



ESCUELA TÉCNICA SUPERIOR DE INGENIEROS
INFORMÁTICOS

UNIVERSIDAD POLITÉCNICA DE MADRID

Modelos de aprendizaje automático sobre el juego del club Movistar Estudiantes

TRABAJO FIN DE MÁSTER

MÁSTER UNIVERSITARIO EN INTELIGENCIA ARTIFICIAL

AUTOR: Ander Alcón Ferreira
TUTOR/ES: Concha Bielza Lozoya y
Pedro Larrañaga Múgica

2018

AGRADECIMIENTOS

Agradecer a mis padres todo el cariño que me han mostrado siempre, además del esfuerzo que han tenido que hacer para que yo pudiera realizar el máster. A mi hermano, por ayudarme a ser quien soy y regalarme esos momentos tan bonitos e inolvidables. A Henar, por darme todo su amor, estar a mi lado y hacerme disfrutar de la vida como sólo ella sabe hacerlo. A Daki.

También me gustaría agradecer la oportunidad que me han proporcionado Pedro Larrañaga y Concha Bielza, además del apoyo incondicional y la ayuda que he recibido por su parte. A Marco y Sergio por su inagotable paciencia y amabilidad, dispuestos a brindarme su ayuda en todo momento sin poner ni una sola mala cara.

A José Ángel Samaniego por su colaboración y amabilidad, mostrándose dispuesto a ayudar siempre que lo he necesitado. Al club Movistar Estudiantes, por proporcionarme los medios para poder desarrollar este trabajo.

El trabajo ha sido parcialmente apoyado por el gobierno regional de Madrid a través del proyecto S2013/ICE-2845-CASI-CAM-CM.

RESUMEN

La liga ACB es una de las mejores ligas profesionales de baloncesto a nivel mundial, incluyendo alguno de los mejores equipos de todo Europa. El nivel de competitividad de la liga es muy alto, mostrando un nivel de entropía de Shannon muy similar al de la National Basketball Association (NBA) de los Estados Unidos. Con un nivel tan alto de competitividad, ser capaz de maximizar el rendimiento del equipo y los jugadores es fundamental para poder desmarcarse del resto de equipos presentes en la liga.

Con este objetivo, en esta memoria se realiza una modelización del juego ofensivo y defensivo del club Movistar Estudiantes utilizando redes bayesianas que se aprenden de datos reales recogidos a lo largo de la temporada 2017-18. Se analizan las relaciones probabilísticas entre las variables que se han identificado como predictoras de las victorias y derrotas basándose en los trabajos sobre la materia realizados previamente. Asimismo, la interpretación de la red bayesiana generada proporciona información muy útil sobre los comportamientos positivos y negativos que pueden estar afectando al resultado de los partidos. El tipo de juego realizado por el equipo se encuentra categorizado en cuatro grandes grupos: bloqueo directo, bloqueo indirecto, perímetro poste y juego rápido. Para cada tipo de juego, se predicen los puntos esperados basándose en el tipo de tiro realizado por el jugador (tiro libre, tiro de dos o tiro triple). Esto permite escoger la mejor selección de tiro para cada tipo de juego, maximizando los puntos esperados ofensivamente y minimizando los esperados defensivamente.

Adicionalmente, se ha utilizado una regresión lineal regularizada *ridge* para evaluar el rendimiento en base a los puntos esperados de cada jugador en diferentes situaciones del encuentro, permitiendo escoger individualmente el quinteto de jugadores que mejor se adapta a cada momento del partido. Para obtener más información del modelo, se han incluido parejas de jugadores para medir su rendimiento cuando se encuentran en la cancha al mismo tiempo. Además, como en determinados momentos del partido las acciones que se toman pueden decidir el resultado final, se ha incluido la variable *clutch time* para ver cómo se manejan los jugadores bajo presión.

Este estudio es capaz de representar de forma precisa el juego del club Movistar Estudiantes, además de mostrar la eficacia de los jugadores en distintas situaciones del partido, lo que permite ajustar en cada momento el quinteto de juego a las necesidades del equipo.

SUMMARY

ACB league is one of the best professional basketball leagues around the world, including some of the best teams across Europe. The level of competitiveness in the league is very high, showing a Shannon entropy level similar to the National Basketball Association (NBA) of the United States. With such high level of competitiveness, being able to maximize the performance of the team and the players could lead to a significant difference over the rest of the teams in the league.

With this purpose, a modelization of both offensive and defensive games of the ACB team Club Movistar Estudiantes is made using Bayesian networks. An analysis of the probabilistic relationships among several team variables that have been identified as relevant predictors of wins and losses in previous works is performed. Additionally, the interpretation of the networks provides useful information about positive and negative behaviors that are affecting the outcome of the games. Four different types of plays are defined to classify the offensive and defensive situations: pick, pick away, perimeter post and fast break. For each type of play, the expected points are calculated based on the shot made by the player (free throw, three point shot and two point shot). This allows to choose the best option for each type of play, maximizing the expected points offensively and minimizing them defensively.

In addition, a *ridge* regularized linear regression is used to evaluate the performance of each team player in different situations of the game, allowing to choose individually which set of players should be playing on each sequence of the game. To extract more information from the model, pairs of players are also evaluated to see how well these combinations work in the field at the same time. Furthermore, as there are moments in the game where the chosen decision can affect the outcome of the game, the variable clutch time is included to see the behavior of the players under pressure.

This study is able to represent precisely how the club Movistar Estudiantes plays, and shows the effectiveness of the players in different situations of the game, allowing to choose the best line-up according to the needs of the team.

BIBLIOGRAFÍA

Índice de figuras	ix
Índice de cuadros	xi
1. INTRODUCCIÓN	1
1.1. Aprendizaje automático en el deporte	2
1.2. Baloncesto: Análisis de rendimiento	3
1.3. Objetivos	4
1.4. Planificación de la tesis	5
2. METODOLOGÍAS	8
2.1. Redes bayesianas	8
2.2. Regresión lineal con regularización	11
2.3. Random forest	13
2.4. Red neuronal H2O	13
2.5. K-nearest neighbors	14
2.6. Test de Friedman y post-hoc	14
3. MODELIZACIÓN DEL ESTILO DE JUEGO	16
3.1. Modelo ofensivo	18
3.1.1. Tipo de juego	18
3.1.1.1. Bloqueo directo	19
3.1.1.2. Bloqueo indirecto	21
3.1.1.3. Juego rápido	23
3.1.1.4. Perímetro poste	24
3.1.1.5. Comparación de los tipos de juego	25
3.1.2. Periodo	26
3.1.2.1. Periodo 1	28
3.1.2.2. Periodo 2	29
3.1.2.3. Periodo 3	30
3.1.2.4. Periodo 4	31
3.1.2.5. Comparación de periodos	32
3.2. Modelo defensivo	33

3.2.1.	Tipo de juego	33
3.2.1.1.	Bloqueo directo	34
3.2.1.2.	Bloqueo indirecto	36
3.2.1.3.	Juego rápido	37
3.2.1.4.	Perímetro poste	38
3.2.1.5.	Comparación de los tipos de juego	40
3.2.2.	Periodo	41
3.2.2.1.	Periodo 1	42
3.2.2.2.	Periodo 2	43
3.2.2.3.	Periodo 3	44
3.2.2.4.	Periodo 4	45
3.2.2.5.	Comparación de los periodos	46
3.2.2.6.	Comparación de los resultados obtenidos al crear las redes	47
4.	MODELIZACIÓN DE LOS JUGADORES	50
4.1.	Estructuración de los datos	51
4.2.	Rendimiento ofensivo	52
4.2.1.	Resultados	53
4.2.2.	Comparación	56
4.3.	Rendimiento defensivo	60
4.3.1.	Resultados	60
4.3.2.	Comparación	63
4.4.	Rendimiento Plus-Minus	67
4.4.1.	Resultados	67
4.4.2.	Comparación	70
5.	CONCLUSIONES Y TRABAJO FUTURO	75
5.1.	Conclusiones	75
5.1.1.	Modelización del estilo de juego	75
5.1.2.	Modelización de los jugadores	76
5.1.3.	Conclusiones generales	77
5.2.	Trabajo futuro	78
6.	BIBLIOGRAFÍA	80

Índice de figuras

1.	Red bayesiana ofensiva por tipo de juego.	19
2.	Red bayesiana ofensiva con bloqueo directo.	20
3.	Red bayesiana ofensiva con bloqueo indirecto.	22
4.	Red bayesiana ofensiva con juego rápido.	23
5.	Red bayesiana ofensiva con perímetro poste.	24
6.	Red bayesiana ofensiva por periodo.	27
7.	Red bayesiana ofensiva en el primer periodo.	28
8.	Red bayesiana ofensiva en el segundo periodo.	29
9.	Red bayesiana ofensiva en el tercer periodo.	30
10.	Red bayesiana ofensiva en el cuarto periodo.	32
11.	Red bayesiana defensiva por tipo de juego.	34
12.	Red bayesiana defensiva con bloqueo directo.	35
13.	Red bayesiana defensiva con bloqueo indirecto.	36
14.	Red bayesiana defensiva con juego rápido.	38
15.	Red bayesiana defensiva con perímetro poste.	39
16.	Red bayesiana defensiva por periodo.	42
17.	Red bayesiana defensiva en el primer periodo.	43
18.	Red bayesiana defensiva en el segundo periodo.	43
19.	Red bayesiana defensiva en el tercer periodo.	44
20.	Red bayesiana defensiva en el cuarto periodo.	45
21.	Ejemplo de la estructura de la base de datos.	52
22.	Coefficientes de regresión del modelo ofensivo obtenidos mediante regresión lineal <i>ridge</i>	53
23.	Variación de error cuadrático medio entre usar (rojo) y no usar (amarillo) las combinaciones de jugadores y la variable <i>clutch time</i> (rendimiento ofensivo).	58
24.	Resultados del rendimiento ofensivo del test de Friedman y análisis post-hoc entre las regresiones en términos de ECM.	59
25.	Resultados del test de Friedman y el análisis post-hoc entre la regresión <i>ridge</i> y los métodos de aprendizaje automático para el rendimiento ofensivo en términos de ECM.	59
26.	Coefficientes de regresión del modelo defensivo obtenidos mediante regresión lineal <i>ridge</i>	61

27.	Diferencia de error cuadrático medio entre usar (rojo) y no usar (amarillo) las combinaciones de jugadores y la variable <i>clutch time</i> (rendimiento defensivo).	65
28.	Test de Friedman y análisis post-hoc entre las regresiones para el ECM del rendimiento defensivo.	65
29.	Test de Friedman y análisis post-hoc entre la regresión <i>ridge</i> y los métodos de aprendizaje automático para el ECM del rendimiento defensivo.	66
30.	Coefficientes de regresión del modelo <i>plus-minus</i> obtenidos mediante regresión lineal <i>ridge</i> .	68
31.	Diferencia de error cuadrático medio entre usar (rojo) y no usar (amarillo) las combinaciones de jugadores y la variable <i>clutch time</i> (rendimiento <i>plus-minus</i>).	72
32.	Test de Friedman y análisis post-hoc entre las regresiones para el ECM del rendimiento <i>plus-minus</i> .	72
33.	Test de Friedman y análisis post-hoc entre la regresión <i>ridge</i> y los métodos de aprendizaje automático para el ECM del rendimiento <i>plus-minus</i> .	73

Índice de cuadros

1.	Equivalencias de las variables de las redes.	18
2.	Variaciones de las probabilidades de las variables por tipo de juego en el modelo ofensivo.	26
3.	Variaciones de las probabilidades de las variables por periodo en el modelo ofensivo.	33
4.	Variaciones de las probabilidades de las variables por tipo de juego en el modelo defensivo.	40
5.	Variaciones de las probabilidades de las variables por periodo en el modelo defensivo.	46
6.	Correspondencia entre abreviaturas y nombres de los métodos de aprendizaje de la red.	47
7.	Comparación de los métodos de aprendizaje de la red utilizados.	48
8.	Equivalencia entre abreviaturas y los nombre originales.	56
9.	Comparación de la regresión lineal <i>ridge</i> con métodos MLR para el rendimiento ofensivo.	57
10.	Resultados obtenidos sin combinaciones de jugadores ni <i>clutch time</i> para el rendimiento ofensivo.	57
11.	Comparación de la regresión lineal <i>ridge</i> con métodos MLR para el rendimiento defensivo.	64
12.	Resultados obtenidos sin combinaciones de jugadores ni <i>clutch time</i> para el rendimiento defensivo.	64
13.	Comparación de la regresión lineal <i>ridge</i> con métodos MLR para el rendimiento <i>plus-minus</i>	71
14.	Resultados <i>plus-minus</i> obtenidos sin combinaciones de jugadores ni <i>clutch time</i>	71

1. INTRODUCCIÓN

El deporte es un fenómeno global que tiene un gran impacto en la sociedad. Concretamente, el baloncesto es uno de los deportes más practicados a nivel mundial, alcanzando un grado de popularidad que supera a otros deportes de gran calado como el fútbol americano (Miller et al. (2017)). Este alto índice de popularidad ha permitido desarrollar competiciones de máxima exigencia donde se miden los mejores jugadores y equipos a nivel profesional. Dentro de las competiciones profesionales existentes, la National Basketball Association (NBA) de los Estados Unidos es sin duda la más importante y popular que existe. La NBA ha conseguido crear una mezcla entre deporte, espectáculo y medios de comunicación que la han convertido en la liga de baloncesto referente, atrayendo a los mejores jugadores de baloncesto de la actualidad (Andrews (1999)). Conseguir atraer a estos jugadores permite que la NBA continúe creciendo, ya que el hecho de poseer a estos jugadores hace que más y más aficionados acudan a ver los partidos o los sigan a través de los medios de comunicación (Humphreys y Johnson (2017)).

Si pasamos al continente europeo, la liga profesional española (ACB) es considerada la segunda mejor liga mundial existente, únicamente por detrás de la ya mencionada NBA. La liga ACB ha mostrado un nivel de competitividad perfectamente comparable al de la NBA, teniendo un nivel de entropía Shannon muy similar (de Saá Guerra et al. (2012)). A pesar de ser la segunda mejor liga mundial, la mayoría de equipos que la componen disponen de un presupuesto limitado, por lo que no pueden permitirse fichar a los mejores jugadores disponibles en el mercado. A la hora de contratar jugadores, tienen que pujar contra equipos europeos que disponen de financiaciones altísimas, lo que reduce las probabilidades de hacerse con dichos jugadores. Por tanto, los equipos se tienen que contentar con fichar a jugadores de perfiles más bajos.

Por tanto, la base del éxito de estos equipos es conseguir maximizar su rendimiento, para conseguir distinguirse del resto de equipos que componen la liga. De esta forma, pueden lograr llegar a cotas que a priori no parecerían factibles. Para cumplir este objetivo, es necesario analizar en profundidad todas las estadísticas recopiladas del equipo, para ser capaces de extraer información relevante que permita alcanzar el máximo potencial. Hasta hace relativamente poco tiempo, era muy difícil sacarle provecho a los datos deportivos, ya que no se disponía de las herramientas y capacidades necesarias para poder analizar la información que se extraía de los partidos. El surgimiento del Big Data ha traído consigo una mejora de la capacidad de computo necesaria para poder aplicar técnicas de aprendizaje automático en el deporte (Rein y Memmert (2016)), permitiendo descubrir información relevante entre todos los datos y estadísticas que se extraen de los jugadores y consiguiendo sacar el máximo beneficio de los recursos disponibles. Para conseguir obtener conocimiento de los datos a través de las técnicas de aprendizaje automático, es necesario escoger el enfoque que más se adapte al problema que se quiere resolver, ya que esto va a permitir maximizar el beneficio obtenido (Stein et al. (2017)).

1.1. Aprendizaje automático en el deporte

Son numerosos los casos en los que se ha logrado aplicar exitosamente el aprendizaje automático en el ámbito deportivo. Se van a exponer varios trabajos de interés, ordenados por el deporte en el que se han realizado. Havard y Oyvind (2000) realizaron una estimación del nivel de calidad de los equipos que componen la máxima competición de fútbol inglesa (Premier League), además de realizar una predicción sobre los resultados de los partidos utilizando un modelo de Poisson independiente. Kampakis (2016) desarrolló un modelo para lograr predecir las lesiones del equipo Tottenham Hotspurs FC de la Premier League, así como el tiempo de recuperación estimado del jugador en caso de haber sufrido una lesión. Para ello, utilizó una gran variedad de algoritmos de aprendizaje, entre los que destacan las máquinas de vectores soporte (SVMs), redes neuronales y el algoritmo random forest. Además, utilizó técnicas de selección de atributos para reducir la dimensionalidad del modelo.

Sipko (2015) propuso un modelo de aprendizaje supervisado utilizando regresión logística y redes neuronales con el objetivo de predecir el resultado de los partidos de tenis. Para definir el modelo, utilizó datos del historial de partidos de los jugadores, además de una combinación de atributos que incluían la fatiga y las lesiones de los jugadores.

En fútbol americano, Lee y Kitani (2016) trataron de predecir la trayectoria que iba a seguir un jugador a la hora de atacar el área del rival. Con este fin, utilizaron un proceso de decisión de Markov para predecir dicha trayectoria.

Macdonald (2012) desarrolló una variante de la estadística *adjusted plus-minus* basada en una regresión lineal *ridge*, para poder evaluar el grado de aportación de los jugadores de hockey hielo en sus respectivos equipos. Para construir el modelo, no tuvo en cuenta la fuerza de los oponentes y de los compañeros, tratando a todos los jugadores por igual y omitiendo la posición del jugador en el equipo.

En baloncesto, Staw y Hoang (1995) trataron de demostrar que el orden en el que son escogidos los jugadores del *draft* afecta a las oportunidades que se le ofrecen en el equipo al que llegan, produciéndose una clara distinción entre los jugadores de la parte alta y baja de la selección. El estudio que realizaron demostró que los jugadores que habían sido elegidos en posiciones más altas disponían de más oportunidades para jugar y confirmar sus cualidades, además de permanecer más tiempo en el equipo. Observaron que hay una tendencia de los equipos a seguir confiando en jugadores que han sido seleccionados en posiciones altas, pese a que éstos no estén teniendo un desempeño positivo. Moxley y Towne (2015) utilizaron un *growth mixture model* (GMM) para predecir el éxito que iba a tener un jugador de baloncesto al incorporarse a la NBA, calculando su potencial y la regularidad que demostraría a lo largo de su trayectoria profesional. Deshpande y Jensen (2016) desarrollaron un modelo bayesiano de regresión lineal para realizar una estimación de la aportación

de los jugadores de la NBA en base al sueldo que percibían, para reflejar si el rendimiento del jugador se correspondía con el salario que estaba percibiendo. Loeffelholz et al. (2009) trataron de predecir el resultado de los partidos de la NBA utilizando distintos tipos de redes neuronales. Además, realizaron una selección de atributos para mejorar el rendimiento de las redes. Su aproximación consiguió optimizar los resultados predichos por los expertos de la NBA.

1.2. Baloncesto: Análisis de rendimiento

Para abordar el problema de maximización del rendimiento del equipo es necesario detectar entre la gran cantidad de estadísticas que se recogen en los partidos de baloncesto cuáles aportan verdadera información. Zak et al. (1979) fue uno de los primeros en determinar qué estadísticas se encuentran más relacionadas con el desenlace final del partido. Tras analizar el potencial real y la producción de los equipos de la NBA en la temporada 1976-1977, concluyó que las estadísticas más relevantes son acierto de tiro, acierto de tiros libres y los rebotes obtenidos por el equipo. Años más tarde, Mai (2004) analizó los partidos de la liga profesional femenina asiática y demostró que la disminución de las faltas personales y pérdidas de balón cometidas por el equipo incrementaba la probabilidad de ganar el partido. Kubatko et al. (2007) consideró que las estadísticas más determinantes en el resultado del partido son el número de tiros realizados, pérdidas de balón, porcentaje de tiros libres y rebotes ofensivos. Pan (2010) añadió el número de tapones realizados por el equipo como otra variable con gran poder predictivo.

Por otro lado, se ha tratado de determinar la aportación individual de un jugador, para poder detectar cuáles son los jugadores más valiosos del equipo y poder tomar decisiones en los partidos a partir de ese conocimiento. Heeren (1988) creó una de las primeras estimaciones del rendimiento de un jugador, otorgando un valor numérico a su aportación individual. Más adelante, Berri (1999) trató de combinar las estadísticas individuales de los jugadores con el número de victorias de los equipos, para posteriormente otorgarles una valoración individual sobre su grado de aportación en el desenlace de los encuentros. Hollinger (2005) propuso la estadística PER (Player Efficiency Rating), la cual muestra la valoración global del jugador en un único coeficiente. Esta estadística tuvo una gran repercusión en la NBA, siendo una de las valoraciones más utilizadas y respetadas. La medida estima que el rendimiento medio de un jugador es de 15 puntos, valor contra el que se comparan los jugadores para ver si su rendimiento ha estado por encima o por debajo de esa media. El PER se calcula a través de una fórmula ponderada en la que se tienen en cuenta los tiros de dos, tiros libres, triples, rebotes, asistencias, robos, tapones y los resultados de acciones negativas (pérdidas, faltas personales y tiros fallados).

Además de las mencionadas anteriormente, otra de las medidas más utilizadas a la hora de otorgar valoraciones es la estadística *plus-minus*, que fue aplicada por primera vez en la década de 1950 por el equipo de Hockey canadiense Montreal

Canadiens. La medida fue poco a poco adoptada por el resto de equipos de la liga, pasando a ser oficialmente recogida por todos los equipos en la liga profesional de Hockey (NHL) estadounidense en la temporada 1967-68. Posteriormente, se produjo una extrapolación a una gran variedad de deportes, entre los que se encuentra el baloncesto. Su objetivo es devolver una medida que muestre la diferencia entre los puntos obtenidos y recibidos por el equipo cuando el jugador se encuentra en la cancha. De esta forma, es posible cuantificar el impacto del jugador en el partido, ya que refleja cómo se comporta el equipo cuando él está en cancha y permite adaptar el quinteto dependiendo de la situación del partido. Esta estadística fue rápidamente incorporada a la NBA, pasando a ser recogida y evaluada en todos los equipos que componían la liga.

Pese a que los métodos de evaluación del rendimiento propuestos anteriormente funcionaban de forma eficaz, Winston (2009) demostró que los valores que se otorgaban a los jugadores sugerían comportamientos inapropiados, como por ejemplo, indicar a los jugadores con malos porcentajes de acierto que tiraran más a canasta para intentar mejorar sus ratios de eficacia. Para resolver este problema, propuso la estadística *adjusted plus-minus* (APM), la cual tenía en cuenta qué jugadores se encuentran jugando y con qué compañeros para evaluar su rendimiento. Con este fin, utilizaba una regresión lineal para realizar las estimaciones de los valores de aportación reales de los jugadores. Niemi (2010) extendió el APM separando las jugadas en ofensivas y defensivas, tratándolas de forma independiente y devolviendo un resultado individual para cada una de ellas. Tras realizar este cálculo, las combinaba en una estimación global final de la aportación del jugador utilizando un modelo jerárquico. Sill (2010) propuso una mejora del APM aplicando una regularización *ridge* a la regresión lineal, además de añadir conocimiento previo sobre los jugadores. Este conocimiento previo permitía ajustar en mayor medida los modelos, ya que éstos partían con una base de conocimiento inicial. Omidiran (2011) proporcionó una interpretación del APM teniendo en cuenta la dispersión de los datos y utilizando también una regresión con penalización.

1.3. Objetivos

El objetivo de esta tesis fin de máster es realizar una modelización del juego del club Movistar Estudiantes, equipo que pertenece a la liga profesional de baloncesto ACB. Se pretende comprender el estilo de juego del equipo, tratando de obtener información que permita optimizar su rendimiento, tanto ofensiva como defensivamente. Se van a realizar dos aproximaciones, una para analizar el juego global del equipo y otra para descubrir la aportación real de cada uno de los jugadores que componen la plantilla en el desarrollo de los partidos.

Para conseguir interpretar el estilo de juego del equipo, se realiza una modelización de los juegos ofensivo y defensivo utilizando redes bayesianas. Para cada una de las redes, se analizan las relaciones probabilísticas entre las variables que se han identificado como predictoras de las victorias y derrotas por trabajos que se han

realizado previamente en la materia. Esto nos permite encontrar comportamientos y tendencias que están afectando al equipo tanto de forma positiva como negativa. A su vez, el modelo puede ayudar al entrenador encontrar nuevas variantes del juego que ayuden a mejorar los resultados obtenidos. Este análisis muestra las carencias y virtudes del equipo, permitiendo al entrenador modificar el estilo de juego para sacar el máximo provecho y aumentar el número de victorias.

Con el fin de dar una evaluación individual de los jugadores, se ha utilizado una regresión lineal *ridge* para crear el modelo. Se han creado tres modelos distintos: uno ofensivo, uno defensivo y uno que refleja la aportación *plus-minus* del jugador. Además de devolver el impacto de cada jugador, se han añadido combinaciones de los jugadores de la plantilla para mostrar su rendimiento cuando se encuentran al mismo tiempo en la cancha, permitiendo encontrar parejas que funcionan muy bien juntos, o, en su defecto, parejas que no se complementan bien. Además, al haber situaciones de partido en las que las decisiones que se toman afectan directamente al resultado final, se ha añadido la variable *clutch time* para ver cuál es el rendimiento de los jugadores bajo presión.

Este trabajo aporta dos métodos novedosos a la hora de analizar el rendimiento del club Movistar Estudiantes. El análisis por tipo de juego y periodo que proporcionan las redes bayesianas no se ha realizado antes en la literatura, por lo que se trata de un nuevo enfoque que ayuda a vislumbrar cómo afectan el tipo de juego y el periodo en los distintos aspectos del juego del equipo y en las relaciones con el resto de variables. Por otro lado, el uso de la regresión lineal *ridge* para determinar el rendimiento de los jugadores fue introducido por Sill (2010), pero en su modelo no incluía combinaciones de jugadores y la variable psicológica *clutch time*. Este trabajo demuestra que la inclusión de estas nuevas variables ayuda a obtener un mejor modelo en términos de error cuadrático medio, además de otorgar un extra de información que resulta muy útil para la toma de decisiones de los entrenadores.

Gracias a este estudio, es posible representar de forma precisa el estilo de juego del club Movistar Estudiantes, permitiendo al entrenador tomar decisiones tácticas para sacar el máximo provecho de los partidos, potenciando las virtudes del equipo y minimizando sus carencias. Además, se muestra el rendimiento de los jugadores en distintas situaciones del partido, permitiendo escoger en todo momento cuáles son los jugadores que mejor se adaptan a la situación actual y satisfacer así las necesidades del equipo.

1.4. Planificación de la tesis

A continuación se muestra un desglose de la estructuración de la tesis:

- Capítulo 1: Contiene una introducción sobre el análisis del baloncesto, trabajo previo realizado en la materia y los objetivos del estudio.
- Capítulo 2: Se introducen las metodologías utilizadas en el desarrollo del proyecto.

- Capítulo 3: Se modeliza el estilo de juego del equipo utilizando redes bayesianas, y se analizan las relaciones probabilísticas creadas entre las variables de la red.
- Capítulo 4: Se realiza una modelización del rendimiento de los jugadores utilizando una regresión lineal *ridge*, proporcionando una valoración ofensiva, defensiva y *plus-minus* equivalente al impacto del jugador en la cancha.
- Capítulo 5: Se incluyen las conclusiones obtenidas tras el desarrollo de este estudio y el posible trabajo futuro a realizar en la materia.

2. METODOLOGÍAS

A continuación se realiza una breve introducción a las dos principales metodologías utilizadas en el desarrollo del proyecto: redes bayesianas y regresión lineal.

2.1. Redes bayesianas

Pearl (1988) definió las redes bayesianas como modelos probabilísticos que combinan teoría de grafos y probabilidad, siendo capaces de modelar de forma eficiente las distribuciones de probabilidad de un grupo de variables $\mathcal{X} = \{X_1, X_2, \dots, X_n\}$. Las redes bayesianas modelizan dependencias e independencias condicionales existentes entre las variables de \mathcal{X} , y están compuestas por dos partes: Un grafo acíclico dirigido (DAG) y un conjunto de parámetros (Koller y Friedman (2009)). El grafo acíclico dirigido representa como nodos las variables aleatorias del conjunto \mathcal{X} , y los arcos entre las variables representan dependencias probabilísticas. Los parámetros de las redes bayesianas reflejan la distribución de probabilidad de cada variable $X_i \in \mathcal{X}$ condicionada a sus nodos padres (con arcos hacia X_i). La siguiente fórmula muestra el cálculo de la probabilidad conjunta, realizando el producto de las probabilidades de cada variables dados sus padres

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | \mathbf{Pa}(X_i))$$

A la hora de aprender una red bayesiana a partir de un conjunto de datos, existen dos métodos principales de aprendizaje:

- Aprendizaje de estructuras basado en pruebas de independencia: Se trata de un método bastante intuitivo, que interpreta las redes bayesianas como una representación de independencias. Está basado en el algoritmo Inductive Causation (IC) diseñado por Verma y Pearl (1991), el cual proporciona un entorno para aprender el DAG de una red bayesiana utilizando los tests de independencia condicional y tratando de reflejar de la mejor forma posible las independencias existentes en la red (Meek (1995); Spirtes et al. (2000); Cheng et al. (2002)), seleccionando el modelo que mejor las representa. El problema de este tipo de aprendizajes es su sensibilidad a los fallos en los tests de independencia individuales, ya que es suficiente que una de estas pruebas devuelva una respuesta incorrecta para confundir a la red en el procedimiento de construcción. Dentro de los tests más utilizados en este tipo de aprendizaje de estructuras, se encuentran los algoritmos *parents and children* (PC) (Spirtes et al. (2000)) y *light mutual min* (LMM) (Mahdi y Mezey (2013)). Estas técnicas no utilizan una medida explícita de ajuste para probar la verosimilitud entre los datos y la red, si no que usan tests estadísticos para obtener el esqueleto de la red y orientar los arcos recuperando las estructuras convergentes ($X_u \rightarrow X_v \leftarrow X_w$) de la red.

