Licenciatura en Negocios Digitales

Materia: Análisis de Datos

Profesor: Vicente Humberto Monteverde

2025

CARRERA: Licenciatura en Negocios Digitales

ASIGNATURA: Análisis de Datos

COMPOSICIÓN DEL EQUIPO DOCENTE: Vicente Humberto Monteverde

1. FUNDAMENTACIÓN DE LA MATERIA:

Así como un árbol se nutre de sus raíces para crecer y florecer, el análisis de datos se fundamenta en la necesidad de extraer conocimiento valioso a partir de la información disponible. En un mundo cada vez más impulsado por los datos, esta materia proporciona las bases conceptuales y las herramientas metodológicas esenciales para transformar datos brutos en *insights* significativos que permitan la toma de decisiones informadas en diversos campos del saber y la práctica profesional. Al comprender los fundamentos del análisis de datos, los estudiantes podrán desarrollar una visión crítica sobre la información, identificar patrones, tendencias y relaciones, y contribuir a la solución de problemas complejos.

2. OBJETIVOS:

Al finalizar la materia, se espera que el estudiante sea capaz de:

- **Comprender los conceptos fundamentales** del análisis de datos, incluyendo los tipos de datos, las etapas del proceso analítico y la importancia de la calidad de los datos (las raíces del conocimiento).
- **Aplicar técnicas descriptivas básicas** para resumir y visualizar conjuntos de datos, obteniendo una visión general de sus principales características (las primeras hojas que exploran el entorno).
- **Utilizar herramientas de software estadístico** para realizar análisis exploratorios y descriptivos de datos (las ramas que permiten un mayor alcance).
- **Interpretar los resultados de los análisis** y comunicarlos de manera efectiva, utilizando diferentes formatos y visualizaciones (los frutos del árbol, el conocimiento compartido).
- **Desarrollar un pensamiento crítico** sobre el uso y la interpretación de los datos en diferentes contextos (la solidez del tronco que sostiene el conocimiento).
- **Identificar los principios éticos** asociados al manejo y análisis de datos (el cuidado del entorno donde crece el árbol).
- **Preprocesamiento de los datos:** Aprender procesos de ETL, limpieza, imputación y análisis de los datos.
- **Aprendizaje Automático:** Introducción al uso de los Algoritmos, tipos de algoritmos, uso básico de ellos.

3. CONTENIDOS MÍNIMOS:

- Introducción al Análisis de Datos: Conceptos básicos, tipos de datos, el proceso de análisis de datos.
- Organización y Presentación de Datos: Tablas de frecuencia, distribuciones, gráficos estadísticos.
- Estadística Descriptiva : Medidas de tendencia central, dispersión, forma. Tablas de contingencia, medidas de asociación, gráficos de dispersión.
- Variables Aleatorias y Distribuciones de Probabilidad: Tipos de variables, distribuciones discretas y continuas fundamentales.
- Preprocesamiento de Datos: Procesos ETL, limpieza e imputación de datos.
- Aprendizaje Automático: Introducción a los Algoritmos, aplicación básica de machine learning.

4. UNIDADES TEMÁTICAS Y BIBLIOGRAFÍA:

Modulo 1:

- ¿Qué es el análisis de datos? Importancia y aplicaciones.
- Tipos de datos: cualitativos y cuantitativos, niveles de medición.
- El proceso de análisis de datos: desde la recolección hasta la interpretación.
- Calidad de los datos: fuentes de error y estrategias de limpieza inicial.
- Introducción a software estadístico (Python, Pandas y Numpy).
- **Bibliografía:**
 - **Para una introducción general y tipos de datos:**
 - **"Introducción a la Estadística" Barbara Illosky, Susan Dean-(2022)- Open Stax:** Los primeros capítulos cubren la naturaleza de los datos, la variabilidad y la organización de datos, lo cual es fundamental para este módulo.
 - **"Estadística para Administración y Economía" de David R. Anderson, Dennis J. Sweeney y Thomas A. Williams-(2008)-Cengage Learning:** Los capítulos iniciales proporcionan una buena base sobre los tipos de datos y el proceso de análisis.
 -
 - **Para el proceso de análisis de datos:**
 - **Ciencia de datos desde cero- Principios básicos con Python"-de Joel Grus-(2019)-O'Reilly:** El Capítulo 1 introduce el concepto de ciencia de datos y el flujo de trabajo, aunque el enfoque principal es Python.
 - Reproducible Data Science with Python- [Valentin Danchev-on line-https://valdanchev.github.io/reproducible-data-science-python/intro.html](https://valdanchev.github.io/reproducible-data-science-python/intro.html)
 - **Para calidad de datos:**

