

Maestría en Ciencia de Datos (MCD)

Programación II



Agenda

Programación II



-  Introducción
-  Presnetación de la asignatura
-  Metodología de Evaluación
-  Bibliografía



Introducción

Programación II



Ramón Osvaldo Guardado Medina



Presentación de la asignatura

Programación II

Nombre de la asignatura	Nivel de Formación	Clave de la asignatura
Programación II	Maestría	
Prerrequisitos: Ninguno	Area de Formación: <i>Básica común obligatoria</i>	Departamento: Sistemas de información
Modalidad: Presencial	Tipo de asignatura: Curso - Taller	
Carga Horaria: 60 Hrs.		
Teoría: 40 Hrs.	Práctica: 20 Hrs.	Créditos: 7
Trayectoria de la asignatura		



Presentación

Programación II

“El maestrante en Ciencia de datos, debe conocer los fundamentos de aprendizaje de maquina y el como implementarlos en un lenguaje de programación (Python), así como de las diversas bibliotecas o frameworks disponibles para hacerlo de una manera mas fácil”

“Con este curso el futuro maestro en Ciencia de datos, podrá convertir los datos en información y conocimiento.”



Objetivos del programa

Programación II

“Obtener habilidades de aprendizaje de maquina, la capacidad de decidir que algoritmo es mejor para cada ocasión e implementarlo usando el lenguaje de programación Python con bibliotecas especializadas en el área”



Contenido Temático

Programación II

- **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



Unidad 1 Ciencia de datos con python

Programación II

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

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 Distribución Anaconda

1.7 Visual Studio Code



ANACONDA®



scikit-learn



rpy2



seaborn



Visual Studio Code



matplotlib



SciPy



Unidad II Fundamentos de Machine Learning

Programación II

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*



Unidad II Fundamentos de Machine Learning

Programación II

2.1.2 Modelado predictivo vs modelado explicativo

2.1.3 Algoritmos supervisados

*2.1.3.1 Algoritmos predictivos (**Corrección:** Regresión)*

2.1.3.2 Algoritmos clasificatorios

2.1.4 Algoritmos no supervisados

2.1.5 Conjuntos de: entrenamiento , validación y prueba

2.1.6 Visualización de datos

2.1.6.1 Biblioteca Matplotlib (Se añade: Seaborn)

*2.1.6.2 Orange (**Cambio a:** VS Code)*

*2.1.6.3 Tableau (**Cambio a:** Pycaret)*



Unidad II Fundamentos de Machine Learning

Programación II

2.1.7 Reducción de dimensiones con análisis de componente principal

2.1.8 Evaluación del rendimiento predictivo



Unidad III Algoritmos de aprendizaje supervisado

Programación II

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

3. Algoritmos de aprendizaje supervisado

3.1 Algoritmos predictivos

3.1.1. Regresión lineal múltiple

3.1.2. K vecinos mas cercanos

3.2 Algoritmos de clasificación

3.2.1 Clasificador ingenuo de Bayes

3.2.2 Regresion logistica



Unidad IV Algoritmos de aprendizaje no supervisado

Programación II

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

4. Algoritmos de aprendizaje no supervisado

4.1. Reglas de asociación y filtros colaborativos

4.2. Análisis de clúster

4.2.1 Jerárquico

4.2.1.1 Aglomerativo

4.2.1.2 Dendograma

4.2.2 No jerárquico

4.2.2.1 K-means

4.2.2.2 Diagrama de codo



Unidad V Análisis de redes sociales y minería de datos

Programación II

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

5. Análisis de redes sociales y mineria de texto

5.1 Análisis de redes sociales

5.1.1 Grafos

5.1.2 Vértices y aristas

5.1.3 Grafos dirigidos y no dirigidos

5.1.4 Grafos ponderados o pesados

5.1.5 Métricas de grafos



Unidad V Análisis de redes sociales y minería de datos

Programación II

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*



Unidad V Análisis de redes sociales y minería de datos

Programación II

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



Actividades Prácticas

Programación II

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

Metodología

Programación II

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



Evaluación

Programación II

• <i>Control de lectura</i>	25 %
• <i>Ejercicios de interpretación</i>	25 %
• <i>Ejercicios en línea</i>	50 %
<i>Total</i>	100 %



Datos de impartición

Programación II

- **Programación II con NRC:** 168463
- **Calendario:** 2023-A.
- **Horario:** Martes de 18:00 a 20:55 hrs.
- **Lugar:** edificio P aula 203.
- **Modalidad:** presencial y transmitida
- **ZOOM:** <https://udg-mx.zoom.us/j/89384917253>



Bibliografía

Programación II

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

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>



Bibliografía

Programación II

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>



Bibliografía

Programación II

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>