- Aprendizaje de estructuras basado en medidas de ajuste y búsqueda: Se define un espacio de hipótesis de los modelos potenciales, es decir, el conjunto de posibles estructuras de la red considerados y una función de puntuación que mide cómo de bien se ajusta la red a los datos proporcionados. El objetivo es encontrar la estructura que maximice la puntuación. El espacio de búsqueda representa el conjunto de todas las redes bayesianas conteniendo un número definido de nodos dado un conjunto de datos. De acuerdo con el número de nodos que contiene el grafo, el número de posibles estructuras que se pueden crear crece exponencialmente. Cuando el número de variables no es muy grande es posible enumerar y evaluar todos los DAGs disponibles y seleccionar el que obtenga un puntaje más alto. Sin embargo, Robinson (1977) demostró que el número de DAGs que contienen n nodos se define a través de la siguiente fórmula

$$f(n) = \sum_{i=1}^n (-1)^{i+1} \binom{n}{i} 2^{i(n-i)} f(n-i) \quad n > 2$$

$$f(0) = 1 \quad , \quad f(1) = 0,$$

lo que demuestra que la búsqueda de la estructura se convierte en un problema NP-Duro (Maxwell (2000)). Debido a que el espacio de búsqueda es demasiado grande, se suelen utilizar algoritmos de búsqueda heurísticos, ya que es imposible evaluar todas las estructuras del espacio de búsqueda. Como este método considera toda la estructura a la vez, es menos sensible a fallos individuales. En cambio, la desventaja de este tipo de aprendizaje de estructura es que presentan un problema de búsqueda que puede que no tenga una solución elegante y eficiente. Entre las funciones de puntuación más relevantes se encuentran el criterio de información bayesiana (BIC) (Schwarz (1978); Rissanen (1986); Barron et al. (1999)), el principio de longitud de descripción mínima (MDL) (Bouckaert (1993); Wai y F. (1994)), equivalencia bayesiana de Dirichlet (BDe) (Heckerman et al. (1995)), el criterio de información Akaike (AIC) (Akaike (1974)) y K2 (Cooper y Herskovits (1991, 1992); Larrañaga et al. (1996a,b)). Uno de los algoritmos más utilizados es el K2 (Cooper y Herskovits (1992)), el cual usa un puntaje de actualización local, lo que le permite ganar más eficiencia ya que sólo recalcula unos pocos valores para obtener el nuevo puntaje del modelo. Otros algoritmos relevantes son el *hill-climbing* (Tsamardinos et al. (2006)) y los algoritmos evolutivos (Wong y Leung (2004)).

Las redes bayesianas representan todas las distribuciones de probabilidad y relaciones de las variables del modelo, pudiendo calcular las probabilidades condicionadas de cualquier variable de la red dada la evidencia de cualquier otra variable. Uno de los principales objetivos de los modelos probabilísticos es ser capaz de responder a consultas sobre probabilidades satisfactoriamente. Con este objetivo, es necesario realizar cierto tipo de razonamiento. Las redes bayesianas son capaces de realizar tres tipos de inferencias:

- Inferencia deductiva: Trata de obtener ciertas conclusiones dada una serie de eventos que han ocurrido.
- Inferencia inductiva: Intenta obtener información sobre las causas de ciertos eventos previamente conocidos.
- Inferencia abductiva: Consiste en obtener las hipótesis más probables dadas una serie de hechos.

En la misma línea, los principales problemas de razonamiento más comunes en redes bayesianas son los siguientes:

- Diagnóstico y predicción: El proceso de inferencia utilizado en el razonamiento inductivo y deductivo se denomina propagación de la probabilidad. Consiste en obtener la probabilidad posterior de un conjunto de variables condicionadas por un conjunto de evidencias.
- *Maximum a posteriori* (MAP): Es un problema abductivo que consiste en encontrar la configuración más probable de un conjunto de variables de la red bayesiana dada una evidencia.
- *Most probable explanation* (MPE): Es un problema abductivo que consiste en buscar la configuración más probable de todas las variables no instanciadas en la red bayesiana dada una evidencia.

Los métodos más deseados a la hora de hacer inferencia son aquellos que permiten obtener el valor exacto de la probabilidad dados los parámetros y la estructura de la red. Este tipo de inferencia se denomina inferencia exacta. Uno de los métodos de inferencia exacta más utilizados es el algoritmo *message passing* (MP) (Pearl (1986)), el cual es muy eficiente para realizar propagación de evidencias en poliárboles. El problema de este método es que sólo funciona para poliárboles, y este tipo de modelos no suelen ser capaces de representar el conocimiento en muchos campos del mundo real. El método más utilizado para realizar este tipo de inferencia es el clustering. Este método compila la red bayesiana usando una técnica de clustering para agrupar los nodos de forma que la estructura final sea un poliárbol, para posteriormente aplicarle el algoritmo MP. A esta segundo tipo de estructura creada se le suele denominar *junction tree*, y fue introducido por Lauritzen y Spiegelhalter (1988).

Si no es posible realizar inferencia exacta debido al tamaño de la red bayesiana, se realiza inferencia aproximada para reducir la complejidad computacional de las inferencias. Pese a lograr esta reducción de complejidad, este tipo de inferencia añade un grado de error a los resultados, ya que se trata de una aproximación de los valores reales. Dentro de los algoritmos de inferencia aproximada, destacan *probabilistic logic sampling* (PLS) (Henrion (1988)), *likelihood weighting* (LW) (Fung y Chang (1990)) y *Markov Chain Monte Carlo* (MCMC).

Gracias a la interpretabilidad y eficacia de las redes bayesianas, han sido utilizadas para múltiples problemas de clasificación supervisada, en ámbitos como la

medicina (Sesen et al. (2013)) y el deporte (Constantinou et al. (2012)). Algunos de los clasificadores bayesianos más utilizados y con mayor repercusión son el naïve Bayes (Minsky (1961)), seminaïve Bayes (Domingos y Pazzani (1997)) y Tree Augmented Naïve Bayes (TAN) (Friedman et al. (1997)). Por otro lado, también han sido aplicadas en problemas de clasificación no supervisada, por ejemplo a la hora de analizar los accidentes de tráfico (Oña et al. (2013)), utilizando algoritmos de clustering como el *Bayesian Hierarchical Clustering* (Heller y Ghahramani (2005)).

2.2. Regresión lineal con regularización

La regresión lineal es un modelo matemático muy utilizado en problemas de predicción de una variable continua. Se utiliza para estudiar la relación lineal entre una variable dependiente Y y una o más variables independientes X (Hastie et al. (2001); Murphy (2012)). La variable dependiente o respuesta Y debe ser continua, mientras que las variables independientes pueden ser continuas, discretas o categóricas. Sólo se debe utilizar una regresión lineal si la relación existente entre las variables es lineal, siendo otros los métodos a utilizar en caso de no linealidad. La notación X_{li} , donde $l = 1, \dots, n$ hace referencia a los datos utilizados, mientras que i representa el número de instancias de $1, \dots, n$ y j el número de variables de $1, \dots, n$. Existen dos tipos de regresión lineal:

- Regresión lineal simple: Estudia la relación lineal entre la variable dependiente Y y una única variable independiente X , permitiendo predecir el valor de Y a partir de X . La regresión lineal simple asume que:

$$Y_i = \beta_0 + \beta_1 X_i + \mathcal{E}_i,$$

donde β_0 es la intersección y β_1 representa la influencia que tienen las variables independientes sobre la variable dependiente, X_i es la variable independiente, Y_i es la variable dependiente y \mathcal{E}_i es el término de error aleatorio que se asume gaussiano, siendo i el índice de la instancia i , (X_i, Y_i) , observada en los datos. El error \mathcal{E}_i recoge todos los factores aleatorios que no son controlables, por lo que es independiente de la variable X_i . Una de las asunciones de la regresión lineal indica que el error \mathcal{E}_i sigue una distribución normal $\mathcal{E}_i \sim N(0, \sigma)$, y asume homecedasticidad.

- Regresión lineal múltiple: En muchos casos, la contribución de una única variable independiente no es suficiente para explicar la variable dependiente Y , por lo que se realiza una regresión lineal múltiple. La hipótesis es:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_n X_{ni} + \mathcal{E}_i,$$

donde las $\beta_i (i = 1, \dots, n)$ son los coeficientes de regresión que representan la influencia que tienen las variables independientes sobre la variable dependiente, X_i son las variables independientes, Y_i es la variable dependiente y \mathcal{E}_i es el

término de error aleatorio que se asume gaussiano. Al igual que en la regresión lineal simple, el error \mathcal{E}_i es independiente de las variables X_i , además de seguir una distribución normal $\mathcal{E}_i \succ N(0, \sigma)$ y asumir homocedasticidad. La regresión lineal múltiple permite estudiar múltiples variables independientes al mismo tiempo, realizando un ajuste de sus coeficientes de regresión para evitar posibles efectos de confusión entre las variables.

La estimación de los parámetros $\beta_i (i = 1, \dots, n)$ se realiza mediante el método de mínimos cuadrados, para poder obtener la recta que mejor se adapte a los datos. Para realizar la estimación se suele utilizar el método de máxima verosimilitud, que al asumir gaussianidad, coincide con mínimos cuadrados:

$$\min_{\beta_0, \beta_1, \dots, \beta_n} \left(\frac{1}{m} \sum_{i=1}^m (y_i - \beta_0 + \beta_1 x_{i1} + \dots + \beta_n x_{in})^2 \right)$$

Los parámetros se calculan a través de la siguiente fórmula:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Las estimaciones de la regresión lineal pueden dar lugar a sobreajuste, debido a que suelen tener poco sesgo y mayor varianza frecuentemente. Con el fin de poder evitar este problema, se utiliza la regularización, para reducir los parámetros y explicar mejor los datos introduciendo un término de penalización. Existen dos regularizaciones principales: la regularización *ridge* y la regularización Lasso, cuyas fórmulas se pueden apreciar a continuación (James et al. (2014)):

Regularización *ridge*:

$$\min_{\beta_0, \beta_1, \dots, \beta_n} \left(\frac{1}{m} \sum_{i=1}^m (y_i - \beta_0 + \beta_1 x_{i1} + \dots + \beta_n x_{in})^2 + \lambda \sum_{j=1}^n \beta_j^2 \right)$$

donde $\lambda \geq 0$ es el coeficiente de penalización que tiene que ser ajustado. La regresión *ridge* reduce los valores absolutos de los coeficientes de regresión, sin que lleguen al valor de cero.

Regularización Lasso:

$$\min_{\beta_0, \beta_1, \dots, \beta_n} \left(\frac{1}{m} \sum_{i=1}^m (y_i - \beta_0 + \beta_1 x_{i1} + \dots + \beta_n x_{in})^2 + \lambda \sum_{j=1}^n |\beta_j| \right)$$

donde $\lambda \geq 0$ es el coeficiente de penalización que tiene que ser ajustado. La regresión Lasso fuerza que la suma de los valores absolutos de los coeficientes de regresión sean menores que un valor fijado, lo que provoca que algunos coeficientes tengan el valor de cero y se descarten dichos valores.

Para medir la bondad de los modelos creados mediante regresión lineal, se va a utilizar el error cuadrático medio (ECM), cuya fórmula se muestra a continuación:

$$ECM = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

Al igual que las redes bayesianas, la regresión lineal se ha aplicado a una gran variedad de ámbitos de forma exitosa. Entre sus múltiples aplicaciones destacan contribuciones a la medicina (Godfrey (1985)), deporte (Nevill et al. (2002)) y visión por computador (Naseem et al. (2010)).

2.3. Random forest

Se trata de un método que proporciona una mejora con respecto al bagging con árboles de clasificación. El algoritmo random forest combina el método bagging y la introducción de aleatoriedad para lograr obtener mejores regresores. Se utilizan árboles en la combinación del predictor (Cutler et al., 2011). De la misma forma que para el método bagging, se parte de un subconjunto obtenido tras realizar la selección de N casos con reemplazamiento de forma aleatoria. En cada nodo del árbol se escogen m variables predictoras al azar del conjunto total de predictoras, para posteriormente elegir la que mejor división proponga del subconjunto m . Cada árbol se construye a partir de un vector aleatorio que puede estar constituido por variables o instancias aleatorias, o incluso una mezcla de las dos. Este método escoge la salida utilizando un sistema de votos en cada árbol, proporcionando una salida final reflejando el valor medio de las votaciones. El algoritmo random forest suele requerir de un alto grado de recursos computacionales, pero los resultados obtenidos suelen ser muy satisfactorios.

2.4. Red neuronal H2O

Las redes neuronales artificiales son modelos computacionales compuesto por conjuntos de neuronas que actúan de forma similar a las neuronas presentes en el cerebro. Las redes neuronales prealimentadas son aquellas redes en las que las conexiones existentes entre los nodos no forman un ciclo. El algoritmo H2O es una red neuronal artificial multicapa prealimentada que es entrenada mediante un gradiente estocástico descendiente utilizando propagación hacia atrás. Esta red puede contener una gran cantidad de capas ocultas. Pese a requerir un alto coste computacional, suelen obtener muy buenos resultados en términos generales.

2.5. K-nearest neighbors

El algoritmo K -nearest neighbors (KNN) es un método no paramétrico que para un nuevo x_0 busca las K observaciones más próximas representadas por N_o . Utilizando estas observaciones como referencia, realiza una estimación $\hat{f}(x_0)$ o \hat{y} realizando la media de todos los valores escogidos ((James et al., 2014)).

$$\hat{f}(x_0) = \frac{1}{K} \sum_{i \in N_o} y_i$$

Es un algoritmo muy sencillo, que pese a su simplicidad es capaz de obtener buenos resultados en un gran número de problemas.

2.6. Test de Friedman y post-hoc

El test de Friedman (Friedman (1937)) es una prueba no paramétrica que se utiliza para ver si existen diferencias entre los grupos que se quieren comparar. Existen dos posibles escenarios: el test rechaza la hipótesis nula, indicando que realmente existen diferencias, o no se produce dicho rechazo y el test muestra que no existen diferencias. Si se produce un rechazo de la hipótesis nula, se utiliza el análisis post-hoc para determinar si las diferencias existentes entre los grupos son significativas o no (García y Herrera (2008)). Se trata de uno de los análisis más utilizados para comparar si los resultados obtenidos son significativamente mejores o no.

3. MODELIZACIÓN DEL ESTILO DE JUEGO

En este capítulo se modeliza el estilo de juego del club Movistar Estudiantes utilizando redes bayesianas. A la hora de hacer el análisis, se han utilizado los datos estadísticos de los partidos de la temporada 2017-18 del club Movistar Estudiantes, proporcionados por el entrenador asistente José Ángel Samaniego. Se incluyen los datos de las tres competiciones que ha disputado el equipo, incluyendo partidos de la liga regular, copa del rey y competiciones europeas. Los datos se encuentran recogidos en un único archivo Excel, que ha servido como base para obtener toda la información necesaria para desarrollar el proyecto. El dataset contiene un total de 15.747 instancias, dentro de las cuales 3.340 hacen referencia a jugadas ofensivas y 4.229 a jugadas defensivas. El resto son instancias denominadas como esfuerzos, que reflejan determinados eventos del partido en los que el equipo realiza acciones o esfuerzos notorios. La información aparece desglosada por cada jugada que ocurre en los partidos, reflejando todas las acciones acontecidas en el intervalo de tiempo desde del inicio hasta el final o la interrupción de la posesión. Estos datos son recogidos a mano por dos miembros del equipo técnico del club, uno recoge las jugadas ofensivas del encuentro y el otro las defensivas. La plantilla está compuesta por quince jugadores, de los cuales se han cogido los que han disputado un mínimo de 500 minutos a lo largo de toda la temporada, contando con un número total final de doce jugadores.

De las 168 variables que se encuentran recogidas en el dataset, se han escogido acierto de tiro, acierto de tiros libres, intentos de dos, intentos de tiro libre, intentos triple, faltas personales, pérdidas y rebotes ofensivos (sólo en el caso de la red ofensiva) como las variables predictoras que otorgan más información, ya que han demostrado previamente su eficacia en la literatura (Zak et al. (1979); Mai (2004); Kubatko et al. (2007); Pan (2010)). Adicionalmente, se han añadido las variables tipo de juego y periodo, las cuales van a actuar como variables clase para ver las relaciones probabilísticas que se crean con respecto al resto de estadísticas. La variable tipo de juego está agrupada en cuatro grandes categorías: bloqueo directo, bloqueo indirecto, perímetro poste y juego rápido. Los bloqueos directos agrupan las jugadas en las que se produce un bloqueo sobre el jugador que tiene la posesión del balón, para lograr una ventaja sobre el defensor rival. En el caso del bloqueo indirecto, la situación es similar, con la única variación de que el bloqueo se realiza sobre un jugador que no tiene el balón. El juego rápido agrupa las jugadas de transiciones ofensivas rápidas, mientras que perímetro poste contiene las jugadas en las que existe un pase desde el perímetro a la zona. En lo que respecta a los periodos, estos se encuentran agrupados en cuatro: primer cuarto, segundo cuarto, tercer cuarto y último cuarto.

Se ha realizado una laboriosa depuración del dataset, ya que la información se encontraba completamente desglosada y no había una guía que indicara a que valor hacía referencia cada variable. Por tanto, ha sido necesario realizar un análisis exhaustivo de cada variable, para poder discernir el significado de cada estadística

recogida. Una vez realizada la identificación, se ha desarrollado un código para extraer las variables predictoras relevantes y organizarlas de una forma estructurada. Posteriormente, se ha dividido el conjunto de datos en dos grupos: grupo ofensivo y grupo defensivo. Las variables se encuentran discretizadas, ya que al ser un dataset desglosado en jugadas, todas las variables excepto tipo de juego y periodo obtienen valores de 1 o 0, dependiendo de si se ha producido el evento o no. Por ejemplo, si una jugada ha terminado en una canasta triple, las variables intento de tres y acierto de tres se mostrarán con un 1, mientras que el resto de estadísticas obtendrán el valor 0.

Se van a crear cuatro redes principales: dos con la variable tipo de juego como variable clase y dos con periodo como variable clase. Debido a que este análisis se centra en describir la influencia del tipo de juego y el periodo sobre el resto de variables, hemos optado por realizar un aprendizaje discriminativo *Greedy Search* (GS) con estas dos variables como variables clase. Los métodos de aprendizaje generativo (por ejemplo, una búsqueda voraz con BDeu como función de puntuación) tienen como objetivo optimizar el ajuste del modelo a la distribución conjunta de los datos. Por tanto, optimizar la distribución a posteriori del tipo de juego dado el resto de variables puede no ser prioritario. De hecho, al realizar ese tipo de aprendizaje la variable tipo de juego aparece más aislada en la red y al intentar predecir su valor a partir del resto de variables la precisión disminuye. De esta forma, el aprendizaje discriminativo ofrece mayor información sobre las relaciones entre el tipo de juego y el resto de variables.

Para realizar las inferencias en la red, se utiliza el método de propagación de evidencia en árboles de unión (Shenoy y Shafer (1990)). Para la creación de las cuatro redes, se ha utilizado la herramienta de software GeNie, que permite crear redes bayesianas de forma sencilla, además de proporcionar una visión de la red muy intuitiva para el usuario. El software puede ser descargado desde este enlace: <https://www.bayesfusion.com/>. Se le ha añadido un conocimiento previo a todas las redes, proporcionándoles las siguientes correspondencias (arcos) entre variables para facilitar su aprendizaje:

- Intentos de dos → Acierto tiros de dos
- Intentos de tres → Acierto tiros de tres
- Intentos de tiro libre → Acierto tiros libres
- Faltas personales → Intentos de tiro libre

Para cada una de las cuatro redes creadas, se ha realizado el cálculo de puntos esperados, lo que ha permitido medir la eficiencia del equipo en cada tipo de juego y periodo. A su vez, se han obtenido los puntos esperados por el tipo de tiro realizado en cada jugada, mostrando cual es la opción de tiro que maximiza los puntos en cada situación. Además, se ha realizado una interpretación de las relaciones probabilísticas creadas entre los nodos, confirmando que las conexiones que se han creado son lógicas

desde el punto de vista baloncestístico. Se han analizado los arcos creados y se han realizado tests de independencias utilizando el Markov Blanket (MB) para ver las independencias condicionales entre las variables.

Para facilitar la interpretación de las redes bayesianas que se muestran en los análisis ofensivos y defensivos, el Cuadro 1 muestra el significado de las variables utilizadas en las redes. En el siguiente enlace <http://cort.as/-7eJO> se pueden descargar las redes bayesianas utilizadas en el proyecto.

Cuadro 1: Equivalencias de las variables de las redes.

Variable	Significado	Estados
TipoJuego	Tipo de juego	BD, BI, JR y PP
BP	Balones perdidos	0 si no ocurre y 1 si ocurre
FR	Faltas recibidas	0 si no ocurre y 1 si ocurre
RO	Rebotes ofensivos	0 si no ocurre y 1 si ocurre
I2	Intentos de dos	0 si no tira y 1 si sí
C2	Canastas de dos	0 si no encesta y 1 si sí
I3	Intentos de tres	0 si no tira y 1 si sí
C3	Canastas de tres	0 si no encesta y 1 si sí
I1	Intentos de tiro libre	0 si no tira, 1 un tiro, 2 dos tiros y 3 tres tiros
C1	Canastas de tiro libre	0 si no encesta, 1 un tiro, 2 dos tiros y 3 tres tiros

3.1. Modelo ofensivo

En esta sección se analiza el juego ofensivo del club Movistar Estudiantes a partir de las dos redes bayesianas creadas, una para analizar las relaciones por tipo de juego y la otra por el periodo en el que se encuentra el partido. Tras crear la primera red, el nodo periodo se encontraba desconectado, siendo ése el motivo por el cual la variable periodo no se encuentra en la red. En la segunda red no se encontraron nodos desconectados. En la construcción de las dos redes bayesianas se ha utilizado un método de validación cruzada, concretamente con diez rodajas.

3.1.1. Tipo de juego

La Figura 1 muestra la red bayesiana del juego ofensivo obtenida utilizando el tipo de juego como variable clase, en la cual se pueden apreciar las relaciones probabilísticas que se han creado entre las variables. Si se analizan los arcos más relevantes que se han creado, se puede ver que existe una relación directa entre tipo de juego y el porcentaje de canastas de dos. Parece ser que dependiendo del tipo de juego que se despliegue, varía la efectividad de los tiros de dos. A su vez, existe una relación directa entre el tipo de juego y los intentos de tres, lo que parece indicar que dependiendo del juego desplegado se realizan más o menos lanzamientos

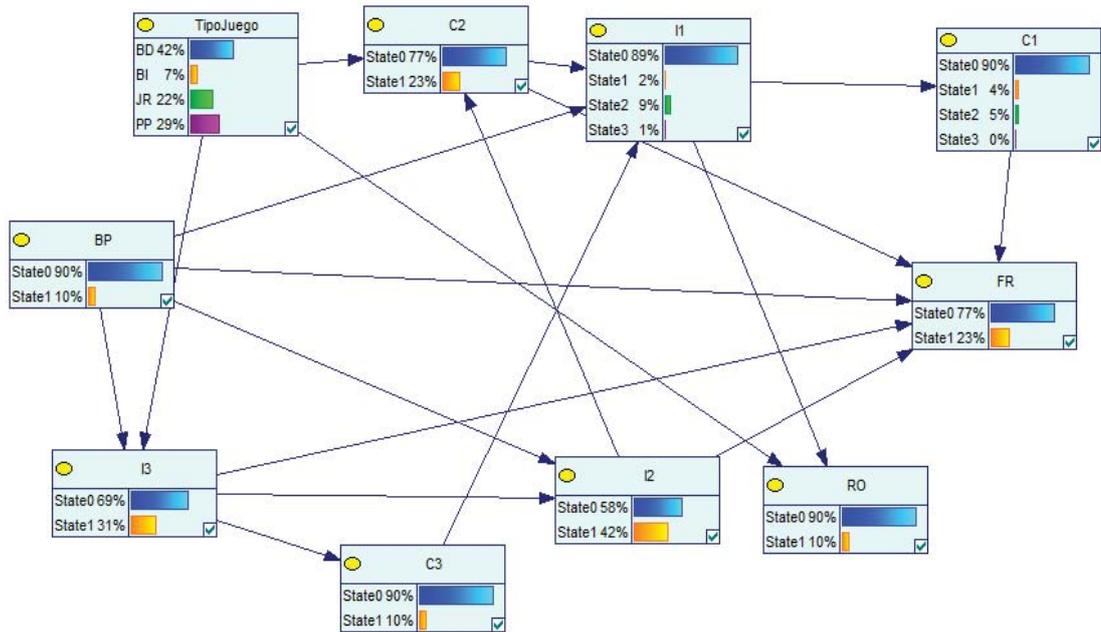


Figura 1: Red bayesiana ofensiva por tipo de juego.

de tres puntos. Hay un arco entre balones perdidos e intentos de dos, intentos de tres, intentos de tiros libre y faltas recibidas. Esto parece indicar que las pérdidas de balón condicionan en gran medida el juego del equipo, ya que si ocurren la jugada acaba automáticamente sin que se pueda lanzar a canasta o recibir una falta. Se han analizado las independencias condicionales más significativas de la red. El porcentaje de triples es independiente del tipo de juego desplegado si se conocen los intentos de tres realizados. Parece ser que si se conocen los triples intentados por el equipo, el tipo de juego no influye en el porcentaje de acierto que obtienen. Ocurre algo similar con el porcentaje de tiros libres, ya que si se conocen los tiros libres intentados, el tipo de juego no influye en el porcentaje de acierto obtenido. Los intentos de dos y de tres no influyen sobre los rebotes ofensivos obtenidos si se conoce el tipo de juego que ha realizado el equipo. Las faltas recibidas son independientes de los intentos de tiro libre si se conoce el porcentaje de tiros libres convertidos.

En el siguiente punto se va a realizar un análisis exhaustivo del modelo por cada tipo de juego.

3.1.1.1. Bloqueo directo

La Figura 2 muestra la red bayesiana generada teniendo como evidencia que el tipo de juego escogido ha sido bloqueo directo. El uso del tipo de juego bloqueo directo influye positivamente en la probabilidad de intentar un triple (31% \rightarrow 42%). Esto se puede deber a que al existir un bloqueo directo, el jugador que tiene la

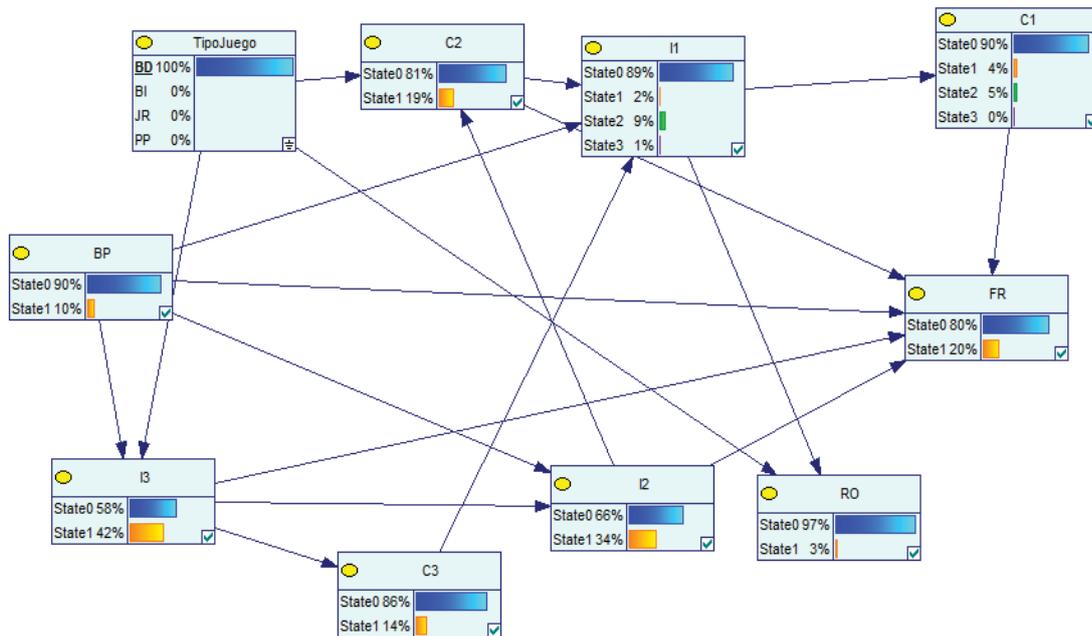


Figura 2: Red bayesiana ofensiva con bloqueo directo.

bola obtiene una ventaja sobre el rival, lo que lleva a que sea capaz de realizar un tiro de tres con relativa comodidad. Este hecho parece aumentar positivamente los puntos esperados del equipo, ya que se aumentan los intentos de tres sin disminuir la probabilidad de convertirlos. Al elevarse la probabilidad de intentos de triple y disminuirse la de intentos de tiros de dos (42% \rightarrow 36%), es posible que se esté priorizando el lanzamiento triple por encima de buscar el *Pick & Roll* con el jugador que realiza el bloqueo. El porcentaje de tiro de dos se mantiene, pese a que disminuye la probabilidad de realizar un intento de dos. Al no atenuarse el porcentaje de dos, puede ser positivo para el equipo finalizar los bloqueos directos más a menudo en tiros de dos. Hay una ligera disminución de la probabilidad de recibir una falta (23% \rightarrow 20%), pudiendo deberse al aumento de tiros de tres por encima de acabar la jugada cerca del aro rival, donde es más probable recibir una falta. La disminución de la probabilidad de conseguir un rebote ofensivo puede estar ligada a que los pívots suelen ser los encargados de realizar los bloqueos directos, con lo que se pierde presencia en la zona y con ello probabilidades de conseguir un rebote ofensivo (10% \rightarrow 3%).

