



Instituto Tecnológico de Costa Rica, Escuela de Computación, Bases de Datos II
Prof. Rodrigo Núñez N, proyecto II – 20%

Jugadores revelación

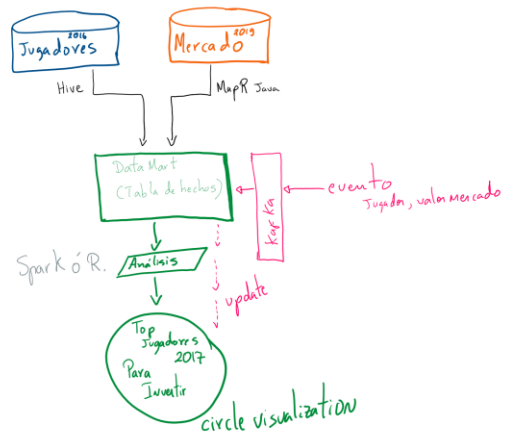
Usted ha sido designado para hacer un viaje en el tiempo para asesorar a los inversionistas de los equipos de futbol europeo en cuales jugadores invertir durante el 2017.

Para lograrlo podrá llevarse al pasado una base de datos que posee el [valor de mercado de los jugadores actualizada al 2019](#), esta base de datos le llamaremos *mercado*.

Usted viajará hasta octubre 2016, donde contará con los datos actualizados de rendimientos de cada jugador, resultados de los partidos, resultados de los clubes, capacidades y habilidades de los jugadores [desde el año 2008 al 2016](#), a esta base de datos la llamaremos *jugadores*.

Para realizar la predicción deberá hacer lo siguiente:

- Enfocarse en los jugadores cuya nacionalidad sean de un solo país de Europa, el de su elección
- Extraer la información que considere importante de los jugadores y el valor que tendrán en el 2019, tome en cuenta en los jugadores las tasas de crecimiento tanto de sus habilidades y records al 2016, como también la tasa de crecimiento de su valor de mercado
- Diseñe un datamart reducido sumalizando la información que considere valiosa para lograr el resultado que requieren los inversores
- Proceda a hacer la extracción, transformación y cargada de la información en el datamart usando la siguiente arquitectura:



- La información de la base de datos de jugadores será procesada usando hive, los datos de mercado serán procesados con MapR programados en Java
- El datamart puede ser un archivo o base de datos en el ecosistema de hadoop
- Dado que el datamart ya será una síntesis más compacta y directa para hacer el análisis final puede ser procesada con mucha más facilidad por algún otro medio, ya sea leyéndola directamente de hadoop o bien exportando el archivo
- Utilice la información del valor del mercado en el 2019 como un punto de referencia de pronóstico del valor de los jugadores a ese año y si ellos para ese año van creciendo o decreciendo su valor
- Utilice la información de las habilidades, los records deportivos entre otros para determinar si el jugador va de subida o de bajada, también puede usar datos como los rendimientos del club
- El modelo de análisis de datos que implemente, los datos que tome en cuenta para diseñar el modelo de análisis son a su criterio, no es necesario que tome en cuenta todas las variables o campos, puede ser más focalizado



- El modelo de análisis lo puede implementar ya sea en pyspark o con lenguaje R
- Los datos de salida del modelo de análisis van a requerir:
 - Jugador, club en el 2016, el valor pronosticado que va a tener en el 2017
- Visualice dicha información utilizando el siguiente [html template](#) de tal forma que los círculos de nivel superior serán el club y los de nivel interno los jugadores. El tamaño del círculo del club se determina por la sumatoria del valor pronosticado 2017 del mercado de sus jugadores, y el tamaño del círculo de cada jugador por el valor pronosticado de cada uno. Es posible ver en los círculos el nombre del club y del jugador.
- El proyecto requiere análisis, selección de datos, encontrar comportamiento en el tiempo y diseñar un cálculo que permita establecer un pronóstico de los jugadores para el siguiente año
- Para validar los resultados del modelo, observe que tantos aciertos tuvo con respecto a ciertos jugadores en la vida real de como fue el crecimiento de dichos jugadores que usted calculó

Otros aspectos

- El proyecto se puede realizar en parejas
- Cualquier sospecha de copia anulará el trabajo
- Fecha de entrega con cita de revisión sábado 16 de enero
- **Opcional por 20 puntos:** implementar un flujo por medio de **Kafka** donde se pueda producir un evento que actualice el valor de mercado de un jugador específico, dicho evento produce que se actualice el datamart y que se vuelva a ejecutar el análisis de los datos, actualizando con ello el pronóstico y la visualización