

Guía docente

230337 - AIPRAC - Resolución de Problemas con Inteligencia Artificial: un Enfoque Práctico

Última modificación: 11/04/2025

Unidad responsable: Escuela Técnica Superior de Ingeniería de Telecomunicación de Barcelona

Unidad que imparte: 739 - TSC - Departamento de Teoría de la Señal y Comunicaciones.

Titulación: GRADO EN INGENIERÍA DE TECNOLOGÍAS Y SERVICIOS DE TELECOMUNICACIÓN (Plan 2015). (Asignatura optativa).

GRADO EN INGENIERÍA ELECTRÓNICA DE TELECOMUNICACIÓN (Plan 2018). (Asignatura optativa).

Curso: 2025

Créditos ECTS: 2.0

Idiomas: Catalán

PROFESORADO

Profesorado responsable: ENRIQUE MONTE MORENO

Otros: Primer quadrimestre:
ENRIQUE MONTE MORENO - 11

CAPACIDADES PREVIAS

Programación Python y buenos conocimientos en las asignaturas de "Cálculo Vectorial", "Señales y Sistemas" y "Probabilidad y Estadística"

REQUISITOS

"Cálculo Vectorial" y "Probabilidad y Estadística"

METODOLOGÍAS DOCENTES

En cada sesión de clase se dedicará una hora a explicar la base de datos con la que se trabajará, y las propiedades estadísticas y geométricas de la misma. A continuación, considerando las propiedades de la base de datos, se justificará la elección de un método de inteligencia artificial adecuado para la misma. Las dos horas siguientes se dedicarán a trabajar con la base de datos, utilizando un 'notebook' de Jupyter en lenguaje Python.

OBJETIVOS DE APRENDIZAJE DE LA ASIGNATURA

El objetivo de la asignatura es dotar a los alumnos de herramientas para analizar las propiedades de una base de datos relacionada con una aplicación real, y decidir cuál es la técnica de inteligencia artificial más adecuada para resolver el problema.

HORAS TOTALES DE DEDICACIÓN DEL ESTUDIANTADO

Tipo	Horas	Porcentaje
Horas grupo grande	6,0	12.00
Horas grupo pequeño	14,0	28.00
Horas aprendizaje autónomo	30,0	60.00

Dedicación total: 50 h

CONTENIDOS

Análisis Exploratorio de Datos (EDA) y Aplicación del Algoritmo K-Means para Predecir Ventas con el Dataset CarSeats

Descripción:

La práctica utiliza la base de datos CarSeats, un conjunto de datos simulados que recopila información sobre las ventas de sillas de coche para niños en 400 tiendas, extraído del libro An Introduction to Statistical Learning. Esta base incluye variables como las ventas unitarias, el precio de los competidores, los ingresos medios de la zona, el presupuesto de publicidad, la población, la edad media, la educación, la ubicación urbana/rural y si la tienda está en Estados Unidos, entre otras. El objetivo es realizar un Análisis Exploratorio de Datos (EDA) para comprender las características del conjunto de datos y aplicar el algoritmo k-means para agrupar los datos y predecir ventas, identificando factores clave que influyen en los resultados comerciales.

Objetivos específicos:

Los objetivos específicos de la práctica son explorar el dataset CarSeats mediante un Análisis Exploratorio de Datos (EDA) para identificar patrones, correlaciones y posibles anomalías en las variables relacionadas con las ventas de sillas de coche para niños. Además, se pretende aplicar el algoritmo k-means para agrupar las tiendas en clústeres según características como el precio, la publicidad o la ubicación, y utilizar estos clústeres para estimar las ventas. Finalmente, se busca evaluar el rendimiento del modelo con métricas como el error cuadrático medio (MSE) y ofrecer recomendaciones para optimizar las estrategias comerciales basadas en los resultados obtenidos.

