

GRADO EN CIENCIA DE DATOS



VNIVERSITAT
DE VALÈNCIA

TRABAJO FIN DE GRADO

ANÁLISIS DE LAS COMPETENCIAS DEL
FUTBOLISTA ÉLITE DE LA ACADEMIA DE LOS
CASOS DE ÉXITO EN FÚTBOL PROFESIONAL

AUTOR:
PERE CAIMARI FUSTER

TUTOR:
VALERO LAPARRA PÉREZ MUELAS



VNIVERSITAT
DE VALÈNCIA [Q] Escola Tècnica Superior
d'Enginyeria ETSE-UV

TRABAJO FIN DE GRADO

ANÁLISIS DE LAS COMPETENCIAS DEL FUTBOLISTA ÉLITE DE LA ACADEMIA DE LOS CASOS DE ÉXITO EN FÚTBOL PROFESIONAL

AUTOR: PERE CAIMARI FUSTER

TUTOR: VALERO LAPARRA PÉREZ MUELAS

Declaración de autoría:

Yo, Pere Caimari Fuster, declaro la autoría del Trabajo Fin de Grado titulado “Análisis de las competencias del futbolista Élite de la Academia de los casos de éxito en fútbol profesional” y que el citado trabajo no infringe las leyes en vigor sobre propiedad intelectual. El material no original que figura en este trabajo ha sido atribuido a sus legítimos autores.

Valencia, 18 de junio de 2025

Fdo: Pere Caimari Fuster

Resumen:

El fútbol es un deporte que apasiona a millones de personas entre las cuales me encuentro y la ciencia de datos me ha permitido acercarme a él desde un punto de vista más analítico y observador. En este ámbito hemos intentado destacar la importancia de las academias en cuanto a la transmisión de valores y el ahorro de dinero en fichajes al no tener que ir a buscarlos en un mercado cada vez más caro y competitivo donde el talento se centra en unos pocos clubes. Entender cuáles son los factores que potencian a los jugadores hasta el profesionalismo es clave en este proceso y es aquí donde se enmarca este trabajo.

Empezamos el trabajo entendiendo el contexto de la Academia VCF, que es el talento en el fútbol, cuáles son sus dimensiones y que se considera éxito en este campo. Una vez definidas las bases teóricas del estudio nos planteamos cuál era la información que queríamos obtener y cuál teníamos a nuestro alcance, para ello integramos datos de WyScout y Transfermarkt para construir una base de datos que nos permita resolver las cuestiones que nos planteamos en base a la metodología establecida. El siguiente reto se trata de entender y construir los modelos necesarios para dar respuestas a las preguntas iniciales y encontrar los resultados que nos confirman o rechazan nuestras hipótesis. Finalmente tras el análisis de los resultados y contrastarlos con opiniones de expertos en la materia hemos llegado a una serie de conclusiones que junto a las limitaciones del estudio y el trabajo futuro cierran el trabajo.

Abstract:

Football is a sport that millions of people are passionate about, and data science has allowed me to approach it from a more analytical and observational point of view. In this area we have tried to highlight the importance of academies in terms of transmitting values and saving money on transfers by not having to look for them in an increasingly expensive and competitive market where talent is concentrated in a few clubs. Understanding what are the factors that drive players to professionalism is key in this process and this is where this work is framed.

We started the work by understanding the context of the VCF Academy, what is talent in football, what are its dimensions and what is considered successful in this field. Once we had defined the theoretical basis of the study, we decided what information we wanted to obtain and what we had at our disposal. To do this, we integrated data from WyScout and Transfermarkt to build a database that would allow us to resolve the questions we had raised based on the established methodology. The next challenge is to understand and build the models needed to answer the initial questions and find the results that confirm or reject our hypotheses. Finally, after analysing the results and contrasting them with the opinions of experts in the field, we have reached a series of conclusions which, together with the limitations of the study and future work, bring the work to a close.

Resum:

El futbol és un esport que apassiona milions de persones entre les quals me trob i la ciència de dades m'ha permès acostar-m'hi des d'un punt de vista més analític i observador. En aquest àmbit hem intentat destacar la importància de les acadèmies quant a la transmissió de valors i l'estalvi de diners en fitxatges en no haver dánar a buscar-los en un mercat cada cop més car i competitiu on el talent se centra en uns quants clubs. Entendre quins són els factors que potencien els jugadors fins al professionalisme és clau en aquest procés i és aquí on sémmarca aquesta feina.

Comencem el treball entenent el context de la Acadèmia VCF, que és el talent al futbol, quines són les seves dimensions i que es considera èxit en aquest camp. Un cop definides les bases teòriques de l'estudi ens vam plantejar quina era la informació que volíem obtenir i quina teníem a la nostre disposició, integrem dades de WyScout i Transfermarkt per construir una base de dades que ens permeti resoldre les qüestions que ens plantegem en base a la metodologia establerta. El repte següent es tracta d'entendre i construir els models necessaris per donar respostes a les preguntes inicials i trobar els resultats que ens confirmen o rebutgen les nostres hipòtesis. Finalment, després de l'anàlisi dels resultats i contrastar-los amb opinions d'experts en la matèria, hem arribat a una sèrie de conclusions que juntament amb les limitacions de l'estudi i el treball futur tanquen el treball.

Agradecimientos:

En primer lugar quería agradecer a mi hermano Joan por inculcarme los valores del mundo del fútbol, entender que la vida es un proceso de subidas y bajadas y que somos tan fuertes por cuantas veces nos levantamos y no por cuantas caemos. A mis padres por apoyarme en todo momento y respaldarme sin dudar en ningún momento acerca de mi y a mis amigos por hacerme ver su admiración y pasión por del deporte.

En segundo lugar a Álvaro Ramos por tutorizarme y corregirme en todo momento hacia un camino de profesionalidad, respeto y hacerme ver la importancia de representar a un club tan grande como el Valencia Club de Fútbol S.A.D. También a mis compañeros (Vicent Gilabert, Nando Ros, Javi Ortega y Ángel Murciano) por enseñarme una nueva forma de entender el fútbol y por consiguiente la vida.

Por último pero no menos importante a Valero Laparra por creer siempre de forma ciega en mis ideas, aceptar el reto en todo momento, orientarme y sacar lo mejor de mi como científico de datos.

Pere Caimari Fuster

Índice general

1. Introducción	19
1.1. Contexto y justificación	19
1.2. Objetivos	22
2. Marco Teórico	23
2.1. Introducción al marco teórico	23
2.2. Academia del Valencia CF	23
2.3. Identificación y definición del talento en el fútbol	24
2.4. Estudios previos sobre factores de éxito en academias deportivas	27
2.5. Criterios para la consideración de una competición como profesional	28
3. Herramientas y técnicas de análisis de datos	31
3.1. Fuentes de datos y técnicas de extracción	31
3.1.1. WyScout	31
3.1.2. Transfermarkt	32
3.1.3. Web scraping	32
3.2. Introducción a la inteligencia artificial y aprendizaje automático	32
3.3. Modelos de clasificación utilizados	33
3.3.1. Regresión logística	33
3.3.2. Random Forest	34
3.3.3. Máquinas de vectores de soporte (SVM)	34
3.3.4. Redes neuronales artificiales	35
3.4. Ajuste de hiperparámetros y métricas de evaluación	36
4. Metodología	39
4.1. Construcción de la base de datos final	39
4.2. Variables analizadas	40
4.3. Descripción de la muestra	41
5. Análisis de datos y resultados	45

5.1. Resultados	45
5.1.1. Configuraciones óptimas de los modelos	45
5.1.2. Rendimiento de los modelos	46
5.1.3. Importancia de las variables	48
5.2. Discusión	51
6. Conclusiones y proyección futura	53
6.1. Conclusiones	53
6.2. Limitaciones del estudio	54
6.3. Trabajo Futuro	55
A. Anexos	57
A.1. Listado de Competiciones Consideradas Profesionales	57
A.2. Variables originales	59
A.3. Preguntas del formulario	63
Bibliografía	65

Índice de figuras

3.1. Plataformas de datos futbolísticos utilizadas.	32
4.1. <i>Distribución de jugadores según éxito profesional.</i> Fuente: Elaboración propia a partir de los datos de la base de datos.	42
4.2. <i>Distribución de jugadores según posición de juego.</i> Fuente: Elaboración propia a partir de los datos de la base de datos.	43
4.3. <i>Distribución de jugadores por edad.</i> Fuente: Elaboración propia a partir de WyScout.	43
5.1. <i>Variables más importantes de la SVM.</i> Fuente: Elaboración propia.	48
5.2. <i>Curvas ROC de los diferentes modelos.</i> Fuente: Elaboración propia.	49
5.3. <i>Matrices de confusión de los diferentes modelos.</i> Fuente: Elaboración propia.	50
5.4. <i>Variables más importantes de la Regresión Logística.</i> Fuente: Elaboración propia.	50
5.5. <i>Variables más importantes del Random Forest con todas las variables.</i> Fuente: Elaboración propia.	51

Capítulo 1

Introducción

1.1. Contexto y justificación

La pasión por el fútbol es algo que comparto con millones de personas alrededor del mundo y, precisamente por ello, este deporte despierta emociones intensas y crea conexiones profundas entre los aficionados. Hoy en día, el mundo del fútbol profesional refleja cómo ha evolucionado hacia una gran industria, con ingresos que provienen de la televisión, patrocinios, entradas a los estadios y mucho más. Las ligas más emblemáticas de Europa, como la Premier League, LaLiga y otras, no solo ofrecen espectáculo, sino que también son la aspiración de los jugadores más talentosos.

En los últimos años, el modelo económico del fútbol ha evolucionado significativamente. La inyección de capital por parte de inversores privados, fondos de inversión y empresas ha permitido a algunos clubes incrementar sus recursos y mejorar sus infraestructuras. Sin embargo, esta tendencia también ha generado una creciente desigualdad económica entre equipos, lo que ha llevado a una concentración de talento en los clubes con mayor capacidad financiera. Las normativas financieras, como el Fair Play Financiero de la UEFA, buscan equilibrar la competencia y garantizar la sostenibilidad de los clubes, aunque su efectividad sigue siendo objeto de debate.

Estos hechos, junto a la pandemia de COVID-19, han llevado a muchos clubes, especialmente aquellos con menos recursos, a replantear sus modelos económicos en busca de una mayor sostenibilidad a largo plazo. La creciente desigualdad financiera, la volatilidad de los ingresos y el aumento de los costes han evidenciado la necesidad de estrategias que garanticen estabilidad más allá de los resultados deportivos.

En este contexto, las academias de fútbol profesionales han adquirido una importancia creciente, no solo como fuente de talento deportivo, sino también como una estrategia económica para los clubes. La formación y promoción de jugadores jóvenes permite a las entidades reducir costos en fichajes y, en algunos casos, obtener ingresos significativos mediante la venta de futbolistas a otros equipos. Esto resalta la necesidad de un enfoque basado en el desarrollo de competencias clave en los jugadores de élite, un aspecto fundamental para la competitividad y sostenibilidad del fútbol profesional.

Debido a ello, las academias de fútbol desempeñan un papel crucial en la formación de jugadores de élite, proporcionando una estructura organizada para el desarrollo deportivo y personal de los jóvenes talentos. En un entorno donde la competencia es cada vez más exigente, estas instituciones se convierten en el punto de partida ideal para que los clubes

puedan identificar, nutrir y preparar a futuros profesionales desde edades tempranas.

En este sentido, el verdadero desafío de las academias va más allá de descubrir talento: consiste en formar futbolistas completos, capaces de rendir al más alto nivel tanto dentro como fuera del campo. Para ello, cada academia aplica sus propios métodos y filosofías de trabajo, aunque casi todas coinciden en la importancia de aspectos transversales como la nutrición, la gestión emocional o la toma de decisiones, claves en el crecimiento integral del jugador.

En términos económicos, las academias representan una inversión estratégica para los clubes. Desarrollar talento propio puede significar una reducción en los costos de fichajes y, al mismo tiempo, generar ingresos mediante la venta de jugadores formados en la cantera. Ejemplos como La Masia del FC Barcelona, históricamente la academia del Ajax o la cantera del Benfica han demostrado cómo un enfoque sólido en el desarrollo juvenil puede traducirse en éxito tanto deportivo como financiero.

Otro aspecto clave de las academias es su capacidad para proporcionar una transición estructurada al fútbol profesional. Los jugadores formados en estos entornos suelen tener una mayor adaptación a la cultura y filosofía del club, facilitando su integración en los primeros equipos. Además, estas instituciones ofrecen oportunidades a jóvenes de distintas procedencias, democratizando el acceso al fútbol profesional y permitiendo que el talento, más allá de los recursos económicos familiares, pueda ser detectado y desarrollado.

Por consiguiente, las academias de fútbol no solo son esenciales para la formación de jugadores de élite, sino que también representan una estrategia clave para la estabilidad y sostenibilidad del fútbol profesional. La inversión en el desarrollo juvenil garantiza una base sólida para el futuro del deporte y refuerza la identidad y competitividad de los clubes en el escenario global.

El fútbol moderno exige un conjunto de competencias clave que distinguen a los jugadores de élite y les permiten destacar en el más alto nivel competitivo. Estas competencias no solo incluyen habilidades técnicas y tácticas, sino también capacidades físicas, cognitivas y psicológicas que optimizan el rendimiento individual y colectivo en el terreno de juego.

A nivel general y para no extendernos en demasia, vamos a dividir al futbolista en 4 grandes grupos:

- Físico

- En el fútbol moderno, el aspecto físico es clave porque influye directamente en el rendimiento, la resistencia y la prevención de lesiones. La velocidad, la fuerza, la agilidad y la capacidad aeróbica permiten a los jugadores mantener una alta intensidad durante todo el partido, responder rápidamente a cambios en el juego y soportar el contacto físico. Además, la preparación física optimiza la recuperación y la durabilidad en una temporada exigente, lo que marca la diferencia en la élite.

- Técnico

- El aspecto técnico en el fútbol se refiere al dominio del balón y la capacidad de realizar movimientos específicos de manera eficiente. Es fundamental para el desarrollo de futbolistas profesionales, ya que permite un rendimiento óptimo, da una ventaja competitiva, facilita la adaptación a diferentes estilos de juego

y ayuda a prevenir lesiones. Una buena técnica es la base para desarrollar otras áreas del juego, como la táctica, la física y la mental.

■ Táctico

- El aspecto táctico en el fútbol engloba la comprensión y aplicación de estrategias colectivas e individuales dentro del juego. Su importancia radica en la capacidad del jugador para posicionarse correctamente, interpretar el juego, anticipar movimientos y adaptarse a distintos sistemas y situaciones. Un futbolista con inteligencia táctica maximiza su rendimiento y el del equipo, optimizando esfuerzos y contribuyendo a la eficacia del planteamiento del entrenador.