A continuación se muestran los puntos esperados en jugadas de bloqueo directo y la rentabilidad de cada tipo de tiro. Para ello, se introduce bloqueo directo como evidencia en la red. Los puntos esperados de una jugada se pueden calcular multiplicando el valor de cada tiro por la probabilidad de convertirlo, sumando los valores resultantes:

- Puntos esperados por bloqueo directo: $2 \cdot (0,19 + 0,05) + (3 \cdot 0,14) + 0,04 = 0,94$

Adicionalmente, se calculan los puntos esperados en función del tipo de tiro. Con este objetivo, se introducen bloqueo directo y el tipo de tiro como evidencias a la red. Al igual que en el caso anterior, se multiplica el valor de cada tiro por la probabilidad de convertirlo:

- Puntos por cada tiro de dos realizado: $2 \cdot 0,55 = 1,10$
- Puntos por cada tiro de tres realizado: $3 \cdot 0,33 = 0,99$
- Puntos esperados por cada tiro libre realizado: $1 \cdot (0,59 + 0,17) = 0,76$

En las siguientes secciones se analizan los puntos esperados y puntos por tipo de tiro de forma análoga al análisis realizado en este punto.

El tipo de tiro con mayor rentabilidad es el tiro de dos puntos, con un valor de 1,10 puntos generados por cada tiro intentado, claramente superior al valor obtenido con un intento de tres puntos (0,99 puntos esperados por jugada). Cuando se realiza un bloqueo directo, suelen aparecer dos ventajas claras. La primera es la posibilidad de realizar un tiro de tres (o incluso de dos) mientras se produce el cambio defensivo del equipo rival, y la segunda la ventaja que se genera cuando el pívot del equipo se queda con un defensor de menor estatura. Por tanto, si se produce un cambio de tendencia hacia priorizar el tiro de dos que se genera tras el bloqueo como recurso para finalizar la jugada de bloqueo directo, se podría producir un aumento en la cantidad de puntos anotada por el equipo, ya que se trata del tipo de finalización con mayor rentabilidad.

3.1.1.2. Bloqueo indirecto

La Figura 3 muestra la red bayesiana generada teniendo como evidencia que el tipo de juego escogido ha sido bloqueo indirecto. El uso del tipo de juego bloqueo indirecto supone un aumento muy significativo de la probabilidad de intentar un triple (31 % \rightarrow 45 %). Esto se puede entender debido a que se suelen realizar bloqueos indirectos para liberar al tirador del equipo y que pueda realizar un tiro triple cómodo. Consecuentemente, se disminuye la probabilidad de realizar un tiro de dos (42 % \rightarrow 32 %), ya que al crearse situaciones con tiros de tres tan abiertos, es infrecuente que el jugador se decida a realizar un tiro de dos puntos, jugada que en caso de acabar exitosamente proporcionaría menos puntos que un triple. Aun así, cabe destacar que a diferencia del bloqueo directo, en el bloqueo indirecto la disminución de la probabilidad de realizar un intento de dos viene acompañada de una atenuación en el porcentaje de tiros de dos (54 % \rightarrow 47 %). Esto indica que los tiros de dos que se realizan no tienen tanta efectividad como en otro tipo de jugada. Sería interesante comprobar si esto se debe a que las jugadas con bloqueo indirecto suelen terminar en tiros de dos lejanos o si realmente esos fallos se producen en tiros más cercanos al aro rival. Por otro lado, también se reduce la probabilidad de recibir

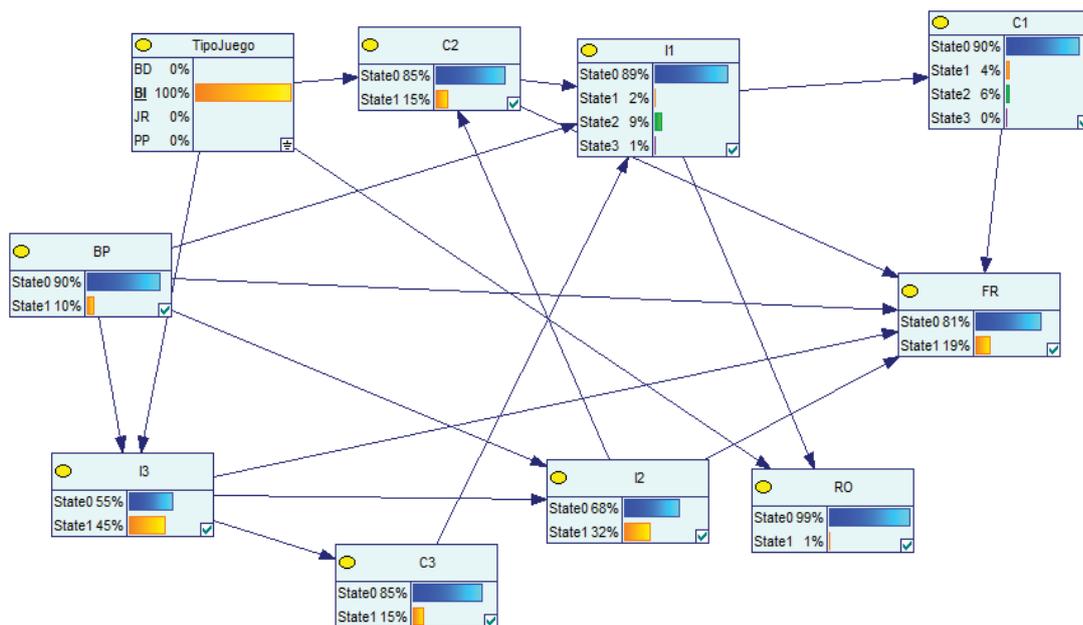


Figura 3: Red bayesiana ofensiva con bloqueo indirecto.

una falta (23% \rightarrow 19%), lo cual puede deberse a que se realizan muy pocas faltas en los lanzamientos de tres. Como ocurría con el bloqueo directo, la probabilidad de conseguir rebotes ofensivos disminuye (10% \rightarrow 1%), lo que puede deberse a que el pívot suele poner los bloqueos, por lo que abandona la zona para realizarlos. En consecuencia, el equipo pierde presencia en la zona y es más difícil hacerse con un rebote.

A continuación se muestran los puntos esperados en jugadas de bloqueo indirecto y la rentabilidad de cada tipo de tiro:

- Puntos esperados por bloqueo indirecto: $2 \cdot (0,15 + 0,06) + (3 \cdot 0,15) + 0,04 = 0,91$
- Puntos por cada tiro de dos realizado: $2 \cdot 0,47 = 0,94$
- Puntos por cada tiro de tres realizado: $3 \cdot 0,33 = 0,99$
- Puntos esperados por cada tiro libre realizado: $1 \cdot (0,59 + 0,17) = 0,76$

El tipo de tiro con mayor rentabilidad es el tiro de tres puntos, con un valor de 0,99 puntos generados por cada tiro intentado. El tiro de dos se encuentra claramente por debajo de este valor, con un valor de 0,94 puntos esperados por jugada. Por tanto, parece favorable continuar con la tendencia de priorizar los tiros de tres por encima del resto de tiros en bloqueos indirectos, ya que proporcionan la mayor rentabilidad de todas. Aun así, es recomendable no abusar de este estilo de jugada, ya que el equipo rival puede interiorizar los movimientos que se realizan para intentar adelantarse y robar el balón.

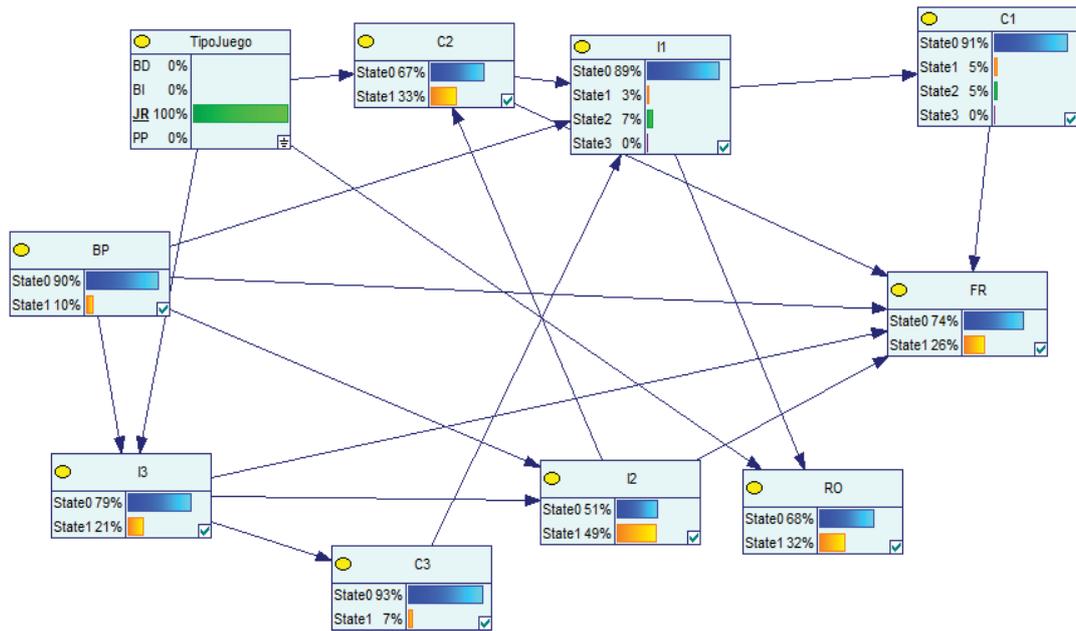


Figura 4: Red bayesiana ofensiva con juego rápido.

3.1.1.3. Juego rápido

La Figura 4 muestra la red bayesiana generada teniendo como evidencia que el tipo de juego escogido ha sido juego rápido. En juego rápido, se puede apreciar que se reduce la probabilidad de realizar un tiro de tres (31% \rightarrow 21%), y aumenta significativamente la probabilidad de realizar un tiro de dos (42% \rightarrow 49%). Normalmente, este tipo de jugadas suelen ser contraataques, en los que se busca llevar el juego al campo rival lo antes posible y finalizar sin consumir demasiado tiempo del reloj de posesión. Ligado a este aumento, se aprecia una subida sustancial en la probabilidad de convertir el tiro de dos (54% \rightarrow 67%), por lo que se trata de un tipo de jugada que suele acabar en canasta. Es interesante apreciar que la probabilidad de conseguir un rebote ofensivo aumenta drásticamente (10% \rightarrow 32%), lo cual puede ser consecuencia de que la defensa suele estar más desorganizada cuando se produce un ataque rápido, siendo muy probable que a los defensores no les de tiempo a llegar y establecerse defensivamente. Esto último también puede influir en el aumento observado en la probabilidad de recibir una falta (23% \rightarrow 26%), lo cual podría indicar que se suelen cometer dos tipos de faltas: o bien faltas antes de que se haya montado el contraataque para parar la jugada, o bien faltas cometidas debido al posicionamiento indebido a la hora de defender un tiro del atacante.

A continuación se muestran los puntos esperados en jugadas de juego rápido y la rentabilidad de cada tipo de tiro:

- Puntos esperados por juego rápido: $2 \cdot (0,33 + 0,05) + (3 \cdot 0,21) + 0,05 = 1,02$

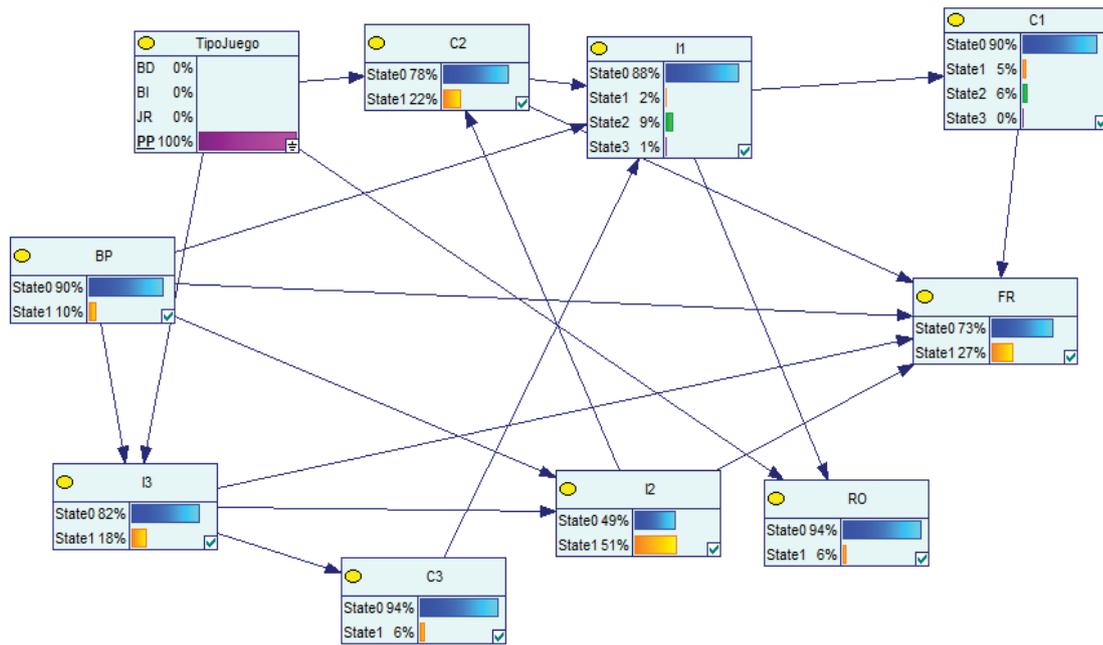


Figura 5: Red bayesiana ofensiva con perímetro poste.

- Puntos por cada tiro de dos realizado: $2 \cdot 0,67 = 1,34$
- Puntos por cada tiro de tres realizado: $3 \cdot 0,33 = 0,99$
- Puntos esperados por cada tiro libre realizado: $1 \cdot (0,59 + 0,17) = 0,76$

El tipo de tiro con mayor rentabilidad es el tiro de dos puntos, con un valor de 1,34 puntos generados por cada tiro intentado. Existe una clara diferencia con respecto a los puntos esperados con intentos de tres (0,99 por jugada), por lo que en este caso no cabe duda de que la mejor decisión es acabar la jugada con un intento de dos. Aun así, sería interesante variar de vez en cuando la finalización, para conseguir pillar por sorpresa al rival y evitar que éste se anticipe y consiga robar el balón.

3.1.1.4. Perímetro poste

La Figura 5 muestra la red bayesiana generada teniendo como evidencia que el tipo de juego escogido ha sido perímetro poste. Dentro del uso del tipo de juego perímetro poste, se aprecia una disminución de la probabilidad de realizar un intento de tres (31% \rightarrow 18%). Por contra, hay un aumento muy considerable de la probabilidad de realizar un intento de dos (42% \rightarrow 51%). Esto se puede deber a que en este tipo de jugadas se juega con el pívot en el poste, por lo que lo más normal suele ser que la jugada finalice con un intento de dos del pívot. Sin embargo,

este aumento considerable de la probabilidad de realizar un tiro de dos no va acompañado de un aumento en la probabilidad de conseguir encestar el tiro, sino todo lo contrario (54% \rightarrow 43%). Esto podría indicar que estos tiros están generalmente mejor defendidos. Por este motivo, no parece recomendable priorizar los tiros de dos en perímetro poste, ya que el aumento del número de tiros realizados no viene acompañado con un aumento del porcentaje de acierto. Sería interesante estudiar alternativas en este tipo de juego, para conseguir sacar el mayor rendimiento de la posesión. Hay un aumento en la probabilidad de recibir una falta (23% \rightarrow 27%), posiblemente porque en la zona suele haber más contacto y se cometen más faltas. Por contra, la probabilidad de conseguir un rebote ofensivo disminuye (10% \rightarrow 6%), lo que se podría deber a que la defensa se encuentra más organizada en este tipo de jugadas y es más difícil conseguir atrapar el balón.

A continuación se muestran los puntos esperados en jugadas de perímetro poste y la rentabilidad de cada tipo de tiro:

- Puntos esperados por perímetro poste: $2 \cdot (0,22 + 0,06) + (3 \cdot 0,06) + 0,05 = 0,79$
- Puntos por cada tiro de dos realizado: $2 \cdot 0,43 = 0,86$
- Puntos por cada tiro de tres realizado: $3 \cdot 0,33 = 0,99$
- Puntos esperados por cada tiro libre realizado: $1 \cdot (0,59 + 0,17) = 0,76$

El tipo de tiro con mayor rentabilidad es el tiro de tres puntos, con un valor de 0,99 puntos generados por cada tiro intentado. Los puntos esperados por cada tiro de dos realizado se encuentran muy por debajo, con un valor de 0,86. El equipo está priorizando finalizar este tipo de jugada con un tiro de dos, lo que le está llevando a obtener un valor menor de puntos por jugada. Un cambio de tendencia hacia el tiro de tres podría ser beneficioso para la producción ofensiva del equipo.

El Cuadro 2 engloba las variaciones de las probabilidades que se han producido dado el tipo de juego del equipo en el modelo ofensivo.

3.1.1.5. Comparación de los tipos de juego

Tras analizar detenidamente los resultados obtenidos de la red bayesiana de tipo de juego, se han obtenido las siguientes conclusiones:

- El tipo de juego que mayor cantidad de puntos esperados genera es el juego rápido, obteniendo 1,02 puntos por posesión. Al tratarse de un juego basado en transiciones rápidas, el hecho de conseguir pillar a la defensa rival desorganizada parece ser clave para obtener un mayor número de canastas. Un crecimiento en la cantidad de jugadas finalizadas rápidamente debería traer consigo un aumento considerable de los puntos obtenidos por el equipo.

Cuadro 2: Variaciones de las probabilidades de las variables por tipo de juego en el modelo ofensivo.

Estadística	General	BD	BI	JR	PP
I2	42 %	34 %	32 %	49 %	51 %
C2	23 %	19 %	15 %	33 %	22 %
I3	31 %	42 %	45 %	21 %	18 %
C3	10 %	14 %	15 %	7 %	6 %
BP	10 %	10 %	10 %	10 %	10 %
FR	23 %	20 %	19 %	26 %	27 %
RO	10 %	3 %	1 %	32 %	6 %
I1	2 %	2 %	2 %	3 %	2 %
	9 %	9 %	9 %	7 %	9 %
	1 %	1 %	1 %	0 %	1 %
C1	4 %	4 %	4 %	5 %	5 %
	5 %	5 %	6 %	5 %	6 %
	0 %	0 %	0 %	0 %	0 %

- Parece que el tipo de juego perímetro poste es el que menos beneficio proporciona al equipo, consiguiendo únicamente 0,79 puntos. Al tratarse de una jugada realizada contra defensas organizadas, resulta más difícil conseguir encestar, ya que tampoco se producen bloqueos que puedan permitir obtener una ventaja sobre el defensor y disponer de un tiro más claro. En consecuencia, sería recomendable disminuir el número de jugadas de perímetro poste, o en su defecto cambiar la selección de tiros que se realiza para intentar aumentar la esperanza de puntos.
- En lo que respecta al bloqueo directo, aunque no se estén consiguiendo registros tan buenos como con el juego rápido, se está consiguiendo obtener un valor de puntos esperados bastante alto con 0,94 puntos por posesión. Si se prioriza el tiro de dos por encima de los intentos de triple, es posible que se consiga aumentar la producción ofensiva y se saque más partido a cada jugada.
- Por último, el bloqueo indirecto está consiguiendo obtener 0,91 puntos por jugada, cerca de la rentabilidad obtenida por los bloqueos directos. Aunque no es un valor extremadamente alto, está contribuyendo a tener una buena producción ofensiva, sobre todo desde la línea de tres puntos. Si no se abusa mucho del bloqueo indirecto, la producción ofensiva puede verse claramente beneficiada, ya que se trata de una jugada que puede ser fácilmente predecible.

3.1.2. Periodo

La Figura 6 muestra la red bayesiana del juego ofensivo obtenida utilizando el periodo como variable clase. Si se analizan los arcos que se han creado en la red, se

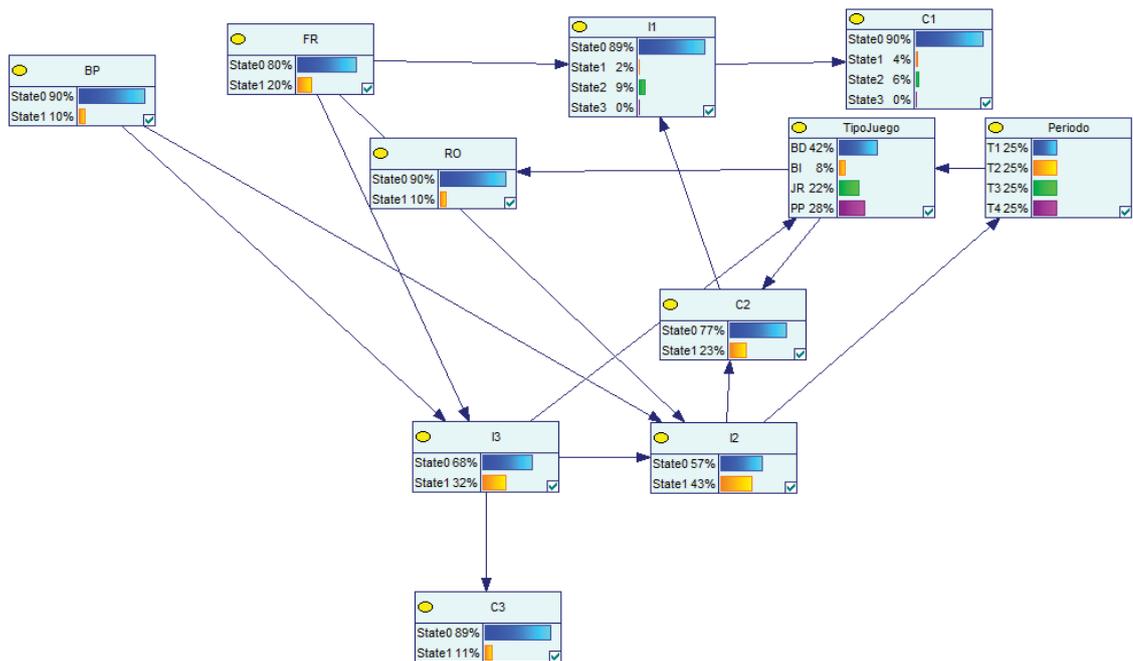


Figura 6: Red bayesiana ofensiva por periodo.

puede apreciar que el periodo no influye en los porcentajes de tiros de dos y de tiros libres, pero sí en el tipo de juego que realiza el equipo. Además, hay una relación directa entre periodo y los intentos de dos, lo que parece indicar que dependiendo del periodo, el equipo aumenta o disminuye los tiros de dos puntos realizados. Por otro lado, el tipo de juego influye directamente en la probabilidad de convertir un tiro de dos. Si se analizan las independencias condicionales entre las variables, se puede apreciar que los porcentajes de acierto de dos, de tres y de tiro libre son independientes del periodo en el que se encuentra el partido si se conoce el tipo de juego desplegado por el equipo, los intentos de dos y los intentos de tres realizados. El porcentaje de triple obtenido por el equipo es independiente del resto de variables de la red si se conocen los intentos de tres realizados, lo que parece indicar que el resto de variables no influyen en el acierto triple del equipo. Análogamente, pasa lo mismo con el porcentaje de acierto de tiros libres, que se muestra independiente del resto de variables de la red conociendo los tiros libres intentados. Los rebotes ofensivos son independientes del resto de variables de la red si se conoce el tipo de juego desplegado por el equipo. Esto parece indicar que únicamente el tipo de juego que se realiza en cada jugada parece afectar a los rebotes ofensivos que consigue el equipo.

A continuación se muestran los resultados obtenidos tras realizar un análisis por periodo, observando las relaciones probabilísticas más interesantes en cada uno de los cuatro cuartos del partido.

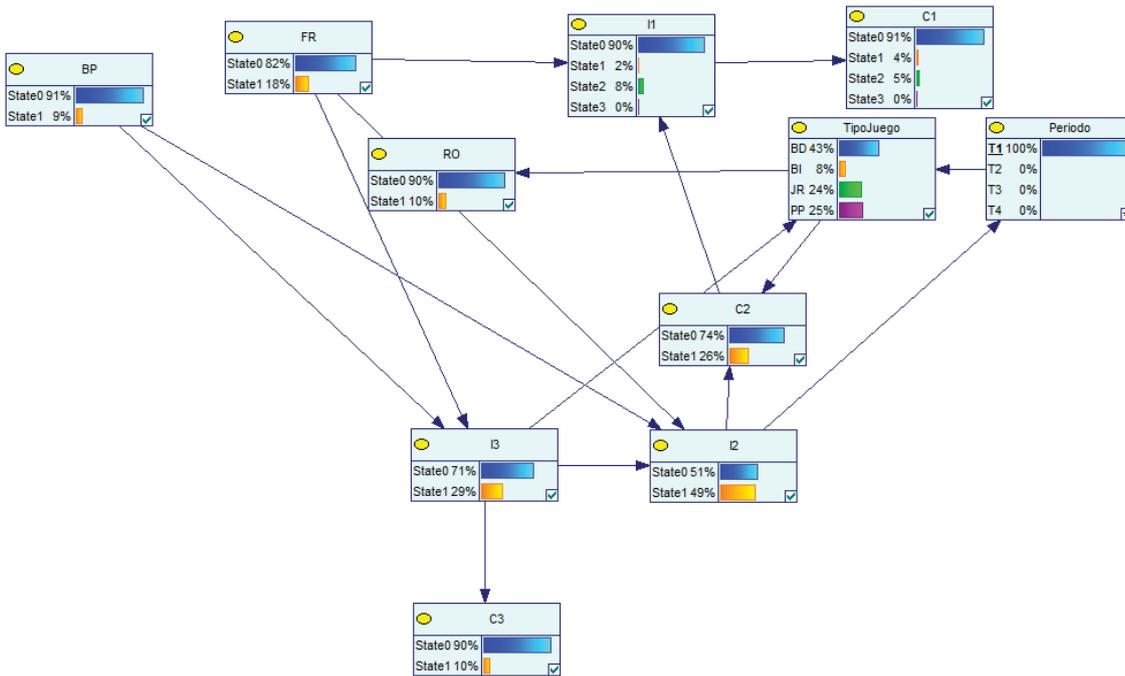


Figura 7: Red bayesiana ofensiva en el primer periodo.

3.1.2.1. Periodo 1

La Figura 7 muestra la red bayesiana generada teniendo como evidencia que el periodo en el que se encuentra el partido es el primero. El tipo de juego más utilizada en este primer cuarto es el bloqueo directo (43%), lo que parece indicar que al principio de los partidos el equipo suele comenzar elaborando más las jugadas, buscando las ventajas que le proporciona realizar un bloqueo. A su vez, la probabilidad de realizar una jugada de juego rápido o perímetro poste también es alta (24% y 25% respectivamente), lo que parece mostrar que se están aprovechando las oportunidades para correr con el balón, además de intentar jugar con los pivots en la zona rival. No resulta extraño que la probabilidad de realizar una jugada de bloqueo indirecto sea mucho menor (8%), ya que es un tipo de juego de la que no conviene abusar para que el equipo rival no se acostumbre y sea capaz de anticiparse a los movimientos. Disminuye la probabilidad de perder el balón, señal de que se está gestionando bien la posesión del balón cometiendo pocos errores y minimizando los riesgos. Hay un aumento en la probabilidad de realizar un tiro de dos (43% \rightarrow 49%). La probabilidad de realizar un intento de tres disminuye (32% \rightarrow 29%), pero el porcentaje de acierto se mantiene, lo que parece mostrar que pese a realizar menos tiros de tres, los tiros están siendo mejor escogidos, trayendo consigo un aumento en la efectividad. La probabilidad de recibir una falta disminuye (20% \rightarrow 18%). Para intentar mejorar el rendimiento ofensivo en este primer periodo, podría ser beneficioso tratar de tirar más triples, ya que el modelo muestra que hay una mejora en los porcentajes pese

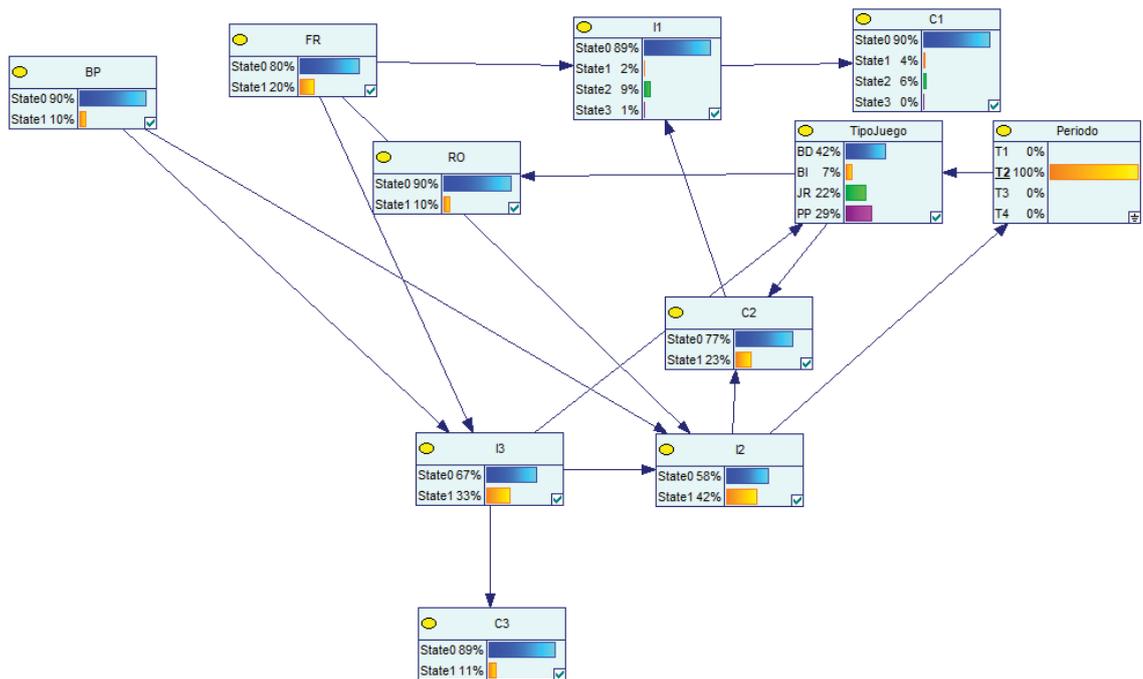


Figura 8: Red bayesiana ofensiva en el segundo periodo.

a la disminución de intentos. También podría ser interesante subir la intensidad de los ataques, tratando de recibir un mayor número de faltas e ir a la línea de tiros libres.

