

Guía docente de la asignatura

Minería de Datos: Preprocesamiento y Clasificación

Fecha última actualización: 23/07/2021
Fecha de aprobación por la Comisión Académica: 26/07/2021

Máster

Máster Universitario en Ciencia de Datos e Ingeniería de Computadores

MÓDULO

Módulo de Nivelación de Conocimientos

RAMA

Ingeniería y Arquitectura

CENTRO RESPONSABLE DEL TÍTULO

Escuela Internacional de Posgrado

Semestre

Primero

Créditos

4

Tipo

Optativa

Tipo de enseñanza

Presencial

PRERREQUISITOS Y/O RECOMENDACIONES

Sin requisitos previos.

BREVE DESCRIPCIÓN DE CONTENIDOS (Según memoria de verificación del Máster)

- Regresión no Lineal. Modelos avanzados.
- Árboles de Decisión, Bagging, Boosting y Random Forest. Algoritmos para la obtención de Reglas.
- Máquinas soporte vectorial (SVM).
- Preprocesamiento de Datos. Discretización, Datos Perdidos, Reducción de la Dimensionalidad. Selección de Modelos Lineales y Regularización. Análisis de Componentes Principales (PCA).

COMPETENCIAS

COMPETENCIAS BÁSICAS

- CB6 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.



- CB7 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio.
- CB8 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios.
- CB9 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.
- CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

COMPETENCIAS GENERALES

- CG01 - Capacidad de acceso y gestión de la información
- CG02 - Capacidad de análisis y síntesis
- CG03 - Capacidad de organización y planificación
- CG04 - Capacidad emprendedora
- CG05 - Capacidad para tomar decisiones de forma autónoma
- CG07 - Motivación por la calidad
- CG08 - Capacidad para trabajar en equipo

COMPETENCIAS ESPECÍFICAS

- CE03 - Capacidad para la aplicación de técnicas y metodologías que permitan abordar desde nuevas perspectivas los problemas de interés, gracias a la disponibilidad de las plataformas de computación y comunicación con altos niveles de prestaciones.
- CE04 - Capacidad de análisis de aplicaciones en ámbitos de biomedicina y bioinformática, optimización y predicción, control avanzado, y robótica bioinspirada, tanto desde el punto de vista de los requisitos para una implementación eficaz de los algoritmos y las técnicas de computación que se usan para abordarlas, como de las características deseables en las arquitecturas donde se ejecutan
- CE05 - Capacidad para modelar y resolver problemas reales o académicos mediante técnicas de ciencia de datos
- CE06 - Capacidad para modelar y resolver problemas reales o académicos mediante tecnologías inteligentes o de inteligencia computacional.
- CE08 - Conocer algunas de las principales aplicaciones de la ciencia de datos y de las tecnologías inteligentes

COMPETENCIAS TRANSVERSALES

- CT01 - Ser consciente de la importancia del desarrollo sostenible y demostrar sensibilidad medioambiental.
- CT02 - Ser consciente del derecho a la no discriminación y al acceso universal al conocimiento de las personas con discapacidad.

RESULTADOS DE APRENDIZAJE (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.

PROGRAMA DE CONTENIDOS TEÓRICOS Y PRÁCTICOS

TEÓRICO

- Tema 1. Modelos no lineales.
- Tema 2. Árboles de decisión. Multiclasificadores.
- Tema 3. Descomposición de problemas multiclase.
- Tema 4. Aprendizaje de reglas.
- 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: Visualización
- Sesión 5: Máquinas de soporte vectorial
- Sesión 6: Preprocesamiento

BIBLIOGRAFÍA

BIBLIOGRAFÍA FUNDAMENTAL

- “An Introduction to Statistical Learning with Applications in R”, Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, Springer, 2013



- “The Elements of Statistical Learning: Data Mining, Inference, and Prediction”, Trevor Hastie, Robert Tibshirani, Jerome Friedman. Second Edition, Springer, 2009.
- “Introduction to Data Mining”, Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Pearson, 2013.

BIBLIOGRAFÍA COMPLEMENTARIA

- "Data Preprocessing in Data Mining". Salvador García, Julián Luengo, Francisco Herrera. Springer, 2015.
- "Data Preparation for Data Mining". Dorian Pyle. Morgan Kaufmann, 1999.
- "Data Mining: Concepts and Techniques. Jiawei Han, Jian Pei, Micheline Kamber, 2011."

ENLACES RECOMENDADOS

Sitio web del Máster Universitario Oficial en Ciencia de Datos e Ingeniería de Computadores:
<http://masteres.ugr.es/datcom/>

Libro básico

<http://www-bcf.usc.edu/~gareth/ISL/>

Libros Avanzados

<http://statweb.stanford.edu/~tibs/ElemStatLearn/>

<http://www-users.cs.umn.edu/~kumar/dmbook/index.php>

Enlaces Adicionales

<http://guidetodatamining.com/>

<http://www.dataminingbook.info/pmwiki.php/Main/BookDownload>

METODOLOGÍA DOCENTE

- MD01 Lección magistral/expositiva
- MD02 Resolución de problemas y estudio de casos prácticos
- MD03 Prácticas de laboratorio
- MD04 Seminarios
- MD05 Análisis de fuentes y documentos
- MD06 Realización de trabajos en grupo
- MD07 Realización de trabajos individuales

EVALUACIÓN (instrumentos de evaluación, criterios de evaluación y porcentaje sobre la calificación final)

EVALUACIÓN ORDINARIA



La evaluación ordinaria de la asignatura consta de dos pruebas.

La primera, que constituye un 50% de la nota consistirá en la realización de un trabajo práctico. Se le entregará a cada alumno una base de datos y se le pedirá que aplique las técnicas de preprocesamiento y clasificación vistas en clase que considere oportunas, para obtener los mejores resultados posibles. La evaluación de la práctica tendrá en cuenta los resultados obtenidos y la calidad de la documentación aportada en la que se describirá el trabajo realizado.

La segunda prueba, que constituye el 50% de la nota consistirá en una defensa oral del trabajo realizado, así como de los conocimientos básicos adquiridos durante el curso.

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.

INFORMACIÓN ADICIONAL

La participación activa en clase podrá ser tenida en cuenta en el sistema de evaluación continua.