■ Mental

- El aspecto mental en el fútbol se refiere a la capacidad de los jugadores para controlar sus emociones, pensamientos y actitudes en el campo y fuera de él. Esto incluye la concentración, la confianza, la motivación, la resiliencia, la disciplina, el manejo de la presión y la capacidad de trabajar en equipo. Es la fortaleza mental que permite a los jugadores rendir al máximo bajo cualquier circunstancia.

Uno de los puntos más críticos y complicados de las carreras de los futbolistas es el paso del fútbol formativo al fútbol competitivo, este suele producirse entre los 18 y 21 años, una edad muy determinante para su futuro profesional en la que deben desarrollar cualidades como la sensatez o la madurez para ser resilientes ante las dificultades que se vayan encontrando y constantes en el trabajo. Además de la predisposición de cada uno por separado, también intervienen una serie de factores que son incontrolables, como las lesiones, la situación familiar del futbolista, el contexto del equipo/liga y muchos otros que afectan a la carrera del deportista.

Llegar a convertirse en futbolista profesional es un camino lleno de obstáculos, donde no siempre basta con talento y trabajo. Son muchos los casos de jugadores que, a priori, lo tienen todo para alcanzar la élite, pero que ven truncado su sueño por una lesión inesperada, una decisión poco acertada al cambiar de equipo o incluso por circunstancias personales que terminan afectando a su rendimiento. Y el reto no acaba ahí. Incluso aquellos que logran dar el salto al fútbol profesional deben enfrentarse a la dificultad de mantenerse en lo más alto, expuestos a los mismos riesgos que, en cualquier momento, pueden poner fin a una trayectoria que parecía consolidada.

Por todo esto, se considera la última etapa formativa la más compleja de todas, en la que el jugador pasa a tener una exigencia directa sobre sus decisiones y el entorno en el que compite se valora únicamente el resultado y su actuación.

En este contexto, cobra especial relevancia la necesidad de definir qué entendemos por éxito en el fútbol formativo. Tradicionalmente, dicho éxito ha estado vinculado a la consecución de un contrato profesional y a la participación en ligas reconocidas, ya que constituyen una validación objetiva del rendimiento del jugador y del trabajo de la academia. Esta definición no solo permite evaluar la eficacia de los procesos de formación, sino que también aporta criterios medibles para analizar la transición efectiva entre el fútbol base y el profesional, elemento central en este trabajo de investigación.

1.2. Objetivos

El fútbol, más que un deporte, es un reflejo de la dedicación, el esfuerzo y el talento que se cultiva desde edades tempranas. En este sentido, la academia del Valencia CF ha sido un semillero de jóvenes promesas que, con el paso del tiempo, buscan dar el salto al fútbol profesional. El objetivo de este trabajo es indagar en las competencias clave que han permitido a algunos de estos futbolistas de élite alcanzar su sueño. Al identificar los patrones comunes en los casos de éxito, se busca proporcionar a los técnicos y profesionales del club herramientas prácticas para influir de manera positiva en el desarrollo de los jóvenes talentos. La meta última de este análisis es clara: que más jugadores de la academia logren llegar al fútbol profesional, contribuyendo, además, a la sostenibilidad económica del club al generar un retorno a través de jugadores formados internamente, reduciendo la dependencia de fichajes externos y asegurando una rentabilidad en el largo plazo.

Otros objetivos particulares que se plantean en este estudio son:

1. Condensar de la mejor forma posible el rendimiento de un jugador

Es posible capturar el rendimiento de un jugador en un partido o a lo largo de una temporada mediante la recogida de variables objetivas, lo realmente complicado es identificar para cada contexto en específico cuáles son las más relevantes que nos ayuden a describir de la mejor forma posible cómo se desenvuelve un jugador sobre el terreno de juego.

2. Integrar diferentes fuentes de datos

Uno de los objetivos más técnicos de este trabajo es la integración de datos de diferentes fuentes, formatos y dimensiones. Durante el proyecto vamos a trabajar con datos estructurados como registros de pases, tiros, etc., datos semiestructurados como los físicos del GPSs o los informes y no estructurados como las imágenes y los vídeos individuales. Pensar en una forma de cómo juntar todos esos datos en un formato adecuado para su estudio es ya un gran reto de por sí que vamos a tratar de resolver.

3. Aprovechamiento de IA y ML en entornos formativos

Muchos de los procesos que se llevan a cabo dentro de las diferentes academias de los clubes profesionales generan una gran cantidad de datos a los que no se les saca partido o no se sabe como almacenarlos y estructurarlos. Uno de mis objetivos con este proyecto es dar a conocer cómo una correcta gestión y almacenaje de dicha información puede llevar en estudios y trabajos de valor tan provechoso como el que queremos desarrollar.

4. Exposición precisa y concisa de los logros obtenidos

Dado que muchos profesionales en el ámbito futbolístico no están familiarizados con el análisis de datos y la estadística, es fundamental simplificar la información que se comparte. Esto asegura una comunicación efectiva y un entendimiento común entre todas las partes.

Capítulo 2

Marco Teórico

2.1. Introducción al marco teórico

Las academias de fútbol juvenil se han consolidado como pilares fundamentales en el desarrollo de futbolistas de élite, actuando como conductos esenciales para nutrir el talento hacia los clubes profesionales y las selecciones nacionales. En un panorama futbolístico cada vez más competitivo y profesionalizado, la inversión y la atención dedicada a la estructuración del desarrollo de jóvenes promesas dentro de estos sistemas académicos ha experimentado un crecimiento significativo. Por consiguiente, resulta imprescindible establecer un marco teórico robusto que permita comprender las complejidades inherentes a estas instituciones y los múltiples factores que contribuyen a su eficacia en la formación de jugadores de alto rendimiento.

El presente análisis se centrará en tres temas principales e interconectados que son cruciales para la comprensión del desarrollo del fútbol juvenil: el caso específico de la academia del Valencia CF, la definición y los métodos de identificación del talento en el fútbol, y el amplio espectro de investigaciones previas que han explorado los factores determinantes del éxito dentro de las academias deportivas. Cada uno de estos temas ofrece perspectivas y conocimientos únicos que, al ser examinados de manera conjunta, pueden proporcionar una comprensión más profunda y matizada del proceso de formación de futbolistas de élite.

2.2. Academia del Valencia CF

La elección de la academia del Valencia CF como objeto de estudio se justifica por su consistente reconocimiento como una de las canteras más prolíficas y exitosas de Europa. Su trayectoria a lo largo del tiempo y su metodología de trabajo, que ha sido reconocida a nivel nacional e internacional, la convierten en un caso de estudio valioso para analizar las prácticas efectivas en el desarrollo del fútbol juvenil [España, 2024, SEE, 2024].

1. Historia de la academia

El Valencia CF cuenta con una larga tradición en el desarrollo de futbolistas, habiendo producido una generación tras otra de jugadores de alto nivel que han alcanzado el éxito en competiciones de primer orden. Nombres como David Silva, Isco y Jordi Alba representan ejemplos notables del impacto formativo de esta academia a lo

largo de los años [España, 2024]. Este legado confirma la capacidad del club para formar talento con proyección internacional de manera sostenida.

2. Estructura, Filosofía y Metodología de la academia

La academia del Valencia adopta una estructura circular que sitúa al jugador en el centro del proceso formativo. Esta concepción se aleja del modelo jerárquico tradicional, promoviendo una colaboración multidisciplinar entre entrenadores, preparadores físicos, psicólogos, nutricionistas, analistas y otros departamentos de soporte [CF, 2024]. Esta estructura facilita una atención individualizada y refleja una visión holística del desarrollo deportivo y personal.

La filosofía que guía este enfoque se resume en el lema “Educar personas, formar futbolistas”, destacando el compromiso del club con la formación integral del joven atleta [CF, 2024]. Este enfoque se apoya en valores fundamentales como el Sentimiento, el Coraje y la Germanor (hermandad), que se transmiten explícitamente a través del Proyecto de Valores de la Academia. Estos elementos no solo refuerzan la identidad institucional, sino que también promueven aspectos como el sentido de pertenencia, la cohesión grupal y la motivación individual [Centre, 2024b].

El modelo metodológico del Valencia CF se considera un referente tanto a nivel nacional como internacional, basándose en un estilo de juego ofensivo, dinámico e intenso en la recuperación del balón [Centre, 2024b]. Este modelo se adapta a la etapa de desarrollo del jugador, integrando componentes técnicos, tácticos, físicos y psicológicos. A ello se suma el programa “VCF Educa”, que busca extender la formación no solo al ámbito deportivo, sino también al académico y ético, beneficiando a jugadores y entrenadores [Centre, 2024a].

3. Éxitos de la academia

Los resultados tangibles del trabajo de la academia se reflejan tanto en el número de jugadores que alcanzan el primer equipo como en aquellos que compiten en las principales ligas europeas. Informes del CIES Football Observatory posicionan al Valencia CF entre las cinco mejores academias de Europa por número de futbolistas formados en las grandes ligas [OneFootball, 2024]. Además, la academia ha sido reconocida por la Asociación Europea de Clubes (ECA) por su impacto social a través de diversas campañas de acción comunitaria [SEE, 2024].

Estas distinciones, tanto cuantitativas como cualitativas, no solo refuerzan la efectividad del enfoque pedagógico y técnico de la academia, sino que también la validan como un modelo sostenible de desarrollo de talento futbolístico.

Además, este trabajo se ha presentado a varios integrantes de la Academia del Valencia CF de diferentes departamentos como Análisis y Rendimiento, Gestión del Talento, Departamento de Porteros, Internacionales y Dirección Deportiva. Sus valoraciones fueron positivas y mostraron interés en el estudio, preguntando por su posible aplicación dentro de la academia.

2.3. Identificación y definición del talento en el fútbol

La identificación y desarrollo del talento en el fútbol son procesos complejos que implican múltiples etapas: detección, identificación, desarrollo y selección. Cada una de estas

fases requiere un enfoque meticuloso para garantizar que los jugadores con habilidades y competencias sobresalientes alcancen el éxito competitivo.

Tradicionalmente, la identificación del talento se ha basado en la observación de cualidades físicas, fisiológicas y de habilidad. Sin embargo, estudios recientes sugieren que factores como el contexto social, el entorno deportivo y la personalidad del individuo también desempeñan un papel crucial en el desarrollo de un futbolista talentoso.

1. Definición general del talento en el fútbol

La noción de talento en el fútbol evoca diversas imágenes y percepciones iniciales, a menudo ligadas a la habilidad innata o a la brillantez en el terreno de juego. Sin embargo, la literatura científica ofrece definiciones más precisas y multifacéticas de este concepto.

En el contexto específico del fútbol, [Reilly et al., 2000] definen la identificación del talento como el proceso de reconocer a participantes actuales que poseen el potencial para convertirse en jugadores de élite. Esta definición, ampliamente citada en la literatura, subraya la importancia de la “potencialidad”, como elemento central del talento. [Nádori, 1993] describe el talento como una facultad o conjunto de facultades específicas que superan la media, pero que aún necesitan manifestarse plenamente. Para los deportes de equipo como el fútbol, [Omeñaca, 2000] amplía la definición de talento incorporando elementos como el pensamiento operativo, la inteligencia en la situación de juego y la creatividad.

La comprensión del talento ha evolucionado desde una visión centrada en las habilidades innatas hacia una perspectiva que reconoce la compleja interacción entre la naturaleza y la crianza. Se ha pasado de considerar el talento como algo estático a entenderlo como un proceso dinámico que se cultiva a través del desarrollo y la adquisición de pericia. La propia definición de talento puede ser contextual, variando según las necesidades específicas y la filosofía de un club o academia. Por lo tanto, la identificación del talento no solo se basa en la evaluación de las habilidades presentes, sino también en la proyección del potencial futuro dentro de un entorno específico.

2. Dimensiones y desarrollo del talento

La evaluación integral del talento en el fútbol requiere considerar diversas dimensiones interconectadas que contribuyen al rendimiento global de un jugador [Sarmento et al., 2018, Gómez-Díaz et al., 2023]. Estas dimensiones, aunque analizadas individualmente, operan de forma sinérgica. La dimensión técnica abarca el dominio del balón y la ejecución eficiente de acciones como el pase, el regate y el disparo [Unnithan et al., 2021, Larkin and O'Connor, 2023]. La dimensión táctica se refiere a la comprensión del juego, la toma de decisiones estratégicas y el posicionamiento [Sáez de Villarreal et al., 2022, Performance, 2023, Larkin and O'Connor, 2023]. La dimensión física engloba atributos como la velocidad, fuerza y resistencia, fundamentales para la ejecución técnica y táctica [Gonçalves et al., 2021, Simplifaster, 2023, RREM, 2023]. Finalmente, la dimensión cognitiva incluye la percepción, la anticipación y la inteligencia de juego, cruciales para procesar información y tomar decisiones efectivas [Unnithan et al., 2021, Gómez-Díaz et al., 2023, Larkin and O'Connor, 2023, MDPI, 2024]. Además de estas, los factores psicosociales como la motivación y la resiliencia son también relevantes [Sarmento et al., 2018].

El desarrollo del talento en el fútbol es un proceso multifacético que se articula en etapas como la detección, identificación, desarrollo y selección [Larkin and O'Connor, 2017, Till and Baker, 2020]. Diversos factores potencian este desarrollo. La calidad, intensidad y especificidad del entrenamiento son primordiales [Williams and Reilly, 2000]. El entorno del jugador, incluyendo la academia, el apoyo familiar y el contexto socio-cultural, ejerce una influencia significativa [Sarmento et al., 2018, Woods et al., 2020, Ribeiro et al., 2024]. Los factores psicológicos como la motivación intrínseca y la fortaleza mental son esenciales [Gledhill et al., 2017, Henriksen et al., 2010]. Los aspectos biológicos relacionados con la maduración también juegan un papel [Vaeyens et al., 2008]. La calidad de los métodos pedagógicos y la relación entrenador-jugador son determinantes [Côté et al., 2007]. La interacción equilibrada de todos estos elementos es crucial para la formación integral de futbolistas de élite, donde un enfoque holístico y personalizado en las academias es fundamental para el éxito del proceso [Henriksen et al., 2010, Reeves et al., 2018, Côté et al., 2007, Larkin and O'Connor, 2017].

A partir de esta base teórica, resulta pertinente revisar los estudios previos centrados en academias deportivas, especialmente aquellos que han investigado cómo estas estructuras organizadas implementan estrategias de desarrollo del talento y qué factores han demostrado ser más determinantes en la transición hacia el fútbol profesional.