A continuación se muestran los puntos esperados para el primer periodo:

- Puntos esperados en el periodo 1: $2 \cdot (0,26 + 0,05) + (3 \cdot 0,11) + 0,04 = 0,99$

La esperanza de puntos en el primer cuarto es de 0,99, un valor bastante alto. Este hecho parece indicar que el equipo empieza los partidos en buena forma, consiguiendo anotar con relativa frecuencia. El hecho de que los jugadores estén frescos al inicio del partido puede estar influyendo en que se comiencen de forma óptima los partidos.

3.1.2.2. Periodo 2

La Figura 8 muestra la red bayesiana generada teniendo como evidencia que el periodo en el que se encuentra el partido es el segundo. Al igual que en el primer cuarto, el tipo de juego que predomina es bloqueo directo (42%), lo que parece ir mostrando una tendencia a utilizar esta jugada como la base principal del juego del equipo. Se produce un aumento en la probabilidad de realizar una jugada de perímetro poste (25% \rightarrow 29%), en detrimento del juego rápido (24% \rightarrow 22%). La probabilidad de realizar una jugada de bloqueo indirecto disminuye ligeramente con respecto al primer cuarto (7%). Se disminuye la probabilidad de realizar un intento

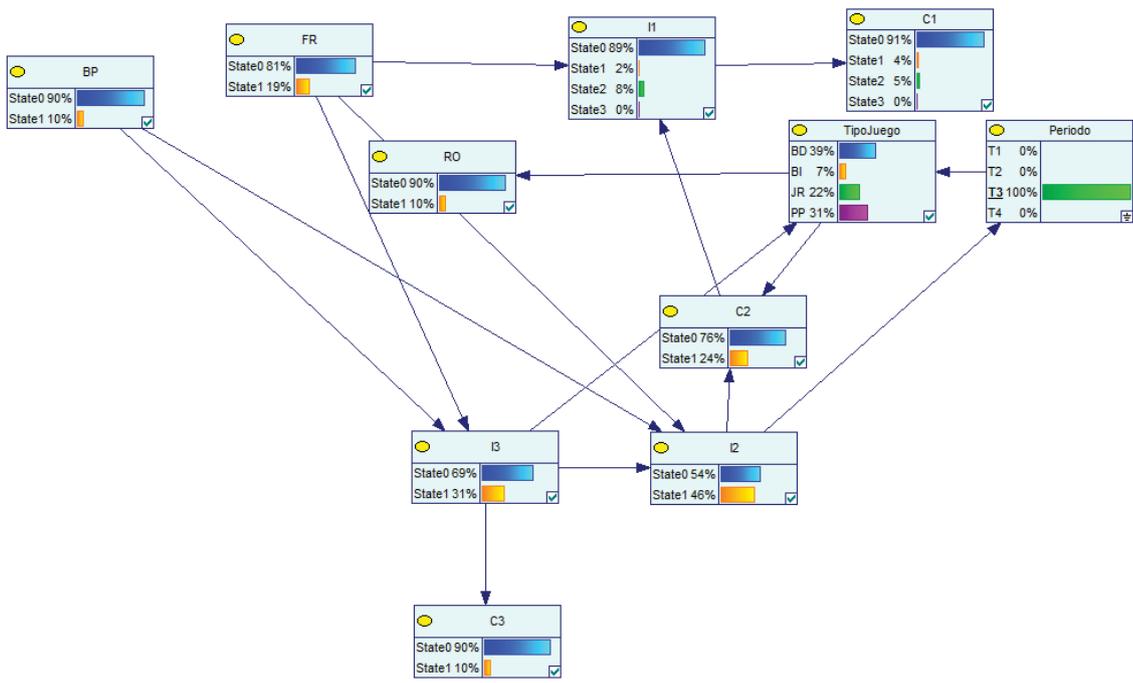


Figura 9: Red bayesiana ofensiva en el tercer periodo.

de dos (43% → 42%), y se mantiene la probabilidad de realizar un intento de tres, al igual que el porcentaje de acierto.

A continuación se muestran los puntos esperados para el segundo periodo:

- Puntos esperados en el periodo 2: $2 \cdot (0,22 + 0,05) + (3 \cdot 0,10) + 0,04 = 0,86$

La esperanza de puntos en el segundo cuarto es de 0,86, reflejando una clara bajada de rendimiento en comparación con el primer periodo (0,99 → 0,86). Como se pudo apreciar en el análisis por tipo de juego realizado anteriormente, perímetro poste es el tipo de juego que menos esperanza de puntos tenía (0,79 puntos esperados por jugada), por lo que el aumento de este juego puede estar penalizando el ataque del equipo. Si se prioriza el juego rápido en el cuarto, se podría lograr una mejora en la producción ofensiva, corrigiendo la tendencia negativa que se aprecia en este segundo cuarto.

3.1.2.3. Periodo 3

La Figura 9 muestra la red bayesiana generada teniendo como evidencia que el periodo en el que se encuentra el partido es el tercero. Bloqueo directo sigue siendo el tipo de juego más utilizado, pese a sufrir una disminución de probabilidad con respecto a los dos anteriores cuartos (42% → 39%). Continuando con el incremento experimentado en el segundo cuarto, la probabilidad de realizar una

jugada de perímetro poste sube levemente (29 % \rightarrow 31 %). Juego rápido aumenta (22 % \rightarrow 23 %) y bloqueo indirecto se mantiene (7 %). Se mantiene la probabilidad de perder el balón, y la probabilidad de recibir una falta disminuye (20 % \rightarrow 19 %). Aumenta la probabilidad de realizar un intento de dos (43 % \rightarrow 46 %), y disminuye la probabilidad de realizar un intento de tres (32 % \rightarrow 31 %). Este hecho parece estar directamente relacionado con el aumento de las jugadas perímetro poste, ya que en ellas predominaba el tiro de dos que mostraban grados muy bajos de efectividad. Pese a la disminución de la probabilidad de realizar un intento de tres, el porcentaje de acierto se mantiene.

A continuación se muestran los puntos esperados para el tercer periodo:

- Puntos esperados en el periodo 3: $2 \cdot (0,4 + 0,05) + (3 \cdot 0,10) + 0,04 = 0,92$

La esperanza de puntos en el tercer cuarto es de 0,92, consiguiendo remontar con respecto al anterior cuarto (0,85 \rightarrow 0,92), aunque sigue sin llegar a los registros obtenidos en el primer periodo. Se sigue apreciando una tendencia alcista en el uso de la jugada perímetro poste, siendo la jugada que menor rendimiento ofensivo proporciona al equipo. Si se consigue reconducir este aumento al tipo de juego juego rápido, sería posible conseguir un incremento en los puntos conseguidos por el equipo, ya que se trata de la jugada con mayor esperanza de puntos (1,02).

3.1.2.4. Periodo 4

La Figura 10 muestra la red bayesiana generada teniendo como evidencia que el periodo en el que se encuentra el partido es el cuarto. Los tipos de juego más utilizados vuelven a ser los que han ido predominando en todos los cuartos: bloqueo directo (43 %) y perímetro poste (29 %). La probabilidad de realizar una jugada de bloqueo indirecto vuelve a aumentar levemente hasta los registros mostrados en el primer cuarto (8 %), mientras que juego rápido sufre una nueva disminución (23 % \rightarrow 21 %). Aumenta la probabilidad de perder el balón (10 % \rightarrow 11 %) y la probabilidad de recibir una falta (20 % \rightarrow 21 %), lo que podría ser una consecuencia directa del aumento de la tensión que supone estar en el periodo decisivo del encuentro. Disminuye la probabilidad de realizar un intento de dos (46 % \rightarrow 37 %), y la probabilidad de ejecutar un intento de tres aumenta (31 % \rightarrow 36 %) . Pese a aumentarse la probabilidad, el porcentaje de acierto triple se mantiene. Esto parece indicar que se está manteniendo la efectividad desde la línea de tres, pese a aumentar la frecuencia de lanzamiento. Por tanto, un aumento en la probabilidad de realizar un tiro de tres podría traer consigo una mejora sustancial de los puntos conseguidos, debido al alto grado de efectividad que se obtiene en este periodo.

A continuación se muestran los puntos esperados para el cuarto periodo:

- Puntos esperados en el periodo 4: $2 \cdot (0,2 + 0,06) + (3 \cdot 0,12) + 0,04 = 0,92$

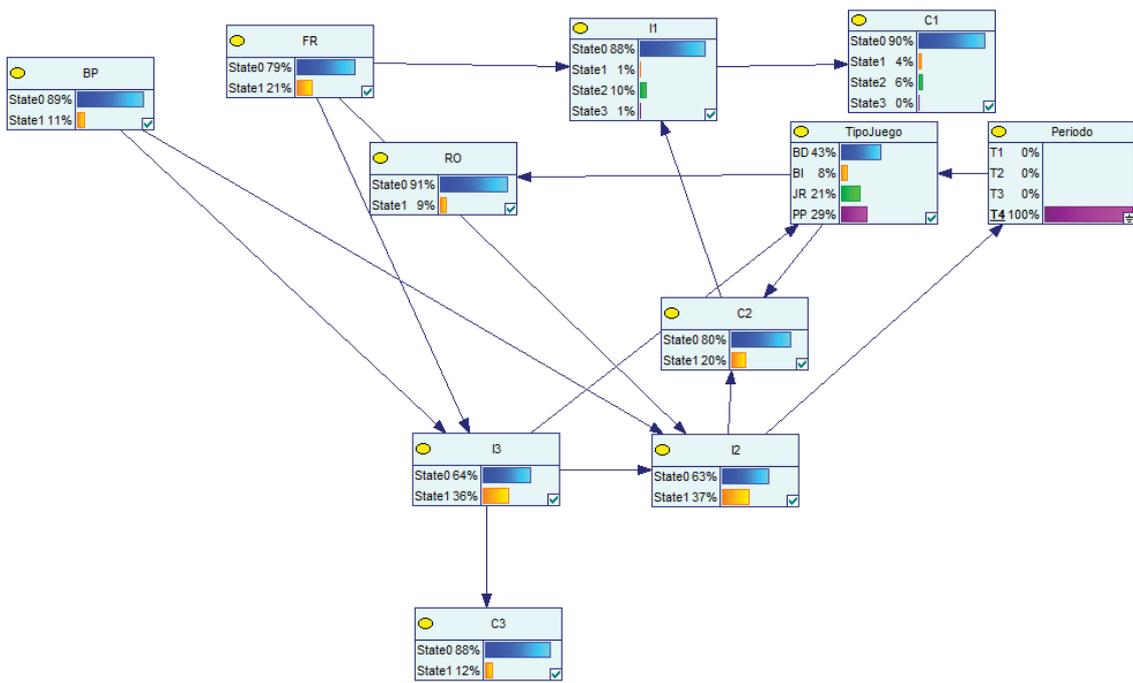


Figura 10: Red bayesiana ofensiva en el cuarto periodo.

La esperanza de puntos en el último cuarto es de 0,92, mismo registro que en el tercero. Al igual que en el anterior, perímetro poste sigue teniendo una influencia demasiado grande, eclipsando a otras jugadas con mejor esperanza de puntos como el juego rápido y el bloqueo indirecto. Aun así, el valor obtenido sigue siendo superior al segundo cuarto, pero se encuentra lejos del primer periodo, el cual ha demostrado ser el mejor de todos en cuanto a puntos esperados por jugada.

El Cuadro 3 engloba las variaciones de las probabilidades que se han producido dado el periodo en el modelo ofensivo.

3.1.2.5. Comparación de periodos

Tras analizar la información proporcionada por la red bayesiana en cada uno de los periodos, se puede apreciar que puede haber una relación directa entre el quinteto que se encuentra en juego en cada uno de los cuartos con el porcentaje de triples que se consigue. Los dos cuartos en los que ese porcentaje es superior son el primero y el último, cuartos en los que suelen estar en la cancha el quinteto titular del equipo. Cabe destacar que el periodo en el que hay menos puntos esperados es el segundo, lo cual podría estar ligado a los jugadores que se encuentran jugando en ese momento. El periodo con mayor índice de puntos esperados es el primero, lo cual parece indicar que el equipo comienza bien los partidos. Sería interesante comprobarlo analizando los puntos que se consiguen en los primeros cuartos y el quinteto que se encuentra jugando. Se puede apreciar que hay una disminución del juego rápido en el segundo cuarto, periodo que genera el menor índice de puntos

Cuadro 3: Variaciones de las probabilidades de las variables por periodo en el modelo ofensivo.

Estadística	General	Periodo 1	Periodo 2	Periodo 3	Periodo 4
TipoJuego	42 %	43 %	42 %	39 %	43 %
	8 %	8 %	7 %	7 %	8 %
	22 %	24 %	22 %	22 %	21 %
	28 %	25 %	29 %	31 %	29 %
I2	43 %	49 %	42 %	46 %	37 %
C2	23 %	26 %	23 %	24 %	20 %
I3	32 %	29 %	33 %	31 %	36 %
C3	11 %	10 %	11 %	10 %	12 %
BP	10 %	9 %	10 %	10 %	11 %
FR	20 %	18 %	20 %	19 %	21 %
RO	10 %	10 %	10 %	10 %	9 %
I1	2 %	2 %	2 %	2 %	1 %
	9 %	8 %	9 %	8 %	9 %
	0 %	0 %	1 %	0 %	1 %
C1	4 %	4 %	4 %	4 %	4 %
	6 %	5 %	6 %	5 %	6 %
	0 %	0 %	0 %	0 %	0 %

esperados. Como ya se ha comentado, el tipo de juego que menos puntos esperados generaba en el anterior modelo era el perímetro poste, mientras que el que más generaba era el juego rápido.

3.2. Modelo defensivo

Al igual que en la sección anterior para el modelo ofensivo, en esta sección se analiza el juego defensivo del club Movistar Estudiantes a partir de las dos redes bayesianas creadas, una para analizar las relaciones por tipo de juego y la otra por el periodo en el que se encuentra el partido. Como también ocurrió en el caso de la red ofensiva, al crear la red con tipo de juego como variable clase el nodo periodo se encontraba desconectado, por lo que en este caso también se eliminó de la red. En la segunda red no se encontraron nodos desconectados. En la construcción de las dos redes bayesianas se ha utilizado el mismo método de validación cruzada.

3.2.1. Tipo de juego

La Figura 11 muestra la red bayesiana del juego defensivo obtenida utilizando el tipo de juego como variable clase, en la cual se pueden distinguir las relaciones probabilísticas que se han creado entre las variables. Al analizar los arcos de la red

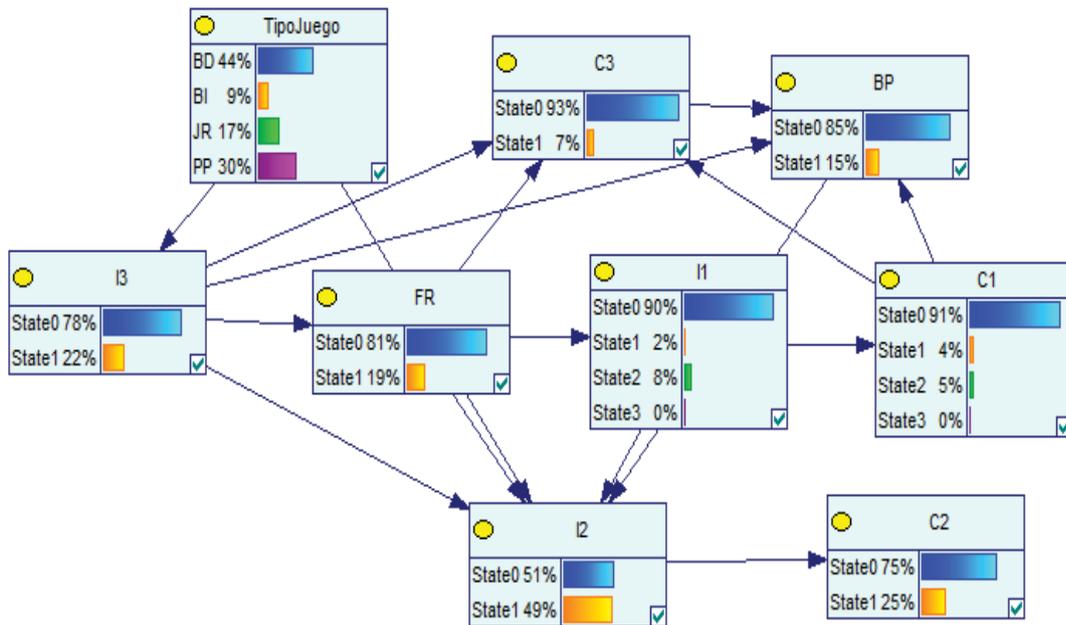


Figura 11: Red bayesiana defensiva por tipo de juego.

bayesiana, se puede apreciar que el tipo de juego tiene una relación directa con el número de tiros de dos y de tres intentados por el equipo rival, pese a no tener una influencia directa sobre sus porcentajes de acierto. Como el porcentaje de tiros es independiente del tipo de juego, se obtienen los mismos puntos esperados por tiro en cada jugada de los cuatro periodos. Si se analizan las independencias condicionales, se puede apreciar que el porcentaje de tiros de dos es independiente del resto de variables del modelo si se conocen los intentos de dos realizados. Como se ha podido intuir al examinar los arcos, el tipo de juego es independiente de los porcentajes de tiros de dos y de tres obtenidos por el equipo si se conocen los intentos de dos y de tres realizados respectivamente.

Al igual que para el modelo ofensivo, a continuación se muestra un desglose por tipo de juego de los puntos más interesantes de la red.

3.2.1.1. Bloqueo directo

La Figura 12 muestra la red bayesiana generada teniendo como evidencia que el tipo de juego escogida ha sido bloqueo directo. Si es éste el tipo de juego utilizada por el equipo contrario, aumenta sus probabilidades de realizar un intento de tres (22% \rightarrow 26%). Además, se produce un aumento en el porcentaje de acierto triple, suponiendo una fuente fiable de puntos para el rival. A su vez, disminuye la probabilidad de que el tiro que realice el contrario sea de dos puntos (49% \rightarrow 47%), al igual que su porcentaje de acierto. Esto demuestra que si la jugada de bloqueo

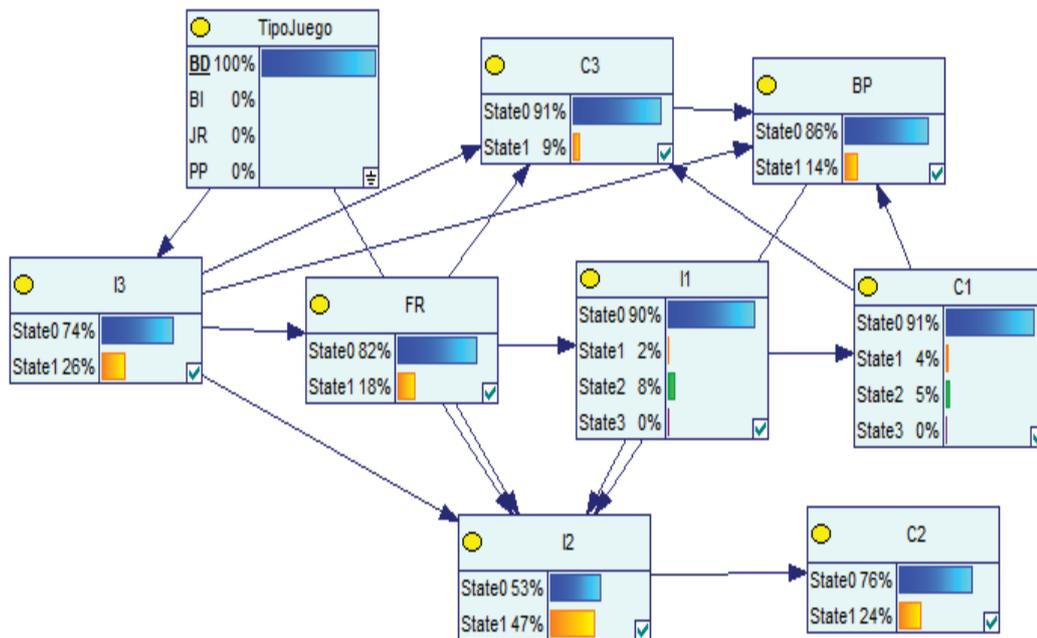


Figura 12: Red bayesiana defensiva con bloqueo directo.

directo del rival termina en un intento de dos, es más probable que no consigan la canasta. Teniendo esto en cuenta, podría ser interesante que la defensa priorizara el salir a la ayuda tras el bloqueo directo, realizando un cambio defensivo e impidiendo que el jugador rival realice un tiro de tres puntos en condiciones favorables. Por otro lado, disminuye ligeramente la probabilidad de cometer una falta en defensa (19% \rightarrow 18%), que a su vez se ve reflejado en la disminución de la probabilidad de intentar tiros libres. Esta atenuación podría deberse al hecho de que la defensa se muestra menos intensa en este tipo de juego, permitiendo a los jugadores rivales realizar tiros más cómodos.

A continuación se muestran los puntos esperados en jugadas de bloqueo directo y la rentabilidad de cada tipo de tiro:

- Puntos esperados por bloqueo directo: $2 \cdot (0,24 + 0,05) + (3 \cdot 0,09) + 0,04 = 0,89$
- Puntos por cada tiro de dos realizado: $2 \cdot 0,51 = 1,02$
- Puntos por cada tiro de tres realizado: $3 \cdot 0,34 = 1,02$
- Puntos esperados por cada tiro libre realizado: $1 \cdot (0,60 + 0,17) = 0,77$

Al no haber una relación directa entre el tipo de juego y los porcentaje de tiro, no existe una variación en ninguno de los cuartos sobre la rentabilidad de elegir un tiro de dos, de tres o tiro libre. Cabe destacar que el equipo rival consigue 0,89 puntos por cada jugada que realiza de bloqueo directo, un valor que está por debajo de los punto ofensivos esperados con este tipo de juego por el equipo ($0,89 < 0,94$). Si se

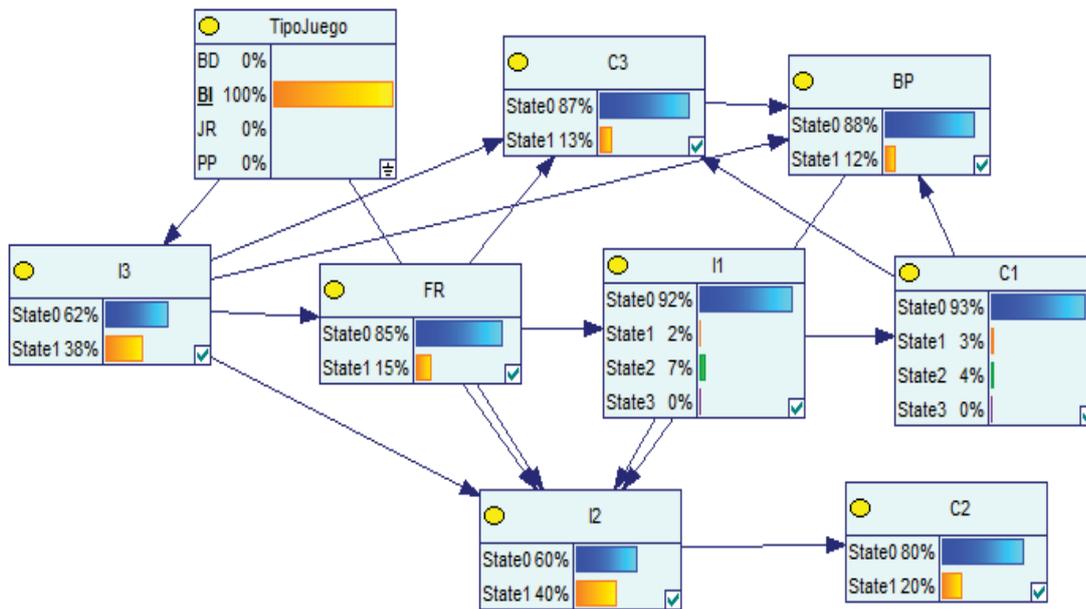


Figura 13: Red bayesiana defensiva con bloqueo indirecto.

consigue apretar algo más en defensa, es posible que se reduzca en mayor medida los puntos que consigue el rival.

3.2.1.2. Bloqueo indirecto

La Figura 13 muestra la red bayesiana generada teniendo como evidencia que el tipo de juego escogida ha sido bloqueo indirecto. Si éste es el tipo de juego utilizada por el equipo contrario, la probabilidad de que termine en un intento de tres aumenta considerablemente (22 % \rightarrow 38 %). Normalmente, cuando se produce un bloqueo indirecto, el defensor se queda clavado en el bloqueo, siendo muy difícil que recupere la distancia que le separa del jugador contrario al que estaba defendiendo. Para que el jugador rival no disponga de un tiro excesivamente cómodo tras el bloqueo, es fundamental que el compañero del defensor haga un cambio defensivo automático y acompañe al jugador para impedir su lanzamiento. Similar a lo que ocurría con las jugadas de bloqueo directo, la probabilidad de que se realice un tiro de dos disminuye (49 % \rightarrow 40 %), así como la probabilidad de convertir el tiro. De esta forma, si se consigue realizar los cambios defensivos a tiempo y defender los tiros de tres, la producción ofensiva del rival se verá mermada en este tipo de juego. La probabilidad de cometer una falta disminuye (19 % \rightarrow 15 %), junto a la probabilidad de recuperar el balón (15 % \rightarrow 12 %). Esto se puede entender debido al gran número de intentos de tres que realiza el rival, ya que en estos casos no se suele cometer falta y es más difícil recuperar el balón.

A continuación se muestran los puntos esperados en jugadas de bloqueo indirecto y la rentabilidad de cada tipo de tiro:

- Puntos esperados por bloqueo indirecto: $2 \cdot (0,20 + 0,04) + (3 \cdot 0,13) + 0,03 = 0,90$
- Puntos por cada tiro de dos realizado: $2 \cdot 0,51 = 1,02$
- Puntos por cada tiro de tres realizado: $3 \cdot 0,34 = 1,02$
- Puntos esperados por cada tiro libre realizado: $1 \cdot (0,60 + 0,17) = 0,77$

Los puntos esperados del rival con bloqueo indirecto son 0,90 por jugada, un valor ligeramente por encima de los conseguidos mediante bloqueo directo (0,89 puntos esperados por jugada). Parece ser que al equipo le cuesta más defender este tipo de juego, en la que el rival intenta aprovecharse de una ventaja generada tras la realización de un bloqueo para conseguir un tiro cómodo, normalmente desde la línea de tres. Para conseguir minimizar la efectividad rival, sería necesario mejorar la comunicación entre los jugadores, realizando cambios defensivos inmediatos cuando se vean superados por el bloqueo realizado por el contrario. De esta forma, sería posible llegar a los tiros del atacante con mayor rapidez, dificultando los lanzamientos y logrando que el jugador realice el tiro con un defensor encima.

3.2.1.3. Juego rápido

La Figura 14 muestra la red bayesiana generada teniendo como evidencia que el tipo de juego escogido ha sido juego rápido. Si éste es el tipo de juego utilizado por el equipo contrario, la probabilidad de que realice un intento de tres puntos se reduce notablemente (22 % \rightarrow 14 %), al igual que el porcentaje de acierto. Por el contrario, la probabilidad de que la jugada rival termine con un intento de dos aumenta (49 % \rightarrow 54 %). Al tratarse de un tipo de juego más dinámico y veloz, es posible que el equipo contrario intente finalizar las jugadas rápidamente, consiguiendo en la gran mayoría de los casos encestar antes de que la defensa se encuentre organizada. Para poder evitarlo, sería interesante diseñar alguna jugada de transición defensiva, de forma que le permita al equipo bajar a defender estructuradamente y evitar recibir puntos fáciles. El aumento de la probabilidad de cometer una falta (19 % \rightarrow 21 %) puede estar ligado al mismo hecho de que la defensa no se encuentra correctamente organizada a la hora de defender y cometa una mayor cantidad de faltas. Se produce un ligero aumento de la probabilidad de recuperar el balón (15 % \rightarrow 16 %) lo que puede deberse al hecho de que al tratarse de ataques rápidos, el rival pueda cometer el descuido de perder el balón con mayor facilidad.

