

MINERÍA DE DATOS: PREPROCESAMIENTO Y CLASIFICACIÓN

Minería de Datos: Preprocesamiento y clasificación

- **Página web:** <http://masteres.ugr.es/datcom/>
- **Material del curso:** <http://prado.ugr.es>
- **Profesores:**
 - García López, Salvador;
 - Luengo Martín, Julián;
 - Molina Cabrera, Daniel;
- **Tutorías:**
<https://decsai.ugr.es/docencia/profesorado>

Ciencia de Datos

CONTENIDO FUNDAMENTAL:

Introducción a la Ciencia de Datos

Software y Lenguajes de Programación:

- Lenguajes Python y R

- Paquetes de R para el Análisis de Datos

Aprendizaje Supervisado: Clasificación y Regresión:

- Modelos Estadísticos (Regresión lineal, Regresión no lineal, KNN,...)

- Modelos de la Inteligencia Artificial (Árboles de Decisión, Reglas, Random Forest, SVM,...)

- Algoritmos avanzados (Deep Learning,...)

- Problemas (Clasificación no Balanceada,...)

- Uso de Datos Masivos

Aprendizaje no Supervisado:

- Reglas de Asociación

- Clustering

- Modelos Básicos

- Uso de Datos Masivos

Análisis Estadístico de Experimentos

Preprocesamiento de Datos

Detección de Anomalías

Big Data

- Técnicas (Cloud computing, Hadoop, MongoDB, Hive)

- Herramientas (Pig, Spark,...)

Ciencia de Datos

Introducción a la
Programación para
Ciencia de Datos

CONTENIDO FUNDAMENTAL:

Introducción a la Ciencia de Datos

Software y Lenguajes de Programación:

Lenguajes Python y R

Paquetes de R para el Análisis de Datos

Aprendizaje Supervisado: Clasificación y Regresión:

Modelos Estadísticos (Regresión lineal, Regresión no lineal, KNN,...)

Modelos de la Inteligencia Artificial (Árboles de Decisión, Reglas, Random Forest, SVM,...)

Algoritmos avanzados (Deep Learning,...)

Problemas (Clasificación no Balanceada,...)

Uso de Datos Masivos

Aprendizaje no Supervisado:

Reglas de Asociación

Clustering

Modelos Básicos

Uso de Datos Masivos

Análisis Estadístico de Experimentos

Preprocesamiento de Datos

Detección de Anomalías

Big Data

Técnicas (Cloud computing, Hadoop, MongoDB, Hive)

Herramientas (Pig, Spark,...)

Ciencia de Datos

Introducción a la Ciencia de Datos

CONTENIDO FUNDAMENTAL:

Introducción a la Ciencia de Datos

Software y Lenguajes de Programación:

Lenguajes Python y R

Paquetes de R para el Análisis de Datos

Aprendizaje Supervisado: Clasificación y Regresión:

Modelos Estadísticos (Regresión lineal, Regresión no lineal, KNN,...)

Modelos de la Inteligencia Artificial (Árboles de Decisión, Reglas, Random Forest, SVM,...)

Algoritmos avanzados (Deep Learning,...)

Problemas (Clasificación no Balanceada,...)

Uso de Datos Masivos

Aprendizaje no Supervisado:

Reglas de Asociación

Clustering

Modelos Básicos

Uso de Datos Masivos

Análisis Estadístico de Experimentos

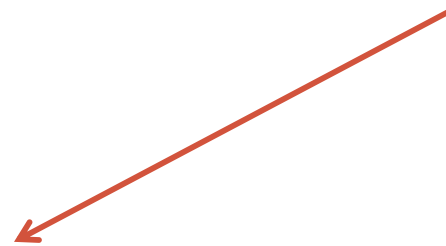
Preprocesamiento de Datos

Detección de Anomalías

Big Data

Técnicas (Cloud computing, Hadoop, MongoDB, Hive)

Herramientas (Pig, Spark,...)



Ciencia de Datos

Minería de Datos:
Preprocesamiento y
clasificación

CONTENIDO FUNDAMENTAL :

Introducción a la Ciencia de Datos

Software y Lenguajes de Programación:

Lenguajes Python y R

Paquetes de R para el Análisis de Datos

Aprendizaje Supervisado: Clasificación y Regresión:

Modelos Estadísticos (Regresión lineal, Regresión no lineal, KNN,...)

Modelos de la Inteligencia Artificial (Árboles de Decisión, Reglas, Random Forest, SVM,...)

Algoritmos avanzados (Deep Learning,...)

Problemas (Clasificación no Balanceada,...)

Uso de Datos Masivos

Aprendizaje no Supervisado:

Reglas de Asociación

Clustering

Modelos Básicos

Uso de Datos Masivos

Análisis Estadístico de Experimentos

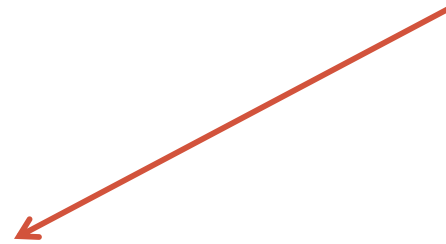
Preprocesamiento de Datos

Detección de Anomalías

Big Data

Técnicas (Cloud computing, Hadoop, MongoDB, Hive)

Herramientas (Pig, Spark,...)