3. Definición de éxito en el fútbol formativo

Uno de los principales desafíos en la evaluación de los procesos de formación deportiva es establecer qué se entiende por “éxito” en el contexto del fútbol formativo. Este concepto resulta clave para valorar no solo la eficacia de las academias en el desarrollo del talento, sino también para realizar estudios longitudinales que analicen el impacto de las trayectorias formativas en el rendimiento profesional.

En la literatura científica existen diferentes formas de operacionalizar el éxito, que varían según los objetivos del análisis. Algunos estudios adoptan criterios cualitativos como la percepción subjetiva del jugador o del entrenador sobre la progresión alcanzada, mientras que otros emplean métricas objetivas como la firma de un contrato profesional o la participación en ligas de alto nivel.

En este trabajo se adopta el criterio establecido por [Abbott et al., 2018], quienes proponen una definición objetiva y verificable del éxito en la carrera profesional de un futbolista: haber disputado más de 900 minutos en competiciones profesionales oficiales. Esta cifra equivale aproximadamente a diez partidos completos, y permite identificar no solo a los jugadores que han debutado profesionalmente, sino también a aquellos que han logrado una continuidad mínima en ese nivel, lo cual resulta más representativo de una integración efectiva en el fútbol profesional.

Este enfoque aporta un criterio robusto para diferenciar entre aquellos jugadores que han tenido una participación testimonial en el fútbol de élite y aquellos que han conseguido consolidarse al menos temporalmente. Así, el éxito en el fútbol formativo no se limita a alcanzar un debut simbólico, sino que implica un cierto grado de permanencia y protagonismo competitivo en el entorno profesional.

Este criterio será utilizado como referencia en el presente trabajo para clasificar y analizar los casos de éxito en la muestra de jugadores estudiados.

2.4. Estudios previos sobre factores de éxito en academias deportivas

Para comprender mejor las estrategias de éxito en las academias deportivas, es valioso analizar estudios de caso de academias de fútbol de renombre que consistentemente producen talento de élite. A continuación, se examinan algunas de las academias más destacadas y sus factores clave de éxito.

Ajax Ámsterdam

La academia del Ajax es reconocida por su método único de educar y nutrir jóvenes talentos. Implementa una formación consistente en el sistema 4-3-3, reflejando la filosofía del “fútbol total”, introducida por Rinus Michels. Este enfoque enfatiza la libertad de expresión en el campo, permitiendo a los jugadores desarrollar creatividad y adaptabilidad. Además, la academia cuenta con programas integrales para todas las edades, instalaciones de primer nivel y entrenadores especializados. Un estudio de [Larsen et al., 2020] destaca la importancia de un entorno holístico en el desarrollo del talento, resaltando la interacción entre jugadores, entrenadores y personal de apoyo en el Ajax.

FC Barcelona

La Masia, academia del FC Barcelona, es famosa por su estilo de juego basado en la posesión y un enfoque holístico que combina educación académica con entrenamiento futbolístico. Fundada en 1979 por Johan Cruyff, refleja la influencia del modelo del Ajax. Un análisis de [Zehnder, 2016] subraya cómo La Masia ha integrado con éxito la educación formal en su programa, con un 40 % de su equipo juvenil A cursando estudios universitarios, promoviendo así el desarrollo integral de los jugadores.

SL Benfica

La academia del Benfica es reconocida por su enfoque en el desarrollo a largo plazo y su apoyo integral a los jugadores, incluyendo aspectos médicos, físicos y psicológicos. Según un artículo de [Kidd, 2025], el Benfica ha generado aproximadamente 516 millones de euros en transferencias de jugadores formados en su academia entre 2014 y 2023, evidenciando su eficacia en la formación de talentos. Pedro Marques, director técnico de la academia, enfatiza la importancia de la detección de talento y la adaptación a las futuras demandas del fútbol en su metodología de formación.

Real Madrid

La Fábrica, academia del Real Madrid, se centra en una filosofía de excelencia y trabajo en equipo. Un estudio del [Sports, 2025] indica que, en los últimos 20 años, La Fábrica ha formado a 166 jugadores que han llegado a las principales ligas europeas, superando a otras academias de renombre como La Masia del Barcelona. Este éxito se atribuye a una inversión significativa en infraestructura y una red de ojeadores eficaz.

Manchester United

La academia del Manchester United tiene un legado en la producción de jugadores destacados. Un estudio de [Richardson et al., 2016] investiga el sistema de desarrollo de jugadores del club, resaltando la importancia de modelos de desarrollo de talento específicos y la adaptación a las necesidades individuales de los jugadores. Este enfoque ha permitido al club mantener una cantera competitiva y productiva a lo largo de los años.

Conclusiones de los Estudios Previos y Transición a la Metodología

El análisis de estudios previos sobre academias deportivas de élite, como las del Ajax, FC Barcelona, SL Benfica, Real Madrid y Manchester United, revela la convergencia en factores clave de éxito: filosofías de juego definidas, estructuras formativas integrales, metodologías de entrenamiento específicas y una atención holística al desarrollo del jugador. Estas investigaciones subrayan la complejidad del proceso de formación y la necesidad de un enfoque multidimensional para nutrir el talento. Partiendo de esta comprensión teórica y de los modelos exitosos identificados, la siguiente etapa de este trabajo se centrará en la descripción detallada de la metodología que se empleará para analizar las competencias de los futbolistas de la Academia VCF, buscando identificar las particularidades y los factores determinantes en su proceso formativo.

2.5. Criterios para la consideración de una competición como profesional

A la hora de delimitar qué competiciones de fútbol pueden considerarse profesionales, es importante señalar que **no existe una definición universal y explícita por parte de los principales organismos internacionales como la FIFA o la UEFA**. Sin embargo, sí se recogen ciertos criterios implícitos que permiten establecer una clasificación razonada.

1. Definiciones institucionales

La **FIFA**, en su *Reglamento sobre el Estatuto y la Transferencia de Jugadores* (2023), define al **jugador profesional** como:

“Aquel jugador que tiene un contrato por escrito con un club y percibe más por su actividad futbolística de lo que gastaría en ella.” [FIFA, 2023]

Por su parte, la **UEFA**, a través del sistema de licencias para clubes, introduce una definición operativa de **club profesional**:

“Un club que participa en una competición organizada por una liga nacional que considera al menos a uno de sus equipos como profesional, lo que implica la existencia de contratos laborales para jugadores y personal técnico.” [UEFA, 2024]

Ambos organismos ponen el foco en la **naturaleza contractual de los jugadores y del cuerpo técnico**, así como en el cumplimiento de ciertos estándares estructurales y financieros por parte de los clubes.

2. Consideraciones clave

A partir de lo anterior, se pueden establecer las siguientes **condiciones necesarias para considerar una competición como profesional**:

- La mayoría de los clubes participantes tienen jugadores con **contrato profesional**.
- La liga está **reconocida oficialmente** por la federación nacional correspondiente.

- Existe una **infraestructura reglamentaria y financiera** adecuada (licencias, auditorías, instalaciones, etc.).
- Los equipos pueden participar en **competiciones UEFA o CONMEBOL**, que exigen profesionalización.

3. Competiciones consideradas profesionales

En base a los criterios anteriores y a datos actualizados sobre la profesionalización de ligas, las competiciones consideradas como profesionales en el marco de este estudio se listan detalladamente en el Apéndice A.1. Esta selección incluye ligas y copas nacionales con participación regular en competiciones internacionales, infraestructura profesional y contratos laborales para jugadores, lo que justifica su inclusión como entornos plenamente profesionalizados.

Capítulo 3

Herramientas y técnicas de análisis de datos

3.1. Fuentes de datos y técnicas de extracción

El presente capítulo detalla los recursos y metodologías empleados para la adquisición, procesamiento y análisis de los datos que fundamentan este estudio. A lo largo de esta sección, se explicarán en primer lugar las **fuentes de datos** principales utilizadas, se describirán las **técnicas de extracción** aplicadas para recopilar la información relevante, y se presentarán las **herramientas de programación** empleadas. Finalmente, se expondrán los **modelos de clasificación** utilizados en la fase de modelado predictivo, junto con las **métricas de evaluación** seleccionadas para valorar su rendimiento.

Para la realización de este estudio se ha utilizado el lenguaje de programación **Python**, una herramienta ampliamente utilizada en el ámbito de la ciencia de datos y el aprendizaje automático. Python permite la manipulación, análisis y visualización de grandes volúmenes de datos gracias a su ecosistema de librerías, entre las que destacan **pandas**, **numpy**, **matplotlib**, **seaborn**, **scikit-learn**, **xgboost** y **tensorflow**.

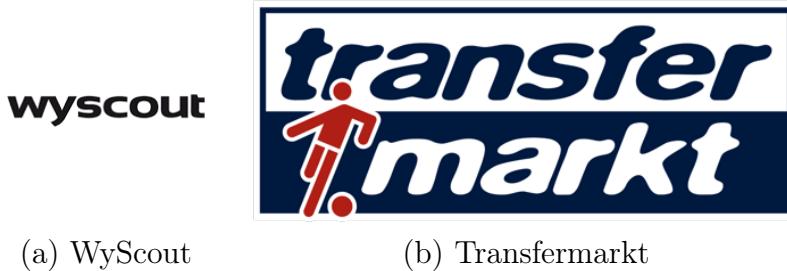
Las dos fuentes de datos principales han sido WyScout y Transfermarkt. Ambas plataformas son ampliamente reconocidas en el ámbito del análisis de datos futbolísticos y han sido seleccionadas por su fiabilidad, accesibilidad y la amplitud de la información que ofrecen. En la figura 3.1 podemos ver cómo son las estructuras de las páginas web.

3.1.1. WyScout

WyScout es una plataforma profesional de análisis de rendimiento utilizada por clubes, entrenadores, scouts y analistas de todo el mundo. A través de su base de datos, se ha accedido a información detallada sobre aspectos técnicos y tácticos de los jugadores analizados. Entre los datos extraídos se encuentran registros individuales como tipos de pase, conducciones, duelos, pérdidas o acciones defensivas, que permiten caracterizar el estilo de juego y las competencias técnicas de cada futbolista. WyScout ha sido esencial para construir una visión cuantitativa de los perfiles de rendimiento durante su etapa formativa.

3.1.2. Transfermarkt

Por otro lado, Transfermarkt ha sido utilizado como fuente para el análisis longitudinal de las trayectorias profesionales de los jugadores. Esta plataforma ofrece información sobre partidos disputados, minutos jugados, clubes por los que ha pasado cada futbolista y la categoría de las competiciones en las que ha participado.



(a) WyScout

(b) Transfermarkt

Figura 3.1: Plataformas de datos futbolísticos utilizadas.

3.1.3. Web scraping

Para extraer información desde Transfermarkt se ha utilizado la técnica de **web scraping**, que consiste en la recopilación automatizada de datos desde páginas web mediante scripts que acceden al contenido HTML de las mismas. Se empleó la librería `BeautifulSoup` junto con `requests`, lo que permitió navegar entre tablas, listas y etiquetas HTML para identificar y almacenar de forma sistemática la información relevante. Esta técnica se utilizará para extraer los datos de la sección 4.1.

3.2. Introducción a la inteligencia artificial y aprendizaje automático

Este trabajo incluye una fase de modelado predictivo basada en técnicas de **inteligencia artificial** (IA), entendida como la disciplina que desarrolla sistemas capaces de emular ciertos aspectos del pensamiento humano, como la resolución de problemas, el razonamiento o el aprendizaje [Russell and Norvig, 2010].

Dentro de la IA, el **aprendizaje automático** (*machine learning*) constituye una subárea fundamental que se centra en el desarrollo de algoritmos capaces de aprender automáticamente a partir de los datos, sin ser explícitamente programados para cada tarea [Mitchell, 1997]. En este estudio se aplica el enfoque de **aprendizaje supervisado**, que utiliza un conjunto de entrenamiento con datos de entrada y salidas conocidas para aprender una función que relacione ambas variables.

El objetivo del aprendizaje supervisado es encontrar una función $f : \mathcal{X} \rightarrow \mathcal{Y}$, donde:

- \mathcal{X} : espacio de características (inputs).
- \mathcal{Y} : espacio de etiquetas o salidas (targets).
- f : función que mapea una entrada $x \in \mathcal{X}$ a una salida $y \in \mathcal{Y}$.

En la práctica, se busca minimizar una función de pérdida L entre las predicciones $\hat{y}_i = f(x_i)$ y los valores reales y_i . Una función de pérdida común es el error cuadrático medio (MSE):

$$\mathcal{L} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.1)$$

donde:

- n : número de observaciones en el conjunto de datos.
- y_i : valor real de la variable dependiente para la observación i .
- \hat{y}_i : valor predicho por el modelo para la observación i .

3.3. Modelos de clasificación utilizados

Los modelos de clasificación utilizados en este estudio son los siguientes:

3.3.1. Regresión logística

La regresión logística es un modelo ampliamente utilizado para tareas de clasificación binaria. Aunque su nombre sugiere un enfoque de regresión, en realidad estima la probabilidad de que una observación pertenezca a una clase determinada ($y = 1$), a partir de una combinación lineal de las variables predictoras. Esta combinación se transforma mediante la función logística o sigmoide, que produce valores entre 0 y 1, lo que permite interpretarlos como probabilidades [Hosmer Jr et al., 2013].

Este modelo es especialmente útil cuando el objetivo es comprender la influencia de cada variable independiente sobre la probabilidad de ocurrencia del evento de interés. Además, sus coeficientes pueden interpretarse como log-odds (logaritmos de razones de probabilidades), lo que lo convierte en un modelo interpretable y explicativo.

Sin embargo, la regresión logística tiene limitaciones en contextos donde las relaciones entre las variables independientes y la variable objetivo son no lineales, o cuando existe una alta multicolinealidad entre predictores. Aun así, por su simplicidad, eficiencia y robustez en problemas bien especificados, sigue siendo una técnica de referencia en clasificación binaria.

$$P(y = 1 | \mathbf{x}) = \frac{1}{1 + e^{-z}}, \quad z = \beta_0 + \sum_{j=1}^n \beta_j x_j \quad (3.2)$$

donde:

- $X = (x_1, x_2, \dots, x_n)$: vector de características de entrada.
- β_0 : término independiente o sesgo.
- β_j : coeficiente asociado a la variable x_j .

- z : combinación lineal de las variables predictoras.

Ventajas: modelo simple, rápido e interpretable. **Inconvenientes:** limitado en problemas no lineales complejos.

