



Datos Generales

1. Nombre de la Asignatura	2. Nivel de formación	3. Clave de la Asignatura	
Programación II	Maestría		
4. Prerrequisitos Ninguno	5. Área de Formación Básica común obligatoria	6. Departamento Sistemas de información	
7. Modalidad Presencial	8. Tipo de Asignatura: Curso-Taller		
9. Carga Horaria 60 horas			
Teoría: 40 horas	Práctica: 20 horas	Total: 60	Créditos: 7
10. Trayectoria de la asignatura			

Contenido del Programa

<p>11. Presentación</p> <p>El maestrante en Ciencia de datos, debe conocer los fundamentos de aprendizaje de maquina y el como implementarlos en un lenguaje de programacion (Python), asi como de las diversas bibliotecas o frameworks disponibles para hacerlo de una manera mas facil</p> <p>Con este curso el futuro maestro en Ciencia de datos, podrá convertir los datos en informacion y conocimiento.</p>
<p>12.- Objetivos del programa</p> <p>Objetivo General</p> <p>Obtener habilidades de aprendizaje de maquina, la capacidad de discernir que algoritmo es mejor para cada ocasion e implementarlo usando el lenguaje de programacion Python con bibliotecas especializadas en el area</p>
<p>13.-Contenido</p> <p>Contenido temático</p> <p>Unidad 1 Ciencia de datos con Python: NumPy, Pandas, rpy2 y Anaconda Unidad 2 Fundamentos de aprendizaje de maquina Unidad 3 Algoritmos de aprendizaje supervisado Unidad 4 Algoritmos de aprendizaje no supervisado Unidad 5 Análisis de redes sociales y minería de texto</p> <p>Contenido desarrollado</p> <p>Unidad 1 Bibliotecas de ciencia de datos para Python</p>

Objetivo particular de la unidad: Conocer acerca de bibliotecas de relacionadas con el análisis de datos en Python para crear software de manera más rápida y eficiente
Desarrollo:

1. Bibliotecas de ciencia de datos para Python
 - 1.1. Ciencia de datos
 - 1.2. Biblioteca NumPy
 - 1.3. Biblioteca Pandas
 - 1.4. Biblioteca rpy2
 - 1.5. Biblioteca Scikit-learn
 - 1.6. Distribucion Anaconda
 - 1.7. Orange

Unidad 2 Fundamentos de Machine Learning

Objetivo particular de la unidad: Explorar conceptos básicos de machine learning para identificar los diferentes tipos de algoritmos y seleccionar el adecuado para resolver un caso particular

2. Fundamentos de Machine Learning
 - 2.1. Machine learning
 - 2.1.1. Muestreo
 - 2.1.1.1. Oversampling
 - 2.1.1.2. Undersampling
 - 2.1.2. Modelado predictivo vs modelado explicativo
 - 2.1.3. Algoritmos supervisados
 - 2.1.3.1. Algoritmos predictivos
 - 2.1.3.2. Algoritmos clasificatorios
 - 2.1.4. Algoritmos no supervisados
 - 2.1.5. Conjuntos de: entrenamiento, validacion y prueba
 - 2.1.6. Visualización de datos
 - 2.1.6.1. Biblioteca Matplotlib
 - 2.1.6.2. Orange
 - 2.1.6.3. Tableau
 - 2.1.7. Reducción de dimensiones con análisis de componente principal
 - 2.1.8. Evaluación del rendimiento predictivo

Unidad 3 Algoritmos de aprendizaje supervisado

Objetivo particular de la unidad: Conocer como funcionan los algoritmos predictivos y clasificacion para elegir el mejor según el problema a resolver

Desarrollo

3. Algoritmos de aprendizaje supervisado
 - 3.1. Algoritmos predictivos
 - 3.1.1. Regresion lineal multiple
 - 3.1.2. K vecinos mas cercanos
 - 3.2. Algoritmos de clasificacion
 - 3.2.1. Clasificador ingenuo de Bayes
 - 3.2.2. Regresion logistica

Unidad 4 Algoritmos de aprendizaje no supervisado



Objetivo particular de la unidad: Aprender sobre diferentes algoritmos de aprendizaje no supervisado y como interpretar sus resultados.

