

Anteproyecto

Autor: Helena Rodríguez Rivero.

Tutor: Luis de la Cruz Piris

Titulación: Ingeniería Telemática.

Título: Técnicas de aprendizaje automático aplicadas a la predicción de resultados de carreras de caballos.

1. Introducción

El aprendizaje automático o machine learning es una rama de la inteligencia artificial en la cual, mediante datos recolectados, se pueden identificar patrones y tomar decisiones propias. En los últimos años estas técnicas se encuentran en múltiples sectores, ya que nos ayuda a resolver problemas de la manera más eficiente. Han mostrado su utilidad en el ámbito de la medicina, en la que conociendo los síntomas del paciente podemos anticiparnos a enfermedades, en la detección de rostro, pudiendo así reconocer a una persona a través de una imagen, para el desarrollo de vehículos autónomos y robots, podemos predecir accidentes antes de que se produzcan. Estos son solo unos de los pocos ejemplos en los que se puede aplicar machine learning para facilitarnos nuestro día a día.

Hay tres grandes tipos de aprendizaje automático. El primero de ellos es 'Aprendizaje por refuerzo', consiste en aprender a base de ensayos para concluir una tarea de la forma más acertada. Otro de ellos es el 'Aprendizaje supervisado', trata de entrenar el modelo con datos y deducir una función para obtener una salida óptima. Y por último 'Aprendizaje no supervisado', se distingue del 'Aprendizaje supervisado' en que no existe un entrenamiento previo.

En este trabajo de fin de grado se pretende aplicar técnicas de machine learning con el objetivo de predecir resultados de carreras de caballos. A partir de una lista con múltiples carreras desarrolladas desde 2011 hasta la actualidad, se digitalizarán y se las dará un formato concreto para obtener el conjunto de datos base para la investigación. Después se escogerán diferentes algoritmos de machine learning, que una vez aplicados, se validaran y se observará cual de estos algoritmos es mas óptimo para el trabajo.

2. Objetivos y campo de aplicación

El objetivo principal es desarrollar un sistema basado en machine learning que sea capaz de predecir el resultado de una próxima carrera de caballos. Para abordar el objetivo principal, se han definido un conjunto de objetivos secundarios tales como:

- Recolectar datos: deben encontrarse datos de carreras de los últimos años, estos datos se utilizarán para entrenar el modelo.
- Adaptar los datos: se encontrarán en los datos adquiridos información que se necesitará cambiar, añadir o borrar.
- Estudiar posibles algoritmos o técnicas de machine learning: hay diferentes algoritmos ya implementados para resolver los problemas que nos pueden causar las predicciones.
- Aplicar el algoritmo sobre los datos y entrenar el modelo: para entrenar el modelo adecuadamente debemos ajustar los hiperparámetros del modelo.
- Predecir y validar los resultados obtenidos: obtener la tasa de aciertos de nuestro modelo de aprendizaje.

Consiguiendo todos los objetivos secundarios podremos lograr el objetivo principal.

3. Descripción del trabajo

Preprocesamiento de datos, es uno de los procesos más importantes para un modelo de aprendizaje ya que el modelo se va a entrenar con estos datos y por lo tanto necesitamos que sean claros y útiles. Para ello debemos de recolectar todos los datos y definir que datos van a ser relevantes.