3.3.2. Random Forest

Random Forest es un algoritmo de ensamblado (*ensemble learning*) que combina múltiples árboles de decisión para mejorar la precisión y la capacidad de generalización del modelo [Breiman, 2001]. Cada árbol es entrenado sobre un subconjunto aleatorio de los datos de entrenamiento (mediante *bagging*) y sobre un subconjunto aleatorio de variables en cada división del árbol, lo que introduce diversidad entre los árboles y reduce la correlación entre ellos.

La predicción final del modelo se obtiene a través del voto mayoritario (en clasificación) o del promedio de las salidas (en regresión). Este enfoque reduce significativamente el riesgo de sobreajuste, una debilidad común de los árboles individuales, y proporciona un modelo más robusto frente al ruido y las anomalías en los datos.

Random Forest es capaz de manejar relaciones no lineales, datos con alta dimensionalidad y variables categóricas o numéricas sin necesidad de normalización previa. No obstante, uno de sus inconvenientes es que puede volverse menos interpretable a medida que se incrementa el número de árboles.

Ventajas: robusto al sobreajuste, útil para conjuntos de datos con ruido y relaciones no lineales. **Inconvenientes:** menos interpretable, más costoso computacionalmente que modelos lineales.

3.3.3. Máquinas de vectores de soporte (SVM)

Las máquinas de vectores de soporte son modelos de clasificación que buscan encontrar el hiperplano que separa de forma óptima las instancias de dos clases, maximizando el margen entre ellas. Las observaciones más cercanas al hiperplano, llamadas vectores de soporte, son las que determinan su posición [Cortes and Vapnik, 1995].

SVM es especialmente eficaz en espacios de alta dimensión y es capaz de realizar transformaciones no lineales mediante el uso de funciones *kernel*, que proyectan los datos a un espacio de características de mayor dimensión donde es posible trazar un límite de decisión lineal.

Este modelo es robusto frente a sobreajuste, especialmente cuando se utiliza una buena elección de kernel y regularización. Sin embargo, presenta limitaciones en grandes conjuntos de datos por su alto coste computacional y su escasa interpretabilidad.

$$\min_{\mathbf{w}, b} \frac{1}{2} |\mathbf{w}|^2 \quad \text{sujeto a} \quad y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 \quad \forall i \quad (3.3)$$

donde:

- \mathbf{w} : vector normal al hiperplano.
- b : término independiente.

- \mathbf{x}_i : vector de entrada para el ejemplo i .
- $y_i \in \{-1, 1\}$: etiqueta de clase para el ejemplo i . El símbolo \in indica “pertenece a”, y $\{-1, 1\}$ representa el conjunto de las dos posibles clases.

Ventajas: efectivas en espacios de alta dimensión, robustas con pocos datos. **Inconvenientes:** no escala bien a grandes volúmenes de datos, difícil de interpretar.

3.3.4. Redes neuronales artificiales

Las redes neuronales artificiales son modelos inspirados en la estructura y funcionamiento del cerebro humano. Están compuestas por capas de neuronas artificiales que transforman la información de entrada a través de combinaciones lineales seguidas de funciones de activación no lineales. La arquitectura más básica es la red neuronal *feedforward*, donde la información fluye en una única dirección desde la entrada hacia la salida [Rosenblatt, 1958].

Estas redes tienen la capacidad de aproximar funciones complejas y modelar relaciones altamente no lineales entre variables. El aprendizaje se realiza mediante el algoritmo de **retropropagación** (o *backpropagation*), que ajusta los pesos de las conexiones neuronales minimizando el error entre la salida predicha y la salida real a través del **descenso por gradiente** [Rumelhart et al., 1986].

$$\Delta w = -\eta \cdot \frac{\partial \mathcal{L}}{\partial w} \quad (3.4)$$

donde:

- Δw : cambio aplicado al peso w .
- $\eta \in \mathbb{R}^+$: tasa de aprendizaje (*learning rate*). El símbolo \in indica “pertenece a”, y \mathbb{R}^+ representa el conjunto de los números reales positivos.
- $\frac{\partial \mathcal{L}}{\partial w}$: derivada parcial de la función de pérdida \mathcal{L} con respecto al peso w . El símbolo ∂ se usa para denotar derivadas parciales.

Si bien las redes neuronales ofrecen un gran poder expresivo, requieren grandes cantidades de datos para evitar el sobreajuste y un entrenamiento intensivo en recursos computacionales. Además, a diferencia de otros modelos más simples, su naturaleza de *caja negra* dificulta la interpretación de los resultados, lo cual puede ser un inconveniente en contextos donde la explicabilidad es crítica.

$$\hat{y} = f(\mathbf{W}\mathbf{x} + \mathbf{b}) \quad (3.5)$$

donde:

- \mathbf{x} : vector de entrada.
- \mathbf{W} : matriz de pesos sinápticos.
- \mathbf{b} : vector de sesgos.

- $f(\cdot)$: función de activación (e.g., ReLU, sigmoide, tanh). El símbolo (\cdot) indica que f es una función de lo que esté dentro del paréntesis.
- \hat{y} : salida del modelo. El símbolo $\hat{\cdot}$ se utiliza a menudo para denotar una predicción o una estimación.

Ventajas: gran capacidad para modelar relaciones complejas y no lineales. **Inconvenientes:** requiere gran cantidad de datos y potencia computacional, difícil de interpretar.

3.4. Ajuste de hiperparámetros y métricas de evaluación

La optimización del rendimiento de los modelos de aprendizaje automático no solo depende de la elección del algoritmo adecuado, sino también de la correcta configuración de sus **hiperparámetros**. Estos parámetros no se aprenden directamente de los datos, sino que se establecen antes del proceso de entrenamiento e influyen significativamente en el comportamiento y la capacidad de generalización del modelo.

Técnicas de búsqueda de hiperparámetros

Para la búsqueda de los hiperparámetros óptimos en los modelos empleados en este estudio, se han utilizado diferentes estrategias:

- **Búsqueda en Cuadrícula (*Grid Search*)**: Para la **Regresión Logística**, se optó por una búsqueda en cuadrícula. Este método evalúa sistemáticamente todas las combinaciones posibles dentro de un conjunto discreto de valores predefinidos para cada hiperparámetro [Bergstra and Bengio, 2011]. Dada la menor complejidad y coste computacional de la regresión logística, la búsqueda en cuadrícula se considera eficiente para explorar de forma exhaustiva sus hiperparámetros clave (como *penalty*, *C* y *solver*).
- **Búsqueda Aleatoria (*Randomized Search*)**: Para modelos como **Random Forest** y las **Máquinas de Vectores de Soporte (SVM)**, que poseen un número de hiperparámetros moderado y una mayor complejidad computacional, se empleó la búsqueda aleatoria [Bergstra and Bengio, 2012]. Esta estrategia muestrea aleatoriamente un número fijo de combinaciones de hiperparámetros dentro de los rangos especificados. Es más eficiente para explorar espacios de búsqueda de dimensión moderada, permitiendo cubrir un rango más amplio de valores con el mismo número de evaluaciones, lo cual es especialmente útil cuando la influencia de algunos hiperparámetros es más significativa que la de otros.

Métricas de evaluación

La elección de las métricas de evaluación es fundamental para obtener una comprensión completa del rendimiento del modelo, especialmente en problemas de clasificación con desequilibrio de clases. En este estudio, se utilizaron varias métricas para evaluar la capacidad predictiva de los modelos:

- **F-beta Score:** Para la selección y ajuste de hiperparámetros, se ha seleccionado el **F-beta Score** como métrica principal para la decisión entre los hiperparámetros [Powers, 2011]. Esta métrica armónica combina la **precisión** (*precision*) y la **exhaustividad** (*recall*). La precisión mide la proporción de identificaciones positivas correctas sobre el total de identificaciones positivas, mientras que la exhaustividad mide la proporción de identificaciones positivas correctas sobre el total de casos positivos reales. La fórmula del F-beta Score se define como:

$$F_{\beta} = (1 + \beta^2) \cdot \frac{\text{Precision} \cdot \text{Recall}}{(\beta^2 \cdot \text{Precision}) + \text{Recall}} \quad (3.6)$$

En nuestro contexto, la clase de "éxito profesional" es minoritaria y su correcta identificación es de suma importancia, buscando minimizar los falsos negativos. Por ello, se ha establecido un valor de $\beta = 2$, que confiere el doble de importancia al *recall* que a la *precision*, asegurando que se identifique correctamente a la mayoría de los casos de éxito potenciales.

- **Accuracy (Exactitud):** El **Accuracy** o exactitud mide la proporción de predicciones correctas sobre el total de observaciones [Provost et al., 1998]. Aunque intuitivo, puede ser una métrica engañosa en conjuntos de datos desequilibrados, ya que un alto Accuracy podría deberse simplemente a una buena predicción de la clase mayoritaria. Por esta razón, no se utilizó como métrica principal para la optimización.

$$\text{Accuracy} = \frac{\text{Verdaderos Positivos} + \text{Verdaderos Negativos}}{\text{Total de Observaciones}} \quad (3.7)$$

- **AUC (Area Under the Receiver Operating Characteristic Curve):** El **AUC** es una métrica robusta para evaluar el rendimiento de modelos de clasificación binaria, especialmente en presencia de desequilibrio de clases [Fawcett, 2006]. El valor AUC, que varía de 0 a 1, evalúa el rendimiento general del modelo en todos los posibles umbrales, representando la capacidad discriminativa del modelo para distinguir entre las clases positivas y negativas.

Capítulo 4

Metodología

4.1. Construcción de la base de datos final

La muestra analizada en este estudio se compone de datos de jugadores jóvenes que han participado en competiciones del fútbol español de tercera y cuarta categoría en los últimos años. El proceso de elaboración del conjunto de datos se realizó en varias etapas estructuradas:

En primer lugar, se descargaron los datos desde la plataforma WyScout de todos los jugadores menores de 23 años que participaron al menos 450 minutos en las siguientes competiciones y temporadas: Segunda División B (2018–2019, 2019–2020, 2020–2021), Primera RFEF (2021–2022, 2022–2023, 2023–2024, 2024–2025) y Segunda RFEF (2022–2023, 2023–2024, 2024–2025). De cada jugador se extrajeron todas las variables disponibles en la base de datos de WyScout.

Posteriormente, se realizó una reclasificación de las posiciones específicas proporcionadas por WyScout, agrupándolas en cuatro grandes bloques funcionales para facilitar la comparación de perfiles de rendimiento similares. La Tabla 4.1 detalla las 21 posiciones específicas originales y su correspondiente asignación a las cuatro categorías generales: porteros (GK), defensas (DEF), centrocampistas (MC) y delanteros (DEL). Esta reducción se basó en la función principal de los jugadores en el campo, buscando una mayor homogeneidad analítica al comparar perfiles de rendimiento similares dentro de cada bloque funcional.

Dentro de las posiciones orginiales de WyScout también habían para un mismo jugador diferentes combinaciones de ellas. Es decir, un jugador podía tener como posición: RB, RCB. Para resolver esta problemática se optó por quedar solo con la primera posición que aparece dentro de esa variable (que es la más representativa).

Posiciones Originales	Posiciones Finales
RB, LB, RCB, LCB, CB, LWB, RWB	DEF
RCMF, LCMF, AMF, RAMF, DMF, LAMF, LD MF, RD MF	MC
CF, LW, RW, RWF, LWF	DEL
GK	GK

Cuadro 4.1: *Mapeo de posiciones originales de WyScout a bloques generales.*

Fuente: Elaboración propia.

En el caso de jugadores duplicados —es decir, aquellos que aparecían en más de una temporada o competición— se optó por conservar únicamente la observación más reciente disponible, con el objetivo de reflejar el estado de desarrollo más actual de cada jugador.

Para vincular la trayectoria profesional de los jugadores, se implementó un proceso de web scraping con el fin de obtener sus identificadores únicos en Transfermarkt. A través de estos identificadores, se construyó una tabla con todas las competiciones en las que había participado cada jugador, incluyendo los minutos disputados por temporada. Esta información permitió calcular el total acumulado de minutos en categorías consideradas como profesionales que hemos considerado anteriormente y, siguiendo el criterio establecido por [Abbott et al., 2018], se etiquetó como jugador “exitoso” aquel que acumuló 900 o más minutos en dichas competiciones.

Uno de las principales dificultades encontradas en esta parte ha sido hacer el mapeo de los nombres de una base de datos a otra, como no estaban estandarizados había algunos casos en los que al buscar el nombre de un jugador con el formato de WyScout en Transfermarkt no lo encontraba o causaba algunos problemas. La Tabla 4.2 siguiente muestra cuáles eran algunas de las diferencias entre los nombres de las diferentes bases de datos que causaban estos inconvenientes.

Nombre en WyScout	Nombre en Transfermarkt
E. Boyomo	Enzo Boyomo
A. Dionkou	Alpha Dionkou
Retu	Alberto Retuerta
Urko González	Urko González de Zárate

Cuadro 4.2: *Mapeo de nombres de WyScout a Transfermarkt.*

Fuente: Elaboración propia.

Como resultado de este proceso de recolección, filtrado, integración y limpieza de datos, se conformó un conjunto final compuesto por 1.536 observaciones (jugadores únicos) y 116 columnas (variables), que constituye la base analítica del presente estudio. La lista completa de variables se encuentra en el Apéndice A.2.

4.2. Variables analizadas

Para la selección de variables relevantes se optó por una estrategia combinada basada en el conocimiento experto. En concreto, se diseñó y distribuyó un **formulario estructurado** dirigido al Área de Intentidad conformada por todos los responsables y miembros de la Academia del VCF

El objetivo del cuestionario fue *identificar las variables más significativas en el seguimiento y la evaluación del rendimiento de jóvenes futbolistas*, tanto desde una perspectiva técnica como contextual. El formulario constaba de tres secciones diferenciadas:

1. **Perfil general del jugador**, donde se exploraban factores básicos como la posición, edad o situación contractual, así como la importancia relativa de las dimensiones técnico-tácticas, físicas y mentales en el desarrollo del futbolista.
2. **Evaluación por bloques temáticos**, agrupando las variables en cinco categorías clave:

- Participación y rendimiento general (ej. minutos jugados, goles, asistencias)
 - Indicadores defensivos (ej. duelos defensivos, entradas, interceptaciones)
 - Indicadores ofensivos y de desequilibrio (ej. regates, centros, duelos ofensivos)
 - Toma de decisiones y visión de juego (ej. pases progresivos, jugadas clave, xA)
 - Preguntas abiertas para incorporar conocimiento no estructurado
3. **Reflexión final**, incluyendo preguntas cualitativas sobre variables imprescindibles, sobrevaloradas, y las metodologías actuales de evaluación utilizadas dentro de la Academia.

