschka, S., & Mirjalili, V. (2020). *Python machine learning: Machine learning and deep learning with Python, Scikit-Learn, and TensorFlow 2* (3rd ed.). Packt Publishing.
- VanderPlas, J. (2022). *Python data science handbook: Essential tools for working with data*. O'Reilly Media.
- Müller, A. C., & Guido, S. (2021). *Introduction to machine learning with Python: A guide for data scientists*. O'Reilly Media.
- Grus, J. (2023). *Data science from scratch: First principles with Python* (2nd ed.). O'Reilly Media.
- McKinney, W. (2022). *Python for data analysis: Data wrangling with Pandas, NumPy, and Jupyter* (3rd ed.). O'Reilly Media.
- Géron, A. (2022). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems (3rd ed.). O'Reilly Media.

	Nombre del documento: Formato de	Código: TecNM-AC-PO-007-02
	Programa de Estudio de asignatura de Especialidad	Revisión: 0
	Referencia a la Norma ISO 9001:2015 8.3, 8.3.1	Página 26 de 29

- Raschka, S., & Mirjalili, V. (2020). Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2 (3rd ed.). Packt Publishing.
- Albon, C. (2020). Machine Learning with Python Cookbook: Practical Solutions from Preprocessing to Deep Learning. O'Reilly Media.
- Müller, A. C., & Guido, S. (2020). Introduction to Machine Learning with Python: A Guide for Data Scientists. O'Reilly Media.
- Aggarwal, C. C. (2022). Machine Learning for Text (2nd ed.). Springer.

Artículos Académicos

- Chen, X., & Zhang, Y. (2023). Teaching data science with Python: A comprehensive review of tools and pedagogical approaches. Journal of Computing in Higher Education, 35(2), 245-267.
- Kumar, R., & Singh, M. (2022). Análisis comparativo de bibliotecas Python para ciencia de datos: NumPy, Pandas y Matplotlib. International Journal of Data Science and Analytics, 14(3), 156-178.
- López-González, E., & Martínez-Castro, N. (2021). Metodologías efectivas para la enseñanza de Python en ciencias de datos. Revista Iberoamericana de Educación en Tecnología, 82(1), 23-45.
- Ferreira, F., Oliveira, M., & Pedro, L. (2023). Data visualization in decision-making: How data analytics and visualization tools are reshaping business intelligence. Journal of Business Analytics, 5(2), 215–229. <https://doi.org/10.1080/2573234X.2023.2101234>
- Li, H., & Li, Y. (2022). Enhancing the effectiveness of data visualizations with Matplotlib and Seaborn for real-time data analysis in Python. International Journal of Data Science and Analytics, 10(3), 145-160. <https://doi.org/10.1007/s41060-021-00235-9>
- Shah, A., & Zhang, T. (2021). Evaluating the impact of visual aesthetics in data visualization: Insights from using Matplotlib and Seaborn in exploratory data analysis. Journal of Computational and Graphical Statistics, 30(1), 112-126. <https://doi.org/10.1080/10618600.2021.1892887>
- Zhou, W., & Song, F. (2021). Statistical data visualization in Python: A comparative study of Seaborn and other Python libraries. Journal of Open Research Software, 9, 25. <https://doi.org/10.5334/jors.351>
- Nguyen, K. P., & Cao, M. (2020). Enhancing analytical comprehension through data visualization: An application of Seaborn and Matplotlib in higher education data sets. Computers & Education, 155, 103919. <https://doi.org/10.1016/j.compedu.2020.103919>
- Benato, F., & Pandini, C. (2023). Advances in machine learning data preparation: Assessing pre-processing strategies with Python's Scikit-Learn. Journal of Data Science and Analytics, 15(2), 109-122. <https://doi.org/10.1007/s41060-023-00345-8>
- Huyen, C., & Jensen, D. (2022). Comparison of supervised and unsupervised learning techniques in healthcare data analysis. Journal of Healthcare Informatics Research, 6(1), 245-261. <https://doi.org/10.1007/s41666-022-00126-9>