- **"Practical Data Quality" de Robert Hawker-(2023)-Packt** Ofrece una visión profunda sobre la calidad de los datos, sus dimensiones y estrategias para mejorarla. (Puede ser más avanzado, pero útil como referencia).
- Artículos y recursos en línea sobre "data cleaning best practices".
- 10 tipos de errores comunes en Python y cómo resolverlos-Collins Kariuk.
- **Para introducción a Python con Pandas:**
 - **"Python for Data Analysis" de Wes McKinney-(2022)-O'Reilly:** Los primeros capítulos son una excelente introducción a Pandas, la manipulación de datos y las estructuras de datos fundamentales (Series y DataFrames).
 - **"Pandas cookbook" Mathew Harrison-(2024)-Packt:** Una guía práctica y accesible para aprender Pandas desde cero.

Modulo 2

- Tablas de frecuencia y distribuciones de frecuencia.
- Representaciones gráficas para datos cualitativos: gráficos de barras, circulares.
- Representaciones gráficas para datos cuantitativos: histogramas, polígonos de frecuencia, ojivas, diagramas de tallo y hojas, boxplots.
- Principios de visualización efectiva de datos.
- Practica con numpy.

Bibliografía:

🔗 Para tablas y distribuciones de frecuencia:

- **"Mathematical & Statistic"-Wackerly,Mendehall y Scheaffer-(2008)-Cengage Learning:** Este libro cubrirá estos temas en detalle en los capítulos dedicados a la estadística descriptiva.
- **Para representaciones gráficas:**
 - **"The Functional Art: An Introduction to Information Graphics and Visualization" de Alberto Cairo:** Un libro muy accesible sobre los principios de una visualización de datos efectiva.
 - **"Introduction to Data Visualization & Stortelling"-Jose Berengues y Marybeth Sandell-(2018)-Typesetting:** Se enfoca en cómo comunicar eficazmente insights a través de visualizaciones.
- **Para práctica con NumPy:**
 - **"Numpy-Beginners guide"-Ivan Idris-(2015)-Packt:** Buen manual de Numpy.
 - **"Python for Data Analysis" de Wes McKinney:** El Capítulo 4 introduce NumPy, que es la base para Pandas y muchas otras bibliotecas de ciencia de datos en Python.
 - **La documentación oficial de NumPy:** Es un recurso excelente y detallado.

Modulo 3

- Medidas de tendencia central: media, mediana, moda. Propiedades y aplicaciones.
- Medidas de dispersión: rango, varianza, desviación estándar, coeficiente de variación. Interpretación y comparación.
- Medidas de posición: cuartiles, deciles, percentiles.
- Medidas de forma: asimetría y curtosis.
- Detección de valores atípicos (outliers).
- Tablas de contingencia y distribuciones conjuntas.
- Medidas de asociación para datos cualitativos: chi-cuadrado, coeficiente de contingencia.
- Diagramas de dispersión y correlación lineal.
- Covarianza y coeficiente de correlación de Pearson: cálculo e interpretación.
- Aplicación de bibliotecas gráficas, Matplotlib, seaborn, plotly y plotly express.