Las respuestas obtenidas permitieron filtrar y priorizar aquellas métricas que los técnicos consideran más representativas del potencial y rendimiento real de un jugador en formación. A partir de este análisis cualitativo, se seleccionó un conjunto reducido de variables clave que se utilizarán en los modelos posteriores.

Las variables seleccionadas tras el proceso de consenso con los expertos de la Academia del Valencia CF son las siguientes: *partidos jugados*, que indica el número total de encuentros disputados; *minutos jugados*, como reflejo de la continuidad y la confianza del cuerpo técnico; *duelos defensivos por 90 minutos*, que mide la frecuencia con la que el jugador disputa balones en tareas defensivas; y el *porcentaje de duelos defensivos ganados*, que representa su efectividad en dichas acciones. Se incluye también el número de *intercepciones por 90 minutos*, que recoge la capacidad de anticipación para cortar pases rivales; las *entradas por 90 minutos*, como acciones directas para recuperar la posesión; y la *posesión conquistada tras entrada*, que valora no solo la acción defensiva sino también su desenlace exitoso. En el juego aéreo, se consideran los *duelos aéreos por 90 minutos* y el *porcentaje de duelos aéreos ganados*, como medida de su rendimiento en disputas por alto. En cuanto al perfil ofensivo y asociativo, se valoran los *regates por 90 minutos* y su *porcentaje de éxito*, como indicador del desequilibrio individual; los *pases por 90 minutos* y su *precisión*, reflejo de su implicación y acierto en la circulación; los *pases progresivos por 90 minutos*, que miden su capacidad para avanzar líneas; y la *precisión en los pases al último tercio*, relevante para evaluar su impacto en zonas de peligro. Por último, se recogen las *jugadas clave por 90 minutos*, que incluyen acciones que generan ocasiones manifiestas de gol; los *desmarques por 90 minutos*, como movimientos sin balón para generar opciones de pase; y la *precisión de desmarques*, entendida como el porcentaje de movimientos que acaban en recepción efectiva del balón.

4.3. Descripción de la muestra

Como se detalló anteriormente, la base de datos final para este estudio se compone de 1.536 jugadores jóvenes que participaron en competiciones del fútbol español de tercera y cuarta categoría. A continuación, se presentan las principales características de la muestra:

1. Distribución de la variable objetivo:

La Figura 4.1 muestra cómo se distribuyen los jugadores entre las categorías “exitoso” y “no exitoso”, en función de haber acumulado o no al menos 900 minutos en competiciones profesionales. Se observa un **desequilibrio significativo**, ya que el 81.5 % de los jugadores (1.252) no alcanzaron dicho umbral, frente al 18.5 % (284

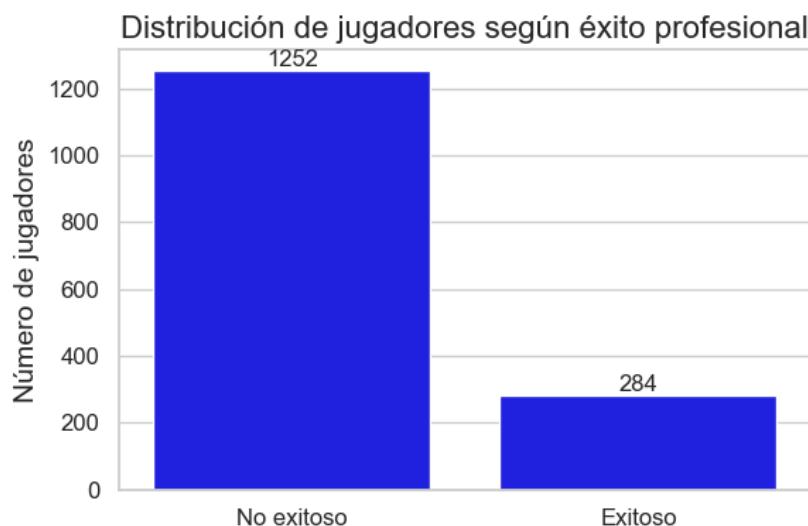


Figura 4.1: *Distribución de jugadores según éxito profesional.*

Fuente: Elaboración propia a partir de los datos de la base de datos.

jugadores) que sí lo hicieron. Esta desproporción plantea desafíos para los modelos de clasificación, que tienden a favorecer la clase mayoritaria.

2. Justificación del uso de técnicas de *oversampling*:

Dado el fuerte desbalance en la variable objetivo, se aplicaron técnicas de *oversampling* para aumentar la representación de la clase minoritaria en el conjunto de entrenamiento. Esto permite a los modelos aprender patrones más robustos y no sesgarse hacia la clase dominante. Las técnicas utilizadas fueron:

- **SMOTE:** Generación de nuevas muestras sintéticas interpolando entre vecinos cercanos de la clase minoritaria.
- **ADASYN:** Variante adaptativa de SMOTE que genera más muestras donde la clase minoritaria está más escasamente representada.
- **ROS (Random Over-Sampling):** Duplicación aleatoria de muestras existentes de la clase minoritaria.

Estas técnicas se aplicaron únicamente al conjunto de entrenamiento para evitar introducir sesgos en la evaluación final del modelo.

3. Distribución por posición:

La Figura 4.2 muestra la distribución de posiciones en la muestra, excluyendo a los porteros, ya que el análisis se centra en jugadores de campo. La posición más representada es la de defensa (631 jugadores), seguida de delantero (494) y centrocampista (411).

4. Distribución por edad:

La Figura 4.3 representa la distribución de edades en la muestra. La mayoría de los jugadores se concentran en el rango de 20 a 23 años, siendo las edades más comunes los 22 y 23 años (388 y 411 jugadores, respectivamente). Las edades más jóvenes (16, 17 y 18 años) tienen una representación escasa, con solo 3, 14 y 49 jugadores, respectivamente. Esto puede explicarse por el criterio de inclusión (mínimo de minutos disputados), que es más difícil de alcanzar en etapas tempranas del desarrollo.

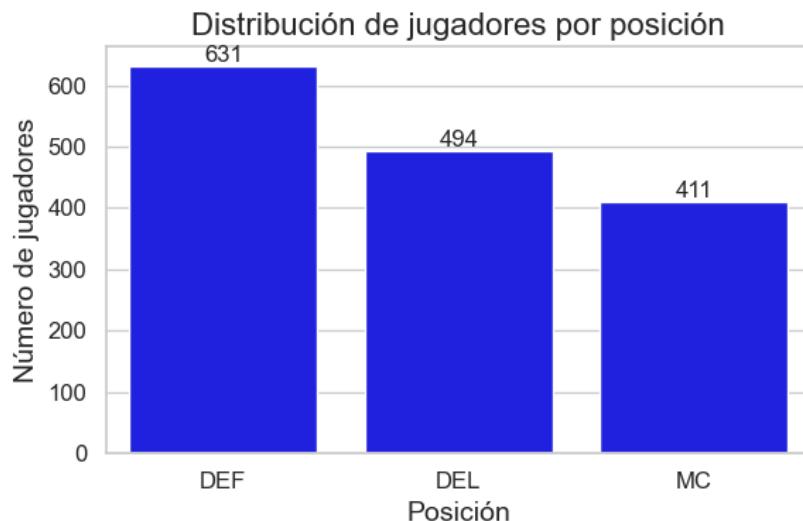


Figura 4.2: *Distribución de jugadores según posición de juego.*
Fuente: Elaboración propia a partir de los datos de la base de datos.

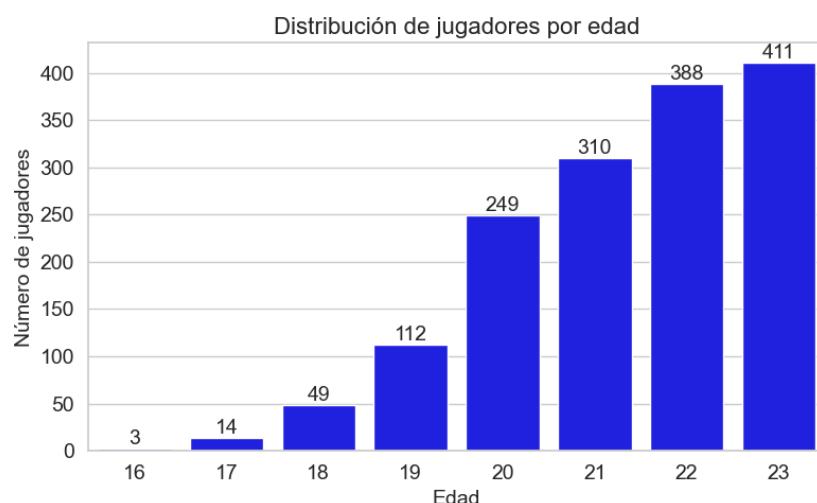


Figura 4.3: *Distribución de jugadores por edad.*
Fuente: Elaboración propia a partir de WyScout.

Esta concentración en edades cercanas a los 23 años indica que la muestra incluye principalmente jugadores en una fase clave de su trayectoria juvenil, donde el rendimiento en competiciones semiprofesionales puede marcar el acceso a categorías superiores. Por ello, resulta especialmente relevante el análisis en esta franja de edad para predecir el éxito profesional.

Capítulo 5

Análisis de datos y resultados

5.1. Resultados

Esta sección presenta los resultados obtenidos de la aplicación de los modelos de clasificación al problema de la predicción del éxito profesional en futbolistas. Se detallarán las configuraciones óptimas de los modelos y los conjuntos de datos, seguidos de la evaluación de su rendimiento mediante las métricas seleccionadas.

5.1.1. Configuraciones óptimas de los modelos

La selección de la mejor combinación de modelo, técnica de remuestreo de datos e hiperparámetros se realizó basándose en la métrica F2-Score, previamente justificada en la sección 3.4 dada su relevancia para priorizar la identificación de la clase minoritaria (jugadores con éxito profesional). Para abordar el desequilibrio inherente en el conjunto de datos, se experimentó con técnicas de sobremuestreo como SMOTE, ROS y ADASYN, además de la utilización del conjunto de datos original sin remuestreo.

Las combinaciones óptimas de conjunto de datos e hiperparámetros para cada modelo son las siguientes:

- **Regresión Logística:** Se obtuvo el mejor rendimiento con la aplicación de la técnica de sobremuestreo **SMOTE** (Synthetic Minority Over-sampling Technique). Los hiperparámetros óptimos encontrados fueron:

- `C = 0.01`
- `penalty = 12`
- `solver = liblinear`

- **Random Forest:** Este modelo alcanzó sus mejores resultados al combinarlo con el conjunto de datos original. Los hiperparámetros finales son:

- `n_estimators = 500`
- `min_samples_split = 5`
- `min_samples_leaf = 1`
- `max_features = None`

- `max_depth = 50`
 - `bootstrap = False`
- **Random Forest con todas las variables:** Este modelo alcanzó su configuración óptima al combinarse con la técnica de sobremuestreo **ADASYN** (Adaptive Synthetic Sampling). Sus hiperparámetros elegidos fueron:
 - `n_estimators = 200`
 - `min_samples_split = 5`
 - `min_samples_leaf = 1`
 - `max_features = log2`
 - `max_depth = None`
 - `bootstrap = False`
- **Máquinas de Vectores de Soporte (SVM):** Curiosamente, el modelo SVM mostró el mejor rendimiento utilizando el **conjunto de datos Original** (sin aplicar técnicas de remuestreo). Los hiperparámetros seleccionados fueron:
 - `C = 0.1`
 - `kernel = linear`
 - `gamma = auto`
 - `degree = 1`
- **Red Neuronal Perceptrón Multicapa (MLP):** El MLP se optimizó al utilizarse con la técnica de sobremuestreo **ADASYN**. Sus hiperparámetros configurados fueron:
 - `activation = tanh`
 - `alpha = 0.001`
 - `hidden_layer_sizes = (128, 64)`
 - `learning_rate_init = 0.01`

Estas configuraciones representan la base para la evaluación de rendimiento de los modelos en la siguiente sección, buscando maximizar la capacidad de detección de jugadores con éxito profesional.

5.1.2. Rendimiento de los modelos

Una vez determinadas las configuraciones óptimas para cada algoritmo, se procedió a evaluar su rendimiento utilizando las métricas definidas. Es importante recordar que el objetivo principal es la correcta identificación de los jugadores con éxito profesional, una clase minoritaria, lo que hace que métricas como el F2-Score y el AUC sean especialmente relevantes. A continuación, se detallan los resultados específicos para cada modelo.

Regresión Logística

El modelo de Regresión Logística, optimizado con la técnica de sobremuestreo SMO-TE, demostró ser el más efectivo para cumplir con el objetivo principal del estudio: la detección de casos de éxito profesional. Concretamente, este modelo fue capaz de identificar correctamente el **71.4 % de los casos de éxito** (recall), un valor que coincidió con su **accuracy general** tal y como podemos ver en el gráfico 5.3a, lo que indica un equilibrio notable en su capacidad de acierto para ambas clases. Además, presentó un valor de **AUC moderadamente elevado**, y su curva ROC se mantuvo consistentemente alejada de la diagonal, aspecto que se puede apreciar en la imagen 5.2a, lo que sugiere una buena capacidad discriminativa general.

Random Forest

Como hemos visto, hemos entrenado dos Random Forest diferentes, esto lo hemos hecho para ver cuáles eran las variables que nos seleccionaba el modelo al cual le pasamos todo el conjunto original de variables, aspecto que debatiremos más adelante. A nivel de resultados del modelo con las mismas variables que los demás: en términos de exactitud no fue el mejor modelo ya que presentó un valor de **AUC de 0.62** tal y como podemos ver en la figura 5.2b además de que la forma de la curva no se acerca a la esquina superior izquierda propia de los modelos con mejores resultados. Sin embargo, cuando nos fijamos en su matriz de confusión de la gráfica 5.3b vemos cómo consigue acertar en una medida medianamente considerable la clase minoritaria comparando con los demás modelos. No está al nivel de la regresión logística en este aspecto, pero mejora a la SVM y el MLP. Esto es debido a su gran capacidad predictora y el tratamiento de clases desbalanceadas.

Máquinas de Vectores de Soporte (SVM)

El modelo de Máquinas de Vectores de Soporte (SVM) presentó un **rendimiento notablemente bajo** en este estudio. Su principal limitación fue que **predijo la gran mayoría de las instancias como casos de no éxito**, resultando en una nula detección de casos de éxito (recall de 0 %) que se puede apreciar en el gráfico 5.3c. Esto provocó que su valor de **AUC fuera bajo (0.5)** y que su curva ROC tuviera una forma poco discriminativa tal y como se observa en la figura 5.2c, llegando a cruzar la diagonal en algunos puntos, lo que es indicativo de un rendimiento no superior al de un clasificador aleatorio. La SVM fue entrenada con el conjunto de datos original, sin remuestreo.

