

Data: 24/05/2020

Assignatura: Models Lineals – L2

Autors: Guillem Miralles | Joan Llavata | Miguel Payá



# TREBALL DE PREDICCIÓ DEL MILLOR QUINTET DE LA TEMPORADA EN L'NBA



GUILLEM MIRALLES  
JOAN LLAVATA  
MIGUEL PAYÁ



## ÍNDEX DE CONTINGUTS:

1.	RESUMEN: OBJECTIUS I RESULTATS	2
1.1.	OBJECTIU:	3
1.2.	RESULTATS	3
2.	INTRODUCCIÓ AL BANC DE DADES:	3
3.	PROCÈS:	6
3.1.	PROCÈS CREACIÓ DE MODEL AMB DADES CONEGUDES	6
3.1.1.	MÈTODE LDA	8
3.1.2.	MÈTODE QDA	8
3.1.3.	MÈTODE GLM	8
3.2.	PROCÈS CREACIÓ DE MODEL AMB LES DADES D'AQUEST ANY	9
4.	RESULTATS I CONCLUSIONS	10
4.1.	RESULTATS MODEL ON CONEGUEM LES DADES I CONCLUSIONS	10
4.2.	RESULTATS MODEL PREDINT LES DADES D'AQUEST ANY I CONCLUSIONS	16
5.	CONCLUSIONS FINALS	16

## 1. RESUMEN: OBJECTIUS I RESULTATS

-El millor quintet de l'any és una condecoració anual atorgada per l'NBA als millors jugadors de la temporada. La votació la realitza un grup de periodistes i locutors esportius dels Estats Units i el Canadà. L'equip ha estat escollit a cada temporada de l'NBA, des de la seua inauguració el 1946. El premi consta de tres quintets compostos per un total de 15 jugadors, cinc a cada equip. Originalment es componia de dos equips, però el 1988 s'augmentà a tres.

Els jugadors reben cinc punts per cada vot en el primer equip, tres punts per cada vot en el segon equip, i un punt per cada vot en el tercer equip. Els cinc jugadors amb major nombre de punts totals entren en el primer equip, amb els següents cinc jugadors integren el segon equip i el mateix amb el tercer. Existeix una restricció per posició. En cada votació de 5 jugadors (de cada quintet), es voten 2 jugadors que siguin guard, al nostre data.frame "PG" i "SG"; els altres 2 que siguin forward, "SF" i "PF" i l'últim jugador que siga center "C".

Bàsicament són els 15 millors jugadors de la temporada. Ens fixarem en les estadístiques que fan aquests 15 jugadors (en totes les temporades), i amb aquesta informació, tenint les estadístiques de tots els jugadors de les últimes 35 temporades, intentarem saber quin estaran al quintet i quins no dels últims anys.



## 1.1. OBJECTIU:

-El nostre objectiu és realitzar un model que siga capaç de predir aquests 15 jugadors. Per a fer açò hem de fer un model de regressió logística múltiple on, segons les estadístiques de cada jugador, el model li done una probabilitat de pertànyer als millors quintets. Hem de tindre en compte la restricció que sols poden haver determinats jugadors de cada posició. Tenim una base de dades, explicada més específicament al punt 2. INTRODUCCIÓ AL BANC DE DADES on tindrem tota la informació per a crear aquest model. Comprovarem els resultats obtinguts pel model en algunes de les últimes temporades on ja sabem els resultats, posteriorment, ho provarem amb les estadístiques que tenim fins ara per a predir els quintets d'aquest any.

## 1.2. RESULTATS:

-Obtenim un bon model amb el mètode GLM, que prediu amb una gran precisió els jugadors que hi ha al quintet que volem saber. Sols podem valorar amb seguretat els resultats del model que prediu els valors per a les dades dels anys 2015, 2016 i 2017. De forma subjectiva amb el que nosaltres sabem sobre la competició, també valorem molt positivament els resultats del quintet d'aquest any, encara que no podem comprovar-ho.

## 2. INTRODUCCIÓ AL BANC DE DADES:

- La nostra informació per a la realització d'aquest model es divideix en dos bases de dades:
  - Una base de dades on tenim tota la **informació desde 1980 fins al 2017** amb totes les estadístiques del jugador per cada temporada. A més li introduïm la variable quintet que ens aportara informació sobre si el jugador està o no al quintet de la temporada.
  - I altra on tenim la informació de les **dades de l'actual temporada**.