	Nombre del documento: Formato de		Código: TecNM-AC-PO-007-02
	Programa de Estudio de asignatura de Especialidad		Revisión: 0
	Referencia a la Norma ISO 9001:2015	8.3, 8.3.1	Página 27 de 29

- Chen, X., & Zhao, Y. (2021). Data splitting and model validation techniques in machine learning applications: A review. *ACM Computing Surveys*, 53(6), 1-33. <https://doi.org/10.1145/3428295>
- Larose, D. T., & Larose, C. D. (2020). Using Scikit-Learn for reproducible machine learning research in social sciences. *International Journal of Data Science*, 5(3), 140-157. <https://doi.org/10.1016/j.ijds.2020.05.013>
- Wang, J., & Blei, D. (2020). Supervised vs. unsupervised learning for feature selection: Practical applications in computational biology. *BMC Bioinformatics*, 21(1), 92. <https://doi.org/10.1186/s12859-020-3457-6>
- Chen, C., Li, O., Tao, D., Barnett, A., Rudin, C., & Su, J. K. (2021). This looks like that: Deep learning for interpretable image recognition. *Journal of Machine Learning Research*, 22(1), 1-33.
- Masubuchi, T., Kikkawa, T., & Nakamura, K. (2022). Performance evaluation of machine learning classifiers with a novel AUC-based metric. *Pattern Recognition Letters*, 154, 45-50.
- Zhang, W., Zhang, C., & He, Y. (2023). Hyperparameter optimization in machine learning: A comprehensive review. *IEEE Access*, 11, 10140-10152.
- Liu, Z., Yang, X., Sun, J., & Zhu, S. (2020). Comparative performance analysis of classification algorithms on imbalanced credit scoring data. *Expert Systems with Applications*, 159, 113-119.
- Zhao, Z., & Wang, Y. (2023). An empirical study on ROC-AUC optimization in deep neural networks. *Neurocomputing*, 518, 32-42.

Recursos en Línea

- Python Software Foundation. (2024). Python documentation. <https://docs.python.org/3/>
- Project Jupyter. (2023). Jupyter documentation. <https://jupyter.org/documentation>
- NumPy. (2024). NumPy documentation. <https://numpy.org/doc/>
- pandas. (2024). pandas documentation. <https://pandas.pydata.org/docs/>
- Matplotlib. (2023). Matplotlib documentation. <https://matplotlib.org/stable/>
- Matplotlib Documentation: Página oficial de documentación de Matplotlib, con tutoriales y ejemplos de uso para gráficos básicos y avanzados. URL: <https://matplotlib.org/stable/contents.html>
- Seaborn Documentation: Documentación oficial de Seaborn, que incluye guías para crear gráficos estadísticos avanzados, como heatmaps, pairplots y boxplots. URL: <https://seaborn.pydata.org/>
- Towards Data Science - Effective Data Visualization with Matplotlib and Seaborn (2022): Artículo en Towards Data Science sobre técnicas avanzadas de visualización en Python, con Matplotlib y Seaborn, y consejos para mejorar la efectividad visual. URL: <https://towardsdatascience.com/effective-data-visualization-with-matplotlib-and-seaborn-2022>
- Python Graph Gallery: Sitio web con una gran variedad de ejemplos y tutoriales para crear gráficos en Python, con un enfoque en Matplotlib y Seaborn, ideal para aprender visualización a través de ejemplos interactivos.