Bibliografía:

- **Para medidas de tendencia central, dispersión y forma:**
- Los libros de estadística introductoria (Moore et al., Anderson et al.) tendrán capítulos dedicados a la estadística descriptiva univariada.
- **Para tablas de contingencia y medidas de asociación:**
- Estos temas también se cubren en los capítulos de estadística descriptiva bivariada de los libros de estadística introductoria.
- **Para diagramas de dispersión y correlación:**
- Ídem, dentro de la estadística descriptiva bivariada.
- **Para aplicación de bibliotecas gráficas (Matplotlib, Seaborn, Plotly y Plotly Express):**
- **"Python Data Science Handbook"-Jake VanderPlas-(2017)-O'Reilly:** Los Capítulos 4 y 5 cubren Matplotlib y Seaborn de manera exhaustiva.
- **La documentación oficial de Matplotlib, Seaborn, Plotly y Plotly Express:** Son recursos muy valiosos con ejemplos y explicaciones detalladas.
- **"Data Visualization With Python: Create an Impact With Meaningful Data Insights Using Interactive and Engaging Visuals"-Mario Dobler y Tim Grobmann-(2019)-Packt:**
- Realmente una gran guía de Visualización con descripciones de estadística descriptiva en forma grafica.
- Tutoriales en línea (como los de Kaggle, Medium, y las propias páginas de las bibliotecas) suelen ser muy útiles para aprender a usar estas herramientas de visualización.

Modulo 4

- Procesos ETL (Extracción, Transformación, Carga): Introducción y conceptos clave.
- Limpieza de datos:
 - Manejo de valores faltantes (imputación, eliminación).
 - Identificación y tratamiento de ruido y errores.
 - Detección y tratamiento de valores atípicos (revisión más profunda).
- Transformación de datos:
 - Escalado y normalización.
 - Discretización y categorización.
 - Creación de nuevas variables (ingeniería de características básica).

Bibliografía:

- **Para procesos ETL y limpieza de datos:**
- **“Practical Data Cleaning”-Lee Baker-(2019)-Chi Squared:** Muy buen manual de Data Cleaning, desde recolección, limpieza, imputación e integración de datos.
- **Para manejo de valores faltantes, ruido, errores y outliers:**
- Los libros mencionados anteriormente suelen tener secciones dedicadas a estos temas dentro del preprocesamiento.
- **Para transformación de datos (escalado, normalización, discretización, ingeniería de características básica):**
- Ídem, dentro de los capítulos de preprocesamiento y, en algunos casos, en libros más avanzados de machine learning en las secciones de preparación de datos.

Modulo 5

- Introducción al Aprendizaje Automático: Conceptos básicos, tipos de aprendizaje (supervisado, no supervisado, reforzamiento).
- Algoritmos de Aprendizaje Supervisado (Introducción y Aplicación Básica):
 - Regresión Lineal Simple y Múltiple (revisión desde una perspectiva de modelado).
 - Clasificación: Regresión Logística (introducción), Árboles de Decisión (conceptos básicos).

Bibliografía:

- **Para conceptos básicos y tipos de aprendizaje:**
- **“Machine Learning with Python”-Ethan Williams-(2019)-Ethan Williams:** El libro probablemente cubre los fundamentos del aprendizaje automático y cómo implementar varios algoritmos utilizando el lenguaje de programación Python.
- **“Machine Learning for Beginners”- Ethem Mining-(2019):** El libro "Machine Learning for Beginners" de probablemente esté diseñado como una **introducción accesible y amigable** al campo del aprendizaje automático. El libro también toca temas importantes como la **preparación de datos**, la **evaluación de modelos** (métricas de rendimiento), y quizás una breve

introducción a la **implementación** de modelos.

- **Modulo 6**

- Evaluación de Modelos (Introducción): Métricas básicas para clasificación y regresión.
- Aplicaciones básicas de Machine Learning en problemas sencillos.

