# Práctica 2.

# Predicción de quinielas mediante el paradigma CBR

***Grupo: 3***

* Raúl Vicente Bueno Sevilla
* Álvaro Pérez Liaño
* Xavier Gallofré Nieva

## Estructura de los casos

|  |  |  |
| --- | --- | --- |
| **Atributos** | | |
| **Nombre** | **Descripción** | **Utilidad** |
| **Id** | Identificador |  |
| **Temporada** | Temporada en la que se juega el partido. | Mientras más alejado esté la temporada menos peso tiene. |
| **Equipo local.** | Nombre del equipo local. |  |
| **Equipo visitante** | Nombre del equipo visitante. |  |
| **Puntos equipo local** | Puntuación que lleva el equipo local. |  |
| **Puntos equipo visitante** | Puntuación que lleva el equipo visitante. |  |
| **Partidos ganados equipo local** | Partidos que ha ganado el equipo local. | Permiten ver la racha que lleva cada equipo. |
| **Partidos empatados equipo local** | Partidos que ha empatado el equipo local. |
| **Partidos perdidos equipo local** | Partidos que ha perdido el equipo local. |
| **Partidos ganados equipo visitante** | Partidos que ha ganado el equipo visitante. |
| **Partidos empatados equipo visitante** | Partidos que ha empatado el equipo visitante. |
| **Partidos perdidos equipo visitante** | Partidos que ha perdido el equipo visitante. |
| **Posición equipo local** | Posición en la liga del equipo local. |  |
| **Posición equipo visitante** | Posición en la liga del equipo visitante. |  |
| **Goles a favor equipo local** | Goles a favor del equipo local. | Permiten ver la moral que tiene cada equipo. |
| **Goles en contra equipo local** | Goles en contra del equipo local. |
| **Goles a favor equipo visitante** | Goles a favor del equipo visitante. |
| **Goles en contra equipo visitante** | Goles en contra del equipo visitante. |
| **Resultado equipo local** | Resultado del partido del equipo local. |  |
| **Resultado equipo visitante** | Resultado del partido del equipo visitante. |  |

|  |  |  |
| --- | --- | --- |
| **Solución** | | |
| **Nombre** | **Descripción** | **Utilidad** |
| **Id** | Identificador. |  |
| **Solución** | Resultado. | Almacenar 1, X o 2. |

## Fuentes de conocimiento para obtener los casos

**http://www.marca.com/estadisticas/futbol/primera/AAAA\_AA/jornada\_XX/**

Donde:

* AAAA\_AA es la temporada
* XX es el número de la jornada

Tenemos información con una plantilla idéntica para temporadas desde el año 2000, por lo que podemos acceder con esas webs a los datos que nos interesan.

Ejemplo: <http://www.marca.com/estadisticas/futbol/primera/2000_01/jornada_37/>

## Funciones de similitud locales y globales

|  |  |  |
| --- | --- | --- |
| **Locales** | | |
| **Nombre** | **Valor** | **Explicación** |
| **Id** |  |  |
| **Temporada** | Intervalo(13) | Cogemos únicamente las temporadas desde el 2000. |
| **Equipo local.** | Igualdad() |  |
| **Equipo visitante** | Igualdad() |  |
| **Puntos equipo local** | Intervalo(126) | Son 38 jornadas (42 en 2ª) por 3 puntos cada partido. |
| **Puntos equipo visitante** | Intervalo(126) |
| **Partidos ganados equipo local** | Intervalo(42) | Hay 42 jornadas en 2ª división, mientras que en 1º hay sólo 38. Cogemos el valor mayor de los 2, ya que el error es muy pequeño. |
| **Partidos empatados equipo local** | Intervalo(42) |
| **Partidos perdidos equipo local** | Intervalo(42) |
| **Partidos ganados equipo visitante** | Intervalo(42) |
| **Partidos empatados equipo visitante** | Intervalo(42) |
| **Partidos perdidos equipo visitante** | Intervalo(42) |
| **Posición equipo local** | Intervalo(22) | Cogemos el número de equipos de 2ª, que son más. |
| **Posición equipo visitante** | Intervalo(22) |
| **Goles a favor equipo local** | Intervalo(130) | Suponemos que no habrá mas de 100 goles a favor o en contra por temporada. |
| **Goles en contra equipo local** | Intervalo(100) |
| **Goles a favor equipo visitante** | Intervalo(130) |
| **Goles en contra equipo visitante** | Intervalo(100) |
| **Resultado equipo local** | Intervalo(10) | Cada equipo no meterá más de 10 goles en un partido. |
| **Resultado equipo visitante** | Intervalo(10) |

|  |  |
| --- | --- |
| **Global** | |
| Average() | Hacemos la media de los atributos. |

## Pesos de los atributos para el algoritmo kNN

|  |  |  |
| --- | --- | --- |
| **Nombre** | **Valor** | **Explicación** |
| **Id** |  |  |
| **Temporada** | 0.3 | Le damos importancia a la temporada mientras más cerca esté de la actual. |
| **Equipo local.** | 0.02 | No nos importa tanto que equipos sean los que han obtenido el resultado. Es más importante la posición de ambos y su relación de goles. |
| **Equipo visitante** | 0.02 |
| **Puntos equipo local** | 0.2 | Importa la distancia en puntos de ambos equipos en la liga. |
| **Puntos equipo visitante** | 0.25 |
| **Partidos ganados equipo local** | 0.1 | La relación de victorias/derrotas de cada equipo es importante. Los partidos empatados tienen poca importancia, en cambio. |
| **Partidos empatados equipo local** | 0.025 |
| **Partidos perdidos equipo local** | 0.05 |
| **Partidos ganados equipo visitante** | 0.15 |
| **Partidos empatados equipo visitante** | 0.025 |
| **Partidos perdidos equipo visitante** | 0.01 |
| **Posición equipo local** | 0.3 | La posición en la que están ambos equipos es relevante. |
| **Posición equipo visitante** | 0.3 |
| **Goles a favor equipo local** | 0.2 | La racha de ambos equipos es importante.  En el equipo local es más relevante que tenga una mala racha, mientras que en el visitante es mejor que tenga una buena racha. |
| **Goles en contra equipo local** | 0.4 |
| **Goles a favor equipo visitante** | 0.4 |
| **Goles en contra equipo visitante** | 0.2 |
| **Resultado equipo local** | 0.25 | Los resultados de los partidos son importantes. |
| **Resultado equipo visitante** | 0.2 |

## Método de selección de los mejores casos

## Usamos 5NN, ya que usar valores por encima de un cierta α, no nos asegura que podamos obtener un resultado.

Tampoco tiene mucho sentido coger más de 5 casos, aunque si en la práctica vemos que son necesarios podemos ampliarlos.

## Método de predicción

Usamos la mayoría ponderada, ya que es mucho más relevante la similitud entre los casos que el número de casos a favor o en contra.

En este caso nos interesa más que se den condiciones muy similares al partido que estamos considerando.

## Medida de confianza

La medida de confianza que utilizamos es **la media de la similitud entre los casos favorables frente a la suma de la similitud de los k casos considerados**.