Red Neuronal Perceptrón Multicapa (MLP)

El modelo de Red Neuronal Perceptrón Multicapa (MLP) mostró resultados con **similaridades al Random Forest** en cuanto a su comportamiento predictivo. Si bien, como se aprecia en la figura 5.3d, consiguió detectar algunos casos de éxito, no logró identificar a la mayoría. La precisión en la clase de "éxito profesional" fue del **28 %**, compensada por un **accuracy general del 73 %**, lo que nuevamente indica una tendencia a clasificar la mayoría de los jugadores como "no éxito". Consecuentemente, su valor de **AUC no fue óptimo (0.66)** y su curva ROC tampoco presentaron la forma deseada tal y como se ve en la imagen 5.2d, reflejando una capacidad discriminativa moderada. El principal problema que podría explicar este rendimiento subóptimo es la posible necesidad de una

mayor cantidad de datos para un entrenamiento efectivo, ya que, al tratarse de un conjunto de datos relativamente pequeño y además desbalanceado, el modelo podría no haber aprendido lo suficiente para detectar patrones complejos de la clase minoritaria. Además, la arquitectura utilizada, con dos capas ocultas y un número considerable de neuronas (128 y 64 respectivamente), implica una gran cantidad de pesos que, en un escenario de datos limitados, podrían no haberse ajustado de forma óptima durante el proceso de aprendizaje, llevando a un sobreajuste o a una generalización deficiente.

Todos estos resultados se pueden comparar de forma sencilla y directa en la tabla 5.1.

Cuadro 5.1: Resultados de las Métricas de Evaluación por Modelo

Modelo	Accuracy	Precision	Recall	AUC	F2-Score
Regresión Logística	0.7143	0.7143	0.3571	0.7143	0.396
Random Forest	0.759	0.392	0.379	0.62	0.381
SVM	0.8182	0.0	0.0	0.5	0.0
MLP	0.734	0.28	0.29	0.66	0.287
Random Forest Completo	0.844	0.64	0.321	0.82	0.350

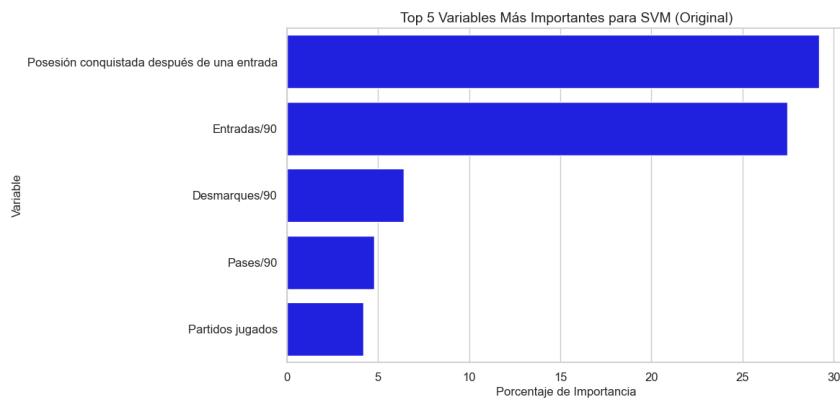


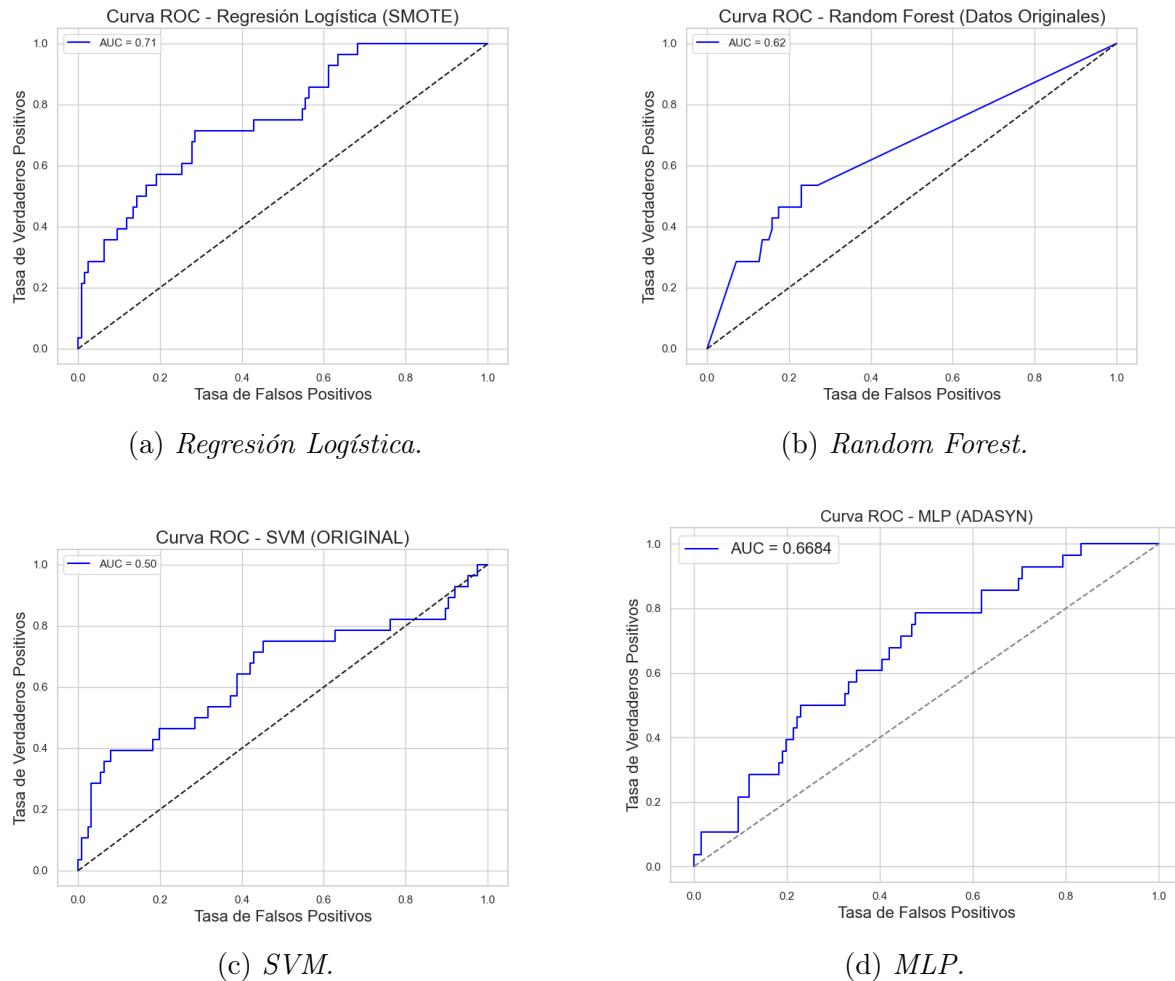
Figura 5.1: *Variables más importantes de la SVM.*

Fuente: Elaboración propia.

5.1.3. Importancia de las variables

Las variables que la Regresión Logística identificó como más importantes para la predicción fueron la **precisión en los pases**, los **desmarques cada 90 minutos** y los **regates cada 90 minutos** tal y como se observa en la imagen 5.4. Estos hallazgos sugieren que atributos técnicos específicos son determinantes en la predicción del éxito según este modelo. Además, coincide con las ideas que tenían los expertos de la materia en cuanto a que los jugadores necesitan dominar con un nivel muy alto estas acciones técnicas para poder llegar a convertirse en futbolistas profesionales.

Las características que Random Forest con todas las variables detectó como más importantes fueron, como ver en la figura 5.5, el **peso**, la **altura** y las **aceleraciones cada 90 minutos**. La prominencia de variables físicas en este modelo, a diferencia de la Regresión Logística, sugiere que Random Forest pudo haber capturado patrones diferentes o que el ADASYN influyó en la forma en que el modelo ponderó las características. La

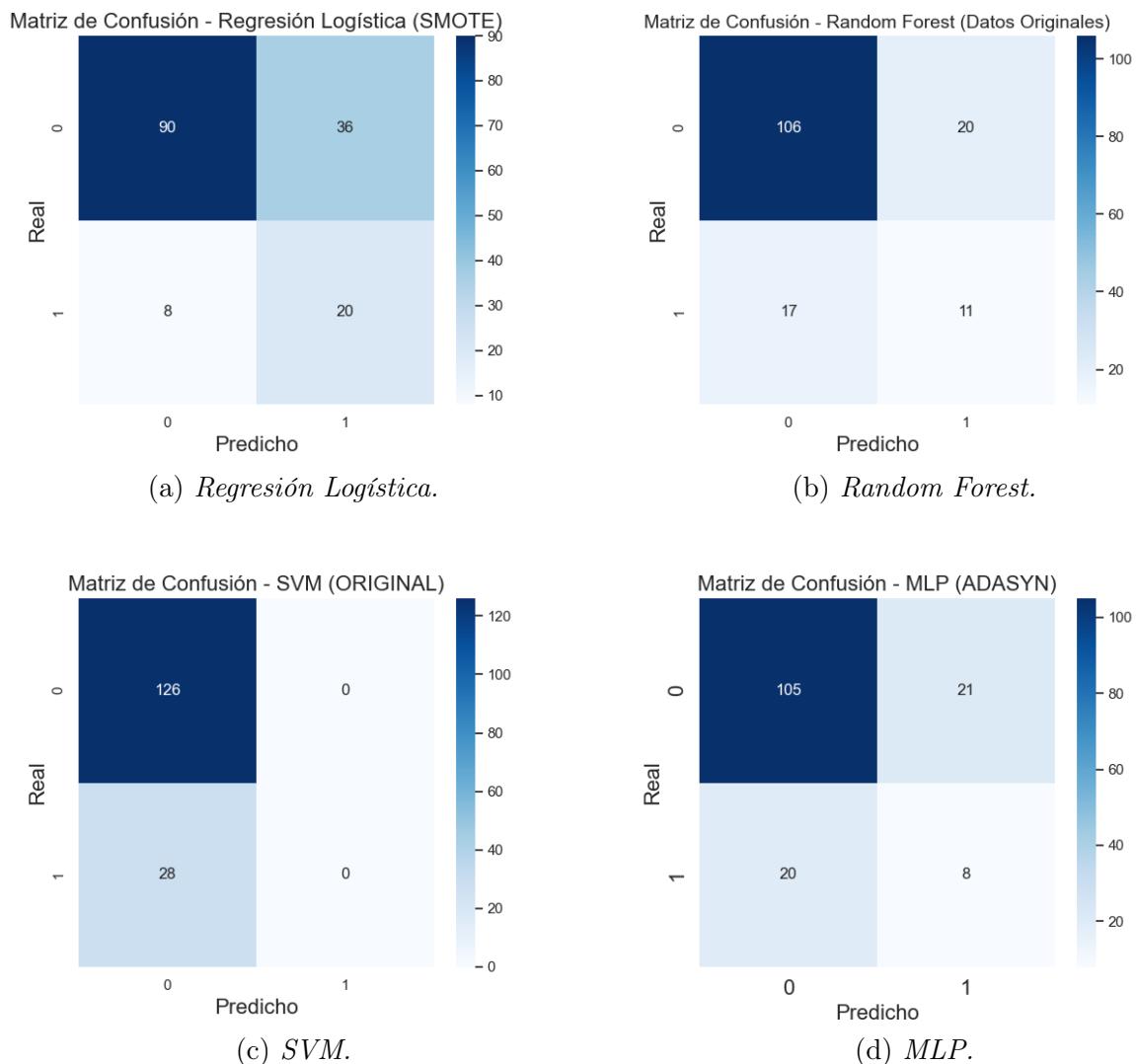
Figura 5.2: *Curvas ROC de los diferentes modelos.*

Fuente: Elaboración propia.

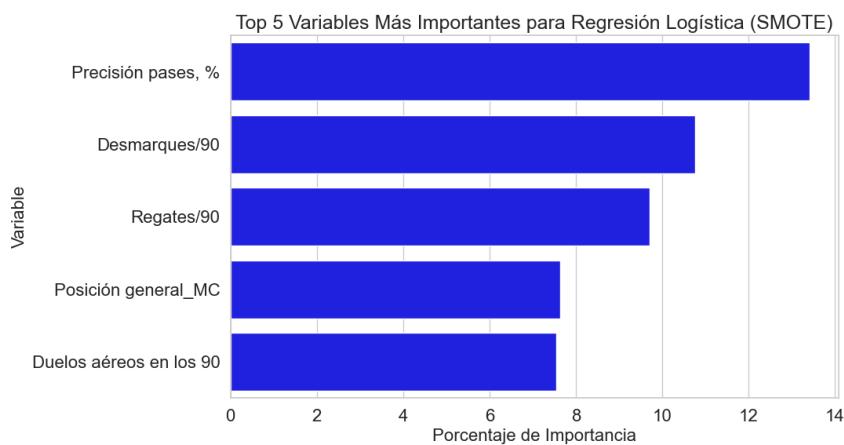
relación entre el peso, la altura y las aceleraciones sugiere que los futbolistas deben ser cada vez más atletas y combinar estas dos facetas para poder llegar al éxito deportivo; esto implica no solo que deben ser pesados, sino controlar la proporción entre una y otra variable, ya que tener índices físicos descompensados es perjudicial para ellos, tal y como nos respondieron los integrantes del Área de Preparación Física.

En cuanto a la importancia de las variables encontrada por la SVM, destacaron de forma muy diferenciada tal y como vemos en el gráfico 5.1 la **posesión conquistada tras una entrada** y las **entradas por cada 90 minutos**, lo que sugiere que para este modelo, las métricas defensivas centradas en la recuperación del balón fueron las más influyentes, aunque no se tradujo en una buena capacidad predictiva general. Aún así, esto puede estar dándonos a entender que la capacidad defensiva de los jugadores está muy relacionada con el número de jugadores que llegan al fútbol profesional. Es decir, tal y como hemos visto en el gráfico 4.2, la mayoría de los jugadores que llegan son defensas, por lo tanto, tener unos buenos índices defensivos ayuda a los jugadores a acercarse al éxito deportivo.

Además, se compararon las 20 variables que el modelo Random Forest entrenado con todas las características originales había seleccionado como las más importantes con las otras 20 que los expertos consideraron para el estudio y posteriormente se aplicaron a los

Figura 5.3: *Matrices de confusión de los diferentes modelos.*

Fuente: Elaboración propia.