Bibliografía:

- **Para evaluación de modelos:**
 - Los libros de introducción al aprendizaje automático (James et al., Géron) tienen capítulos dedicados a la evaluación de modelos de clasificación y regresión, incluyendo métricas como precisión, recall, F1-score, RMSE, etc.
- **Para aplicaciones básicas de Machine Learning:**
 - Buscar ejemplos y casos de estudio sencillos en línea o en los últimos capítulos de los libros introductorios de machine learning puede ser muy útil. Plataformas como Kaggle tienen notebooks introductorios con aplicaciones prácticas.
- **Para la biblioteca de Machine Learning:**
- **“Machine Learning with Python”-Ethan Williams-(2019)-Ethan Williams:** El libro probablemente cubre los fundamentos del aprendizaje automático y cómo implementar varios algoritmos utilizando el lenguaje de programación Python.
- **“Simplifying Machine Learning with Pycaret”-Giannis Tolios-(2023)-Learnpub:** El libro busca simplificar y acelerar el proceso de aprendizaje automático utilizando las capacidades de PyCaret, convirtiéndolo en un recurso valioso para personas que buscan un enfoque práctico y eficiente para las tareas de aprendizaje automático en Python.

Recomendaciones Generales:

- **Prioriza libros que utilicen Python, Pandas y Numpy:** directamente para los módulos donde se menciona explícitamente.
- **No dudes en complementar con recursos en línea:** La documentación oficial de las bibliotecas (Pandas, NumPy, Matplotlib, Seaborn, Scikit-Learn, Pycaret), tutoriales en plataformas como Kaggle, Coursera, edX y Medium pueden ser muy valiosos.

Links:

<https://www.python.org/>

<https://pandas.pydata.org/>

<https://numpy.org/>

<https://matplotlib.org/>

<https://plotly.com/>

<https://plotly.com/python/plotly-express/>

<https://seaborn.pydata.org/>

<https://altair.com/>

<https://scipy.org/>

<https://scikit-learn.org/stable/>

<https://pycaret.org/>

5. METODOLOGÍA Y SISTEMA DE EVALUACIÓN:

La metodología de enseñanza-aprendizaje se centrará en un enfoque **activo y participativo**, buscando que el estudiante construya su propio conocimiento a través de:

- **Clases Teórico-Prácticas:** Se abordarán los conceptos fundamentales del análisis de datos mediante la exposición teórica, complementada con ejemplos concretos y la exploración de sus diversas aplicaciones. La ejercitación práctica y el estudio de casos reales permitirán a los estudiantes aplicar activamente las técnicas aprendidas.
- **Desarrollo de Habilidades con Cuadernos Colab:** Se potenciará el manejo de herramientas informáticas esenciales para el análisis de datos a través del trabajo práctico con cuadernos Colab. Estas sesiones interactivas facilitarán la implementación de los métodos estudiados en un entorno computacional real.
- **Profundización y Debate Bibliográfico:** La lectura y discusión de bibliografía especializada fomentarán una comprensión más profunda de los temas y estimularán el desarrollo de un pensamiento crítico y reflexivo sobre las diferentes perspectivas en el campo del análisis de datos.
- **Aprendizaje Colaborativo y Foro de Consultas:** Los foros son de debate sencillo y suscripción forzosa. Se promoverá la sinergia grupal a través de trabajos colaborativos en la resolución de problemas de análisis de datos, fomentando tanto la autonomía individual como el intercambio de conocimientos. El foro virtual actuará como un espacio dinámico para que los estudiantes formulen consultas y se brinden apoyo mutuo.

- **Práctica Continua con Cuadernos Integrados:** Los estudiantes dispondrán de cuadernos Colab que integran tanto la teoría como ejercicios prácticos. Esta metodología permitirá una práctica constante y autónoma, con la posibilidad de resolver los problemas planteados y acceder a explicaciones detalladas dentro del mismo entorno de aprendizaje.

Sistema de Evaluación:

La evaluación será **continua y formativa**, buscando valorar el proceso de aprendizaje del estudiante y su capacidad para aplicar los conocimientos adquiridos. Se considerarán los siguientes aspectos:

- **Participación activa en clase:** Contribuciones significativas a las discusiones y resolución de ejercicios.
- **Evaluaciones por modulo:** Verificación de la comprensión de los conceptos y la aplicación de las técnicas.
- **Examen parcial y recuperatorio :** Aplicación de los conocimientos adquiridos a un problema de análisis de datos concreto, incluyendo la presentación de resultados e interpretaciones.
- **Examen final (integrador):** Evaluación global de los contenidos de la materia,.