



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN



FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN
FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

Minería de datos

Avance I - PIA

Daniel de la Rosa Coss
Juan Alfredo Cantú Zavala
Wendy Olivia Bazúa Corrales



1. TÍTULO DE LA BASE DE DATOS:

a) Nombre con el que se encuentra en la página encontrada.

FIFA 20 complete player dataset

b) URL de la página

<https://www.kaggle.com/stefanoleone992/fifa-20-complete-player-dataset>

2. DESCRIPCIÓN DE LOS DATOS:

a) Que tipo de datos son (tablas, imágenes, encuestas, etc)

Tablas con datos de tipo texto y número que indican características específicas de los jugadores que van desde datos personales hasta el nivel de habilidades con las que cuenta.

b) Descripción de las columnas (Que datos se describen en cada columna. Mediciones, tiempo, texto largo, etc.)

- sofifa_id: número único en fifa (id del jugador) -
- player_url: url del jugador en sofifa - link
- short_name: nombre corto del jugador. - texto
- long_name: nombre largo del jugador - texto
- age: edad del jugador - número.
- dob: fecha de nacimiento - fecha
- height_cm: altura en centímetros del jugador. - número
- weight_kg: peso en kilogramos del jugador. - número
- nationality: nacionalidad del jugador. -texto
- club: club para el que juega el jugador. -texto
- overall: calificación del jugador en el fifa. -número
- potential: calificación potencial del jugador. -número
- value_eur: valor en euros del jugador. -número
- wage_eur: salario en euros del jugador. -número
- player_positions: posiciones del jugador - texto.
- preferred_foot: pierna preferida del jugador - texto
- international_reputation: atributo de reputación internacional - número
- weak_foot: atributo del pie debil - número
- skill_moves: atributo de calidad de movimiento - número
- work_rate: atributo de productividad - texto
- body_type: tipo de cuerpo - texto
- real_face: si se tiene o no escaneada la cara real del jugador - booleano
- release_clause_eur: cláusula de rescisión en euros - número

- player_tags: etiquetas del jugador - texto.
- team_position: posicion del jugador en el equipo - texto.
- team_jersey_number: número del jugador - número
- loaned_from: que club prestó al jugador - texto.
- joined: fecha de llegada al club - fecha.
- contract_valid_until: fecha de vencimiento del contrato (año) - número.
- nation_position: posicion del jugador en su selección nacional - texto.
- nation_jersey_number: número del jugador en su selección nacional - número.
- pace: nivel del ritmo del jugador - número.
- shooting: nivel del tiro del jugador - número.
- passing: nivel de pase del jugador - número.
- dribbling: nivel de dribles del jugador - número.
- defending: nivel de defensa del jugador - número.
- physic: nivel del físico del jugador - número.
- gk_diving: nivel de lanzadas del portero - número.
- gk_handling: nivel de manejo con las manos del portero - número.
- gk_kicking: nivel de patada del portero -número.
- gk_reflexes: nivel de reflejos del portero - número.
- gk_speed: velocidad del portero - número:
- gk_positioning: nivel de posicionamiento del portero - número.
- player_traits: rasgos del jugador - texto.
- attacking_crossing: nivel de cruce del jugador - número.
- attacking_finishing: nivel de finalización del jugador - número.
- attacking_heading_accuracy: nivel de cabeceo del jugador - número.
- attacking_short_passing: nivel de pase corto del jugador - número.
- attacking_volleys: nivel de tiro del jugador - número.
- skill_dribbling: nivel de habilidad de dribles del jugador - número.
- skill_curve: nivel de habilidad de curva del jugador - número.
- skill_fk_accuracy: nivel de habilidad para cobrar tiros libres del jugador - número.
- skill_long_passing: nivel de habilidad de pase largo del jugador - número.
- skill_ball_control: nivel de habilidad de control del balón del jugador - número.
- movement_acceleration: nivel de aceleración del jugador - número.
- movement_sprint_speed: nivel de sprint del jugador - numero.
- movement_agility: nivel de agilidad del jugador - numero.
- movement_reaction: nivel de reacción del jugador - número.

- movement_balance: nivel de balance del jugador - número.
- power_shot_power: poder de tiro del jugador - número.
- power_jumping: poder de salto del jugador - número.
- power_stamina: poder de estamina del jugador - número.
- power_strength: fuerza del jugador - numero.
- power_long_shots: nivel de tiros largos del jugador - número.
- mentality_aggression: nivel de mentalidad del jugador - número.
- mentality_interceptions: nivel de intercepciones del jugador - número.
- mentality_positioning: nivel de colocación del jugador - número.
- mentality_vision: nivel visión del jugador - número.
- mentality_penalties: nivel de cobro de penales del jugador - número.
- mentality_composure: nivel de compostura del jugador - número.
- defending_marking: nivel de defensa del jugador - número.
- defending_standing_tackle: nivel de entrada de pie del jugador - número.
- defending_sliding_tackle: nivel de barrida del jugador - número.
- goalkeeping_diving: nivel de lanzadas del portero - número.
- goalkeeping_handling: nivel de manejo con las manos del portero - número.
- goalkeeping_kicking: nivel de patada del portero -número.
- goalkeeping_reflexes: nivel de reflejos del portero - número.
- ls: nivel del jugador en la posición de delantero por la izquierda - texto.
- st: nivel del jugador de centro delantero - texto.
- rs: nivel del jugador de delantero por derecha -texto.
- lw: nivel del jugador de extremo izquierdo - texto.
- lf: nivel del jugador como volante por la izquierda - texto.
- cf: nivel del jugador como mediocampista ofensivo - texto.
- rf: nivel del jugador volante por derecha - texto.
- rw: nivel del jugador como extremo derecho-texto.
- lam: nivel del jugador de mediocampista por derecha - texto.
- cam: nivel del jugador de mediocampista central - texto.
- ram: nivel del jugador como medio por derecha - texto.
- lm: nivel del jugador como mediocampista por izquierda - texto.
- lcm: nivel del jugador como mediocentro por izquierda -texto.
- cm:nivel del jugador como mediocentro - texto.
- rcm:nivel del jugador como medio centro por derecha - texto.
- rm: nivel del jugador como medio por derecha - texto.