Figura 5.4: *Variables más importantes de la Regresión Logística.*

Fuente: Elaboración propia.

modelos. Entre un conjunto y otro coinciden el **30 %** de las variables siendo las siguientes:

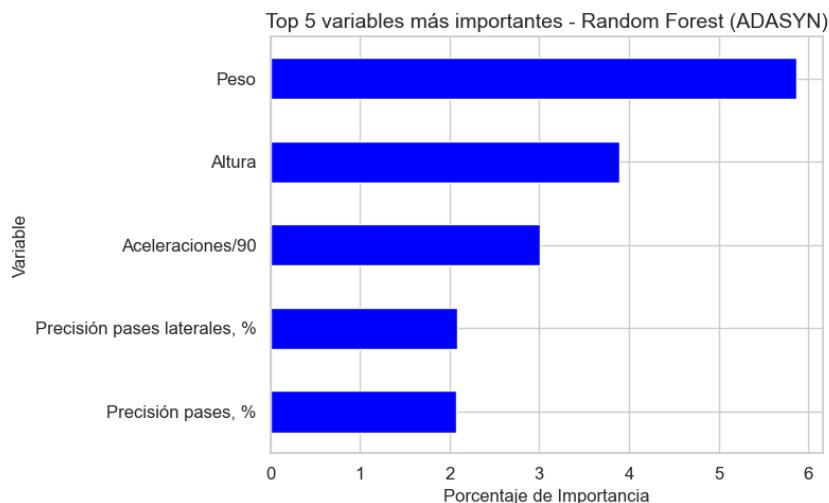


Figura 5.5: *Variables más importantes del Random Forest con todas las variables.*

Fuente: Elaboración propia.

Duelos aéreos cada 90', Precisión en pases, Pases progresivos cada 90', Desmarques cada 90', Precisión en desmarques y Regates cada 90'. Esta coincidencia nos indica que nuestros modelos han conseguido captar la necesidad técnica y la necesidad de ganar duelos que los expertos habían contrastado en los formularios.

5.2. Discusión

Los hallazgos sobre la importancia de las variables revelan perspectivas diversas según el modelo, lo que invita a una reflexión profunda sobre las capacidades que realmente se valoran en el camino hacia el fútbol profesional.

Énfasis en lo técnico vs. lo físico-atlético

La **Regresión Logística** subraya la relevancia de cualidades técnicas refinadas como la **precisión en los pases**, los **desmarques cada 90 minutos** y los **regates cada 90 minutos**. Esta prevalencia de variables técnicas en un modelo que ha demostrado ser eficaz en la detección de éxitos profesionales concuerda plenamente con las ideas de los expertos consultados: un dominio de estas acciones a un alto nivel es fundamental para los futbolistas de élite.

Por otro lado, el modelo Random Forest destacó de manera significativa el **peso**, la **altura** y las **aceleraciones cada 90 minutos**. Esta sobrevaloración de aspectos físicos no solo refuerza la observación de la *creciente necesidad de convertir a los futbolistas en atletas*, tal y como se destacó en la sección 2, sino que también subraya que sin estas capacidades atléticas, los jugadores no pueden competir eficazmente al máximo nivel. La importancia del peso y la altura, en particular, sugiere que la combinación de características antropométricas y funcionales (aceleraciones) es crítica. La reflexión de los preparadores físicos acerca de la proporción entre estas variables es crucial, ya que un físico descompensado puede ser perjudicial. Este modelo, entrenado con ADASYN, podría haber enfatizado la necesidad de una base física sólida para sobresalir en un entorno

competitivo, complementando las habilidades técnicas.

La sobrevaloración de métricas derivadas (xG, xA)

Un patrón recurrente observado en todos y cada uno de los modelos es la **sobrevaloración de variables derivadas** de otras, como los xG (goles esperados) o xA (asistencias esperadas). Esta tendencia revela una limitación intrínseca de estos indicadores en entornos formativos. En etapas de desarrollo, donde el **error es un aspecto inherente y constante** en el aprendizaje de los jugadores, la aplicación de métricas de rendimiento orientadas a resultados puros puede ser engañoso. Los xG y xA , si bien útiles en el fútbol profesional para evaluar el rendimiento de los equipos y jugadores en contextos de alta eficiencia, pueden sacar de contexto el proceso de desarrollo individual. Un jugador joven podría tener un xG bajo simplemente porque está asumiendo riesgos necesarios para su crecimiento o porque su contexto de equipo no favorece la creación de oportunidades "limpias", no necesariamente por falta de potencial.

El jugador como sistema complejo: Un desafío para los modelos

La discusión con expertos ha enfatizado la **necesidad de entender al jugador como un todo**, desde la perspectiva de los sistemas complejos, y no como la mera suma de sus partes. Esto implica que las cualidades (técnicas, físicas, psicológicas, tácticas) no operan de forma aislada, sino en interacción dinámica para resolver situaciones de juego. Aunque los modelos han identificado variables importantes, no hemos sido capaces de detectar esta idea como tal en nuestros resultados. La falta de una clara preponderancia de unas variables sobre otras, o de interacciones complejas detectables directamente, podría sugerir que, o bien los modelos propuestos no son suficientemente complejos para capturar esta naturaleza sistémica del jugador, o que en el contexto de los datos disponibles, no existe un subconjunto reducido de variables que por sí mismas sean abrumadoramente más importantes que el resto. Esto podría llevar a la interpretación de que **no existen variables individuales que dominen por encima de las demás** de forma universal, sino que el éxito emerge de una combinación equilibrada y adaptativa de múltiples habilidades.

Capítulo 6

Conclusiones y proyección futura

En este capítulo vamos a tratar de exponer de la forma más clara posible cuáles han sido las conclusiones a las que hemos llegado tras analizar los resultados del capítulo anterior, abordar las limitaciones de nuestro estudio con el fin de ver cuáles han sido las barreras que nos hemos encontrado y proponer las posibles líneas de investigación futuras o caminos a explorar para posteriores estudios.

6.1. Conclusiones

Tras analizar los resultados y contrastarlos con las opiniones y respuestas de los integrantes del Área de Identidad de la Academia VCF, estas son algunas de las conclusiones a las que hemos llegado:

- Nuestros modelos han conseguido encontrar paralelismos entre sus resultados y las preguntas iniciales que nos planteábamos al inicio del estudio. Las respuestas de los expertos en la materia coinciden en grandes rasgos con las de nuestros modelos. Principalmente en la participación continuada y sostenida a través del tiempo en los equipos de las academias o categorías cercanas al profesionalismo representadas mediante las variables de *Partidos jugados* o *Minutos jugados*. Estos factores les hace enfrentarse a diferentes situaciones y contextos que son cruciales para crear esa mejoría como deportistas de élite.
- Nuestros resultados obtenidos en la **Regresión Logística** (que es el modelo que ha obtenido un mayor valor de la métrica F-Beta Score tal y como hemos visto en la sección 5.1.2) destacan en gran medida los aspectos técnicos como la *Precisión en los pases*, *Regates/90* y *Desmarques/90*. Esto nos indica que el aspecto técnico sigue siendo fundamental para que los jugadores lleguen al fútbol profesional donde se van a encontrar a deportistas capaces de llevar a cabo acciones de mucho nivel en este sentido y deben estar preparados para lidiar con estos oponentes.
- En el caso concreto de los jugadores de la Academia del Valencia CF no hemos podido llegar a conclusiones significativas ya sea debido al tamaño de la muestra o que en ella faltan jugadores que en los últimos años han llegado al primer equipo (Diego López, Javi Guerra, Cristhian Mosquera, Fran Pérez o César Tárrega) ya que estos jugadores competían en 3ra RFEF y 2a RFEF en las temporadas 2021-2022 y 2022-2023. Además, al tratarse de Marco Camús como el único caso de

éxito en nuestro banco de datos, esto puede que haya provocado un desbalance suficientemente significativo como para no llegar a las conclusiones que deseábamos.

En resumen, el estudio nos ha servido para conocer de primera mano cuáles son los aspectos fundamentales que llevan a los jugadores de las academias o en etapas de formación al éxito deportivo. Nuestras conclusiones nos indican que la capacidad técnica, física y la continuidad en el tiempo son variables fundamentales para dicho objetivo; además, contrastando con las opiniones del Área de Identidad, hemos podido comprobar cómo nuestro estudio tiene un valor complementario que ayuda a tomar decisiones más objetivas y descubrir vías aún no exploradas en entornos deportivos formativos.

6.2. Limitaciones del estudio

El presente estudio, a pesar de su rigurosidad metodológica y los hallazgos significativos obtenidos, presenta una serie de **limitaciones** que deben ser consideradas para una correcta interpretación de sus resultados y para futuras investigaciones. Estas limitaciones se agrupan en relación con el tamaño y la composición de la muestra, la naturaleza de las variables analizadas, los desafíos técnicos en la recopilación de datos y consideraciones metodológicas específicas:

1. **Tamaño y temporalidad de la muestra:** La recopilación de datos se limitó a las últimas siete temporadas disponibles en WyScout (2018-2019 a 2024-2025 para algunas ligas), con la notable ausencia de información para las temporadas 2021-2022 y 2022-2023 de Segunda RFEF, un periodo que podría haber sido crucial para el desarrollo de ciertos jugadores. Además, el estudio se centró en futbolistas menores de 23 años con al menos 450 minutos disputados, excluyendo categorías juveniles y cadetes, cuyo análisis podría revelar patrones de desarrollo temprano. La limitación geográfica a jugadores en el sistema de ligas español también restringe la generalización de los hallazgos a otros contextos futbolísticos con estructuras de desarrollo diferentes.
2. **Limitación a variables técnicas:** El análisis se basó fundamentalmente en variables técnicas proporcionadas por WyScout, dejando de lado dimensiones cruciales e importantes del rendimiento futbolístico como los atributos físicos (e.g., velocidad, resistencia, potencia), los atributos psicológicos (e.g., mentalidad, resiliencia, liderazgo) y las capacidades tácticas (e.g., comprensión del juego, inteligencia posicional, toma de decisiones). Esta dependencia de una representación parcial del jugador podría limitar la capacidad predictiva del modelo y ofrecer una visión incompleta de los factores que realmente influyen en el éxito profesional.
3. **Possibles sesgos en los datos de WyScout:** Si bien WyScout es una fuente de datos profesional y ampliamente utilizada, no se puede descartar la existencia de posibles sesgos en la recopilación o el registro de ciertas métricas. La forma en que se capturan los eventos en un partido y cómo se agregan las estadísticas podría introducir sutiles desviaciones o errores en los datos que, aunque pequeños, podrían afectar la fiabilidad de las conclusiones del estudio. Una comprensión detallada de la metodología interna de recolección de datos de la plataforma sería necesaria para evaluar el alcance de estos posibles sesgos.

4. Limitaciones técnicas y metodológicas: La construcción de la base de datos enfrentó desafíos técnicos, especialmente en la integración de nombres entre WyScout y Transfermarkt, lo que podría haber introducido errores en la identificación de las trayectorias profesionales. La dependencia de Transfermarkt para la obtención de los minutos en categorías profesionales también supone una vulnerabilidad ante posibles interrupciones o errores en esta fuente externa. Desde una perspectiva metodológica, la definición de éxito profesional basada en un umbral de minutos jugados es operativa pero simplifica una realidad compleja, omitiendo otros indicadores relevantes. Finalmente, la decisión de incluir únicamente jugadores con más de 450 minutos jugados por temporada introduce un posible sesgo de selección. Al excluir a jugadores con menor participación, se podría estar pasando por alto a talentos con trayectorias de desarrollo no lineales o aquellos que alcanzaron el éxito de forma tardía, favoreciendo la inclusión de perfiles más consolidados en el periodo analizado.

6.3. Trabajo Futuro

Las limitaciones identificadas en el presente estudio abren interesantes vías para futuras investigaciones que permitan profundizar en la comprensión del desarrollo del talento futbolístico y superar las restricciones encontradas. Entre las principales líneas de trabajo futuro se destacan:

- 1. Ampliación del espectro de variables:** La inclusión de datos relativos a las condiciones físicas (como velocidad, potencia y resistencia), los atributos mentales (como la resiliencia, la capacidad de liderazgo y la inteligencia emocional) y las capacidades tácticas (como la comprensión del juego, el posicionamiento y la toma de decisiones) permitiría construir modelos predictivos más robustos y holísticos. La integración de estas dimensiones clave del rendimiento futbolístico, que en este estudio no pudieron ser consideradas debido a la disponibilidad de datos, enriquecería significativamente la capacidad de identificar y predecir el éxito profesional.
- 2. Análisis comparativo:** Sería de gran valor extender el estudio a un marco comparativo. Analizar y contrastar el desarrollo de competencias entre diferentes canteras dentro de un mismo país o, aún más ambicioso, entre diferentes países con distintas metodologías de formación y estructuras de ligas juveniles, podría revelar qué competencias son priorizadas y mejor desarrolladas en cada contexto. Esta perspectiva comparada ofrecería información valiosa para la optimización de los programas de desarrollo de talento.
- 3. Integración de los porteros:** Dada la especificidad de su rol y las métricas particulares que definen su rendimiento, los porteros fueron excluidos del presente estudio para evitar inconsistencias en los modelos diseñados para jugadores de campo. Un trabajo futuro podría abordar la creación de modelos específicos para la predicción del éxito profesional en esta posición clave, utilizando variables adaptadas a sus funciones y responsabilidades.
- 4. Desarrollo de una infraestructura de gestión de datos en academias:** Actualmente, la información relevante sobre los jugadores (datos técnicos, físicos, mentales, tácticos, informes de scouting, etc.) suele estar dispersa en múltiples fuentes

y formatos, lo que dificulta su acceso, análisis y aprovechamiento a largo plazo. La creación de una plataforma centralizada, escalable y sostenible en el tiempo, que permita integrar todas las fuentes de datos de un mismo club, se erige como un objetivo estratégico fundamental. Esta infraestructura de datos no solo facilitaría la realización de estudios longitudinales y el seguimiento del desarrollo individualizado de los jugadores, sino que también permitiría la identificación temprana de patrones y tendencias, la evaluación de la efectividad de las metodologías de entrenamiento y la creación de modelos predictivos internos adaptados a la filosofía y las necesidades específicas del club.

5. **Análisis del impacto de las lesiones:** Investigar cómo las lesiones sufridas durante las etapas formativas y tempranas de la carrera profesional influyen en la trayectoria y el éxito final de los futbolistas. La inclusión de datos de historial de lesiones podría proporcionar información valiosa para la prevención y la gestión del riesgo.
6. **Estudio del impacto de diferentes estilos de juego y entrenadores:** Analizar si la exposición a diferentes estilos de juego y la influencia de diversos entrenadores en las categorías inferiores tiene un impacto significativo en el desarrollo de competencias específicas y en la probabilidad de alcanzar el éxito profesional.