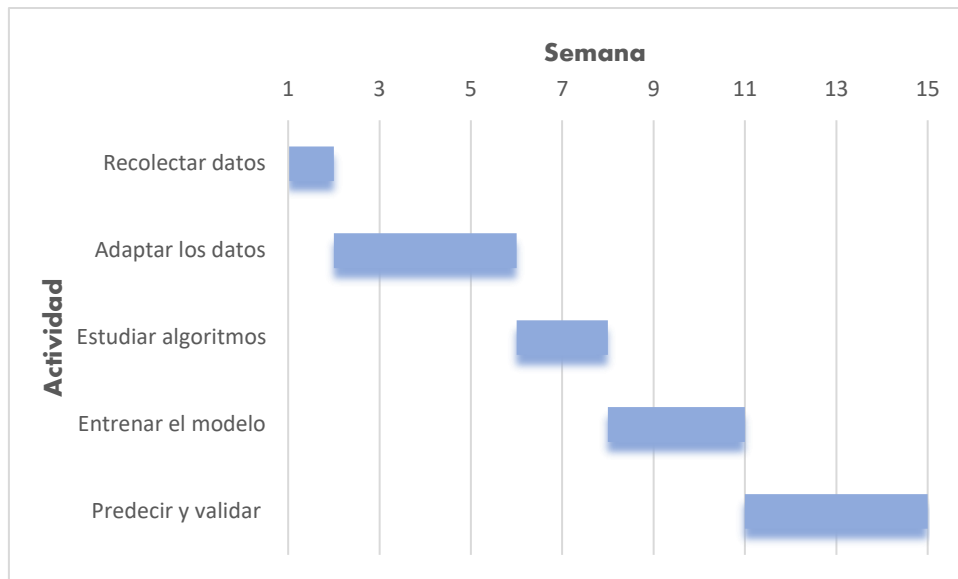
Búsqueda de algoritmos de machine learning, existen diferentes algoritmos de machine learning y saber cuál va a resultar más útil es prácticamente imposible, por ello vamos a comparar varios de estos algoritmos para observar cual se acerca más al resultado real obtenido.

Aplicar los diferentes algoritmos conlleva estudiar cada uno de ellos. Para aumentar el rendimiento de nuestro modelo debemos ajustar los hiperparámetros de cada algoritmo.

Predecir y validar los resultados para comprobar que modelo es mas adecuado para nuestro trabajo.

4. Metodología y plan de trabajo

- Recolectar datos: búsqueda de documentos.
- Adaptar los datos: modificar los documentos obtenidos.
- Estudiar posibles algoritmos o técnicas de machine learning.
- Aplicar el algoritmo sobre los datos y entrenar el modelo.
- Predecir y validar los resultados obtenidos.



Se van a dedicar 20 horas de trabajo por semana.

5. Medios

Se va a utilizar un ordenador portátil ASUS con un procesador Intel(R) Core(TM) i7-7500U, con 8,00 GB de RAM y un sistema operativo de x64. Además, se puede utilizar Google Colab si se necesitan más recursos o el uso de plataformas en la nube si necesitamos alguno de los servicios que ofrece como Amazon Web Service o Google Cloud.

Se van a utilizar librerías de Python, tales como pandas, numpy, scikit-learn y bibliotecas de machine learning gratuitas.

6. Bibliografía

- [1] BBVA, «BBVA NOTICIAS,» 08 11 2019. [En línea]. Available:
<https://www.bbva.com/es/machine-learning-que-es-y-como-funciona/>. [Último acceso: 21 09 2021].
- [2] J. I. Bagnato, «Aprende Machine Learning,» 11 09 2017. [En línea]. Available:
<https://www.aprendemachinelearning.com/7-pasos-machine-learning-construir-maquina/>. [Último acceso: 21 09 2021].
- [3] RTVE, «RTVE.es,» 03 06 2021. [En línea]. Available:
<https://www.rtve.es/rtve/20210603/lototurf-probabilidad-toque-ganar-dinero-estadisticas-carreras-caballos/2094620.shtml>. [Último acceso: 21 09 2021].
- [4] J. Ortiz, «Joanybel Ortiz,» 22 03 2019. [En línea]. Available:
<http://www.joanybelortiz.com/aplicaciones-machine-learning-ejemplos/>. [Último acceso: 27 09 2021].
- [5] AWS, «AWS,» 2021. [En línea]. Available:
https://docs.aws.amazon.com/es_es/sagemaker/latest/dg/algorithms-choose.html.
[Último acceso: 28 09 2021].
- [6] J. M. Heras, «iartificial.net,» 10 10 2020. [En línea]. Available:
<https://www.iartificial.net/librerias-de-python-para-machine-learning/>. [Último acceso: 28 09 2021].