Temario

- **Teórico**

- Tema 1. Modelos no lineales, lineales avanzados y redes neuronales.
- Tema 2. Árboles de decisión y aprendizaje de reglas.
- Tema 3. Multclasificadores y ensembles.
- Tema 4. Descomposición de problemas multiclase.
- Tema 5. Máquinas soporte vectorial (SVM).
- Tema 6. Preprocesamiento de datos.

- **Práctico**

- Sesión 1: Análisis de modelos
- Sesión 2: Regresión logística
- Sesión 3: Árboles y reglas
- Sesión 4: Máquinas de soporte vectorial
- Sesión 5: Visualización
- Sesión 6: Preprocesamiento

Bibliografía

- “The Elements of Statistical Learning: Data Mining, Inference, and Prediction”, Trevor Hastie, Robert Tibshirani, Jerome Friedman. Second Edition, Springer, 2009.
- "Data Preparation for Data Mining". Dorian Pyle. Morgan Kaufmann, 1999.
- "Data Preprocessing in Data Mining". Salvador García, Julián Luengo, Francisco Herrera. Springer, 2015.
- "Data Mining: Concepts and Techniques. Jiawei Han, Jian Pei, Micheline Kamber, 2011.
- “Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems”. Aurélien Géron. O’Reilly, 2019.
- “Hands-On Data Preprocessing in Python: Learn how to effectively prepare data for successful data analytics”. Roy Jafair. Packt, 2022.

Objetivos

- El alumno aprenderá cómo **resolver problemas de clasificación en minería de datos**, aplicando las distintas técnicas conocidas para la etapa de modelado, así como **técnicas de preprocesamiento de datos** en aquellos problemas en los que la etapa de preprocesamiento sea necesaria o recomendable.
- En particular, el alumno:
 - Conocerá y manejará las diferentes técnicas de preprocesamiento de datos existentes para la selección, limpieza, enriquecimiento, reducción y transformación de datos mediante discretización, tratamiento de datos perdidos, reducción de la dimensionalidad, selección y regularización de modelos, etc.
 - Conocerá y manejará técnicas de modelado a partir de datos para clasificación como la regresión no lineal, SVMs, Árboles de Decisión y Random Forest. También tendrá conocimientos básicos sobre técnicas como el Bagging, el Boosting y los algoritmos para la obtención de reglas.
- Al mismo tiempo y de manera práctica el alumno deberá:
 - Saber aplicar las técnicas anteriormente mencionadas para modelar y resolver problemas reales o académicos
 - Conocer las principales herramientas software existentes para modelar y resolver problemas de clasificación a partir de datos.

Cronograma del curso 2024-2025

día	horario	profesor		contenido
21-10-2024 (L)	15:30-18:00	Salvador García	T	Modelos no lineales, lineales avanzados y redes neuronales.
22-10-2024 (M)	18:00-20:30	Daniel Molina	P	Visualización
23-10-2024 (X)	15:30-18:00	Julián Luengo	T/P	Preprocesamiento de Datos.
28-11-2024 (L)	15:30-18:00	Salvador García	T	Árboles de decisión y SVMs.
29-11-2024 (M)	18:00-20:30	Daniel Molina	P	Análisis de modelos no lineales y validación.
30-11-2024 (X)	15:30-18:00	Julián Luengo	T/P	Preprocesamiento de Datos.
04-11-2024 (L)	15:30-18:00	Salvador García	T	Multclasificadores y ensembles. Descomposición.
05-11-2024 (M)	18:00-20:30	Julián Luengo	T/P	Preprocesamiento de Datos.
06-11-2024 (X)	15:30-18:00	Julián Luengo	T/P	Preprocesamiento de Datos.
11-11-2024 (L)	15:30-18:00	Daniel Molina	P	Ensembles, SVMs y Redes Neuronales.
12-11-2024 (M)	18:00-20:30	Julián Luengo	T/P	Preprocesamiento de Datos.
13-11-2024 (X)	15:30-18:00	Julián Luengo	T/P	Preprocesamiento de Datos.

Aula: 0.6-0.7 Edificio Auxiliar de la ETSIIT

Evaluación

- Evaluación Ordinaria:
 - La evaluación de la asignatura se realizará mediante un trabajo práctico.
 - El trabajo se realizará en equipos de 5 miembros.
 - Se dispondrá varias bases de datos, y sobre una de ellas, cada equipo debe aplicar las técnicas de preprocesamiento y clasificación, vistas en clase, que considere oportunas para obtener los mejores resultados posibles ante las condiciones especificadas en la prueba.
 - Se realizará una documentación basada en notebooks y una presentación para evaluar el trabajo en conjunto y el trabajo individual de cada componente del equipo.
 - Cómo muy tarde, el día 1-11-2024 se comunicará a todos los profesores del curso los miembros de cada equipo.
 - Realización del trabajo: del 11-11-2024 al 19-01-2025.
 - Entrega del trabajo en PRADO: 19-01-2025.

Evaluación

- Evaluación extraordinaria
 - La evaluación extraordinaria se realizará siguiendo la misma estructura que la ordinaria.
- Evaluación única final
 - La evaluación única final se realizará en un solo acto académico que incluirá pruebas tanto de tipo teórico como práctico que garanticen que el alumno ha adquirido la totalidad de las competencias descritas en esta guía docente.