En la primera base de dades tenim les estadístiques de cada jugador en una temporada determinada. Les dades són les següents:

- <b>Year:</b> la temporada a la que pertanyen les estadístiques del jugador
- <b>Player:</b> nom del jugador
- <b>Pos:</b> posició que ocupa el jugador
- <b>Age:</b> l'edat del jugador
- <b>Tm:</b> equip al que pertany.
- <b>G:</b> el total de partits que ha jugat.
- <b>GS:</b> el total de partits que ha sortit de titular.
- <b>MP:</b> total de minuts jugats.
- <b>PER:</b> Valoració per minut del jugador en pista.
- <b>TS%:</b> Probabilitat d'anotar un punt per cada tir realitzat
- <b>3PAR:</b> Percentatge de tirs intentats de tres per tirs de camp intentats

Data: 24/05/2020

Assignatura: Models Lineals – L2 (Ciència de Dades)

Autors: Guillem Miralles | Joan Llavata | Miguel Payá



-FTr: Nombre de tirs lliures intentats per tir de camp intentat.
-ORB%: Tant per cent dels rebots ofensius obtesos
-DRB%: Tant per cent dels rebots defensius obtesos
-TRB%: Tant per cent dels rebots totals obtesos
-AST%: De tots els punts anotats de l'equip, tant per cent dels que han segut generats per una assistència del jugador en qüestió
-STL%: Tant per cent de robos de un jugador per jugada defensiva del equip mentre està a pista
-BLK%: Tant per cent de tapons de un jugador per jugada defensiva del equip mentre està a pista
-TOV%: Tant per cent de pèrdues de baló per cada volta que un jugador té el baló
-USG%: Tant per cent de jugades ofensives de l'equip acabades per el jugador en qüestió
-OWS: Victories que ha atorgat aquest jugador a l'equip únicament ofensivament
-DWS: Victories que ha atorgat aquest jugador a l'equip únicament defensivament
-WS: Victories que ha atorgat aquest jugador a l'equip en total.
-WS/48: Victories que ha atorgat aquest jugador a l'equip per partit
-OBPM: Estadística avançada que valora la contribució d'un jugador ofensivament per 100 possessions
-DBPM: Estadística avançada que valora la contribució d'un jugador defensivament per 100 possessions
-BPM: Estadística avançada que valora la contribució d'un jugador per 100 possessions
-VORP: Estadística avançada que trata el valor que aporta un jugador al seu equip.
-FG: total de tirs anotats.
-FGA: total de tirs realitzats.
-FG%: percentatge de tirs anotats.
-3P: total de tirs de 3 anotats.
-3PA: total de tirs de 3 realitzats.
-3P%: percentatge de tirs de 3 anotats.
-2P: total de tirs de 2 anotats.
-2PA: total de tirs de 2 realitzats
-2P%: percentatge de tirs de 2 anotats
-eFG%: Probabilitat d'anotar un punt per cada tir de camp realitzat
-FT: total de tirs lliures anotats.
-FTA: total de tirs lliures realitzats.
-FT%: percentatge de tirs lliures anotats.
-ORB: total de rebots ofensius capturats.
-DRB: total de rebots defensius capturats.
-TRB: total de rebots capturats.
-AST: total de assistències donades.
-STL: total de de robos realitzats.
-BLK: total de tapons col·locats.
-TOV: total de pèrdues de baló.
-PF: total de faltes personals comeses.
-PTS: total de punts anotats.

Data: 24/05/2020

Assignatura: Models Lineals – L2 (Ciència de Dades)

Autors: Guillem Miralles | Joan Llavata | Miguel Payá



+ La variable que introduim nosaltres:

**-quintet:** Si el jugador està al millor quintet de la temporada (1 si si ho està i 0 si no).

-En la nostra segon base de dades (informació de la temporada actual) tenim menys variables que a la primera. Aquestes variables han sigut adaptades perquè siguin idèntiques a les anteriors. Per aquesta qüestió hem de realitzar un model similar al de la base de dades anterior, però menys precís degut a la falta de variables.