Actividades vinculadas:

Jupyter Notebook en python

Dedicación: 3h

Grupo mediano/Prácticas: 3h

Análisis Exploratorio de Datos (EDA) y Regresión Lineal para la Predicción del Tiempo de Supervivencia en Pacientes con Tumores Cerebrales

Descripción:

La práctica utiliza la base de datos "Brain Cancer Data", que contiene información sobre pacientes con tumores cerebrales primarios tratados con radioterapia estereotáctica en el Instituto de Cáncer Masaryk Memorial Brno. Esta base de datos incluye variables como el sexo, el diagnóstico (meningioma, glioma LG, glioma HG, otros), la localización del tumor, el índice de Karnofsky, el volumen tumoral bruto, el método estereotáctico (SRS o SRT), el estado del paciente al final del estudio y el tiempo de supervivencia en meses. El objetivo es realizar un Análisis Exploratorio de Datos (EDA) para comprender las relaciones entre estas variables y el tiempo de supervivencia, y desarrollar un modelo de regresión lineal para predecir este tiempo, identificando factores clave que influyen en la supervivencia.

Objetivos específicos:

Los objetivos específicos de la práctica son explorar la base de datos "Brain Cancer Data" para identificar patrones y relaciones entre variables como el sexo, el tipo de diagnóstico, la localización del tumor, el índice de Karnofsky, el volumen tumoral, el método de radioterapia estereotáctica y el estado del paciente, con el fin de comprender su impacto en el tiempo de supervivencia. A través de un Análisis Exploratorio de Datos (EDA), se pretende analizar la distribución y correlación de estas variables, y posteriormente aplicar una regresión lineal para construir un modelo predictivo que estime el tiempo de supervivencia, ofreciendo una interpretación clara de los factores más influyentes en el pronóstico de los pacientes.

Actividades vinculadas:

Jupyter Notebook en python

Dedicación: 3h

Grupo mediano/Prácticas: 3h

Análisis Exploratorio de Datos (EDA) y Regresión Logística para la Predicción de Enfermedad Coronaria con la Base de Datos South African Heart Disease

Descripción:

Esta práctica utiliza la base de datos South African Heart Disease Data, que recopila información retrospectiva de 462 hombres de una región de alto riesgo de enfermedades cardíacas en el Cabo Occidental, Sudáfrica. El objetivo principal es realizar un Análisis Exploratorio de Datos (EDA) para comprender la relación entre variables como la presión arterial, el consumo de tabaco, el colesterol LDL, la edad y el historial familiar, entre otros, con la presencia de enfermedad coronaria (CHD). Además, se aplica un modelo de regresión logística para crear un indicador probabilístico de CHD y evaluar los factores de riesgo, ofreciendo una interpretación clara de los resultados para la predicción y prevención de esta enfermedad.

Objetivos específicos:

Los objetivos específicos de esta práctica son analizar la base de datos South African Heart Disease Data mediante un Análisis Exploratorio de Datos (EDA) para identificar patrones y correlaciones entre factores de riesgo como la presión arterial, el colesterol LDL, el tabaquismo, la edad y el historial familiar con la enfermedad coronaria (CHD). Además, se busca desarrollar un modelo de regresión logística que genere un indicador probabilístico para predecir la presencia de CHD, interpretar el impacto de cada variable en el riesgo de la enfermedad y proporcionar recomendaciones basadas en los resultados para la prevención y gestión de factores de riesgo cardiovascular.

Actividades vinculadas:

Jupyter Notebook en python

Dedicación: 3h

Grupo mediano/Prácticas: 3h

Análisis Exploratorio de Datos (EDA) y Predicción de Salarios con Regresión Lineal Regularizada Utilizando la Base de Datos Mid-Atlantic Wage Data

Descripción:

La práctica emplea la base de datos Mid-Atlantic Wage Data, que incluye información sobre los salarios y características demográficas, educativas, laborales y de salud de 3000 trabajadores masculinos de la región del Mid-Atlantic de Estados Unidos, recopilada a partir del Suplemento de Marzo de 2011 del Current Population Survey. El objetivo es realizar un Análisis Exploratorio de Datos (EDA) para comprender las relaciones entre variables como edad, educación, tipo de trabajo y estado civil con el salario, y aplicar técnicas de regresión lineal con regularización (Ridge y Lasso) para predecir el logaritmo del salario, analizando también el problema del sobreajuste y la selección de características relevantes.