A continuación se muestran los puntos esperados en jugadas de juego rápido y la rentabilidad de cada tipo de tiro:

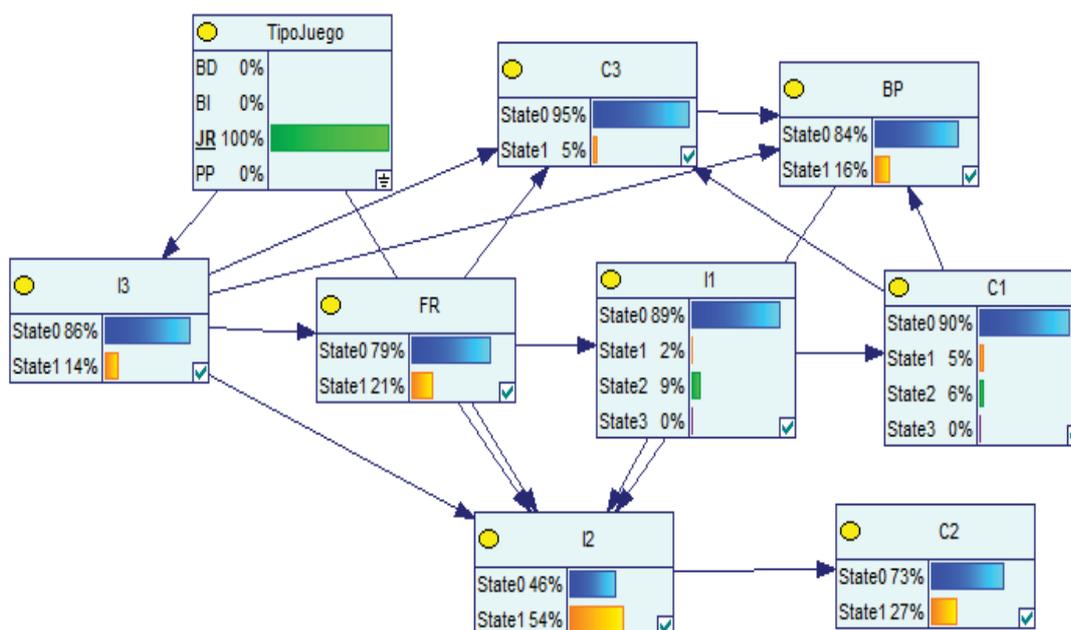


Figura 14: Red bayesiana defensiva con juego rápido.

- Puntos esperados por juego rápido: $2 \cdot (0,27 + 0,06) + (3 \cdot 0,05) + 0,05 = 0,86$
- Puntos por cada tiro de dos realizado: $2 \cdot 0,51 = 1,02$
- Puntos por cada tiro de tres realizado: $3 \cdot 0,34 = 1,02$
- Puntos esperados por cada tiro libre realizado: $1 \cdot (0,60 + 0,17) = 0,77$

La rentabilidad por jugada obtenida por el rival es 0,86, la mínima conseguida hasta el momento. Parece ser que el equipo está siendo capaz de realizar un repliegue eficaz, impidiendo que los jugadores contrarios consigan anotar puntos rápidamente. Como posible línea de mejora, podrían diseñarse jugadas de transición defensiva para intentar optimizar todavía más la defensa, tratando de minimizar al máximo el beneficio que obtiene el equipo rival en el ataque rápido. De esta forma, la defensa estaría en todo momento organizada, libre de debilidades que puedan dar alguna ventaja al contrario.

3.2.1.4. Perímetro poste

La Figura 15 muestra la red bayesiana generada teniendo como evidencia que el tipo de juego escogido ha sido perímetro poste. Si éste es el tipo de juego utilizado por el equipo contrario, la probabilidad de que finalice su jugada con un intento de tres disminuye significativamente (22% \rightarrow 17%), al igual que la probabilidad

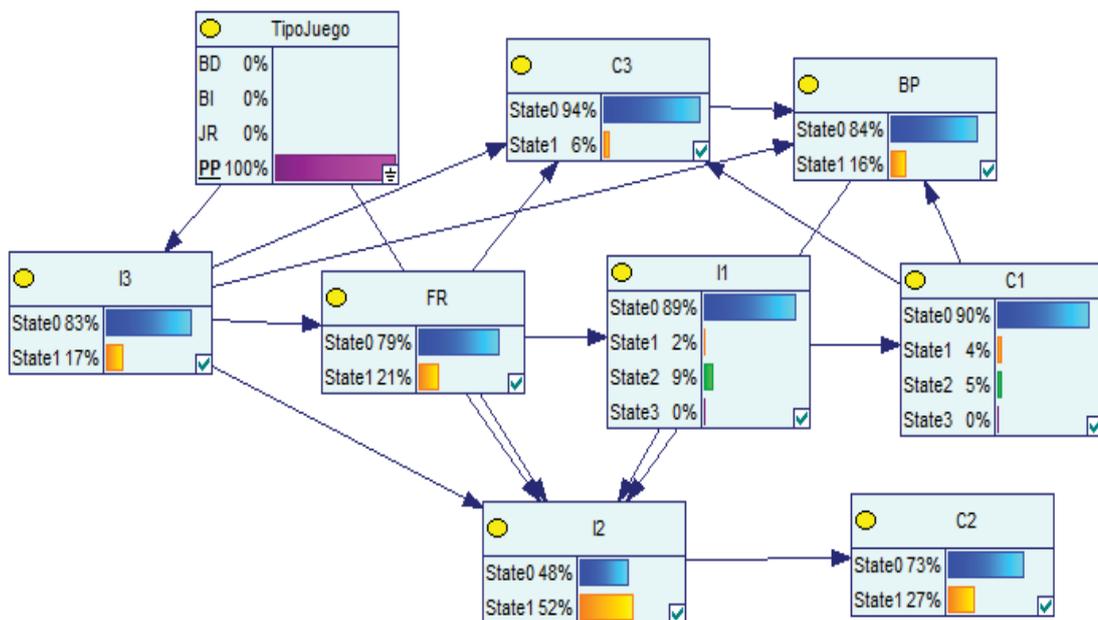


Figura 15: Red bayesiana defensiva con perímetro poste.

de que se convierta el tiro. En cambio, la probabilidad de que el rival realice un tiro de dos aumenta (49 % \rightarrow 52 %), así como su porcentaje de acierto. Este hecho podría deberse a que en este tipo de juego el atacante le da más importancia a que la jugada finalice con un intento de dos del pívot del equipo. El aumento del porcentaje de acierto puede significar que se está permitiendo que el rival consiga tirar a canasta sin demasiada oposición, por lo que sería recomendable aumentar la intensidad defensiva del equipo, especialmente en la zona. Por otro lado, se está produciendo un incremento del número de faltas cometidas por el equipo (19 % \rightarrow 21 %), lo cual indica que las faltas que se están realizando no están sirviendo para parar al rival, ya que no hay una disminución del porcentaje de acierto de sus tiros. Por tanto, sería conveniente escoger mejor las faltas que se cometen para poder minimizar la producción ofensiva del rival. La probabilidad de recuperar el balón aumenta ligeramente (15 % \rightarrow 16 %), lo cual parece indicar que se están defendiendo bien las líneas de pase cuando se produce un pase del rival a la zona.

A continuación se muestran los puntos esperados en jugadas con perímetro poste y la rentabilidad de cada tipo de tiro:

- Puntos esperados por perímetro poste: $2 \cdot (0,27 + 0,05) + (3 \cdot 0,06) + 0,04 = 0,86$
- Puntos por cada tiro de dos realizado: $2 \cdot 0,51 = 1,02$
- Puntos por cada tiro de tres realizado: $3 \cdot 0,34 = 1,02$
- Puntos esperados por cada tiro libre realizado: $1 \cdot (0,60 + 0,17) = 0,77$

Similar al juego rápido, los puntos esperados por cada jugada de perímetro poste son 0,86. Parece ser que en este caso también está siendo posible parar el empuje del rival, consiguiendo defender con acierto los jugadas de perímetro poste. Por contra, la efectividad ofensiva obtenida con este tipo de juego por el equipo es bastante menor ($0,79 < 0,86$), por lo que el rival está siendo capaz de sacar mayor provecho que el Estudiantes. Para conseguir reducir los puntos del contrario, sería interesante realizar defensas de dos contra uno en la zona, poniendo dos defensores sobre el pívot atacante y tratando de cortar sus líneas de pase. Así se conseguiría mantener una presión intensa sobre el jugador, minimizando en gran medida la probabilidad de que consiga realizar un tiro y obstaculizando sus líneas de pase, lo que incrementaría notablemente la probabilidad de que el rival realice una pérdida.

El Cuadro 4 agrupa los cambios de probabilidades dependiendo del tipo de juego en el modelo defensivo.

Cuadro 4: Variaciones de las probabilidades de las variables por tipo de juego en el modelo defensivo.

Estadística	General	BD	BI	JR	PP
I2	49 %	47 %	40 %	54 %	52 %
C2	25 %	24 %	20 %	27 %	27 %
I3	22 %	26 %	38 %	14 %	17 %
C3	7 %	9 %	13 %	5 %	6 %
BP	15 %	14 %	12 %	16 %	16 %
FR	19 %	18 %	15 %	21 %	21 %
	2 %	2 %	2 %	2 %	2 %
I1	8 %	8 %	7 %	9 %	9 %
	0 %	0 %	0 %	0 %	0 %
	4 %	4 %	3 %	5 %	4 %
C1	5 %	5 %	4 %	6 %	5 %
	0 %	0 %	0 %	0 %	0 %

3.2.1.5. Comparación de los tipos de juego

Tras analizar detenidamente los resultados obtenidos de la red bayesiana de tipo de juego, se han obtenido las siguientes conclusiones:

- Bloqueo indirecto es la jugada que más perjudica la defensa del Estudiantes, siendo el tipo de juego con mayor rentabilidad del rival (0,90 puntos esperados por jugada). Para conseguir disminuir esta cifra, sería interesante realizar cambios automáticos en defensa, alternando el defensor tras cada bloqueo realizado por el equipo atacante. De esta forma, se conseguiría tener un hombre sobre el tirador en todo momento, imposibilitando que el contrario consiga tiros cómodos y minimizando su porcentaje de acierto.

- Las jugadas de bloqueo directo también están suponiendo un problema para el equipo, ya que tiene un valor de puntos esperados muy similar al de bloqueo indirecto (0,89 puntos esperados por jugada). Vuelve a mostrarse que el problema puede residir en un mal entendimiento entre los jugadores a la hora de enfrentarse a los bloqueos rivales, desventaja que aprovecha el equipo contrario para obtener beneficio. Si se implementa el sistema de ayudas descrito en el punto anterior, sería posible realizar cambios defensivos automáticos que permitan tener un hombre con cada jugador atacante en todo momento, evitando situaciones de tiro cómodas.
- A diferencia de para el equipo rival, el juego rápido es el tipo de juego que más rentabilidad ofensiva otorgaba al Estudiantes (1,02 puntos esperados por jugada). En este caso, está siendo eficazmente neutralizado por el equipo, consiguiendo uno de los dos rendimientos ofensivos más bajos del contrario (0,86 puntos esperados por jugada). El equipo atacante no está siendo capaz de sacarle el máximo partido a las transiciones rápidas, lo que se puede deber a un eficiente repliegue defensivo por parte del Estudiantes. Si se quiere explotar todavía más esta faceta, se podrían diseñar jugadas específicas de repliegue defensivo, que ayudarán un poco a más a no perder la unión y estructura en ningún momento. También sería posible aumentar el número de faltas, cortando los contraataques del rival antes de que comiencen.
- En lo que respecta a las jugadas de perímetro poste, se está obteniendo un buen rendimiento defensivo, equiparable al obtenido frente al juego rápido (en ambos casos 0,86 puntos esperados por jugada). La realización de defensas de dos contra uno en la zona puede ayudar a que disminuya todavía más la rentabilidad ofensiva del rival en las jugadas de perímetro poste, ya que obstaculizaría en gran medida la posibilidad de realizar un tiro, además de impedir que el rival dé pases con comodidad.

3.2.2. Periodo

La Figura 16 muestra la red bayesiana del juego defensivo obtenida utilizando el periodo como variable clase. Al realizar un primer análisis de los arcos la red, se puede apreciar que el periodo tiene una influencia directa sobre el tipo de juego. A su vez, el tipo de juego influye sobre la probabilidad de realizar intentos de dos y de tres, pero no sobre la probabilidad de lanzar un tiro libre. No hay una relación directa entre periodo y los porcentajes de acierto de los tiros de dos y de tres, y lo mismo le ocurre al tipo de juego. Al analizar las independencias condicionales de la red, se puede apreciar que periodo es independiente del resto de variables del modelo si se conoce el tipo de juego desplegado por el equipo. A su vez, como se ha podido intuir en los arcos, el tipo de juego es independiente de los porcentajes de dos y de tres obtenidos por el equipo si se conocen los tiros de dos y de tres realizados respectivamente. El porcentaje de tiros libres es independiente del resto de variables de la red si se conocen los intentos de tiro libre realizados por el equipo.

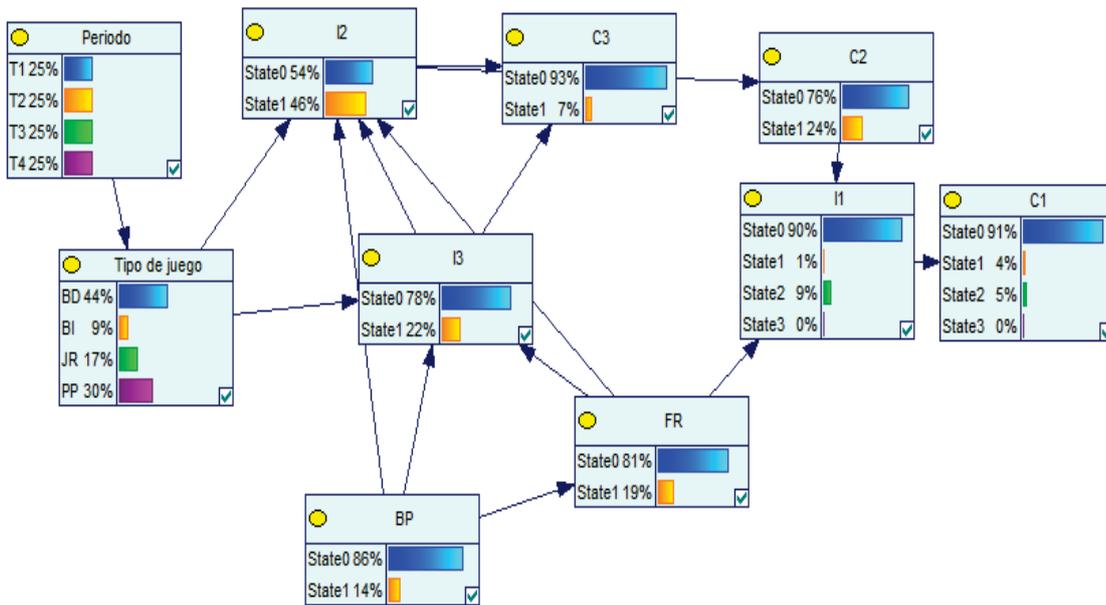


Figura 16: Red bayesiana defensiva por periodo.

Por tanto, se va a realizar un análisis de cómo afecta el periodo al tipo de juego del rival, ejecutando una comparación con el juego ofensivo por periodo del Estudiantes y analizando las diferencias significativas. A continuación se muestran los resultados obtenidos tras realizar el análisis. Al no haber apenas variación del resto de variables dependiendo el periodo debido a las independencias condicionales mostradas anteriormente, sólo se va a analizar la repercusión del periodo al tipo de juego desarrollado por el equipo rival.

3.2.2.1. Periodo 1

La Figura 17 muestra la red bayesiana generada teniendo como evidencia que el periodo en el que se encuentra el partido es el primero. El tipo de juego más utilizado es el bloqueo directo (43%), muy diferenciado del resto de jugadas. Al igual que ocurría con el modelo ofensivo, parece ser que los equipos suelen preferir comenzar con un juego más lento, elaborando las jugadas en vez de precipitarse en ataques rápidos. Esto se ve reflejado también en la alta probabilidad de realizar una jugada de perímetro poste (32%), la cual suele requerir de un mayor tiempo de construcción al jugar con el pívot en la pintura. Es interesante ver que los equipos rivales suelen abusar menos que el Estudiantes del juego rápido en el primer cuarto (16% < 25%). A su vez, la probabilidad de realizar un bloqueo indirecto es ligeramente superior a la probabilidad que obtenía el Estudiantes en ataque (9% > 8%), lo que demuestra que utilizan en mayor medida este recurso. Se aprecia que los rivales llevan un

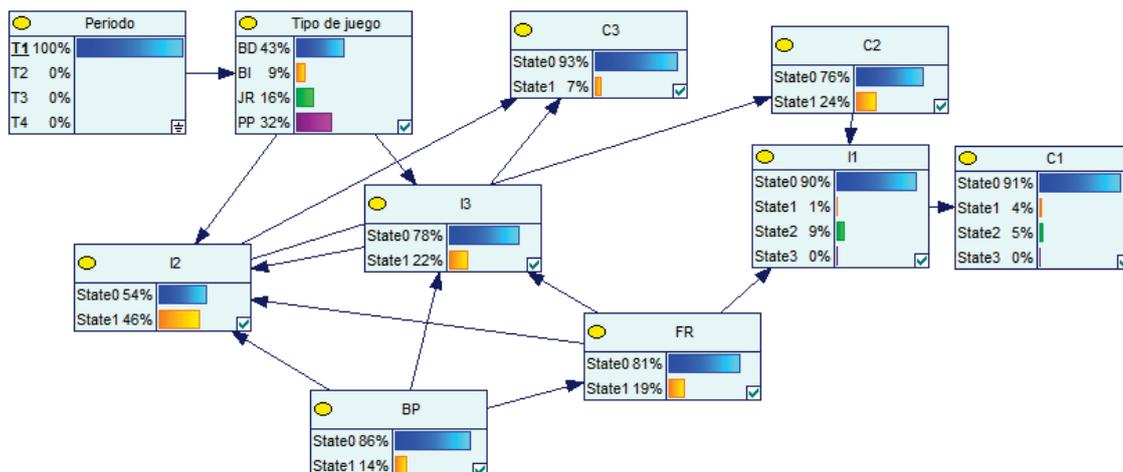


Figura 17: Red bayesiana defensiva en el primer periodo.

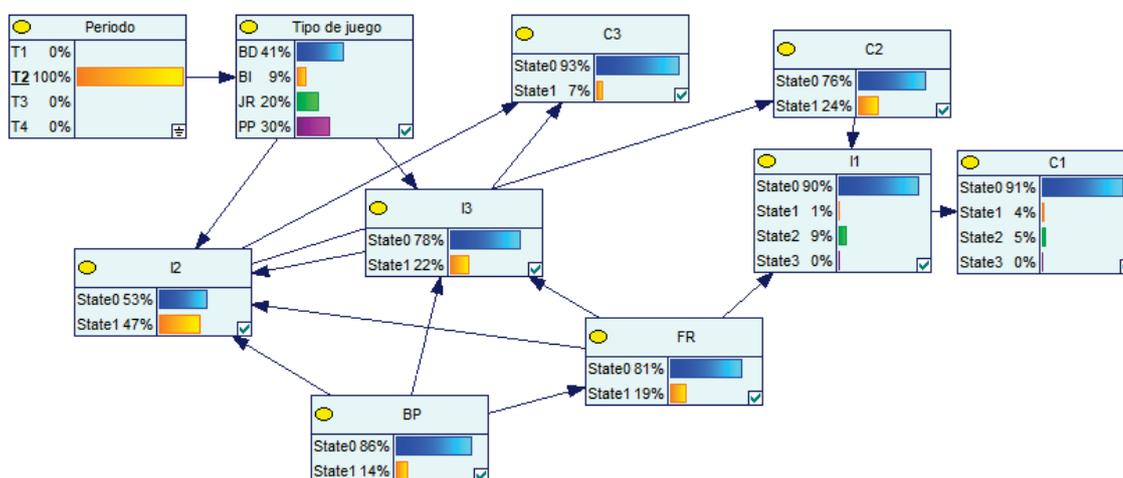


Figura 18: Red bayesiana defensiva en el segundo periodo.

estilo de juego más pausado que el Estudiantes en este primer cuarto, teniendo una probabilidad más alta en las jugadas de bloqueo directo y perímetro poste (43 % > 42 % y 32 % > 25 %, respectivamente).

3.2.2.2. Periodo 2

La Figura 18 muestra la red bayesiana generada teniendo como evidencia que el periodo en el que se encuentra el partido es el segundo. Una vez más el tipo de juego más utilizado es el bloqueo directo, aunque su probabilidad disminuye con respecto al primer cuarto (43 % → 41 %). De la misma forma, perímetro poste sigue siendo la segunda jugada más utilizada, pese a una ligera disminución con respecto al primer periodo (32 % → 30 %). Se produce un aumento en la probabilidad de

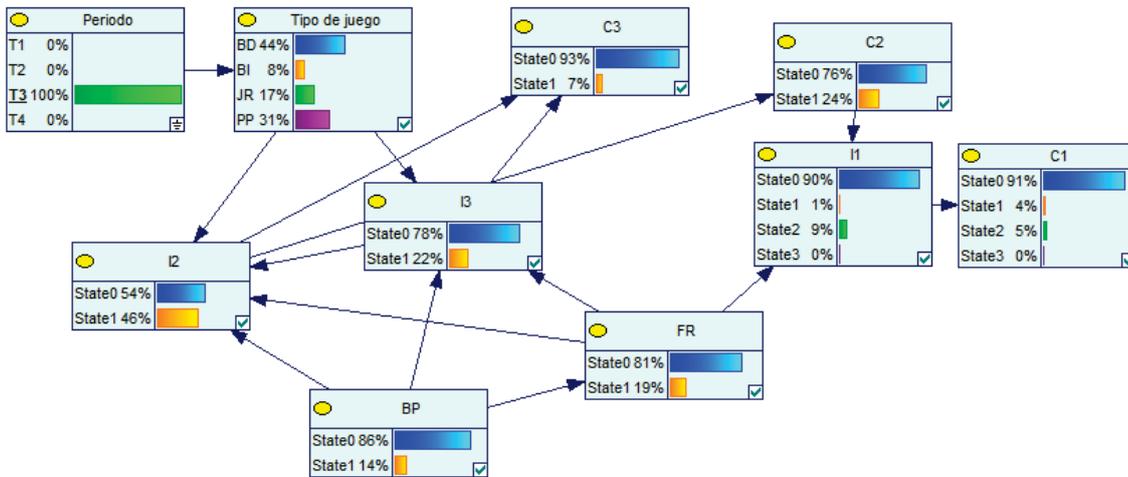


Figura 19: Red bayesiana defensiva en el tercer periodo.

utilizar juego rápido como jugada (16 % \rightarrow 20 %), mientras que la probabilidad de bloqueo indirecto se mantiene estable (9 %). Pese a mantenerse la tendencia del primer cuarto, el aumento del juego rápido puede indicar que el equipo rival mete una marcha más al partido, tratando de acelerarlo y pillar al Estudiantes con la guardia baja en defensa. Aun así, predomina un estilo de juego más pausado, con jugadas que requieren más elaboración. Si se realiza una comparación con el juego ofensivo por periodo Estudiantes, la diferencia obtenida no es tan amplia como en el primer cuarto, devolviendo unos registros para bloqueo directo (41 % < 42 %), bloqueo indirecto (9 % > 8 %), juego rápido (20 % < 22 %) y perímetro poste (30 % > 29 %) muy parecidos. El estilo de juego del Estudiantes es muy similar al de sus rivales en este segundo periodo.

3.2.2.3. Periodo 3

La figura 19 muestra la red bayesiana generada teniendo como evidencia que el periodo en el que se encuentra el partido es el tercero. Siguiendo la tendencia mostrada en los dos cuartos anteriores, los dos tipos de juego predominantes son bloqueo directo y perímetro poste, aumentando ligeramente sus probabilidades con respecto al segundo cuarto (41 % \rightarrow 44 % y 30 % \rightarrow 31 %, respectivamente). A su vez, las probabilidades de realizar bloqueo indirecto o de juego rápido disminuyen (9 % \rightarrow 8 % y 20 % \rightarrow 17 %, respectivamente). En este cuarto, parece que los equipos rivales vuelven a bajar un poco el ritmo, tratando de hacer jugadas más largas y tener más el balón. Si se compara con el juego ofensivo del Estudiantes en este cuarto, se aprecian grandes diferencias en el juego rápido (17 % < 23 %) y bloqueo directo (44 % > 39 %), mientras que en perímetro poste (31 % = 31 %) y bloqueo indirecto (8 % > 7 %) son prácticamente inexistentes. Esto demuestra que el equipo rival opta por un juego más pausado que el Estudiantes, que por el contrario trata

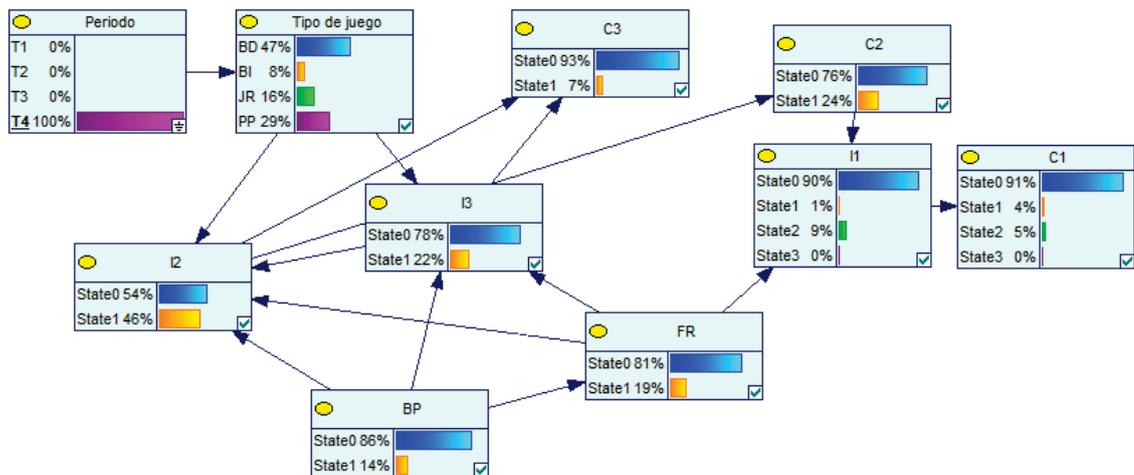


Figura 20: Red bayesiana defensiva en el cuarto periodo.

de sacar partido de jugadas rápidas que le puedan permitir conseguir puntos en el menor tiempo posible.

3.2.2.4. Periodo 4

La Figura 20 muestra la red bayesiana generada teniendo como evidencia que el periodo en el que se encuentra el partido es el cuarto. Como ha ocurrido en los tres periodos anteriores, los dos tipos de juego con mayor probabilidad son bloqueo directo (47%) y perímetro poste (29%). La primera aumenta respecto al cuarto anterior (44% \rightarrow 47%), mientras que la segunda sufre una disminución de su probabilidad (31% \rightarrow 29%). Los otros dos tipos de juego no sufren apenas modificaciones, ocurriendo únicamente una ligera disminución en la probabilidad de realizar juego rápido (17% \rightarrow 16%). En este último cuarto se hace más notoria la predisposición de los equipos rivales a controlar el partido, ya que existe una mayor probabilidad de realizar jugadas que requieren un mayor tiempo de ejecución. Esta diferencia es mayor que la encontrada en cualquier otro periodo. Si se realiza una comparación con el desempeño ofensivo del Estudiantes en este último cuarto, se aprecian diferencias muy significativas en bloqueo directo (47% > 43%) y juego rápido (16% < 20%), aunque no así en los dos tipos de juego restantes: bloqueo indirecto y perímetro poste (8% = 8% y 29% = 29%, respectivamente). En resumen, en este cuarto periodo existen diferencias significativas entre el juego del Estudiantes y los rivales, tendiendo este primero a desplegar un juego más rápido que el rival.

El Cuadro 5 agrupa los cambios de probabilidades dependiendo del periodo en el modelo defensivo.

Cuadro 5: Variaciones de las probabilidades de las variables por periodo en el modelo defensivo.

Estadística	General	Periodo 1	Periodo 2	Periodo 3	Periodo 4
TipoJuego	44 %	43 %	41 %	44 %	47 %
	9 %	9 %	9 %	8 %	8 %
	17 %	16 %	20 %	17 %	16 %
	30 %	32 %	30 %	31 %	29 %
I2	46 %	46 %	47 %	46 %	46 %
C2	24 %	24 %	24 %	24 %	24 %
I3	22 %	22 %	22 %	22 %	22 %
C3	7 %	7 %	7 %	7 %	7 %
BP	14 %	14 %	14 %	14 %	14 %
FR	19 %	19 %	19 %	19 %	19 %
I1	1 %	1 %	1 %	1 %	1 %
	9 %	9 %	9 %	9 %	9 %
	0 %	0 %	0 %	0 %	0 %
C1	4 %	4 %	4 %	4 %	4 %
	5 %	5 %	5 %	5 %	5 %
	0 %	0 %	0 %	0 %	0 %

3.2.2.5. Comparación de los periodos

Tras haber realizado un análisis exhaustivo de la red bayesiana para cada uno de los cuatro periodos de los partidos, a continuación se muestran las conclusiones principales que se han podido obtener:

- En los cuatro periodos predominan los tipos de juego bloqueo directo y perímetro poste, muy por encima de los otros dos tipos de jugada disponibles (bloqueo indirecto y juego rápido).
- Los dos estilos de juego predominantes son más pausados, y suelen necesitar de largas posesiones para ejecutarse eficientemente. Por tanto, se aprecia una tendencia de los rivales hacia lograr dominar el tempo del partido, intentando aprovechar cada posesión lo máximo posible.
- El juego del Estudiantes, en cambio, es más ofensivo, tratando de jugar posesiones cortas que acaben rápidamente en un intento de dos o de tres, tratando de llevar el partido a un nivel de ritmo más elevado.
- El tipo de juego del Estudiantes y el de los equipos a los que se enfrenta es muy distinto, demostrando que el equipo trata de seguir un estilo de juego propio que se desmarca de la línea general que siguen el resto de equipos a los que se ha enfrentado.