<b>-Player:</b> el nom del jugador.
<b>-Tm:</b> equip del jugador
<b>-Pos:</b> la posició que ocupa el jugador
<b>-Age :</b> l'edat del jugador.
<b>-G:</b> el total de partits que ha jugat.
<b>-USG% :</b> Tant per cent de jugades ofensives de l'equip acabades per el jugador en qüestió
<b>-TOV%:</b> Tant per cent de pèrdues de baló per cada volta que un jugador té el baló
<b>-FTA:</b> total de tirs lliures realitzats.
<b>-FT%:</b> percentatge de tirs lliures anotats.
<b>-2PA :</b> total de tirs de 2 anotats
<b>-2P%:</b> total de tirs de 2 anotats
<b>-PA :</b> total de tirs de 3 realitzats
<b>-3P%:</b> percentatge de tirs de 3 anotats eFG.
<b>-TS%:</b> Probabilitat d'antoar un punt per cada tir ralitzat
<b>-PTS:</b> total de punts anotats
<b>-AST%:</b> De tots els punts anotats de l'equip, tant per cent dels que han segut generats per una assistència del jugador en qüestió
<b>-TRB:</b> total de rebots capturats
<b>-AST:</b> total de assistències donades
<b>-STL:</b> Tant per cent de robos de un jugador per jugada defensiva del equip mentre està a pista
<b>-BLK:</b> total de tacons col·locats
<b>-TOV:</b> total de pèrdues de baló



### 3. PROCÉS:

[Còdi complet al .rmd i HTML]

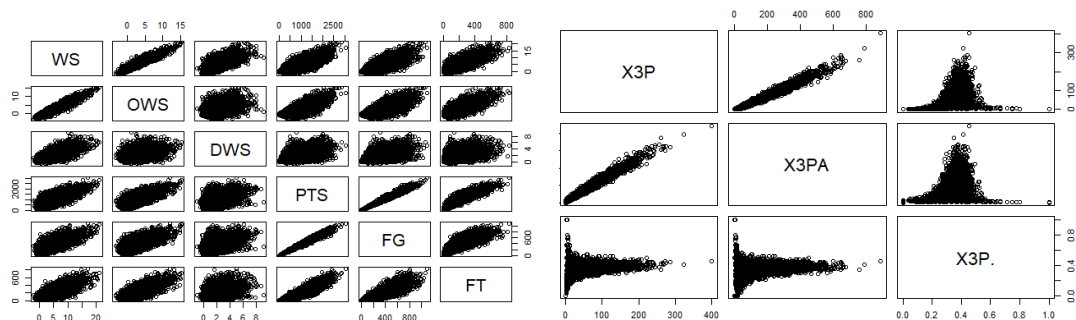
-Anem a dividir el procés en dues parts:

- Per una part tenim la creació del model amb un base de dades que conté totes les estadístiques desde el 1980 fins al 2017. En aquesta, predirem els quintets dels anys 2015, 2016 i 2017. A l'estudi podrem comprovar quins jugadors ha predit correctament, i quins no.
- En la segona part agafarem les dades que hi han aquest any de la competició fins la parada a causa del COVID-19. En aquest no podrem comprovar si el model accerta o no els jugadors ja que encara no s'ha atorgat aquest premi. També treballarem en un data.frame que no conté les mateixes variables (encara que és similar) per tant tindrem que crear altre model.

#### 3.1. PROCÉS CREACIÓ DE MODEL AMB DADES CONEGUES (1980 – 2017)

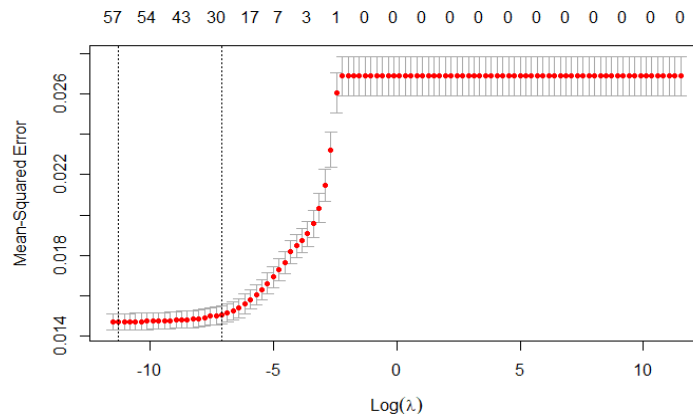
Els passos que hem seguit són els següents:

- **Lectura i neteja del dataframe** (lectura del csv, establir el tipus de variable, canviar valors nuls per 0, eliminar files amb jugadors duplicats ...)
- Introduir a la BD una **variable** anomenada **quintet**, que ens indique si el jugador ha estat o no al millor quintet d'eixa temporada.
- Establim **un conjunt d'entrenament**, i un **conjunt de prova**. El conjunt d'entrenament tindrà les dades del 1980 fins al 2011, i el conjunt de prova les dades de 2011 fins a 2017, més menys **80%-20%**.
- **Visualitzem les dades** i observem que moltes de les variables tenen correlació entre elles o no ens aporten informació relevant. Per tant, al tindre tantes variables predictores, volem realitzar una **regularització** d'aquestes.
  - Utilitzem **diverses eines de visualització** com pairs. Uns exemples on podem explicar aquesta correlació entre variables poden ser aquests:





- Hem observat que hi ha moltes variables correlacionades. Per tant anem a realitzar la regularització de variables, amb la finalitat de reduir la variància de les mateixes. Utilitzem el **mètode Lasso** i comprovem si els resultats obtinguts són ajustats.



(Intercept)	PosPF	PosSF	Age	G	MP	TS.
2.222875e-02	-1.436547e-03	-3.706560e-03	3.249107e-04	7.727267e-04	-1.716349e-04	-4.249438e-02
FTr	STL.	TOV.	USG.	DWS	WS	BPM
-1.607393e-02	-7.022702e-04	2.898649e-04	-2.537874e-04	6.191524e-03	1.821565e-02	-4.986440e-04
VORP	FG	FGA	X2P	FTA	PF	ORB
3.161288e-02	2.369019e-05	1.295914e-04	6.016663e-05	3.683313e-04	-3.212804e-04	-1.507431e-04
DRB	AST	STL	BLK			
1.636141e-04	1.907179e-04	-9.944877e-05	3.400609e-04			

- Observem que aquestes variables són les que el mètode Lasso ens indica que són més explicatives, ja que tenen uns coeficients diferent de 0
- Realitzem un mètode de Regressió Logística Múltiple (**GLM**) amb estes variables per tal de fer el següent **step**. Aquest comando ens indica el millor mètode glm comparant mitjançant la informació de AKAIKE (**AIC**).

```

      Df Deviance   AIC
<none>      911.26  941.26
- Age      1    914.56  942.56
- STL      1    915.38  943.38
- PF       1    917.01  945.01
- ORB      1    918.41  946.41
- TOV      1    920.89  948.89
- BLK      1    924.19  952.19
- USG      1    925.62  953.62
- VORP     1    927.48  955.48
- BPM      1    929.51  957.51
- DWS      1    937.73  965.73
- AST      1    968.08  996.08
- G        1    971.50  999.50
- FGA      1    984.52 1012.52
- WS       1   1107.48 1135.48

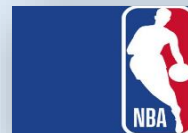
Call: glm(formula = quinteto ~ Age + G + STL. + TOV. + USG. + DWS +
  WS + BPM + VORP + FGA + PF + ORB + AST + BLK, family = "binomial",
  data = bd)

Coefficients:
(Intercept)      Age          G      STL.      TOV.      USG.      DWS
-10.818575    0.045524  -0.133487  -0.220845   0.083679   0.104698   0.447596
WS          BPM          FGA          PF          ORB          AST
 0.885259    0.166501  -0.498484   0.004465  -0.005797   0.004185   0.004953
BLK
 0.008042

Degrees of Freedom: 15359 Total (i.e. Null); 15345 Residual
Null Deviance: 3944
Residual Deviance: 911.3      AIC: 941.3

```

- Veiem com les variables que ens interessen és reduïxen considerablement. Com estem realitzant una regressió logística, en les variables que hem obtingut del punt anterior, realitzem tres models utilitzant tres mètodes diferents que són els que compararem.

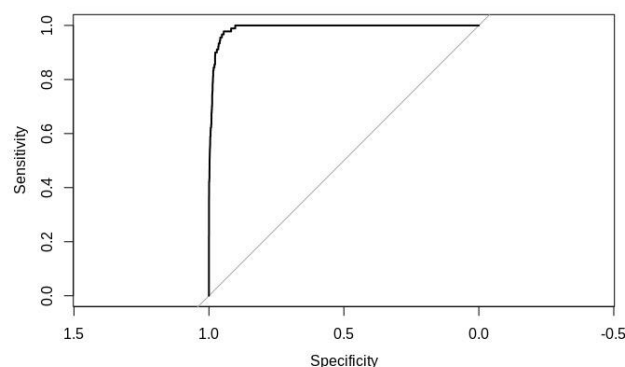


Aquests tres mètodes són: Regressió Logística Múltiple (**GLM**), Anàlisi Discriminant Quadràtic (**QDA**) i Anàlisi Discriminant Lineal (**LDA**). No agafem el mètode KNN perquè ja sabem que els valors veïns no són interessants per predir el pròxim valor.

