Inteligencia para Big Data: métodos y tecnologías

Curso Académico: (2019 / 2020) **Fecha de revisión: 03/05/2019 11:43:38**

Departamento asignado a la asignatura: Departamento de Informática

Coordinador/a: ALER MUR, RICARDO

Tipo: Obligatoria Créditos ECTS : 3.0

Curso : 1 Cuatrimestre : 1

REQUISITOS (ASIGNATURAS O MATERIAS CUYO CONOCIMIENTO SE PRESUPONE)

Conocimientos básicos de programación.

OBJETIVOS

Competencias Básicas
Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación
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
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
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
Aplicar los fundamentos teóricos de las técnicas de recogida, almacenamiento, tratamiento y presentación de información, especialmente para grandes volúmenes de datos, como base para el desarrollo y adaptación de dichas técnicas a problemas concretos
Identificar diferentes técnicas para almacenar, replicar y distribuir grandes cantidades de datos, y diferenciarlas en función de sus características teóricas y prácticas
Identificar las técnicas de análisis de datos mas adecuadas para cada problema y saber aplicarlas para el análisis, diseño y solución de los mismos
Obtener soluciones prácticas y eficientes para problemas de tratamiento de grandes volúmenes de datos, tanto individualmente como en equipo
Sintetizar las conclusiones obtenidas de estos análisis y presentarlas de manera clara y convincente en un entorno bilingüe (español e inglés) tanto por escrito como oralmente
Ser capaz de generar nuevas ideas (creatividad) y de anticipar nuevas situaciones, en los contextos del análisis de datos y de la toma de decisiones
Utilizar habilidades para el trabajo en equipo y para relacionarse con otros de forma autónoma

Competencias Específicas
Identificar y seleccionar las herramientas software adecuadas para el tratamiento de grandes cantidades de datos
Saber diseñar sistemas para el procesamiento de los datos, desde la obtención y filtrado inicial de los mismos, su análisis estadístico, hasta la presentación de los resultados finales
Utilizar técnicas y herramientas de investigación operativa utilizables con datos masivos en procedimientos para su análisis, visualización de sus resultados o dentro de sistemas de apoyo a decisiones
Aplicar los principios básicos y fundamentales del aprendizaje automático al diseño de procedimientos y la mejora de los mismos
Interpretar las especificaciones funcionales encaminadas al desarrollo de aplicaciones basadas en el aprendizaje automático
Identificar la oportunidad de utilizar aprendizaje automático para resolver problemas reales

Realizar el análisis y el diseño detallado de aplicaciones informáticas basadas en el aprendizaje automático

Resultados de aprendizaje
- Conocimientos básicos y fundamentales del aprendizaje automático
- Comprensión de las técnicas básicas de aprendizaje automático
- Utilización de forma práctica de estas técnicas básicas en problemas reales
- Capacidad para analizar las tareas más adecuadas para cada técnica
- Entender cuándo utilizar aprendizaje automático para resolver problemas reales

DESCRIPCIÓN DE CONTENIDOS: PROGRAMA

1. Introducción / conceptos básicos
2. Métodos para la construcción de modelos para clasificáción y regression
   2.1. Vecino más cercano
   2.2. Árboles y reglas
3. Metodología / secuencia de operaciones:
   3.1. Secuencia básica
   3.2. Ajuste de hiperparámetros
   3.3. Evaluación de modelos
4. Métodos para pre-proceso y selección de atributos (filter y wrapper)
5. Métodos basados en conjuntos de modelos:
   5.1. Bagging / Random Forests
   5.2. Boosting / Gradient Boosting
6. Métodos para aprendizaje on-line / procesamiento de flujos de datos
7. Métodos para datos desbalanceados
8. Tecnologías software para aprendizaje automático y datos masivos:
   8.1. Python / Scikit-learn
   8.2. Mapreduce
   8.3. Spark / MLLIB / ML / Pyspark

ACTIVIDADES FORMATIVAS, METODOLOGÍA A UTILIZAR Y RÉGIMEN DE TUTORÍAS

Actividades formativas
Clase teórica
Prácticas de laboratorio
Trabajo en grupo
Trabajo individual del estudiante

Metodologías docentes
Exposiciones en clase del profesor con soporte de medios informáticos y audiovisuales, en las que se desarrollan los conceptos principales de la materia y se proporciona la bibliografía para complementar el aprendizaje de los alumnos.
Lectura crítica de textos recomendados por el profesor de la asignatura:Artículos de prensa, informes, manuales y/o artículos académicos, bien para su posterior discusión en clase, bien para ampliar y consolidar los conocimientos de la asignatura.
Resolución de casos prácticos, problemas, etc.¿ planteados por el profesor de manera individual o en grupo
Elaboración de trabajos e informes de manera individual o en grupo

SISTEMA DE EVALUACIÓN

Peso porcentual del Examen/Prueba Final:	30
Peso porcentual del resto de la evaluación:	70

A) Prácticas durante el curso (70%)
B) Examen final (30%)

Para aprobar la asignatura A+B>=5. El examen final no tiene nota mínima.

BIBLIOGRAFÍA BÁSICA

- Hastie, Tibshirani, Friedman The Elements of Statistical Learning, Springer, 2016

- Holden Karau, Andy Konwinski, Patrick Wendell, Matei Zaharia Learning Spark, O'Reilly, 2015

- Raul Garreta, Guillermo Moncecchi Learning scikit-learn: Machine Learning in Python, Packt Publishing, 2013

RECURSOS ELECTRÓNICOS BÁSICOS

- . Spark: https://spark.apache.org/

- . Scikit learn: http://scikit-learn.org/stable/index.html