3.2.2.6. Comparación de los resultados obtenidos al crear las redes

A continuación se realiza una comparación en términos de eficiencia y área bajo la curva (AUC) de los distintos métodos que se han utilizado a la hora de construir las redes introducidos en el capítulo 2, demostrando que el método Greedy Search con tipo de juego y periodo como variable clase es el que mejores resultados obtiene. Primero, el Cuadro 6 representa las equivalencias entre las abreviaturas y los nombres de los métodos utilizados.

Cuadro 6: Correspondencia entre abreviaturas y nombres de los métodos de aprendizaje de la red.

Abreviatura	Nombre
GS +acc	Greedy Search que maximiza la puntuación de tipo de juego y periodo
GS +Bdeu	Greedy Search que maximiza el Bdeu
PC	Algoritmo PC
TAN	Tree Augmented Naïve Bayes

El Cuadro 7 muestra la comparación realizada entre los métodos de aprendizaje de la red. Se puede apreciar que el algoritmo Greedy Search que maximiza la puntuación es el que mejores resultados obtiene en términos de precisión en todas las redes, mostrando unos registros de AUC muy competitivos. La validación se ha realizado utilizando un método honesto de crossvalidación de 10 rodajas. Cabe destacar que la red bayesiana construida utilizando periodo como variable clase obtiene una precisión muy baja, lo que parece indicar que el periodo en el que se encuentra no influye en gran medida en el resto de variables.

Cuadro 7: Comparación de los métodos de aprendizaje de la red utilizados.

	Red	Algoritmo	Precisión	Área bajo la curva (AUC)
TIPO DE JUEGO	Ofensiva	GS + acc	0.49559	0.64189
		GS + Bdeu	0.46734	0.61764
		PC	0.47485	0.54290
		TAN	0.47132	0.63150
	Defensiva	GS + acc	0.44156	0.51772
		GS + Bdeu	0.43977	0.50473
		PC	0.44049	0.51765
		TAN	0.42647	0.52500
PERIODO	Ofensiva	GS + acc	0.28597	0.55038
		GS + Bdeu	0.23698	0.47676
		PC	0.23654	0.43772
		TAN	0.28420	0.54500
	Defensiva	GS + acc	0.25387	0.48848
		GS + Bdeu	0.22294	0.48517
		PC	0.22726	0.46561
		TAN	0.25171	0.51962

4. MODELIZACIÓN DE LOS JUGADORES

Este capítulo se centra en realizar una modelización del rendimiento de los jugadores que componen la plantilla del club Movistar Estudiantes. Además de la evaluación individual, se mostrarán rendimientos de combinaciones de jugadores que más veces coincidan en la cancha, pudiendo extraer información sobre lo bien o mal que interactúan entre ellos. En la misma línea, se ha incluido un análisis de cómo funcionan los jugadores bajo presión, añadiendo la variable *clutch time* y estudiando el comportamiento de cada jugador en ese espacio de tiempo del partido en el que una sola jugada puede cambiar irremediablemente el resultado final. Esta variable hace referencia al periodo de tiempo dentro de los partidos en el que no se sabe cual va a ser el desenlace del encuentro, ya que éste se encuentra muy igualado. Disponer de jugadores que se mueven bien bajo esta presión permite que el equipo sea capaz de remontar desventajas en estos instantes, o incluso aguantar un marcador positivo sin sucumbir a la precipitación y el nerviosismo.

Para lograr este objetivo, se ha utilizado una regresión lineal *ridge* a la hora de crear el modelo, método que ha sido capaz de reflejar la aportación de los jugadores de forma precisa e interpretable. Se han creado tres modelos distintos: el primero para analizar el rendimiento ofensivo, el segundo para el rendimiento defensivo y el tercero para el balance *plus-minus* del jugador que muestra su aportación global al equipo. De esta forma, es posible ver los puntos fuertes y débiles de cada jugador, así como de ciertas combinaciones de jugadores y el rendimiento individual en situaciones de máxima exigencia en los partidos.

Se ha realizado una comparación del método de regresión lineal *ridge* con otra serie de algoritmos que se han utilizado en diversos estudios de la misma índole para ver cuál obtiene mejores resultados. Primero, se ha realizado una comparación con la regresión lineal Lasso, así como con una regresión lineal sin regularización. Además, se ha analizado el valor de incluir las parejas de jugadores y el *clutch time* en el modelo, realizando una comparación de los tres regresiones mencionadas anteriormente con y sin estas variables. Por último, se han comparado los resultados obtenidos con la red neuronal H2O, el algoritmo random forest y el k-nearest neighbors. En todas las comparaciones realizadas, la regresión logística *ridge* supera a todos los métodos alternativos, mostrando unos valores de error cuadrático medio menores y proporcionando un modelo con mayor interpretabilidad.

Para apreciar si hay diferencias significativas entre los métodos alternativos contra los que se ha realizado la comparación y el método propuesto, se han utilizado el test de Friedman y el análisis post-hoc. En primer lugar, se ha tratado de averiguar si las diferencias entre las tres regresiones lineales que se han comparado son significativas o no. En segundo lugar, se ha realizado el mismo proceso para equiparar los resultados obtenidos con la regresión lineal *ridge* y los algoritmos H2O, Random Forest y k-nearest neighbours, para ver si las diferencias existentes entre los algoritmos son significativas.

4.1. Estructuración de los datos

Al igual que para el capítulo 3, los datos utilizados han sido extraídos del Excel proporcionado por el club Movistar Estudiantes. En este caso, las variables escogidas y la organización son completamente distintas. Se han utilizado los doce jugadores que se habían seleccionado en el primer preprocesado. Se realiza una reducción de la dimensionalidad del dataset al agrupar las jugadas por los quintetos que se encontraban en juego. Para conseguir este objetivo, ha hecho falta preprocesar una nueva serie de variables que identifican el quinteto que se encuentra en juego en cada jugada del partido. Gracias a estos nuevos datos, ha sido posible realizar la agrupación por quintetos mencionada anteriormente, lo que nos ha permitido calcular el número de veces que ha coincidido cada quinteto en el campo al mismo tiempo. Se han calculado los puntos obtenidos por cada uno de los quintetos, agregando las canastas de dos, de tres y tiros libres convertidos y dividiéndolos entre el número de jugadas en las que han estado juntos. Una vez calculados, se han añadido al dataset, obteniendo una nueva primera versión de la base de datos en la que aparecen los distintos quintetos de jugadores que han estado al mismo tiempo en la cancha y los puntos que han conseguido juntos. Los jugadores son variables del dataset, tomando el valor 1 si se encuentran presentes en el terreno de juego y tomando el valor 0 en caso contrario.

Por tanto, la base de datos obtenida muestra el quinteto de jugadores que ha estado en juego, el número de veces que se ha dado esa combinación de jugadores y los puntos de promedio que han conseguido cuando al estar al mismo tiempo en el partido, conformando un total de 15 variables (12 equivalentes a los jugadores del equipo, la variable clutch time, el número de apariciones de cada quinteto y los puntos promedio obtenidos por jugada). El dataset se ha dividido en dos subgrupos: el grupo ofensivo y el grupo defensivo. Además, se ha creado un tercer dataset en el que se encuentran los quintetos que han participado ofensiva y defensivamente. Este último dataset ha servido para calcular los puntos de promedio *plus-minus* de cada quinteto. Utilizando esta nueva base de datos, se ha procedido a realizar el estudio de la aportación de los jugadores a la hora de conseguir puntos, y, en su defecto, el número de puntos que permiten que el rival anote por jugada. La figura 21 muestra cómo se ha estructurado la base de datos.

Con esta base de datos, la regresión lineal *ridge* realiza una estimación de los puntos esperados por cada quinteto, indicando qué jugadores están contribuyendo positivamente y negativamente a conseguir puntos. Los pesos de los quintetos se tienen en cuenta, para darle más valor a los quintetos que se suelen repetir más a menudo a lo largo del tiempo. Con este fin, se pasa la variable pesos a los métodos a la hora de realizar el aprendizaje para que tengan en cuenta este valor.

VARIABLES DE LOS JUGADORES	Clutch Time	Pesos	Media
1 0 0 0 1 0 0 1 0 0 1 1	0	11	1.81
0 0 0 0 1 0 1 1 0 1 1 0	1	28	0.71
0 1 1 0 0 1 0 0 1 0 1 0	0	47	0.94
1 0 0 0 1 1 1 0 0 0 0 1	1	23	0.69

Figura 21: Ejemplo de la estructura de la base de datos.

4.2. Rendimiento ofensivo

Tener un buen rendimiento global de los jugadores es un factor vital para ganar partidos. Pese a que los equipos tratan de tener un juego lo más equilibrado posible, siempre suele haber un aspecto del juego en el que destacan, ya sea en el aspecto ofensivo o defensivo. En el baloncesto europeo, hasta hace pocos años siempre se había mantenido la idea de que el juego defensivo era más importante que el ofensivo. La influencia de la NBA en el baloncesto mundial ha conseguido cambiar esa percepción, ya que defiende la filosofía de crear un juego ofensivo y vistoso que permite obtener marcadores muy altos y brindar un espectáculo emocionante. Pese a este cambio de mentalidad global sobre el estilo de juego dominante, siguen siendo muchos los detractores de priorizar un buen rendimiento ofensivo por encima del defensivo.

Como se ha podido demostrar en el análisis del juego ofensivo del club Movistar Estudiantes mediante redes bayesianas, éste está principalmente basado en lograr obtener puntos de la forma más rápida posible, siendo claro el aumento en el uso del juego rápido por encima del de sus rivales. A su vez, la red ofensiva también ha conseguido reflejar que el equipo se decanta mucho por realizar tiros de tres. Teniendo en cuenta estos dos aspectos como puntos de partida, se puede ver que el equipo tiene una gran dependencia en el acierto ofensivo de sus jugadores, siendo vital para ellos el disponer de un rendimiento ofensivo muy alto para poder imponerse con claridad en los partidos. Por tanto, es indispensable para el equipo detectar cuáles son los factores que le ayudan a lograr este objetivo, así como los aspectos que están penalizando el rendimiento, ya que esto les va a permitir optimizar al máximo sus recursos y ser capaces de llegar a cotas más altas.

En esta sección se realiza un detallado análisis del rendimiento ofensivo del Estudiantes, mostrando los resultados obtenidos con la regresión lineal *ridge* y realizando una comparación con los diversos métodos mencionados anteriormente.

HakansonQ_ClutchTime	CvetkovicQ_ClutchTime	SavaneQ_ClutchTime	SuttonQ_LandsbergQ	CookQ_SuttonQ
0.091929232	0.070938896	0.063894283	0.020193482	0.015673771
VicedoQ_CookQ	PenaQ	SuttonQ	CookQ	VicedoQ_MedleyQ
0.013735756	0.013246001	0.010434627	0.010150969	0.007380395
CookQ_BrownQ	SuttonQ_ClutchTime	BrownQ	CookQ_LandsbergQ	LandsbergQ
0.006969795	0.006215076	0.005389024	0.005152865	0.001969950
ArteagaQ_ClutchTime	BrizuelaQ	CvetkovicQ	CookQ_MedleyQ	SavaneQ
0.000000000	-0.002213789	-0.002219016	-0.002410151	-0.002891032
VicedoQ	VicedoQ_BrownQ	PenaQ_ClutchTime	MedleyQ	LandsbergQ_MedleyQ
-0.003508634	-0.003743816	-0.005451878	-0.006617861	-0.007024454
ArteagaQ	HakansonQ	HakansonQ_VicedoQ	LandsbergQ_ClutchTime	CookQ_ClutchTime
-0.008455850	-0.012739428	-0.018326700	-0.024929824	-0.031416594
MedleyQ_ClutchTime	VicedoQ_ClutchTime	BrownQ_ClutchTime	BrizuelaQ_ClutchTime	
-0.039157712	-0.047712446	-0.059895693	-0.068607354	

Figura 22: Coeficientes de regresión del modelo ofensivo obtenidos mediante regresión lineal *ridge*.

4.2.1. Resultados

La Figura 22 muestra los coeficientes de regresión obtenidos mediante la regresión lineal *ridge* en el modelo ofensivo. Los coeficientes positivos contribuyen eficazmente a aumentar el rendimiento ofensivo del equipo, mientras que los negativos penalizan el rendimiento. Se han eliminado las variables que no tienen ninguna ocurrencia en los datos. Estos casos son combinaciones de jugadores que no han jugado nunca juntos y jugadores que no han participado en ninguna jugada de *clutch time*.

Para las combinaciones de jugadores, se han escogido únicamente aquellas que han estado juntas en la cancha un mínimo de 60 veces, *threshold* establecido para reflejar realmente el valor de las parejas que se encuentran a menudo juntos sin tener en cuenta los casos aislados y puntuales. En el caso del *clutch time*, se muestra la combinación de cada jugador con dicha variable. Se ha seguido el mismo procedimiento en el análisis del rendimiento defensivo y *plus-minus*.

De esta forma, es posible identificar los factores que influyen en el ataque del equipo, así como el grado de influencia que demuestran. Para la máxima información del modelo, se van a detallar las conclusiones obtenidas en tres grandes grupos: jugadores, combinaciones y situaciones del partido.

Las conclusiones obtenidas de los jugadores se agrupan a continuación:

- Sorprende que el jugador que proporciona un mejor rendimiento ofensivo sea Peña. Se trata de un jugador polivalente, que puede alternar las posiciones de escolta, alero y ala-pívot. Pese a no haber disputado un gran número de minutos a lo largo de la temporada, ha conseguido sacar un rendimiento ofensivo muy positivo cuando se ha encontrado en la cancha. Si el equipo necesita obtener puntos con urgencia, es el mejor candidato para conseguirlo.
- Los jugadores Sutton, Cook, Brown y Landsberg son los otros cuatro jugadores que también tienen un rendimiento ofensivo positivo. Cook es el base titular del equipo, con una gran trayectoria europea y muchos minutos a su espalda.

Sutton y Brown son dos de los pívots con mayor impacto en el equipo, el primero con una trayectoria mucho más amplia y el segundo con un futuro muy prometedor. Por último, Landsberg ha sido candidato al MVP (*most valuable player*) de la ACB este año, siendo un pilar fundamental del equipo y disputando una gran cantidad de minutos. No es extraño que estos jugadores sean clave en el rendimiento ofensivo del equipo, ya que son cuatro de los mejores jugadores de la plantilla.

- Al tratarse de jugadores que juegan en distintas posiciones, es posible elaborar un quinteto que contenga a los cinco jugadores con mayor rendimiento ofensivo. El quinteto ofensivo ideal del equipo es el siguiente: Cook (base), Peña (escolta), Landsberg (alero), Sutton (ala-pívot) y Brown (pívot).
- El jugador que penaliza el rendimiento ofensivo en mayor medida es Hakanson. Parece ser que el base suplente del equipo no está consiguiendo plasmar su calidad ofensiva en el campo, causando un impacto negativo en el equipo. Esto puede deberse a su juventud, ya que carece de la experiencia de otros jugadores del equipo.
- Los jugadores Brizuela, Arteaga, Vicedo, Cvetkovic, Savané y Medley también penalizan el rendimiento ofensivo del equipo. En lo que respecta a los cuatro primeros, se trata de jugadores muy jóvenes, con una media de edad que no supera los 23 años. Esta falta de experiencia puede estar influyendo en que su desempeño ofensivo no se encuentre a la altura de sus otros compañeros. Por otro lado, Medley y Savané son jugadores muy veteranos, con muchos años y experiencia a sus espaldas. Puede ser que ya no se encuentren en buena forma debido a su edad, lo que parece estar influyendo en su rendimiento ofensivo.

Las conclusiones obtenidas de las posibles combinaciones de jugadores son las siguientes:

- La combinación que mejores resultados parece dar al equipo es la pareja Sutton-Landsberg. Se trata de dos de los jugadores que mejor rendimiento ofensivo individual obtenían, por lo que parece ser que si juegan juntos al mismo tiempo pueden todavía conseguir un mejor desempeño. Landsberg juega en el exterior, mientras que Sutton juega en la pintura principalmente. Parece ser que existe una muy buena compenetración entre los dos jugadores, ventaja que el equipo debe explotar.
- Cook parece ser el jugador que más ayuda a sus compañeros de equipo a mejorar el rendimiento ofensivo. Sus combinaciones con Sutton, Vicedo, Brown y Landsberg son cuatro de las seis conexiones que proporcionan un beneficio ofensivo al equipo. Es especialmente notorio el caso de la combinación Cook-Vicedo, ya que Vicedo es un jugador que individualmente obtenía un rendimiento ofensivo negativo. Al mismo tiempo, la combinación Vicedo-Medley sorprende, ya que son dos jugadores que por si solos influían negativamente

en el juego ofensivo. Parece ser que el estilo de los dos jugadores se complementa muy bien cuando los dos se encuentran en la cancha al mismo tiempo. Parece vital el papel de Cook en el equipo, demostrando que su sola presencia repercute positivamente en la gran mayoría de sus compañeros.

- La pareja que más penaliza el rendimiento ofensivo es Hakanson-Vicedo, dos jugadores que individualmente también eran perjudiciales ofensivamente. El modelo demuestra que no es conveniente juntar a los dos jugadores en cancha al mismo tiempo, ya que esto afecta notablemente al desempeño ofensivo global del equipo. Al tratarse de dos jugadores muy jóvenes, es posible que no sean capaces de llevar adecuadamente el ritmo del partido, o incluso que sus dos estilos de juego sean incompatibles. Sería beneficioso para el equipo reducir el número de ocasiones en las que se encuentran estos dos jugadores en la cancha al mismo tiempo.
- Las siguientes combinaciones de jugadores también resultan perjudiciales para el rendimiento ofensivo del equipo: Cook-Medley, Vicedo-Brown y Landsberg-Medley. El jugador que más se repite en estas combinaciones es Medley, que pese a complementarse muy bien con Vicedo, para otros jugadores no resulta beneficioso. No es capaz de complementarse bien con Cook, que es el jugador que más consigue aumentar el rendimiento ofensivo de sus compañeros. También vuelve a aparecer Vicedo, que como ya se ha mencionado anteriormente, su combinación con Medley es la única que resulta beneficiosa para el equipo. Es interesante para el equipo tener en cuenta estas parejas, para evitar que estén compartiendo minutos durante un periodo largo de tiempo.

Por último, las conclusiones obtenidas de las situaciones de máxima exigencia de los jugadores se muestran a continuación:

- Sorprendentemente, Hakanson es el jugador que demuestra una mayor templanza en los momentos decisivos del partido, situaciones en las que parece no temblarle la muñeca. Es el jugador con mejor rendimiento ofensivo en los instantes decisivos del partido, convirtiéndose en un jugador fundamental para esos minutos agónicos en los que se decide el resultado.
- Además de Hakanson, Cvetkovic, Savané y Sutton son claves en el *clutch time*, proporcionando un muy buen rendimiento en este tipo de situaciones. Cvetkovic también es un jugador muy joven, pero Savané y Sutton son dos de los jugadores más veteranos de la plantilla, con una larga trayectoria en la que se han enfrentado a muchas situaciones en las que se decide el resultado, lo que les puede estar permitiendo desenvolverse mejor.
- El jugador que peor desempeño demuestra en el *clutch time* es Brizuela. Su inexperiencia puede jugar un papel fundamental en su rendimiento, ya que en su corta carrera profesional no ha podido enfrentarse a muchas situaciones de este tipo.

- Peña, Landsberg, Cook, Medley, Vicedo y Brown también parecen demostrar dificultades a la hora de rendir eficientemente en el clutch time. Es curioso que algunos de los jugadores que no consiguen rendir bien en los momentos decisivos del partido tienen un rendimiento ofensivo muy notorio en situaciones normales del encuentro. Por tanto, es importante que el equipo conozca que jugadores se desenvuelven mejor en este escenario, para poder elegir cuál es el quinteto que va a conseguir el mayor rendimiento y maximizar el beneficio obtenido.

4.2.2. Comparación

Se ha realizado una comparación del método de regresión lineal *ridge* con parejas y *clutch time* propuesto en este estudio con otros tipos de regresión y métodos de *machine learning* en R (MLR), para ver qué método es mejor a la hora de evaluar el rendimiento ofensivo. Por un lado, se ha comparado la regresión del estudio con una regresión lineal Lasso regularizada y una regresión sin regularización, además de realizar una comparación de las regresiones con y sin las combinaciones de jugadores y el *clutch time*. Por otro lado, también se ha realizado una comparación con la red neuronal (H2O), el algoritmo random forest y el k-nearest neighbors. Para medir la efectividad de cada método, se han utilizado como medida el error cuadrático medio, así como el test de Friedman y análisis post-hoc para ver si las diferencias mostradas entre los métodos son significativas. Primero se va a realizar la comparación entre las regresiones y los algoritmos de aprendizaje automático H2O, random forest y k-nearest neighbors, para posteriormente ver las diferencias entre los resultados obtenidos con las combinaciones de jugadores y *clutch time* y sin estas variables.

El Cuadro 8 muestra las equivalencias entre las abreviaturas que se utilizan en el análisis y los nombres originales de los métodos.

Cuadro 8: Equivalencia entre abreviaturas y los nombre originales.

Abreviatura	Nombre
RGR	Regresión lineal <i>ridge</i>
RGL	Regresión lineal Lasso
RF	Random forest
H2O	Red neuronal H2O
KNN	K-nearest neighbors
RGSR	Regresión lineal sin regularización

El Cuadro 9 muestra el valor de error cuadrático medio que tienen los métodos si se incluyen las combinaciones de jugadores y el *clutch time*, mientras que el Cuadro 10 muestra los mismos resultados sin incluir estas variables. A continuación se agrupan las principales conclusiones obtenidas tras la comparación de modelos:

- Se puede apreciar que el método que obtiene un menor error cuadrático medio es la regresión lineal *ridge* que incluye las parejas de jugadores y la variable psicológica *clutch time*, por encima del resto de métodos.
- Se demuestra que se obtiene una disminución del error cuadrático medio si se utiliza una regresión lineal con regularización.
- Si no se incluyen las combinaciones ni la variable *clutch time*, el método que mejor funciona es la regresión lineal Lasso, seguida de cerca por la regresión *ridge*. La disminución del índice de error al incluir las nuevas variables indica que éstas están proporcionando una información adicional al modelo, lo que puede resultar muy útil para el entrenador del equipo.
- El método propuesto en el estudio obtiene un índice de error menor que el resto de clasificadores de aprendizaje automático utilizados en la comparación. Entre ellos, el método con mejores resultados es el random forest, algoritmo que proporciona una difícil interpretación, mientras que el método con peor resultado es la red neuronal H2O.

Cuadro 9: Comparación de la regresión lineal *ridge* con métodos MLR para el rendimiento ofensivo.

Algoritmo	Error cuadrático medio (ECM)
RGR	0.15178 ± 0.3378
RGL	0.15480 ± 0.3384
RF	0.16011 ± 0.3419
H2O	0.28209 ± 0.5244
KNN	0.24305 ± 0.4707

Cuadro 10: Resultados obtenidos sin combinaciones de jugadores ni *clutch time* para el rendimiento ofensivo.

Algoritmo	Error cuadrático medio (ECM)
RGR	0.15400 ± 0.3379
RGL	0.15298 ± 0.3359
RGSR	0.16190 ± 0.3498
RF	0.17361 ± 0.3576
H2O	0.27030 ± 0.4459
KNN	0.27361 ± 0.4496

La Figura 23 muestra la la variación del error cuadrático medio usando las parejas de jugadores y la variable *clutch time* (en color rojo) y no usarlas (en color amarillo). Se puede apreciar que se logra reducir el error cuadrático medio en tres de los cinco

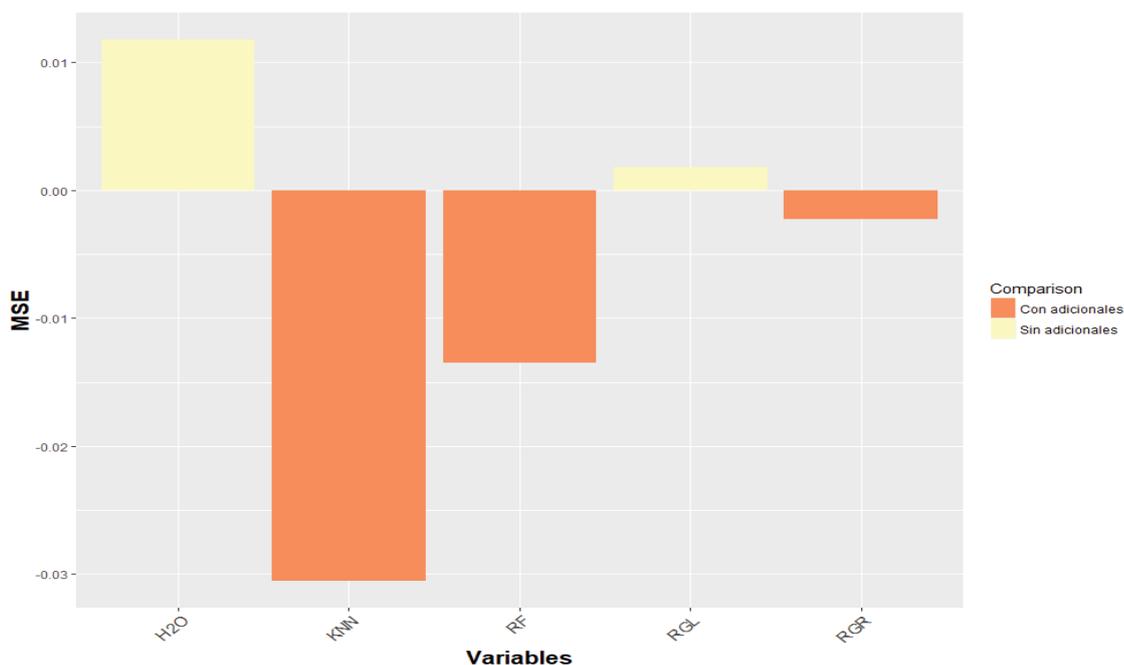


Figura 23: Variación de error cuadrático medio entre usar (rojo) y no usar (amarillo) las combinaciones de jugadores y la variable *clutch time* (rendimiento ofensivo).

algoritmos testados. Esto parecer indicar que la adición de los nuevos atributos logra aportar más información al modelo.

Por tanto, para el rendimiento ofensivo el método que mejores resultados obtiene es la regresión lineal *ridge*. Consigue mejores registros en términos de error cuadrático medio, además de proporcionar una mayor cantidad de información relevante sobre el modelo y los jugadores.

Para ver si estas diferencias que se han demostrado en la comparación de los modelos son realmente significativas, se ha utilizado el test de Friedman y el análisis post-hoc. Primero, se han realizado los tests entre los distintos tipos de regresiones. Los resultados obtenidos se pueden apreciar en la Figura 24. Si se interpreta el resultado, es posible concluir que existen diferencias significativas entre la regresión lineal *ridge* regularizada con combinaciones y *clutch time* y el resto de regresiones. A su vez, no parece haber diferencias relevantes entre la regresión lineal Lasso regularizada, la regresión lineal sin utilizar parejas ni *clutch time* y la regresión sin regularización.

Los resultados de la comparación entre la regresión lineal *ridge* con combinaciones y *clutch time* y los métodos de aprendizaje automático se muestran en la Figura 25. En este caso, se aprecia que no existen diferencias significativas entre el método random forest y la regresión *ridge*. A su vez, sí existe una diferencia relevante entre estos dos métodos y la red neuronal H2O y el algoritmo k-nearest neighbors.

En resumen, tras analizar y comparar el método de regresión lineal *ridge*, es posible concluir que es el que mejores resultados devuelve en términos de error

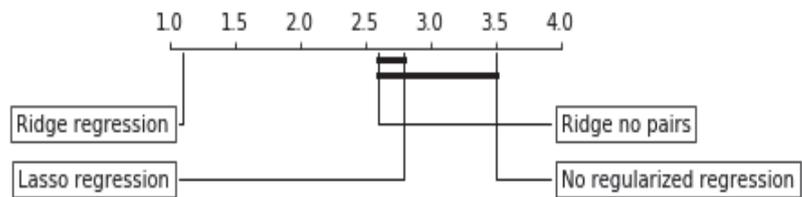


Figura 24: Resultados del rendimiento ofensivo del test de Friedman y análisis post-hoc entre las regresiones en términos de ECM.

