



**UNIVERSIDAD
DE GRANADA**

**Departamento de Ciencias de la
Computación e Inteligencia Artificial**

E.T.S. de Ingenierías Informática y de Telecomunicación

Inteligencia de Negocio

Guion de Prácticas

Práctica 3: Competición en Kaggle

Curso 2024-2025

Grado en Ingeniería Informática
Grado en Ingeniería Informática y Matemáticas
Grado en Ingeniería Informática y Administración y Dirección de Empresas

Práctica 3

Competición en Kaggle

1. Objetivos y Evaluación

En esta tercera práctica de la asignatura Inteligencia de Negocio veremos el uso de métodos avanzados para aprendizaje supervisado sobre una competición de Kaggle de regresión [disponible aquí](#). El estudiante adquirirá destrezas para mejorar la capacidad predictiva del modelo mientras se familiariza con una de las plataformas de competición en ciencias de datos más populares.

La práctica se calificará hasta un **máximo de 2.5 puntos**. La evaluación se dará en parte en función de la posición final (relativa al conjunto de estudiantes participantes) que ocupe el resultado propuesto por el estudiante, con una asignación linealmente proporcional entre 1,666 puntos para la primera posición y 0,333 puntos para el último. Para ser evaluado, no bastará con subir los resultados a Kaggle, se deberá también adjuntar un documento que describa el proceso seguido y justificado por el estudiante para resolver la práctica y demostrar mediante la actividad registrada en Kaggle que ha habido un esfuerzo por mejorar los resultados. En otro caso, el alumno no obtendrá ninguna puntuación en esta práctica.

Sobre la puntuación obtenida en base a la posición, se aplicará un factor corrector $[0,5, 1,5]$ (es decir, se podrá reducir o aumentar hasta un 50 %) en función de la calidad de la documentación presentada y las soluciones abordadas.

2. Descripción del Problema y Tareas

La competición será [Adivinar el precio de un coche usado](#). El objetivo es predecir, dadas las características de un coche utilizado, su precio. En este reto disponemos de una serie de características para cada coche (marca, modelo, año del modelo, kilometraje, tipo de combustible, motor, etc.) que puede ser utilizado para aproximar lo más posible su precio. Dentro de los datos de la competición, los atributos de entrada están descritos [aquí](#).

El conjunto de entrenamiento consta de 188533 instancias y 13 atributos (de los cuales el *Id* del coche no debe ser tenido en cuenta), contiene datos principalmente categóricos, pero

también alguno numérico (en distintas escalas). Para este concurso, hay una única variable objetivo, *price*, que es lo que se desea predecir.

El rendimiento en esta competición se evaluará usando la medida de la raíz del error cuadrático medio (*root-mean-square error*, RMSE) entre los valores predichos y los valores observados.

En esta práctica se permite el uso de cualquier *software*, algoritmo, o lenguaje que el estudiante considere útil. Está terminantemente prohibido usar la clase en los datos de test, en caso de conocerse, para entrenar, configurar o mejorar el modelo predictivo. También se prohíbe que los alumnos compartan entre ellos soluciones de la competición. Cualquier indicio de estas conductas supondrá la anulación de la práctica.

3. Documentación

La documentación explicará las estrategias seguidas y el progreso que se ha ido desarrollando durante la competición. Deberán razonarse brevemente los diferentes pasos tomados demostrando que el alumno comprende los distintos enfoques que ha seguido para resolver el problema. Podrá apoyarse en visualización de datos u otras técnicas de análisis para comprender las propiedades del problema. Se recomienda también añadir en la documentación extractos de los *scripts* para explicar el trabajo realizado. Será obligatorio incluir una tabla que contenga tantas filas como soluciones se han subido a Kaggle incluyendo columnas que resuman cada experimento conteniendo, al menos:

- La fecha y hora de subida a Kaggle.
- La posición que ocupó en ese momento.
- El *score* sobre el conjunto de datos de entrenamiento.
- El *score* obtenido en Kaggle al subir la predicción en test.
- Breve descripción del preprocesado realizado.
- Breve descripción de el/los algoritmo(s) de regresión empleado(s) y configuración de parámetros de esos algoritmos.

La ausencia de esta tabla o una descripción incompleta de la misma supondrá la anulación de la práctica.

Adicionalmente, la segunda página de la documentación (después de la portada y antes del índice) contendrá una captura de pantalla de la tabla *Leaderboard* (que contiene las columnas *Team*, *Score*, *Entires* y *Last*) disponible en la web de Kaggle para cada estudiante. De cada subida realizada a Kaggle se conservará el fichero *.csv* y el *script* en Python o similar usado para ese experimento. Se nombrarán de forma clara y enumerada para poder identificar con facilidad a qué experimento de la tabla corresponde. Este material se entregará

junto a la documentación. El alumno deberá definir como *Team* en Kaggle su primer apellido, segundo apellido y nombre de pila sin espacios terminando con `_UGR_IN`. Por ejemplo: `CasillasBarranqueroJorge_UGR_IN`. La posición que se valorará para la práctica será la mejor de la cuenta `xxx_UGR_IN`, no se aceptará ninguna otra. Se recuerda a los/as estudiantes que Kaggle incorpora mecanismos para detectar múltiples inscripciones en la competición de la misma persona y bloquea las cuentas cuando sucede, impidiendo así seguir compitiendo.

4. Entrega

La participación en la competición en Kaggle finaliza el domingo 7 de enero de 2025. El estudiante deberá también entregar para el domingo 7 de enero de 2025 a las 23:59 una documentación que explique las tareas realizadas y todas las soluciones `.csv` subidas a Kaggle junto con los *scripts* utilizados.

Este material se entregará a través de la web de la asignatura en <https://prado.ugr.es> en un único archivo `zip`. Por ejemplo, la estudiante “María Teresa del Castillo Gómez” subirá el archivo `P3-delCastillo-Gómez-MaríaTeresa.zip`. La documentación, contenida en ese mismo archivo `zip`, tendrá el mismo nombre pero con extensión `pdf`.