Desarrollo:

- 4. Algoritmos de aprendizaje no supervisado
 - 4.1. Reglas de asociación y filtros colaborativos
 - 4.2. Análisis de cluster
 - 4.2.1. Jerárquico
 - 4.2.1.1. Aglomerativo
 - 4.2.1.1.1. Dendograma
 - 4.2.2. No jerárquico
 - 4.2.2.1. K-means
 - 4.2.2.1.1. Diagrama de código

Unidad 5 Análisis de redes sociales y minería de texto

Objetivo particular de la unidad: Aprender sobre análisis de información proveniente de redes sociales para realizar predicciones y clasificaciones exitosas

Desarrollo:

- 5. Análisis de redes sociales y minería de texto
 - 5.1. Análisis de redes sociales
 - 5.1.1. Grafos
 - 5.1.2. Vertices y aristas
 - 5.1.3. Grafos dirigidos y no dirigidos
 - 5.1.4. Grafos ponderados o pesados
 - 5.1.5. Métricas de grafos
 - 5.2. Minería de texto
 - 5.2.1. Introducción a la minería de texto
 - 5.2.2. Tipos de minería de texto
 - 5.2.2.1. Clasificación
 - 5.2.2.2. Extrayendo significado o procesamiento natural de lenguaje
 - 5.2.3. Bolsa de palabras
 - 5.2.4. Matriz de términos-documentos
 - 5.2.5. Preprocesamiento
 - 5.2.6. Creación de tokens
 - 5.2.7. Derivación
 - 5.2.8. Presencia y ausencia vs frecuencia
 - 5.2.9. Frecuencia de términos-Frecuencia inversa de documentos
 - 5.2.10. Latent semantic index (LSI)
 - 5.2.11. Clasificación de documentos con LSI

14. Actividades Prácticas

Ejercicios en línea, ejercicios de interpretación de resultados y controles de lectura

15.- Metodología

Lectura previa, estudios de casos, discusión de temas, cursos en línea, participación de los alumnos.

16.- Evaluación

1) Control de lectura	25%
2) Ejercicios de interpretacion	25%
3) Ejercicio en linea	50%
Total	100%

17.- Bibliografía

1. Shmueli, G. (2020) Data mining for business analytics: Concepts, techniques and applications in Python

Otros materiales

1. Bibliotecas de ciencia de datos

- 1.1. <https://numpy.org>
- 1.2. https://rpy2.readthedocs.io/en/version_2.8.x/
- 1.3. <https://www.anaconda.com>

2. Pandas

- 2.1. <https://pandas.pydata.org>
- 2.2. <https://www.datacamp.com/courses/pandas-foundations>
- 2.3. <https://www.datacamp.com/courses/manipulating-dataframes-with-pandas>
- 2.4. <https://www.datacamp.com/courses/pandas-joins-for-spreadsheet-users>
- 2.5. <https://www.datacamp.com/courses/optimizing-python-code-with-pandas>

3. Aprendizaje de maquina

3.1. General

- 3.1.1. <https://cognitiveclass.ai/courses/data-science-101>
- 3.1.2. <https://www.linkedin.com/learning/machine-learning-and-ai-foundations-value-estimations/welcome>
- 3.1.3. <https://cognitiveclass.ai/courses/data-analysis-python>
- 3.1.4. <https://cognitiveclass.ai/courses/machine-learning-with-python>
- 3.1.5. <https://www.kaggle.com/dansbecker/basic-data-exploration>
- 3.1.6. <https://www.datacamp.com/courses/python-data-science-toolbox-part-1>
- 3.1.7. <https://www.datacamp.com/courses/python-data-science-toolbox-part-2>

3.2. Supervisado

- 3.2.1. <https://www.datacamp.com/courses/supervised-learning-with-scikit-learn>
- 3.2.2. <https://www.datacamp.com/courses/sentiment-analysis-in-python>

3.3. No supervisado

- 3.3.1. <https://www.datacamp.com/courses/unsupervised-learning-in-python>

4. Visualizacion

- 4.1. <https://matplotlib.org>
- 4.2. <https://www.datacamp.com/courses/introduction-to-matplotlib>
- 4.3. <https://cognitiveclass.ai/courses/data-visualization-with-python>
- 4.4. <https://www.tableau.com/es-es/learn/training>

18.- Perfil del profesor

Profesor con nivel mínimo de maestría, de reconocido prestigio, con amplia experiencia en la



docencia, programación y análisis de datos

19.- Nombre de los profesores que imparten la materia

Fernando Carrillo Castro

20.- Lugar y fecha de su aprobación (incluyendo la última actualización)

13 de Enero de 2020

21.- Instancias que aprobaron el programa (Junta Académica y/o Coordinación del programa)

Junta Académica de la MCD y el Departamento de Sistemas de Información