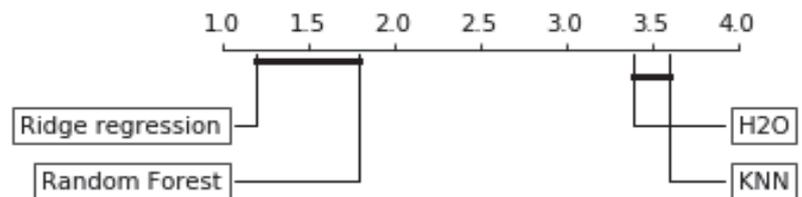


Figura 25: Resultados del test de Friedman y el análisis post-hoc entre la regresión *ridge* y los métodos de aprendizaje automático para el rendimiento ofensivo en términos de ECM.

cuadrático medio. Aunque las diferencias que existen entre este método y el random forest no son significativas, la regresión es fácilmente interpretable, mientras que el random forest es un modelo de muy difícil interpretación. Por tanto, se demuestra que el método escogido en el estudio es el que mejores resultados devuelve y el más interpretable para medir el rendimiento ofensivo de los jugadores del club Movistar Estudiantes.

4.3. Rendimiento defensivo

Aunque se está produciendo un cambio de tendencia en el juego hacia priorizar el rendimiento ofensivo, el rendimiento defensivo sigue siendo un elemento clave a la hora de conseguir un resultado positivo en los partidos. Una buena defensa consigue minimizar los puntos recibidos, lo que permite frenar el ataque rival y poder obtener más fácilmente una ventaja en el marcador. Pese a que el juego defensivo no es muy vistoso (salvo en alguna jugada aislada, como por ejemplo, los taponés), es muy efectivo, ya que se puede conseguir ganar el partido con un desempeño ofensivo no muy elevado si se consiguen detener las acometidas del rival.

La red bayesiana defensiva del club Movistar Estudiantes ha mostrado que el equipo demuestra una mayor debilidad defensiva en las jugadas que contienen bloqueos, lo que parecía indicar que no existe una buena comunicación para salir a las ayudas. Por tanto, al igual que para el rendimiento ofensivo, en esta sección de realiza un detallado análisis del rendimiento defensivo del Estudiantes, para identificar qué factores ayudan y penalizan en este aspecto. Se muestran los resultados obtenidos con la regresión lineal *ridge* y realizando una comparación con los diversos métodos mencionados anteriormente.

4.3.1. Resultados

La Figura 26 muestra los coeficientes de regresión obtenidos mediante la regresión lineal *ridge*. En este caso, los coeficientes positivos indican una contribución defensiva negativa, ya que significa que se está permitiendo al rival anotar un mayor número de puntos. Por el contrario, si éstos son negativos, están consiguiendo un rendimiento defensivo positivo, impidiendo que el equipo contrario consiga anotar.

A continuación se muestran las conclusiones obtenidas de forma detallada, para cada uno de los tres grandes apartados analizables: rendimiento individual, rendimiento por combinaciones de jugadores y rendimiento en las situaciones decisivas del partido.

Seguidamente se muestran las conclusiones obtenidas del rendimiento individual:

- El jugador que consigue el mejor rendimiento defensivo del equipo es Brizuela. Es un jugador joven, en plenitud física, lo que le puede permitir defender con una mayor intensidad y conseguir pasar los bloqueos rivales utilizando su físico. Es el mejor jugador si se quiere secar al rival, por lo que sería conveniente que estuviera en la cancha cuando es necesario parar la ofensiva contraria.
- Los jugadores Hakanson, Cvetkovic, Brown, Sutton y Savané también tiene un desempeño defensivo positivo. Los tres primeros son jugadores muy jóvenes, capaces de realizar un despliegue físico superior que les permite estar activos en defensa y minimizar los puntos conseguidos por el rival. Parece ser que consiguen compensar su falta de experiencia con su capacidad física. Por el contrario, Sutton y Savané son jugadores muy veteranos, con un historial de

PenaQ_ClutchTime	BrizuelaQ_ClutchTime	SuttonQ_ClutchTime	cvetkovicQ_ClutchTime	MedleyQ_ClutchTime
1.946147e-02	1.534161e-02	8.075190e-03	7.819820e-03	7.023912e-03
CookQ_ClutchTime	LandsbergQ_ClutchTime	BrownQ_ClutchTime	LandsbergQ_MedleyQ	HakansonQ_ClutchTime
4.908574e-03	3.253930e-03	2.288201e-03	2.277243e-03	2.138964e-03
SavaneQ_ClutchTime	PenaQ	MedleyQ	CookQ_MedleyQ	VicedoQ_CookQ
1.880218e-03	1.840557e-03	1.789128e-03	1.716580e-03	1.602152e-03
VicedoQ_MedleyQ	LandsbergQ	HakansonQ_MedleyQ	CookQ	ArteagaQ
1.309657e-03	8.095029e-04	7.798982e-04	6.543905e-04	6.534615e-04
VicedoQ	ArteagaQ_ClutchTime	VicedoQ_BrownQ	SuttonQ_LandsbergQ	HakansonQ_SuttonQ
1.318109e-04	0.000000e+00	-2.613029e-05	-7.581448e-05	-1.778127e-04
HakansonQ	CookQ_BrownQ	BrizuelaQ_CookQ	SavaneQ	CookQ_LandsbergQ
-3.375901e-04	-3.421309e-04	-3.668948e-04	-4.446345e-04	-4.843539e-04
LandsbergQ_BrownQ	BrownQ	CookQ_SuttonQ	CvetkovicQ	SuttonQ
-6.082909e-04	-7.312883e-04	-8.198362e-04	-9.583510e-04	-9.826193e-04
HakansonQ_VicedoQ	VicedoQ_SuttonQ	VicedoQ_ClutchTime	BrizuelaQ	
-1.115607e-03	-1.130092e-03	-1.377898e-03	-1.405364e-03	

Figura 26: Coeficientes de regresión del modelo defensivo obtenidos mediante regresión lineal *ridge*.

encuentros notorio a sus espaldas. Pese a que ya no se encuentren en plenitud física por su avanzada edad, consiguen contrarrestar esa debilidad con su experiencia, lo que les permite rendir a un gran nivel defensivo sin necesidad de estar tan en forma como los jugadores más jóvenes de la plantilla.

- Los jugadores que proporcionan un rendimiento defensivo positivo permiten realizar un quinteto de garantías versátil. Hakanson, Brizuela y Cvetkovic se pueden intercambiar las posiciones de base escolta y alero, mientras que Sutton, Brown y Savané se encargan de ocupar las posiciones de ala-pívot y pívot. El entrenador puede probar distintas combinaciones de estos jugadores para mejorar el rendimiento, teniendo en cuenta las sinergias entre los jugadores que se analizan más adelante.
- El jugador que más perjudica el rendimiento defensivo del equipo es Peña. Pese a ser el jugador con mejor rendimiento ofensivo, este rendimiento se desploma drásticamente en defensa. Por tanto, Peña puede realizar un buen papel en el ataque, pero si el equipo necesita defender, es el último jugador que conviene utilizar.
- Los jugadores Vicedo, Arteaga, Landsberg, Cook y Medley también penalizan el rendimiento defensivo del equipo. Los dos primeros son jugadores jóvenes que, pese a tener buenas cualidades físicas, no están siendo capaces de demostrar un desempeño defensivo positivo. En lo que respecta a Landsberg, es un jugador con un claro perfil ofensivo, muy valioso en ataque pero con más carencias a la hora de defender. Cook y Medley son jugadores más veteranos, que puede ser que debido al declive físico asociado a la edad no sean capaces de rendir defensivamente como lo hacían hace años. Teniendo esto en cuenta, el entrenador puede decidir dejar a estos jugadores y a Peña en el banquillo cuando necesite realizar una buena jugada defensiva a toda costa, ya que en caso contrario no va a conseguir frenar el ataque rival.

Las conclusiones que se obtienen a partir de las combinaciones de jugadores son

las siguientes:

- La pareja Vicedo-Sutton es la que mejor rendimiento defensivo proporciona al equipo. Sorprende la presencia de Vicedo, jugador que individualmente no mostraba un buen desempeño. Parece ser que si comparte cancha con un buen defensor como Sutton, consigue imprimir un punto más de intensidad y es capaz de mejorar su capacidad defensiva. Sutton es un defensor individual consumado, por lo que su presencia en la mejor pareja defensiva no es de extrañar.
- Las parejas Vicedo-Hakanson, Cook-Sutton, Landsberg-Brown, Hakanson-Sutton, Sutton-Landsberg y Vicedo-Brown también consiguen obtener un rendimiento defensivo positivo para el equipo. Es interesante apreciar que en cada pareja por lo menos uno de los jugadores es un buen defensor individualmente. Parece ser que la combinación de Hakanson, Sutton o Brown con el resto de sus compañeros está permitiendo aumentar el rendimiento de jugadores que por si solos no son capaces de llegar a unas cotas óptimas defensivas. Por tanto, si el equipo necesita priorizar el rendimiento defensivo en ciertos momentos del partido, es importante tener en cuenta estas parejas, ya que aunque haya algunos jugadores que por si solos no son capaces de defender adecuadamente, mejoran notablemente sus prestaciones cuando se encuentran con ciertos compañeros a su lado.
- La pareja que demuestra los peores registros defensivos al compartir minutos en la cancha es Landsberg-Medley. Landsberg tiene un perfil claramente ofensivo, faceta en la cual destaca notoriamente. Por el contrario, como ya se ha demostrado en el análisis individual de los jugadores, defensivamente no es capaz de rendir al mismo nivel. Por otro lado, Medley también ha obtenido una valoración individual negativa, por lo que al juntar los dos perfiles, no sorprende que sea la peor pareja defensiva del equipo. Para reducir al máximo el impacto negativo que puede tener esta pareja en el rendimiento defensivo, es importante minimizar el tiempo que están juntos en la cancha, buscando combinaciones de jugadores que sí que ayuden a mejorar en defensa.
- Otras parejas que perjudican a la aportación defensiva del equipo son Hakanson-Medley, Vicedo-Medley, Vicedo-Cook y Cook-Medley. Se trata de parejas en la que los jugadores por si solos no tienen un buen rendimiento en defensa, por lo que su combinación tampoco resulta beneficiosa para el equipo. Destaca el hecho de que Medley parece ser el jugador que peores registros obtiene, ya que prácticamente aparece en todas las combinaciones de jugadores negativas. Esto demuestra su ineficacia defensiva, ya que hay un gran número de jugadores que tienen un mal desempeño cuando juegan a su lado y mejoran sus registros al lado de otros jugadores.

Por último, las conclusiones del rendimiento de los jugadores obtenidas a partir de las situaciones de mayor exigencia de los partidos son las siguientes:

- El jugador que mejor defiende en situaciones de máxima tensión es Vicedo. Aunque en situaciones normales no es capaz de tener un buen rendimiento, es capaz de adaptarse bien en los momentos donde el desenlace del partido está en el aire, aumentando la intensidad y ayudando al equipo.
- Es curioso que el único jugador que tiene un rendimiento defensivo positivo en el *clutch time* de todos los jugadores que componen la plantilla es Vicedo. Esto demuestra que el equipo tiene serias dificultades a la hora de defender las jugadas del rival en estos instantes del partido, lo que puede llevarle a perder los partidos que estén en el alambre. Esto puede deberse a una falta de concentración, provocada por el nivel de tensión que se respira en esos momentos.
- Al igual que en el rendimiento defensivo individual, Peña es el jugador que peores registros defensivos muestra. A diferencia que Vicedo, jugador que pese a no defender bien durante el partido se crece en momentos de máxima tensión, Peña mantiene su mal rendimiento.
- Toda la plantilla excepto Vicedo tiene un mal rendimiento defensivo en el *clutch time*, lo que hace sumamente difícil el elegir qué jugadores escoger cuando hay que defender en un momento clave del partido. Es importante que el equipo realice entrenamientos específicos que le ayuden a defender mejor en estas situaciones, ya que se le pueden estar escapando muchos partidos por culpa de esta debilidad.

4.3.2. Comparación

Al igual que se ha hecho para el rendimiento ofensivo, se ha realizado una comparación del método de regresión lineal *ridge* con parejas y *clutch time* propuesto en este estudio con los otros tipos de regresión (regresión lineal Lasso y regresión sin regularización) y los métodos de MLR (H2O, random forest y k-nearest neighbors). A su vez, se ha comparado si se obtienen mejores resultados con y sin las combinaciones y el *clutch time*. Las medidas utilizadas son las mismas que en el caso anterior, error cuadrático medio, test de Friedman y el análisis post-hoc para ver si las diferencias que se aprecian son significativas.

El Cuadro 11 muestra el valor de error cuadrático medio que tienen los métodos si se incluyen las combinaciones de jugadores y el *clutch time*, mientras que el Cuadro 12 muestra los mismos resultados sin incluir estas variables. Al igual que se ha realizado para el modelo ofensivo, a continuación se agrupan las principales conclusiones obtenidas tras la comparación de modelos:

- Una vez más, el método que obtiene un mejor resultado en términos de error cuadrático medio es la regresión lineal *ridge* que incluye las combinaciones y el *clutch time*.

- En la misma línea, se vuelve a ver que la regularización está permitiendo obtener mejores resultados en la regresión lineal.
- Se vuelve a apreciar que la inclusión de las combinaciones de jugadores y la variable psicológica *clutch time* está permitiendo obtener mejores resultados, proporcionando mayor información y ganando interpretabilidad.
- De los métodos de MLR utilizados en la comparación, el random forest vuelve a ser el mejor.
- Si no se incluyen las combinaciones ni el *clutch time*, la regresión lineal *ridge* es la que mejores resultados obtiene.

Cuadro 11: Comparación de la regresión lineal *ridge* con métodos MLR para el rendimiento defensivo.

Algoritmo	Error cuadrático medio (ECM)
RGR	0.12219 ± 0.2981
RGL	0.12229 ± 0.2983
RF	0.12775 ± 0.2884
H2O	0.19090 ± 0.3605
KNN	0.18703 ± 0.3734

Cuadro 12: Resultados obtenidos sin combinaciones de jugadores ni *clutch time* para el rendimiento defensivo.

Algoritmo	Error cuadrático medio (ECM)
RGR	0.12282 ± 0.2991
RGL	0.12341 ± .3001
RGSR	0.12341 ± 0.3061
RF	0.13396 ± 0.2998
H2O	0.16314 ± 0.3366
KNN	0.16990 ± 0.3693

En la Figura 27 se puede apreciar la diferencia de error cuadrático medio entre usar (rojo) y no usar (amarillo) las combinaciones de jugadores y la variable *clutch time*. En este caso, existen diferencias en las dos regresiones (*ridge* y Lasso), así como en el random forest. En las dos regresiones la diferencia es menor, pero esta aumenta en el caso del algoritmo random forest, donde se aprecia una mayor diferencia. Por tanto, se puede concluir que el añadir estas nuevas variables está siendo beneficioso para el modelo, ya que está ayudando en mayor o menor medida a ajustar las predicciones y obtener mayor información, además de proporcionar un grado mayor de interpretabilidad.

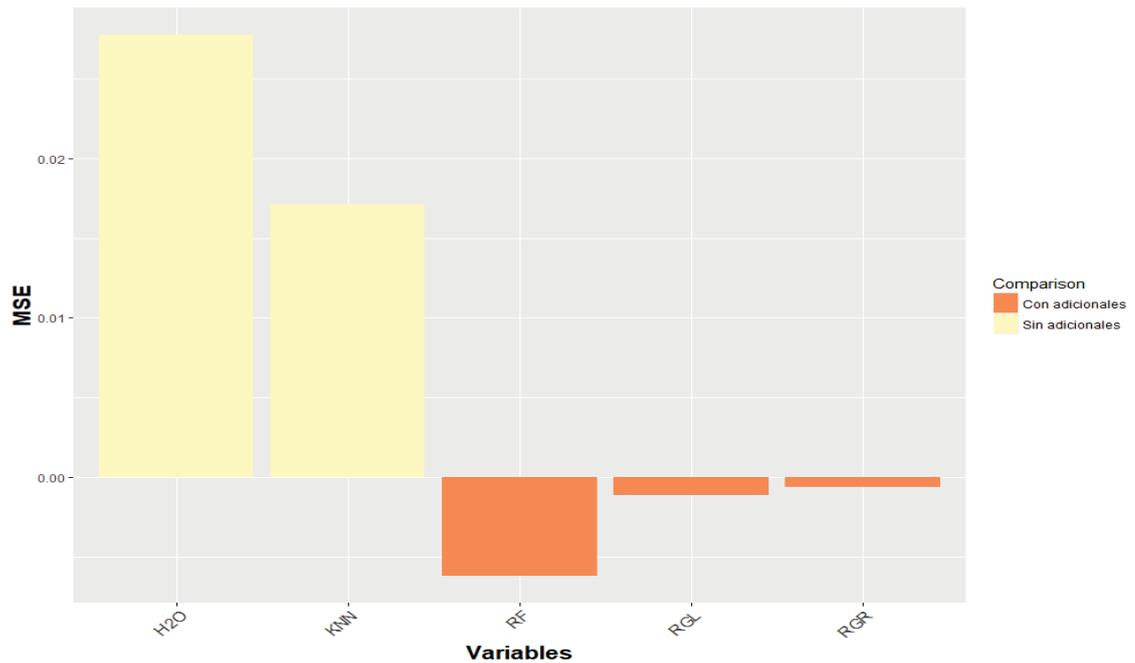


Figura 27: Diferencia de error cuadrático medio entre usar (rojo) y no usar (amarillo) las combinaciones de jugadores y la variable *clutch time* (rendimiento defensivo).

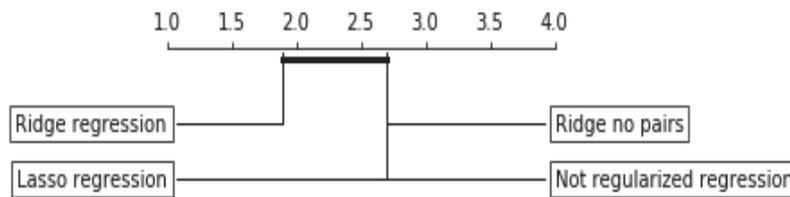


Figura 28: Test de Friedman y análisis post-hoc entre las regresiones para el ECM del rendimiento defensivo.

Por tanto, para el rendimiento defensivo el método que mejores resultados obtiene es la regresión lineal *ridge*, al igual que ha ocurrido en el caso del rendimiento ofensivo. Consigue el mejor registro en términos de error cuadrático medio. Como ya se ha mencionado anteriormente, facilita la interpretación del modelo y permite obtener una mayor cantidad de información relevante.

Al igual que se ha realizado en el modelo ofensivo, para ver si estas diferencias que se han demostrado en la comparación de los modelos son realmente significativas, se ha utilizado el test de Friedman seguido por el análisis post-hoc. Primero, se han realizado los tests entre los distintos tipos de regresiones. Los resultados obtenidos se pueden apreciar en la Figura 28. Si se interpreta el resultado, éste muestra que pese a que la que mejores resultados devuelve es la regresión lineal *ridge*, no hay diferencias significativas entre ninguna de las regresiones.

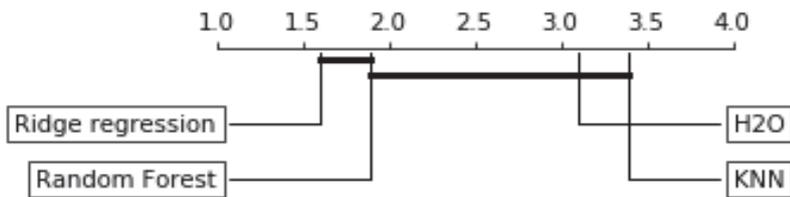


Figura 29: Test de Friedman y análisis post-hoc entre la regresión *ridge* y los métodos de aprendizaje automático para el ECM del rendimiento defensivo.

Los resultados de la comparación entre la regresión lineal *ridge* con combinaciones y *clutch time* y los métodos de aprendizaje automático se muestran en la figura 29. En este caso, el resultado es parecido al de las regresiones, ya que aunque la regresión *ridge* es mejor, no parecen existir diferencias significativas entre los distintos métodos.

En resumen, tras analizar y comparar el método de regresión lineal *ridge*, es posible concluir que es el que mejores resultados devuelve en términos de error cuadrático medio. Aunque no se encuentren diferencias significativas entre los distintos métodos en el rendimiento defensivo, la regresión *ridge* es la que permite obtener mayor información del modelo. Por tanto, se demuestra que el método escogido en el estudio es el que mejores resultados devuelve y el más interpretable para medir el rendimiento defensivo de los jugadores del club Movistar Estudiantes.

4.4. Rendimiento Plus-Minus

La valoración *plus-minus* es una estimación para ver cuál es el rendimiento global del jugador, ya que combina sus esfuerzos ofensivo y defensivo. La estadística *plus-minus* se calcula haciendo una resta entre los puntos que consigue el equipo ofensivamente y los que recibe defensivamente mientras el jugador se encuentra en la cancha. Es una medida muy extendida globalmente, no sólo en el baloncesto, sino en otros deportes como el hockey sobre hielo, baseball y fútbol americano. En lo que respecta al baloncesto, la estadística está muy asentada en la NBA, extendiéndose poco a poco al resto de ligas con mayor relevancia a nivel mundial. Esta valoración es muy útil, ya que permite detectar comportamientos de jugadores que en primer momento pasaban desapercibidos. Por ejemplo, un jugador puede ser muy bueno de cara a canasta, pero a su vez puede estar repercutiendo muy negativamente en la defensa del equipo. Por tanto, mediante esta estadística se pueden detectar estos casos, haciendo posible ajustar los jugadores que deben jugar dependiendo de la situación en la que se encuentra el partido.

En esta sección se realiza un detallado análisis del rendimiento *plus-minus* del Estudiantes, mostrando los resultados obtenidos con la regresión lineal *ridge* y realizando una comparación con los diversos métodos mencionados anteriormente.

4.4.1. Resultados

La Figura 30 muestra los coeficientes de regresión obtenidos mediante la regresión lineal *ridge*. Al igual que en el rendimiento ofensivo, los coeficientes positivos indican una contribución positiva, ya que el jugador está siendo capaz de proporcionar un rendimiento global óptimo al equipo. Por el contrario, si éstos son negativos, están consiguiendo un rendimiento negativo, al no conseguir contribuir de forma eficaz en el equipo.

A continuación se muestran las conclusiones obtenidas de forma detallada, para cada uno de los tres grandes apartados analizables: rendimiento individual, rendimiento por combinaciones de jugadores y rendimiento en las situaciones decisivas del partido.

A continuación se agrupan las conclusiones obtenidas al analizar el rendimiento individual de los jugadores:

- El jugador que tiene el mejor rendimiento global en el equipo es Sutton. Al tratarse de un jugador veterano con una larga trayectoria, esto le puede estar permitiendo rendir a un nivel muy alto. Su rendimiento ofensivo y defensivo ha sido positivo en ambos casos, por lo que no es de extrañar que sea el mejor jugador de la plantilla en términos de eficiencia global. Puede ser que el juego que despliega el club Movistar Estudiantes también esté ayudando a que consiga este rendimiento, ya que puede ser que se adapte a la perfección a su estilo.

CvetkovicQ_ClutchTime	HakansonQ_ClutchTime	VicedoQ_ClutchTime	CookQ_SuttonQ	SuttonQ_LandsbergQ
0.056593291	0.046273613	0.024181832	0.013418732	0.013190500
SuttonQ	SavaneQ_ClutchTime	BrownQ	CookQ_BrownQ	CookQ_LandsbergQ
0.011341304	0.011253389	0.007915516	0.007097320	0.006225615
BrizuelaQ	CvetkovicQ	SavaneQ	CookQ	VicedoQ_CookQ
0.006125698	0.004661014	0.004090896	0.004002493	0.003142875
ArteagaQ_ClutchTime	LandsbergQ	BrownQ_ClutchTime	LandsbergQ_ClutchTime	VicedoQ
0.000000000	-0.001177120	-0.001326796	-0.002726413	-0.002781489
VicedoQ_BrownQ	HakansonQ	PenaQ	CookQ_MedleyQ	SuttonQ_ClutchTime
-0.003034680	-0.007605192	-0.008069793	-0.009725022	-0.010792452
ArteagaQ	MedleyQ	LandsbergQ_MedleyQ	CookQ_ClutchTime	MedleyQ_ClutchTime
-0.011038705	-0.013915892	-0.017417924	-0.025384071	-0.055352129
PenaQ_ClutchTime	BrizuelaQ_ClutchTime			
-0.098190078	-0.101762021			

Figura 30: Coeficientes de regresión del modelo *plus-minus* obtenidos mediante regresión lineal *ridge*.

- Los jugadores Brown, Brizuela, Cvetkovic, Savané y Cook también están consiguiendo un rendimiento global positivo para el equipo. Es una mezcla de jugadores jóvenes y veteranos, que están sabiendo interpretar muy bien el juego del equipo y sacar el máximo provecho de él. Brown ha mostrado un buen desempeño ofensivo y defensivo, por lo que no es extraño que se encuentre en este grupo. Cook, en cambio, ha obtenido un buen rendimiento ofensivo, pero no ha sido así en el ámbito defensivo. De todas formas, parece ser que su rendimiento global sí es positivo para el equipo. Por el contrario, Brizuela, Cvetkovic y Savané no han sido capaces de tener un buen rendimiento ofensivo, carencia que han sabido tapar eficazmente con su desempeño defensivo.
- Una vez más, es posible crear un quinteto de garantías con los mejores jugadores a nivel de rendimiento global. En este quinteto, Cook sería el base, mientras que Brizuela y Cvetkovic se alternarían las posiciones de escolta y alero. A su vez, Brown, Sutton y Savané serían las alternativas para ocupar los puestos de ala-pívot y pívot.
- En el lado opuesto, el jugador que consigue un peor rendimiento global es Medley. El resultado no es sorprendente, ya que tanto en ataque como en defensa su rendimiento no ha sido bueno. Parece ser que este jugador es el que más perjudica al equipo globalmente, por lo que puede ser conveniente reducir los minutos que este jugador disputa en los partidos.
- Los jugadores Landsberg, Vicedo, Hakanson, Peña y Arteaga son el resto de jugadores que no consiguen obtener un rendimiento positivo global. En el caso de Landsberg y Peña, su óptimo rendimiento ofensivo no les ha permitido subsanar sus carencias defensivas, por lo que su aportación global al equipo es negativa. El caso de Hakanson es el contrario, ya que pese a tener un buen desempeño defensivo, su bajo rendimiento ofensivo no le ha permitido tener una aportación global positiva. Vicedo y Arteaga han tenido un desempeño negativo tanto ofensiva como defensivamente, por lo que su presencia en este grupo no es sorprendente.

Las conclusiones que se obtienen a partir de las combinaciones de jugadores son las siguientes:

- La pareja Cook-Sutton es la combinación que obtiene un mayor rendimiento global para el equipo. Sutton es el jugador con mejor rendimiento global, mientras que Cook pese a no ser el mejor, también tiene un rendimiento positivo. Es posible que al tratarse de un base y un pívot, se entiendan muy bien en la cancha y esto les permita realizar jugadas juntos de forma exitosa. Se trata de una ventaja que el equipo debería tener en cuenta, ya que disponer de esta pareja en la cancha al mismo tiempo le permite obtener un muy buen rendimiento global.
- El resto de parejas con un desempeño global positivo son Sutton-Landsberg, Cook-Brown, Cook-Landsberg y Vicedo-Cook. Como ocurría en el apartado ofensivo, Cook parece ser el jugador que hace mejores a sus compañeros, ya que aparece en cuatro de las cinco mejores combinaciones de jugadores. Al tratarse de un base con dilatada experiencia en Europa, no sorprende que pese a su avanzada edad siga siendo capaz de mejorar el rendimiento de sus compañeros. Landsberg y Sutton también parecen ser jugadores clave, ya que aparecen en dos de las cinco mejores parejas. Es interesante que el equipo intente mantener a los jugadores de las distintas parejas con buen rendimiento el máximo número de minutos posible, realizando cambios en base a esta información para poder maximizar el desempeño del equipo.
- Si nos vamos al otro extremo, la pareja Landsberg-Medley es la que obtiene peores resultados. El modelo parece mostrar que pese al rendimiento ofensivo óptimo de Landsberg, al jugar al lado de Medley no se consigue obtener un buen desempeño global. Medley es el jugador que peor rendimiento individual proporciona al equipo, factor que parece estar penalizando a Landsberg. Se trata de una combinación de jugadores que es conveniente evitar, tratando de dar prioridad a otras parejas que ayuden a obtener un mejor rendimiento.
- Las combinaciones Vicedo-Brown y Cook-Medley son las otras dos que proporcionan un rendimiento global negativo. Una vez más, la presencia de Medley está teniendo un impacto negativo en el juego, lo que está llevando incluso a penalizar la efectividad de Cook, el jugador del equipo que más ayuda a sus compañeros a ser mejores. Por otro lado, parece ser que Vicedo y Brown no rinden bien juntos, pese a que el rendimiento global de Brown es positivo.

Por último, las conclusiones del rendimiento de los jugadores obtenidas a partir de las situaciones de mayor exigencia de los partidos son las siguientes:

- El jugador con el mejor rendimiento global en situaciones de máxima exigencia es Cvetkovic. Tras haber mostrado un buen rendimiento ofensivo en el *clutch time* pese a su juventud, su falta de eficacia defensiva no ha impedido que sea el jugador de todo la plantilla que obtiene el mejor desempeño.