Apéndice A

Anexos

A.1. Listado de Competiciones Consideradas Profesionales

En base a los criterios establecidos en la Sección 3 del cuerpo principal de este trabajo, las siguientes competiciones se consideran profesionales para el análisis realizado:

UEFA:

- UEFA Champions League
- UEFA Europa League
- UEFA Conference League

España:

- LaLiga EA Sports
- LaLiga Hypermotion
- Copa del Rey

Inglaterra:

- Premier League
- Championship
- FA Cup
- Community Shield

Alemania:

- Bundesliga
- 2. Bundesliga

- DFB-Pokal

Italia:

- Serie A
- Serie B
- Coppa Italia
- Supercoppa Italiana

Francia:

- Ligue 1
- Ligue 2
- Coupe de France

Portugal:

- Liga Portugal
- Taça de Portugal
- Allianz Cup
- Supertaça

Países Bajos:

- Eredivisie
- TOTO KNVB Beker

Bélgica:

- Jupiler Pro League
- Croky Cup

Otros países europeos:

- Superliga (Dinamarca)
- Allsvenskan (Suecia)
- Eliteserien (Noruega)
- Ekstraklasa (Polonia)
- Fortuna Liga (Chequia)
- Super League (Suiza)

- Süper Lig (Turquía)
- SuperSport HNL (Croacia)
- Super League 1 (Grecia)

CONMEBOL:

- Copa Libertadores
- Copa Sudamericana
- Liga Profesional de Argentina
- Campeonato Brasileiro Série A
- Campeonato Uruguayo
- Primera División de Chile
- Categoría Primera A (Colombia)
- LigaPro (Ecuador)
- Primera División de Paraguay
- Liga 1 (Perú)
- Primera División de Venezuela

A.2. Variables originales

La lista de variables originales del estudio extraídas de WyScout es la siguiente:

- Jugador
- Equipo
- Equipo durante el período seleccionado
- Posición específica
- Edad
- Valor de mercado (Transfermarkt)
- Vencimiento contrato
- Partidos jugados
- Minutos jugados
- Goles
- xG

- Asistencias
- xA
- Duelos/90
- Duelos ganados, %
- País de nacimiento
- Pasaporte
- Pie
- Altura
- Peso
- En prestamo
- Acciones defensivas realizadas/90
- Duelos defensivos/90
- Duelos defensivos ganados, %
- Duelos aéreos en los 90
- Duelos aéreos ganados, %
- Entradas/90
- Posesión conquistada después de una entrada
- Tiros interceptados/90
- Intercepciones/90
- Posesión conquistada después de una interceptación
- Faltas/90
- Tarjetas amarillas
- Tarjetas amarillas/90
- Tarjetas rojas
- Tarjetas rojas/90
- Acciones de ataque exitosas/90
- Goles/90
- Goles (excepto los penaltis)
- Goles, excepto los penaltis/90
- xG/90

- Goles de cabeza
- Goles de cabeza/90
- Remates
- Remates/90
- Tiros a la portería, %
- Goles hechos, %
- Asistencias/90
- Centros/90
- Precisión centros, %
- Centros desde la banda izquierda/90
- Precisión centros desde la banda izquierda, %
- Centros desde la banda derecha/90
- Precisión centros desde la banda derecha, %
- Centros al área pequeña/90
- Regates/90
- Regates realizados, %
- Duelos atacantes/90
- Duelos atacantes ganados, %
- Toques en el área de penalti/90
- Carreras en progresión/90
- Aceleraciones/90
- Pases recibidos /90
- Pases largos recibidos/90
- Faltas recibidas/90
- Pases/90
- Precisión pases, %
- Pases hacia adelante/90
- Precisión pases hacia adelante, %
- Pases hacia atrás/90
- Precision pases hacia atrás, %

- Pases laterales/90
- Precisión pases laterales, %
- Pases cortos / medios /90
- Precisión pases cortos / medios, %
- Pases largos/90
- Precisión pases largos, %
- Longitud media pases, m
- Longitud media pases largos, m
- xA/90
- Asistencias/90.1
- Second assists/90
- Third assists/90
- Desmarques/90
- Precisión desmarques, %
- Jugadas claves/90
- Pases en el último tercio/90
- Precisión pases en el último tercio, %
- Pases al área de penalti/90
- Pases hacia el área pequeña, %
- Pases en profundidad/90
- Precisión pases en profundidad, %
- Ataque en profundidad/90
- Centros desde el último tercio/90
- Pases progresivos/90
- Precisión pases progresivos, %
- Goles recibidos
- Goles recibidos/90
- Remates en contra
- Remates en contra/90
- Porterías imbatidas en los 90

- Paradas, %
- xG en contra
- xG en contra/90
- Goles evitados
- Goles evitados/90
- Pases hacia atrás recibidos del arquero/90
- Salidas/90
- Duelos aéreos en los 90.1
- Tiros libres/90
- Tiros libres directos/90
- Tiros libres directos, %
- Córneres/90
- Penaltis a favor
- Penaltis realizados, %

A.3. Preguntas del formulario

Las preguntas que se pasaron a los expertos de la Academia del Valencia CF son las siguientes:

Perfil general del jugador

- Ordena los siguientes factores según su importancia en el desarrollo de un jugador (1 = menos importante, 4 = más importante): técnico, táctico, físico y mental
- ¿Consideras que la posición específica del jugador influye significativamente en qué variables son más relevantes para evaluar su desarrollo?

Evaluación por bloques temáticos

- ¿Cuáles son las variables más relevantes para valorar el seguimiento del rendimiento competitivo?
 - Partidos jugados
 - Minutos jugados
 - Goles
 - Goles/90
 - Asistencias

- Asistencias/90
 - xG / xA
- ¿Cuáles son más relevantes para evaluar la capacidad defensiva de un jugador?
- Duelos defensivos/90
 - % de duelos defensivos ganados
 - Interceptaciones/90
 - Entradas/90
 - Posición recuperada tras una entrada
 - Faltas/90
 - Tarjetas amarillas y rojas
 - Tiros interceptados/90
 - Duelos aéreos y % ganados
- ¿Qué variables consideras más representativas para valorar el impacto ofensivo de un jugador?
- Regates/90 y éxito %
 - Carreras en progresión/90
 - Aceleraciones/90
 - Duelos ofensivos/90 y % ganados
 - Toques en el área rival
 - Tiros y tiros a puerta %
 - Centros y precisión
- ¿Qué variables consideras más representativas de la capacidad de generar ventajas con balón?
- Pases/90 y precisión total
 - Pases progresivos/90 y precisión
 - Pases al último tercio/90
 - Pases al área de penalti/90
 - xA, Second assists/90
 - Jugadas clave/90
 - Desmarques y precisión

Reflexión final

- ¿Cuáles son las 5 variables más importantes para identificar el potencial de un jugador en formación?
- ¿Hay alguna otra variable o indicador no incluido en este cuestionario que consideres clave en el proceso de evaluación?

- ¿Qué variables consideras sobrevaloradas o poco útiles en este tipo de análisis?
- ¿Qué herramientas o métodos utilizáis actualmente en la Academia para valorar el rendimiento y evolución del jugador? ¿Qué papel juegan las estadísticas?
- En tu experiencia, ¿qué variables de las mencionadas se trabajan mejor en la Academia del Valencia CF?

Bibliografía

- [Abbott et al., 2018] Abbott, A., Button, C., Pepping, G.-J., and Collins, D. (2018). The association between early career performance and long-term career attainment in professional football. *Journal of Sports Sciences*, 36(5):586–594.
- [Bergstra and Bengio, 2011] Bergstra, J. and Bengio, Y. (2011). Algorithms for hyper-parameter optimization. In *Advances in neural information processing systems*, pages 2546–2554.
- [Bergstra and Bengio, 2012] Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13:281–305.
- [Breiman, 2001] Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.
- [Centre, 2024a] Centre, F. T. (2024a). Miguel angulo on developing players and coaches alike. Último acceso: 27 marzo 2025.
- [Centre, 2024b] Centre, F. T. (2024b). Valencia cf academy director luis martínez. Último acceso: 27 marzo 2025.
- [CF, 2024] CF, V. (2024). Vcf academy philosophy. Último acceso: 27 marzo 2025.
- [Cortes and Vapnik, 1995] Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.
- [Côté et al., 2007] Côté, J., Baker, J., and Abernethy, B. (2007). Practice and play in the development of sport expertise. *Handbook of Sport Psychology*, page 184–202.
- [España, 2024] España, L. N. (2024). La cantera del valencia: una de las mejores de españa. Último acceso: 27 marzo 2025.
- [Fawcett, 2006] Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861–874.
- [FIFA, 2023] FIFA (2023). Regulations on the status and transfer of players. Artículo 2: Definición de jugador profesional.
- [Gledhill et al., 2017] Gledhill, A., Harwood, C., and Forsdyke, D. (2017). Psychosocial factors associated with talent development in football: A systematic review. *Psychology of Sport and Exercise*, 31:93–112.
- [Gonçalves et al., 2021] Gonçalves, B. et al. (2021). Physical and physiological demands in football talent identification: A systematic review. *PMC*.

- [Gómez-Díaz et al., 2023] Gómez-Díaz, R. et al. (2023). Key factors influencing talent development in youth football: A systematic literature review. *ResearchGate*.
- [Henriksen et al., 2010] Henriksen, K., Stambulova, N., and Roessler, K. K. (2010). Holistic approach to athletic talent development environments: A successful sailing milieu. *Psychology of Sport and Exercise*, 11(3):212–222.
- [Hosmer Jr et al., 2013] Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied logistic regression*. John Wiley Sons.
- [Kidd, 2025] Kidd, R. (2025). S.l. benfica brings famed academy model to u.s. in search of talent. *Forbes*.
- [Larkin and O'Connor, 2023] Larkin, P. and O'Connor, D. (2023). Player talent identification and development in football. *Apunts Sports Medicine*.
- [Larkin and O'Connor, 2017] Larkin, P. and O'Connor, D. (2017). Talent identification and development in soccer: An australian context. *European Journal of Sport Science*, 17(6):715–723.
- [Larsen et al., 2020] Larsen, C. H., Alfermann, D., and Henriksen, K. (2020). A world class academy in professional football: The case of ajax amsterdam. *International Journal of Sports Science & Coaching*, 15(1):69–78.
- [MDPI, 2024] MDPI (2024). Neuroscientific approaches to cognitive performance in elite athletes. *MDPI*.
- [Mitchell, 1997] Mitchell, T. M. (1997). *Machine learning*. McGraw-hill New York.
- [Nádori, 1993] Nádori, L. (1993). Aptitud específica superior a la media que necesita manifestarse. Fuente no localizada directamente. Consultar bases de datos académicas.
- [Omeñaca, 2000] Omeñaca, J. V. R. (2000). La educación en valores desde los deportes de equipo. estudio de la aplicación de un programa sistémico en un grupo de educación primaria. Último acceso: 27 marzo 2025.
- [OneFootball, 2024] OneFootball (2024). Valencia cf are in the top 5 for academy graduate selection across the five big leagues. Último acceso: 27 marzo 2025.
- [Performance, 2023] Performance, X. S. (2023). Exploring the 4-coactive model: Enhancing performance in sport.
- [Powers, 2011] Powers, D. M. W. (2011). Evaluation: From precision, recall and f-measure to roc, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2(1):37–63.
- [Provost et al., 1998] Provost, F., Fawcett, T., and Kohavi, R. (1998). The case against accuracy estimation for comparing induction algorithms. *Machine learning*, 30(2-3):253–272.
- [Reeves et al., 2018] Reeves, M. J., Roberts, S. J., and Littlewood, M. A. (2018). A scoping review of the role of sport science in the development of youth football players. *Journal of Sports Sciences*, 36(1):1–9.

- [Reilly et al., 2000] Reilly, T. et al. (2000). A multidisciplinary approach to talent identification in soccer. *Journal of Sports Sciences*.
- [Ribeiro et al., 2024] Ribeiro, J., Sarmento, H., Figueiredo, A., and Duarte, R. (2024). Influence of sociocultural and economic factors on football talent development: A review. *Physical Culture and Sport. Studies and Research*, 101(1):47–58.
- [Richardson et al., 2016] Richardson, D., Gilbourne, D., Littlewood, M., and Nesti, M. (2016). Qualitative perspectives on how manchester united football club produces elite youth football players. *Soccer & Society*, 17(6):705–721.
- [Rosenblatt, 1958] Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386.
- [RREM, 2023] RREM (2023). Psychological and physical aspects in sports performance. *Lumen Publishing*.
- [Rumelhart et al., 1986] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.
- [Russell and Norvig, 2010] Russell, S. J. and Norvig, P. (2010). *Inteligencia artificial: un enfoque moderno*. Pearson Educación.
- [Sarmento et al., 2018] Sarmento, H. et al. (2018). The role of psychological and social factors in football talent development: A systematic review. *Apunts Sports Medicine*.
- [SEE, 2024] SEE, V. C. (2024). Valencia cf academy, the fourth best in europe. Último acceso: 27 marzo 2025.
- [Simplifaster, 2023] Simplifaster (2023). The four-coactive model in player preparation: Enhancing football performance.
- [Sports, 2025] Sports, B. (2025). Real madrid has a better academy than barcelona, and these numbers prove it. *BEIN Sports*.
- [Sáez de Villarreal et al., 2022] Sáez de Villarreal, E. et al. (2022). Tactical awareness and decision-making in football: An integrative approach. *PMC*.
- [Till and Baker, 2020] Till, K. and Baker, J. (2020). Developmental pathways in sport: Comparing talented male and female athletes. *Journal of Sports Sciences*, 38(14):1594–1600.
- [UEFA, 2024] UEFA (2024). Uefa club licensing and financial sustainability regulations. Artículo 54: Contratos escritos con jugadores profesionales.
- [Unnithan et al., 2021] Unnithan, V. et al. (2021). Football-specific technical and tactical skills assessment in youth development programs. *PMC*.
- [Vaeyens et al., 2008] Vaeyens, R., Lenoir, M., Williams, A. M., and Philippaerts, R. M. (2008). Talent identification and development programmes in sport: Current models and future directions. *Sports Medicine*, 38(9):703–714.
- [Williams and Reilly, 2000] Williams, A. M. and Reilly, T. (2000). Talent identification and development in soccer. *Journal of Sports Sciences*, 18(9):657–667. Último acceso: 27 marzo 2025.

[Woods et al., 2020] Woods, C. T., Robertson, S., and Sinclair, W. H. (2020). The influence of environment on the development of elite athletes in australian football. *Talent Development Excellence*, 12(1):43–54.

[Zehnder, 2016] Zehnder, E. (2016). Fc barcelona and its exemplary academy la masia. *Egon Zehnder Insights*.