- lwb: nivel del jugador como lateral izquierdo - texto.
- ldm: nivel del jugador como medio defensivo izquierdo - texto.
- cdm: nivel del jugador como medio defensivo centro - texto.
- rdm: nivel del jugador como medio defensivo por derecha - texto.
- rwb: nivel del jugador como lateral por la derecha - texto.
- lb: nivel del jugador como defensa por izquierda - texto.
- lcb: nivel del jugador como defensa central por izquierda - texto.
- cb: nivel del jugador como defensa central - texto.
- rcb: nivel del jugador como defensa central por derecha - texto.
- rb: nivel del jugador como defensa por derecha.

3. JUSTIFICACIÓN DEL USO DE DATOS:

a) Cuales son las características que les llamaron la atención de los datos. Que les hizo querer trabajar con ellos.

De las cosas que más nos gustó de los datos fue que nos parecieron bastante completos para cada jugador, y con todos estos datos tenemos la posibilidad de hacer una búsqueda bastante completa, planteando distintos escenarios en los cuales busquemos jugadores con las características que se ajusten a nuestras necesidades.

b) Que beneficio encuentran en trabajar con esos datos.

Que las habilidades de los jugadores están expresados en números, es decir, esta habilidad tiene una calificación del 1 al 100, entonces tener un número nos parece más útil que tener opiniones en texto ya que eso puede ser más difícil de descifrar que una calificación, el hecho de que esté expresado en número nos permite identificar de una manera más rápida si un jugador cuenta con una característica de su juego bien desarrollada, a que si tuviéramos otro tipo de valores como opiniones de expertos en donde tendríamos que aislar palabras clave como “bueno”, “malo”, etc., aunado a esto como mencionamos en el punto anterior son muchas las características evaluadas de cada jugador lo que nos amplía el panorama para buscar y encontrar a un jugador con las características y habilidades de juego que cumplan con nuestras necesidades.

4. PLANTEAMIENTO DEL PROBLEMA:

a) Problemática

- Objetivo: Ayudar en la toma de decisión de un equipo de fútbol para comprar un jugador respecto a sus necesidades.
- Problema: Hay demasiados jugadores de fútbol y a diversos precios por lo que el equipo no podría hacer una investigación totalmente completa para resolver y no todos los jugadores que buscan en base a sus características podrían quedar fuera de su presupuesto.
- Solución: Desarrollar un programa el cual realice una clasificación de jugadores en base a las características que pide el equipo.

b) Mejoras

Nos gustaría mejorar la manera en que se toman ciertas decisiones en las cuales se toman en cuenta distintas características ya que hay decisiones que podemos tomar resolviendo preguntas de si y no, o bien, buscando una o más características específicas, en este proceso existen distintos factores que pueden llevarnos a tomar una decisión precipitada o en su defecto errónea por ejemplo la cantidad de datos o características que se tengan, errores por omisión (desde no tomar en cuenta un elemento que cumpla con las características que queremos o no considerar uno que sí lo haga), sesgo, etc., sabemos que nosotros como humanos tenemos algunas desventajas si nos comparamos con una computadora la principal es la eficiencia en el procesamiento de datos, en vista de eso nos gustaría usar esa eficiencia a nuestro favor para ayudarnos tomar una decisión que considere todo los elementos con los que contamos y con todas las especificaciones que necesitemos.

5. OBJETIVO FINAL:

a) Principal:

Desarrollar un programa el cual ayude en la toma de decisiones para cuestiones en las que se busca la mejor opción que pueda suplir una necesidad de la mejor manera, no solo ajustándose a un presupuesto, sino que también busque las características deseadas y pueda entregar múltiples opciones.

b) Secundario:

Llevar dicho programa no solo a la cuestión de compra de jugadores sino también a otras áreas en las que igual pueda funcionar y ayudar a las personas como por ejemplo en asesorías financieras ayudando a las personas a comprar un auto que se ajuste no solo a su presupuesto, sino que también tome en cuenta los gustos de las personas.

6. PLANEACIÓN DE LA HERRAMIENTA A UTILIZAR:

a) Tipo de técnica a utilizar:

Dado a que nuestro objetivo es encontrar varios prospectos para el equipo la mejor técnica que se acopla a nuestro objetivo es la técnica de clustering, debido a que nuestra base de datos es grande y tiene demasiadas variables que describen a cada jugador y diferenciándolo de la otra técnica que es clasificación, el clustering es una técnica descriptiva y es lo que buscamos, la descripción de jugadores que se basan en lo que se necesita, mientras que la clasificación es para valores predictivos, cosa que no buscamos.

b) Algoritmo o herramienta:

- Clustering Jerárquico: Los clusters se ordenan jerárquicamente de forma que los objetos que pertenecen a un cluster también pertenecen a su cluster padre.
- Clustering con superposiciones: un objeto puede pertenecer a más de un cluster: Nos pareció bien esta técnica porque para tomar decisiones en ocasiones solo necesitamos saber qué queremos y en vista de esto es tan sencillo como clasificar nuestros elementos en tantos subgrupos, o en este caso clusters, como características que busquemos, por lo que estableceremos clusters que tengan características específicas dentro del cluster principal o cluster padre.