- El resto de jugadores que obtienen un rendimiento global positivo en el *clutch time* son Hakanson, Vicedo y Savané. Al igual que Cvetkovic, Hakanson y Savané habían demostrado previamente su buen rendimiento ofensivo en estas situaciones, lo que les ha permitido tener una aportación global positiva. Por su parte, Vicedo era el único jugador del equipo que demostraba sus dotes defensivas cuando el partido está en el alambre, lo que le ha ayudado a pertenecer a este grupo.
- Por otro lado, el jugador que peor se adapta a estas situaciones es Brizuela. Como se había mostrado tanto en el rendimiento ofensivo como defensivo, Brizuela no es capaz de adaptarse bien a los momentos de *clutch time* del partido. Esto puede deberse a su juventud, ya que todavía tiene una corta carrera deportiva y no tiene la experiencia de otros jugadores en situaciones de extrema exigencia.
- Los jugadores Brown, Landsberg, Sutton, Cook, Medley y Peña tampoco son capaces de rendir a un buen nivel en el *clutch time*. El caso de Brown es el más curioso, ya que pese a tener un buen rendimiento ofensivo, sus carencias defensivas han pesado más, asignándole un rendimiento global negativo. El resto de jugadores habían demostrado su ineficacia en las situaciones de máxima exigencia, por lo que su presencia no resulta una sorpresa.

4.4.2. Comparación

Por último, se ha realizado una comparación del método de regresión lineal *ridge* con parejas y *clutch time* propuesto en este estudio con los otros tipos de regresión (regresión lineal Lasso regularizada y regresión sin regularización) y los métodos de MLR (H2O, random forest y k-nearest neighbors). Igual que en las secciones anteriores, se ha comparado si se obtienen mejores resultados con y sin las combinaciones y el *clutch time*. Las medidas utilizadas son las mismas que para el rendimiento ofensivo y defensivo, error cuadrático medio, el test de Friedman y el análisis post-hoc para ver si las diferencias que se aprecian son significativas.

El Cuadro 13 muestra el valor de error cuadrático medio que tienen los métodos si se incluyen las combinaciones de jugadores y el *clutch time*, mientras que el Cuadro 14 muestra los mismos resultados sin incluir estas variables. Como ya se ha realizado para los dos anteriores modelos, a continuación se agrupan las principales conclusiones obtenidas de la comparación de modelos:

- Se repite la tendencia que se ha seguido hasta ahora, siendo la regresión lineal *ridge* que incluye las combinaciones y el *clutch time* el método que obtiene un mejor resultado en términos de error cuadrático medio.
- A su vez, se vuelve a ver que la regularización está permitiendo obtener mejores resultados en la regresión lineal.

- La regresión lineal Lasso regularizada obtiene unos valores muy cercanos a los obtenidos por la regresión *ridge*, pero no los consigue superar.
- Se vuelve a apreciar que la inclusión de las combinaciones de jugadores y la variable psicológica *clutch time* está permitiendo obtener mejores resultados, proporcionando mayor información y ganando interpretabilidad.
- De los métodos de MLR utilizados en la comparación, el random forest vuelve a ser el que mejores resultados devuelve.
- Si no se incluyen las combinaciones ni el *clutch time*, la regresión lineal *ridge* consigue obtener los mejores resultados.
- La red neuronal H2O y el algoritmo k-nearest neighbors no han conseguido competir con el resto de métodos en términos de error cuadrático medio, obteniendo unos resultados claramente peores.

Cuadro 13: Comparación de la regresión lineal *ridge* con métodos MLR para el rendimiento *plus-minus*.

Algoritmo	Error cuadrático medio
RGR	0.27619 ± 0.5730
RGL	0.27679 ± 0.5750
RF	0.27849 ± 0.5759
H2O	0.42068 ± 0.7520
KNN	0.35549 ± 0.6968

Cuadro 14: Resultados *plus-minus* obtenidos sin combinaciones de jugadores ni *clutch time*.

Algoritmo	Error cuadrático medio
RGR	0.27742 ± 0.5741
RGL	0.27751 ± 0.5765
RGSR	0.28775 ± 0.5905
RF	0.29567 ± 0.5993
H2O	0.34307 ± 0.6428
KNN	0.39104 ± 0.7147

En la Figura 31 se puede apreciar la diferencia de error cuadrático medio entre usar (rojo) y no usar (amarillo) las combinaciones de jugadores y la variable *clutch time*. Al igual que en los anteriores modelos, se vuelven a apreciar diferencias. Pese a que no se puede ver con claridad, existen diferencias mínimas en las dos regresiones (*ridge* y Lasso), mientras que la diferencia se convierte en más notoria para los algoritmos random forest y k-nearest neighbors. Por tanto, una vez más se demuestra

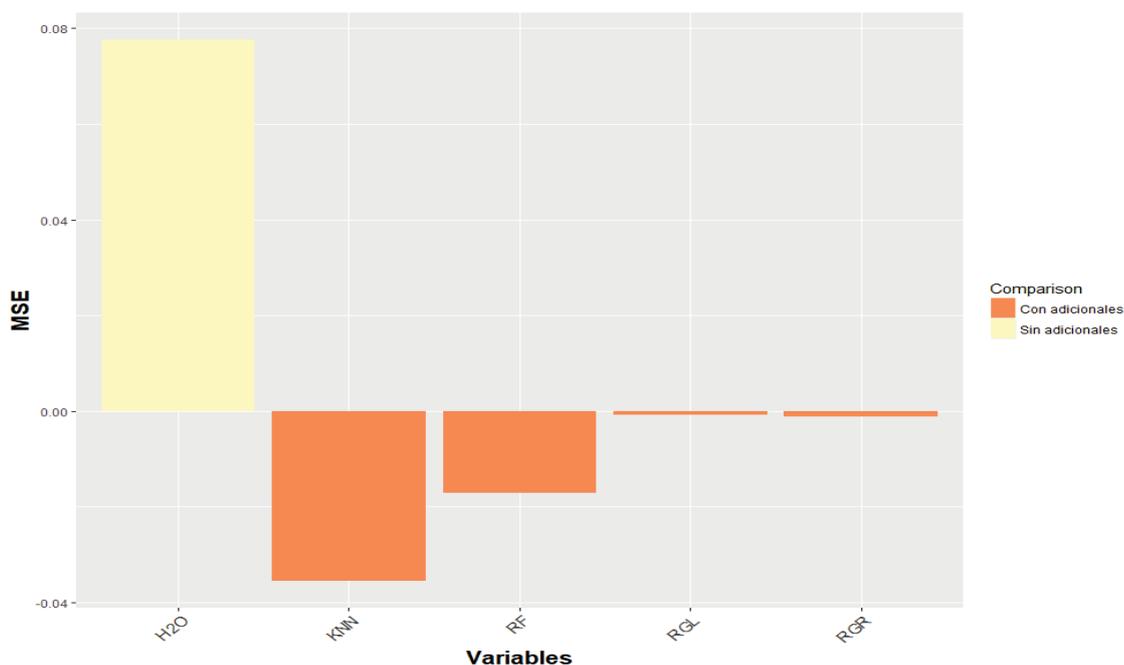


Figura 31: Diferencia de error cuadrático medio entre usar (rojo) y no usar (amarillo) las combinaciones de jugadores y la variable *clutch time* (rendimiento *plus-minus*).

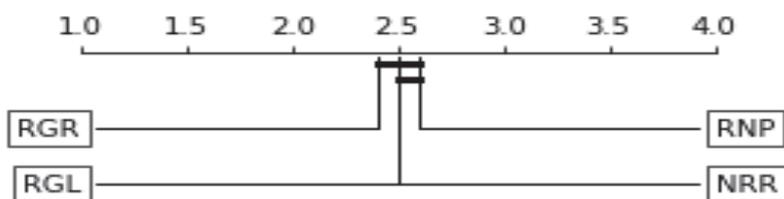


Figura 32: Test de Friedman y análisis post-hoc entre las regresiones para el ECM del rendimiento *plus-minus*.

que el añadir estas nuevas variables está siendo beneficioso para el modelo, ya que está ayudando en mayor o menor medida a ajustar las predicciones y obtener mayor información, además de proporcionar un grado mayor de interpretabilidad.

Por tanto, para el rendimiento *plus-minus* de los jugadores el método que mejores resultados obtiene es la regresión lineal *ridge*, al igual que ha ocurrido en los dos casos anteriores: rendimiento ofensivo y defensivo. Consigue el mejor registro en términos de error cuadrático medio. Como ya se ha mencionado en más de una ocasión, facilita la interpretación del modelo y permite obtener una mayor cantidad de información relevante.

Por último, como se ha realizado en los dos casos anteriores, para ver si estas diferencias que se han demostrado en la comparación de los modelos son realmente significativas, se ha utilizado el test de Friedman seguido por el análisis post-hoc.

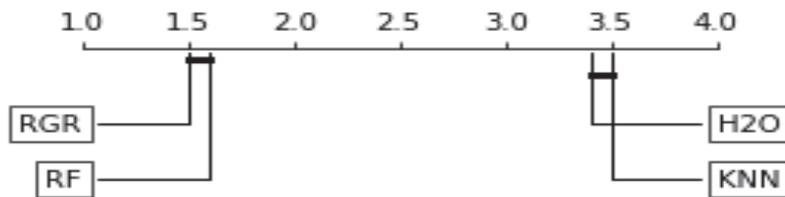


Figura 33: Test de Friedman y análisis post-hoc entre la regresión *ridge* y los métodos de aprendizaje automático para el ECM del rendimiento *plus-minus*.

Primero, se han realizado los tests entre los distintos tipos de regresiones. Los resultados obtenidos se pueden apreciar en la Figura 32. Del mismo modo que ocurría con el modelo de rendimiento defensivo, no existen diferencias significativas entre las regresiones, pese a que la regresión *ridge* devuelve los mejores resultados.

La Figura 33 muestra la comparación entre la regresión lineal *ridge* y los métodos de MLR. En este caso, no existen diferencias significativas entre la regresión y el random forest, así como entre la red neuronal y el k-nearest neighbors. Sin embargo, sí existen diferencias significativas entre los dos primeros métodos y los dos siguientes.

En resumen, tras analizar y comparar el método de regresión lineal *ridge*, es posible concluir que se trata del método que mejores resultados devuelve en términos de error cuadrático medio. Aunque no se han encontrado diferencias significativas entre los distintos métodos en el rendimiento *plus-minus*, la regresión *ridge* es la que permite obtener un mayor grado de interpretabilidad. Por tanto, se demuestra una vez más que el método escogido en el estudio es el que mejores resultados devuelve y el más interpretable para medir el rendimiento *plus-minus* de los jugadores del club Movistar Estudiantes.

5. CONCLUSIONES Y TRABAJO FUTURO

5.1. Conclusiones

Tras analizar en profundidad el juego del club Movistar Estudiantes, en esta sección se agrupan las conclusiones más importantes obtenidas a lo largo del proyecto, tanto del estilo de juego como de los jugadores que componen la plantilla y las conclusiones generales obtenidas.

5.1.1. Modelización del estilo de juego

Las conclusiones principales obtenidas tras la modelización del estilo de juego del equipo son las siguientes:

- El ataque del club Movistar Estudiantes está claramente orientado a implantar un ritmo de juego muy rápido en el partido. El equipo recurre a jugadas rápidas en mucha mayor medida que sus rivales, tratando de acabar las jugadas en tiro en el mínimo tiempo posible. El objetivo de este estilo es tratar de maximizar los puntos que consigue el equipo, siguiendo la filosofía de que cuantos más tiros realice el equipo mayor será la cantidad de puntos que pueda conseguir. Siguiendo en esa línea, también se ha detectado una tendencia a priorizar el tiro de tres puntos por encima del resto. Al tratarse del tipo de lanzamiento que más puntos puede conseguir, un aumento del número de tiros debe traducirse en un incremento de los puntos conseguidos por el equipo.
- Se aprecia una disminución significativa del rendimiento ofensivo del equipo en el segundo periodo de los partidos. Esta disminución puede ser debida a la incorporación de jugadores de rotación, que no poseen las características de los jugadores titulares del equipo. Es interesante para el club Movistar Estudiantes tratar de aumentar la eficiencia de sus jugadores en este cuarto, ya que esta falta de acierto puede ser una losa demasiado grande para remontar en los siguientes dos periodos.
- Las jugadas que implican bloqueos de los equipos rivales son las que más penalizan el rendimiento defensivo del equipo. Parece ser que el Estudiantes no está siendo capaz de dar una respuesta adecuada a estas situaciones, permitiendo que el rival consiga una gran cantidad de puntos. Para tratar de corregir este problema, podría ser interesante que el equipo adoptara una filosofía defensiva de cambios automáticos veloces, lo que le permitiría tener un hombre siempre encima del jugador rival. Esta solución podría ser beneficiosa para el equipo, siempre y cuando exista una buena comunicación entre los jugadores de la plantilla.
- Los rivales del Estudiantes priorizan las jugadas largas, tratando de mantener el control del partido y sin abusar del juego rápido. Defensivamente, el equipo

tiene que dar un paso adelante para minimizar el acierto rival, mejorando la comunicación entre los compañeros y realizando las ayudas necesarias para defender a los rivales. Esto les permitiría mejorar su defensa estática, aspecto fundamental debido al alto uso por parte de los rivales de las jugadas que consumen más tiempo de posesión.

- El club Movistar Estudiantes trata de imprimir su propio sello en la liga, defendiendo un estilo de juego claramente ofensivo, tratando de maximizar al máximo los puntos conseguidos aunque ello le lleva a recibir una mayor cantidad de puntos.

5.1.2. Modelización de los jugadores

Las conclusiones principales obtenidas tras la modelización de los jugadores del equipo son las siguientes:

- El jugador que proporciona un mayor rendimiento ofensivo individual es Peña. Pese a no haber disputado una gran cantidad de minutos, su desempeño en ataque ha sido el más destacable. Por otro lado, el modelo ha demostrado que el jugador que mejor se compenetra con sus compañeros es Cook. Se trata del jugador que más veces aparece en las parejas que optimizan el juego ofensivo del equipo, por lo que resulta clave a la hora de hacer mejores a sus compañeros. Finalmente, se ha podido comprobar que el jugador que mejor adapta su juego ofensivo a las situaciones de *clutch time* es Hakanson. Es un jugador que pese a su juventud, no parece temblarle la muñeca en situaciones decisivas.
- Pese a que es un jugador que se crece en los instantes decisivos del partido, el rendimiento ofensivo individual de Hakanson es el que peores resultados obtiene. Además, cuando él y Vicedo se encuentran juntos en la cancha, el rendimiento del equipo también es negativo, siendo la pareja que peor funciona. En el *clutch time*, el jugador que peor se adapta es Brizuela. Parece ser que en este caso la juventud y la falta de experiencia de los tres pueden ser los factores que estén influyendo en su eficiencia.
- Brizuela es el jugador que obtiene el mejor rendimiento defensivo de todo el equipo, imprimiendo un nivel más de intensidad que sus compañeros. La pareja Vicedo-Sutton está consiguiendo los mejores registros defensivos, complementándose muy bien cuando juegan juntos. Si hablamos de situaciones de *clutch time*, Vicedo es el único jugador que demuestra estar a la altura, siendo el único con un desempeño positivo en estas situaciones.
- El jugador que había mostrado el mejor rendimiento ofensivo (Peña) es ahora el que peores resultados obtiene en el ámbito defensivo, donde no está siendo capaz de estar a la altura. Por otro lado, Medley es el jugador que parece estar influyendo negativamente en la defensa del equipo, ya que aparece en prácticamente todas las combinaciones defensivas que no están rindiendo bien. Aunque

no es el peor jugador defensivo, su presencia sí que disminuye notablemente los registros de sus compañeros. Todo el equipo salvo Vicedo rinden mal en defensa en los momentos de *clutch time*, por lo que parece ser que el equipo tiene una tarea pendiente en esta faceta.

- En lo que respecta al rendimiento *plus-minus*, Sutton es el jugador que mejores registros obtiene, lo que demuestra que su experiencia le puede estar ayudando a rendir a un gran nivel. Cook, Landsberg y Sutton son los jugadores que consiguen maximizar el rendimiento *plus-minus* de sus compañeros, siendo piezas fundamentales dentro del equipo. En el *clutch time*, Cvetkovic es el jugador que mejor se adapta a las situaciones decisivas de los encuentros.
- Medley es el jugador con peores registros *plus-minus*. Además, su combinación con Landsberg es la que proporciona un rendimiento más bajo. Parece ser que la inclusión de Medley en los quintetos está teniendo un efecto negativo, siendo el jugador que menos parece adaptarse al estilo de juego del club Movistar Estudiantes. Por otro lado, Brizuela es el jugador que peor adapta su juego a las situaciones *clutch time* de los partidos.

5.1.3. Conclusiones generales

- El uso de redes bayesianas para modelizar el juego del equipo permite realizar un análisis de las independencias condicionales que se encuentran en la red, lo que proporciona una gran cantidad de información sobre el juego del equipo.
- Las relaciones obtenidas tras analizar las redes bayesianas ofensiva y defensiva son lógicas desde el punto de vista baloncestístico. Además, las conclusiones que se han conseguido se han contrastado con el entrenador asistente José Ángel Samaniego, validando su veracidad.
- Se ha demostrado que la inclusión de las variables de combinaciones de jugadores y *clutch time* ha permitido mejorar los resultados obtenidos. Además, el uso de éstas proporciona un grado más alto de información al entrenador del equipo, ya que se muestra como de bien funcionan los jugadores juntos y como se adaptan a situaciones de máxima exigencia en los minutos de la verdad de los partidos.
- La regresión lineal *ridge* con combinaciones de jugadores y la variable *clutch time* ha obtenido los mejores resultados en términos de error cuadrático medio, por encima del resto de métodos utilizados en la comparación. El método MLR con mejores resultados ha sido el random forest, método mucho menos interpretable que el propuesto en este trabajo.

5.2. Trabajo futuro

Por último, se muestran las ideas de trabajo futuro que han surgido tras la realización de este proyecto:

- Se ha presentado un artículo a la conferencia ECML (*European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*), el cual se encuentra en proceso de aceptación.
- Realizar un análisis de cada tipo de juego, diferenciando entre los distintos tipos de jugadas que se realizan en cada uno de ellos. Una vez realizada esa distinción, se crearía un modelo que permitiera determinar cuál de ellas resulta más determinante en el partido e incrementa la efectividad del equipo.
- Se podría realizar una distinción entre los tipos de tiros que se realizan dentro de los tiros de dos puntos, diferenciando entre bandejas, mates, tiros de media distancia... A su vez, se realizaría el mismo proceso con los tiros de tres, distinguiendo entre las posiciones desde las que se puede lanzar el tiro. De esta forma, se podría observar qué situación de tiro tiene una mayor probabilidad de conseguir puntos.
- Sería interesante comprobar el grado en el que afecta la presencia del quinteto titular en la cancha, ya que los resultados obtenidos tras analizar el modelo de periodos mostraban un decremento en la producción del equipo en el segundo periodo. De esta forma, se podría apreciar la variación de los puntos esperados del equipo con el quinteto dentro y fuera de la cancha.
- Si se analiza como afecta el tipo de juego que se utiliza a la posición del jugador, podría obtenerse información sobre qué estilo de juego beneficia a ciertos jugadores, siendo posible adaptar los sistemas de juego en base a los resultados obtenidos. A su vez, esto permitiría poder elaborar estrategias de juego para cada situación del partido, potenciando a los jugadores exteriores o interiores dependiendo de la situación.
- Se podrían crear modelos más específicos, centrados en potenciar una estadística en concreto. De esta forma, sería posible disponer de quintetos adaptados para cada una de las situaciones que se puedan dar en el partido, obteniendo siempre el máximo beneficio.

6. BIBLIOGRAFÍA

- Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723, 1974.
- Andrews, D. L. Whither the NBA, whither America? *Peace Review*, 11(4):505–510, 1999.
- Barron, A.; Birgé, L., y Massart, P. Risk bounds for model selection via penalization. *Probability Theory and Related Fields*, 113(3):301–413, 1999.
- Berri, D. J. Who is 'most valuable'? Measuring the player's production of wins in the national basketball association. *Managerial and Decision Economics*, 20(8):411–427, 1999.
- Bouckaert, R. Probabilistic network construction using the minimum description length principle. In Clarke, Michael; Kruse, Rudolf, y Moral, Seraffín, editors, *Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 41–48. Springer, 1993.
- Cheng, J.; Greiner, R.; Kelly, J.; Bell, D., y Liu, W. Learning Bayesian networks from data: An information-theory based approach. *Artificial Intelligence*, 137(1-2):43–90, 2002.
- Constantinou, A. C.; Fenton, N. E., y Neil, M. pi-football: A Bayesian network model for forecasting association football match outcomes. *Knowledge-Based Systems*, 36:322–339, 2012.
- Cooper, G. F. y Herskovits, E. A Bayesian method for constructing Bayesian belief networks from databases. In *Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence*, pages 86–94. Morgan Kaufmann Publishers Inc., 1991.
- Cooper, G. F. y Herskovits, E. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9(4):309–347, 1992.
- Cutler, A.; Cutler, D., y Stevens, J. Random forests. In *Ensemble Machine Learning - EML*, volume 45, pages 157–176, 2011.
- de Saá Guerra, Y.; Martín-González, J.; Sarmiento, S.; Rodríguez-Ruiz, D.; García-Rodríguez, A., y García-Manso, J.M. A model for competitiveness level analysis in sports competitions: Application to basketball. *Physica A Statistical and Theoretical Physics*, 391:2997–3004, 2012.
- Deshpande, S. y Jensen, T. Estimating an NBA players impact on his teams chances of winning. *Journal of Quantitative Analysis in Sports*, 12(2):51–72, 2016.
- Domingos, P. y Pazzani, M. On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning*, 29(2):103–130, 1997.

- Friedman, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32(200): 675–701, 1937.
- Friedman, N.; Geiger, D., y Goldszmidt, M. Bayesian network classifiers. *Machine Learning*, 29(2):131–163, 1997.
- Fung, R. M. y Chang, K. Weighing and integrating evidence for stochastic simulation in Bayesian networks. In *Proceedings of the Fifth Annual Conference on Uncertainty in Artificial Intelligence*, pages 209–220, 1990.
- García, S. y Herrera, F. An extension on "Statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons. *Journal of Machine Learning Research*, 9(12):2677 – 2694, 2008.
- Godfrey, K. Simple linear regression in medical research. *New England Journal of Medicine*, 313(26):1629–1636, 1985.
- Hastie, T.; Tibshirani, R., y Friedman, J. *The Elements of Statistical Learning*. Springer, 2001.
- Havard, R. y Oyvind, S. Prediction and retrospective analysis of soccer matches in a league. *Journal of the Royal Statistical Society*, 49(3):399–418, 2000.
- Heckerman, D.; Geiger, D., y Chickering, D. M. Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, 20(3):197–243, 1995.
- Heeren, D. *The Basketball Abstract*. Los Angeles Sports Distributors, 1988.
- Heller, K. A. y Ghahramani, Z. Bayesian hierarchical clustering. In *Proceedings of the 22Nd International Conference on Machine Learning*, pages 297–304, 2005.
- Henrion, M. Propagating uncertainty in Bayesian networks by probabilistic logic sampling. In *Uncertainty in Artificial Intelligence*, volume 5, pages 149 – 163. 1988.
- Hollinger, J. *Pro Basketball Forecast*. Potomac Books Inc, Dulles, VA, 2005.
- Humphreys, B. y Johnson, C. The effect of superstar players on game attendance: Evidence from the NBA. *SSRN Electronic Journal*, 2017.
- James, G.; Witten, D.; Hastie, T., y Tibshirani, R. *An Introduction to Statistical Learning: With Applications in R*. Springer, 2014.
- Kampakis, S. *Predictive Modelling of Football Injuries*. PhD thesis, Cornell University, New York, 2016.
- Koller, D. y Friedman, N. *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning*. The MIT Press, 2009.

- Kubatko, J.; Oliver, D.; Pelton, K., y T Rosenbaum, D. A starting point for analyzing basketball statistics. *Journal of Quantitative Analysis in Sports*, 3:1–24, 2007.
- Larrañaga, P.; Kuijpers, C. M. H.; Murga, R. H., y Yurramendi, Y. Learning Bayesian network structures by searching for the best ordering with genetic algorithms. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 26(4):487–493, 1996a.
- Larrañaga, P.; Poza, M.; Yurramendi, Y.; Murga, R. H., y Kuijpers, C. M. H. Structure learning of Bayesian networks by genetic algorithms: A performance analysis of control parameters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):912–926, 1996b.
- Lauritzen, S. L. y Spiegelhalter, D. J. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society*, 50(2):157–224, 1988.
- Lee, N. y Kitani, K. M. Predicting wide receiver trajectories in american football. *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–9, 2016.
- Loeffelholz, B.; Bednar, E., y Bauer, K. Predicting NBA games using neural networks. *Journal of Quantitative Analysis in Sports*, 5:7–7, 2009.
- Macdonald, B. Adjusted plus-minus for NHL players using ridge regression with goals, shots, fenwick, and corsi. *Journal of Quantitative Analysis in Sports*, 8, 2012.
- Mahdi, R. y Mezey, J. Sub-local constraint-based learning of Bayesian networks using a joint dependence criterion. *Journal of Machine Learning Research*, 14: 1563–1603, 2013.
- Mai, Y. H. The analysis of the statistics of Asian women basketball championships in Sandai. Master thesis, National Taiwan Sport University, 2004.
- Maxwell, D. Learning bayesian networks is np-complete. 112, 2000.
- Meek, C. Causal inference and causal explanation with background knowledge. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 403–410. Morgan Kaufmann Publishers Inc., 1995.
- Miller, R.; Schwarz, H., y Talke, I.S. Forecasting sports popularity: Application of time series analysis. *Academic Journal of Interdisciplinary Studies*, 6(2):75–82, 2017.
- Minsky, M. Steps toward artificial intelligence. In *Computers and Thought*, pages 406–450. McGraw-Hill, 1961.

- Moxley, J. H. y Towne, T. J. Predicting success in the national basketball association: Stability and potential. *Psychology of Sport and Exercise*, 16:128–136, 2015.
- Murphy, K. P. *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.
- Naseem, I.; Togneri, R., y Bennamoun, M. Linear regression for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11):2106–2112, 2010.
- Nevill, A.; Balmer, N.J, y Williams, A. The influence of crowd noise and experience upon refereeing decisions in football. *Psychology of Sport and Exercise*, 3:261–272, 2002.
- Niemi, J. Evaluating individual player contributions in basketball. *Journal of Sport Management*, page 4914,Äi4923, 2010.
- Omidiran, D. A new look at adjusted plus/minus for basketball analysis. MIT Sloan Sports Analytics Conference, 2011.
- Oña, J.; López, G.; Mujalli, R., y Calvo, F. J. Analysis of traffic accidents on rural highways using latent class clustering and bayesian networks. *Accident Analysis and Prevention*, 51:1 – 10, 2013.
- Pan, Y. F. Probit regression model to predict results of NBA basketball games. Master thesis, National Hsinchu University of Education, 2010.
- Pearl, J. Fusion, propagation, and structuring in belief networks. *Artificial Intelligence*, 29(3):241–288, 1986.
- Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., 1988.
- Rein, R. y Memmert, D. Big data and tactical analysis in elite soccer: Future challenges and opportunities for sports science. volume 1410, 2016.
- Rissanen, J. Stochastic complexity and modeling. *The Annals of Statistics*, 14(3): 1080–1100, 1986.
- Robinson, R. W. Counting unlabeled acyclic digraphs. In Little, Charles H. C., editor, *Combinatorial Mathematics V*, pages 28–43. Springer, 1977.
- Schwarz, G. Estimating the dimension of a model. *The Annals of Statistics*, 6(2): 461–464, 1978.
- Sesen, M.; Nicholson, A.; Banares-Alcantara, R.; Kadir, T., y Brady, M. Bayesian networks for clinical decision support in lung cancer care. *PloS One*, 8:e82349, 2013.

- Shenoy, P. y Shafer, G. Axioms for probability and belief-function propagation. 4: 169–198, 1990.
- Sill, J. Improved NBA adjusted +/- using regularization and out-of-sample testing. MIT Sloan Sports Analytics Conference, 2010.
- Sipko, M. Machine learning for the prediction of professional tennis matches. Bachelor’s thesis, Imperial College London, 2015.
- Spirtes, P.; Glymour, C., y Scheines, R. *Causation, Prediction, and Search*. Springer, 2000.
- Staw, B. y Hoang, H. Sunk costs in the NBA: Why draft order affects playing time and survival in professional basketball. *Administrative Science Quarterly*, 40:474, 1995.
- Stein, M.; Janetzko, H.; Seebacher, D.; Jager, A.; Nagel, M.; Holsch, J.; Kosub, S.; Schreck, T.; Keim, D., y Grossniklaus, M. How to make sense of team sport data: From acquisition to data modeling and research aspects. *Data*, 2:2, 2017.
- Tsamardinos, Ioannis; Brown, Laura E., y Aliferis, Constantin F. The max-min hill-climbing bayesian network structure learning algorithm. *Machine Learning*, 65(1):31–78, 2006.
- Verma, T. y Pearl, J. Equivalence and synthesis of causal models. In *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*, pages 255–270. Elsevier Science Inc., 1991.
- Wai, L. y F., Bacchus. Learning Bayesian belief networks: An approach based on the MDL principle. *Computational Intelligence*, 10(3):269–293, 1994.
- Winston, W. L. *How Gamblers, Managers, and Sports Enthusiasts Use Mathematics in Baseball, Basketball, and Football*. Princenton, Indiana University’s Kelley School of Business, 2009.
- Wong, M. y Leung, K. An efficient data mining method for learning Bayesian networks using an evolutionary algorithm-based hybrid approach. *IEEE Transactions on Evolutionary Computation*, 8(4):378–404, 2004.
- Zak, T. A.; Huang, C. J., y Siegfried, J. J. Production Efficiency: The Case of Professional Basketball. *The Journal of Business*, 52(3):379–392, 1979.