	Nombre del documento: Formato de		Código: TecNM-AC-PO-007-02
	Programa de Estudio de asignatura de Especialidad		Revisión: 0
	Referencia a la Norma ISO 9001:2015	8.3, 8.3.1	Página 28 de 29

URL: https://www.python-graph-gallery.com/ • DataCamp - Data Visualization in Python: Curso en línea de DataCamp sobre visualización de datos con Python, que cubre tanto gráficos básicos con Matplotlib como gráficos avanzados con Seaborn. URL: https://www.datacamp.com/courses/data-visualization-in-python • Scikit-Learn Documentation: Página oficial de documentación de Scikit-Learn, que ofrece guías, ejemplos y tutoriales completos para implementar modelos de aprendizaje automático en Python. URL: https://scikit-learn.org/stable/ • KDnuggets - Machine Learning Basics: Supervised vs. Unsupervised Learning (2021): Artículo en KDnuggets que explica los conceptos básicos de aprendizaje supervisado y no supervisado, con ejemplos y aplicaciones. URL: https://www.kdnuggets.com/2021/02/machine-learning-basics-supervised-unsupervised.html • Towards Data Science - Data Preparation for Machine Learning (2022): Artículo en Towards Data Science sobre técnicas de preparación de datos para aprendizaje automático, como la división en conjuntos de entrenamiento y prueba, y el procesamiento de datos faltantes. URL: https://towardsdatascience.com/data-preparation-for-machine-learning-2022 • Analytics Vidhya - Scikit-Learn Tutorial for Machine Learning in Python (2023): Tutorial en Analytics Vidhya sobre cómo usar Scikit-Learn para construir y evaluar modelos de aprendizaje automático en Python. URL: https://www.analyticsvidhya.com/blog/2023/03/scikit-learn-tutorial-for-machine-learning-in-python/ • DataCamp - Machine Learning with Scikit-Learn: Curso en línea de DataCamp que cubre los fundamentos de Scikit-Learn y su aplicación en aprendizaje supervisado y no supervisado. Ideal para aprender de manera práctica. URL: https://www.datacamp.com/courses/machine-learning-with-scikit-learn • Scikit-learn Documentation La documentación oficial de Scikit-learn es una excelente fuente para aprender sobre clasificación, ajuste de modelos, validación cruzada y métricas de evaluación como AUC y precisión. URL: https://scikit-learn.org/stable/ • Kaggle: Machine Learning Competitions and Datasets Kaggle es una plataforma que ofrece competencias y conjuntos de datos en los que los estudiantes pueden aplicar algoritmos de clasificación y evaluar modelos. Incluye tutoriales sobre cómo usar regresión logística, KNN, y Random Forest. URL: https://www.kaggle.com/ • Towards Data Science: Machine Learning Tutorials Towards Data Science es una plataforma que alberga artículos educativos sobre machine learning. Ofrece guías detalladas sobre el uso de modelos de clasificación, la optimización de hiperparámetros y la interpretación de las métricas de rendimiento. URL: https://towardsdatascience.com/
--

	Nombre del documento: Formato de		Código: TecNM-AC-PO-007-02
	Programa de Estudio de asignatura de Especialidad		Revisión: 0
	Referencia a la Norma ISO 9001:2015	8.3, 8.3.1	Página 29 de 29

- **DataCamp: Supervised Learning in Python**
DataCamp ofrece cursos interactivos que cubren conceptos de machine learning supervisado, incluyendo regresión logística, KNN y optimización de modelos. Los estudiantes pueden practicar directamente en la plataforma.
URL: <https://www.datacamp.com/courses/supervised-learning-in-python>
- **Coursera: Machine Learning by Andrew Ng (Stanford University)**
Este es uno de los cursos más populares de machine learning, ofrecido por Andrew Ng, que cubre teoría y práctica sobre clasificación, regresión y métricas de rendimiento como precisión y AUC.
URL: <https://www.coursera.org/learn/machine-learning>
- **DeepLearning.AI: Machine Learning Specialization**
Esta especialización de Coursera abarca temas avanzados como optimización de hiperparámetros y técnicas de validación cruzada, fundamentales para ajustar modelos de clasificación.
URL: <https://www.coursera.org/specializations/machine-learning-introduction>