Objetivos específicos:

Los objetivos específicos de la práctica son explorar la base de datos Mid-Atlantic Wage Data mediante un Análisis Exploratorio de Datos (EDA) para identificar patrones y correlaciones entre variables como la edad, el nivel educativo, el tipo de trabajo y el estado de salud con el salario de los trabajadores. Además, se busca desarrollar modelos predictivos utilizando regresión lineal con regularización Ridge y Lasso para estimar el logaritmo del salario, evaluar el impacto del sobreajuste y determinar qué variables son más relevantes para la predicción, con el fin de proporcionar una herramienta robusta para el análisis económico y laboral.

Actividades vinculadas:

Jupyter notebook en python.

Dedicación: 3h

Grupo mediano/Prácticas: 3h

Análisis Exploratorio de Datos (EDA) y Aplicación de Árboles de Decisión para Predecir Compras de Zumo de Naranja

Descripción:

Esta práctica utiliza una base de datos de 1.070 compras de zumo de naranja de las marcas Citrus Hill y Minute Maid, recopiladas en varias tiendas durante un período determinado. La base contiene información sobre la marca comprada, la semana de compra, la tienda, los precios originales y con descuento, las promociones especiales, la fidelidad de los clientes a cada marca y la diferencia de precio entre las opciones. Mediante el Análisis Exploratorio de Datos (EDA) y el uso de árboles de decisión, se busca comprender los factores que influyen en la decisión de compra de los consumidores y desarrollar reglas simples para predecir el comportamiento de los clientes.

Objetivos específicos:

Esta práctica se centra en la aplicación de los árboles de decisión para discernir la marca de zumo de naranja elegida por el consumidor, basándose en las características inherentes a su compra. A través de una exploración exhaustiva de la base de datos mediante el análisis exploratorio (EDA), se pretende desentrañar las dinámicas subyacentes y las interrelaciones entre las variables. El objetivo primordial es dominar la utilización y la interpretación de los árboles de decisión, con el fin de identificar las dependencias clave que influyen en la decisión de compra y derivar reglas concisas que eluciden el comportamiento de los clientes en relación con las diferentes variables analizadas.

Actividades vinculadas:

Jupyter notebook en python.

Dedicación: 3h

Grupo mediano/Prácticas: 3h

Clasificación de Texto con Máquinas de Vectores de Soporte (SVM)

Descripción:

Base de Datos: Reseñas de películas

Técnica: Máquinas de Vectores de Soporte (SVM)

Objetivos específicos:

Análisis de una base de datos de reseñas de películas. Preprocesa los datos de texto (por ejemplo, elimina palabras vacías, realiza stemming o lematización). Utiliza Máquinas de Vectores de Soporte (SVM) para clasificar las reseñas de películas como positivas, negativas o neutras. Explora diferentes técnicas de procesamiento de texto y evalúa su impacto en el rendimiento del modelo.

Actividades vinculadas:

Jupyter notebook en python

Dedicación: 3h

Grupo mediano/Prácticas: 3h

Última Sesión

Descripción:

Recapitulación de lo aprendido

Dedicación: 2h

Grupo mediano/Prácticas: 2h

SISTEMA DE CALIFICACIÓN

La calificación se basará en el 'notebook' entregado al final de cada sesión de trabajo.



BIBLIOGRAFÍA

Básica:

- James, G.; Witten, D.; Hastie, T.; Tibshirani, R. An Introduction to Statistical Learning : with Applications in R [en línea]. 2nd ed. New York: Springer, 2021 [Consulta: 12/06/2024]. Disponible a : <https://ebookcentral-proquest-com.recursos.biblioteca.upc.edu/lib/upcatalunya-ebooks/detail.action?pq-origsite=primo&docID=6686746>. ISBN 9781071614181.
- Coelho, L.P.; Richert, W.; Brucher, M. Building machine learning systems with Python: explore machine learning and deep learning techniques for building intelligent systems using scikit-learn and TensorFlow [en línea]. 3rd ed. Birmingham, England: Packt, 2018 [Consulta: 17/06/2024]. Disponible a : <https://ebookcentral-proquest-com.recursos.biblioteca.upc.edu/lib/upcatalunya-ebooks/detail.action?pq-origsite=primo&docID=5485017>. ISBN 9781788622226.