



UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICO



Asignatura:

MINERÍA DE DATOS.

Tema:

AVANCE 1 – PROYECTO INTEGRADOR.

Datos de la materia:

Profesora: Mayra Cristina Berrones Reyes.

Grupo 002. Miércoles 19:00hrs-22:00hrs.

Datos del equipo:

EQUIPO 9.

Saúl Arath Hernández Hernández Mat. 1815642

Alison Leonor Roldán Sánchez Mat. 1806488

José Ignacio Ruíz Contreras Mat. 1887848

Fecha: 28/10/20

1. Título de la base de datos a trabajar: *players_stats_by_season_full_details*.

- a) Nombre con el que se encuentra en la página:
Basketball Players Stats per Season – 49 leagues.
- b) URL de la página: <https://www.kaggle.com/jacobbaruch/basketball-players-stats-per-season-49-leagues>

2. Descripción de los datos.

- a) Tipo de datos: 19 columnas tipo entero, 10 columnas tipo cadena de texto, 1 columna tipo decimal y 1 columna de otro tipo no especificado.
- b) Descripción de las columnas (información relacionada a manera individual del jugador):

Parámetro	Descripción
Season	Temporada en que jugó (yyyy – yyyy)
Stage	Escenario en que jugó (International, NBA: Playoffs, Regular_Season)
Player	Nombre completo del jugador
Team	Nombre del equipo
GP	# Juegos jugados
MIN	# Minutos jugados
FGM	# Tiros de campo anotados (2 y 3 puntos)
FGA	# Intentos de tiro de campo (2 y 3 puntos)
3PM	# Tres puntos hechos
3PA	# Intentos de tres puntos
FTM	# Tiros libres realizados
FTA	# Intentos de tiros libres
TOV	# Pérdidas de balón

PF	# Falta personal
ORB	# Rebotes ofensivos
DRB	# Rebotes defensivos
REB	# Rebotes
AST	# Asistencias
STL	# Robos
BLK	# Bloqueos
PTS	# Puntos
birth_year	Año de nacimiento
birth_month	Mes de nacimiento
birth_date	Fecha de cumpleaños
Height	Estatura [Pies]
height_cm	Estatura [Centímetros]
Weight	Peso [Libras]
weight_kg	Peso [Kilogramos]
Nationality	Nacionalidad
high_school	Escuela preparatoria a la que asistió el jugador

3. Justificación del uso de datos.

- a) Las características que más nos llamaron la atención de los datos es que trabaja con varias regiones del mundo, no solo se centra en ligas americanas como lo es la NBA, si no que nos da la posibilidad de analizar alrededor de diferentes regiones las estadísticas de los jugadores de Basketball a lo largo de las temporadas, de igual manera, notamos que cuenta con una gran cantidad de información relacionada al jugador con la que podríamos obtener tanto un perfil ideal del jugador como alguna otra herramienta útil. El factor decisivo que nos hizo querer trabajar con esta base

de datos fue tanto la posible sencillez como los planteamientos de problemas que obtuvimos mediante el contexto que nos brindaba la base de datos, podemos plantear un problema desde diferentes puntos de vista principalmente de manera deportiva.

- b) El principal beneficio que pensamos tener al trabajar con esta base de datos, son las diferentes técnicas de minería que podemos aplicar como mencionaremos en los puntos siguientes, de igual manera, a manera personal uno de los beneficios que podríamos tener es una herramienta con la cual podremos identificar a los jugadores más valiosos o equipos que por estadísticas sean más probables de ganar y así realizar apuestas de una manera más segura, viendo un beneficio fuera de lo personal sería aplicado a la directiva de un equipo profesional en donde mediante el análisis propuesto, podamos establecer un plan de logística con los resultados obtenidos en donde podamos buscar a jugadores de elite en diferentes regiones del mundo y así crear una liga más competitiva o crear más valor sobre el equipo, si lo vemos a modo empresarial, tendría muchas aplicaciones de donde podamos sacar provecho a estos datos.

4. Planteamiento del problema.

- a) Problemáticas:
 - a. Un apostador desea buscar una herramienta con la cual pueda acertar de manera más segura sus decisiones, las apuestas constan sobre que jugador será MVP en la temporada.
 - b. La directiva de un equipo de la NBA desea renovar su plantilla de jugadores, para ello busca a los mejores jugadores, su decisión se basará respecto a los rendimientos de los jugadores en sus equipos actuales.
 - c. Una organización de visorias de jugadores jóvenes de basketball quiere saber de qué escuelas han salido los mejores jugadores de sus respectivas ligas para en un futuro, tener los ojos sobre las futuras promesas y reclutarlos.
 - d. Un multimillonario busca crear una liga de basketball de elite, por lo que necesita saber en qué región sé es más competitivo para establecer su liga en esa región se basa en los resultados individuales de los jugadores respecto a sus estadísticas.

5. Objetivo final.

A) Principal:

Encontrar cual es la región con mejor desempeño en el deporte, sin tomar en cuenta la liga americana NBA, para poder saber qué características debe tener un jugador para estadísticamente hablando, tener un alto nivel.

B) Secundarios:

- a) Poder determinar cuáles son las estadísticas más destacadas dependiendo de la región en la que juega cada deportista.
- b) Averiguar si hay una variable relevante como la estatura, el peso o incluso el año de nacimiento que afecte directamente a las estadísticas en la base de datos.
- c) Corroborar si realmente es un factor tener las mejores estadísticas para poder ganar un campeonato.
- d) Destacar por región la estadística con mayor volumen.

6) Planeación de la herramienta a utilizar.

Para poder comprender los datos y a primera instancia notar si existe alguna tendencia nos apoyaremos con las herramientas de las técnicas de **visualización**.

Ya con una mayor claridad al momento de ver los datos, podremos **clasificarlos**. Utilizaremos dentro de las técnicas de predicción la técnica de árbol de decisión, para poder encontrar cuál de las variables tiene mayor importancia para cierta estadística.

De igual manera de la rama de predicción buscaremos crear un modelo de **regresión** múltiple que al momento de brindar la nacionalidad y las características físicas del jugador podamos determinar el nivel de sus estadísticas.