- Realitzem comparacions entre els models i ens fixem en les següents resultats per escollir el que més ens interese.

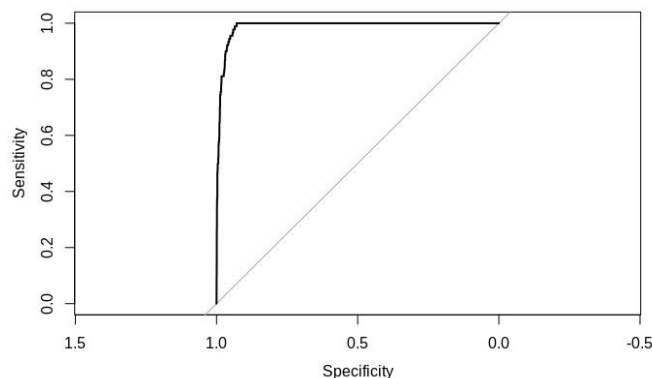
### 3.1.1. MÈTODE LDA

LDA	0	1
0	3422	21
1	47	69
% Acerts	98.08935%	
Àrea sota la corva	0.9901	
IC 95%	0.9862-0.994	



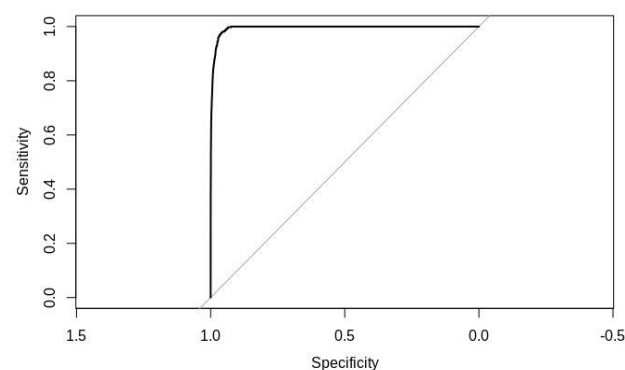
### 3.1.2. MÈTODE QDA

QDA	0	1
0	3296	4
1	173	86
% Acerts	95.02669%	
Àrea sota la corva	0.9881	
IC 95%	0.9841-0.992	



### 3.1.3. MÈTODE GLM

GLM	0	1
0	3462	37
1	7	53
% Acerts	98.7637 %	
Àrea sota la corva	0.9946	
IC 95%	0.9934-0.9958	







### MATRIZ DE CONFUSIÓN

		Estimado por el modelo	
Realidad		Verdadero positivo	Falso negativo
		Falso positivo	Verdadero negativo

Fig.3

• **Escollim el mètode GLM** ja que és el que millor prediu els veritaders positius i negatius. A banda és el que més reduïx els falsos positius (realment el que ens interessa a les errades que fa el model), però els falsos negatius són més alts que als altres models. Podem afirmar que els tres models són bons, però per les anteriors raons ens quedarem amb el GLM.

- Amb el model tenim una taula amb els jugadors que tenen una major probabilitat d'estar al quintet. A aquesta taula li apliquem una **funció** en la qual tenim en compte la **restricció** que **sols poden haver determinats jugadors per cada posició**.
- Finalment **visualitzem els resultats** que comentarem al punt 4.1.RESULTATS.

## 3.2. PROCÉS CREACIÓ DE MODEL AMB LES DADES D'AQUEST ANY

-El procés que hem seguit en aquest cas és molt similar a l'anterior, pràcticament idèntic. En resum, fem els següents passos:

- **Lectura i adequació de les dades** (canviem el nom de les variables per a que siguin iguals a l'altre data.frame, canviar valors nuls per 0 ...)
- Fem una **regressió logística múltiple amb el mètode GLM**. Fem un **step per a escollir les variables d'interés**.
- Torem a observar els diferents mètodes de regressió logística per a elaborar el millor model. De nou ens quedem amb el **mètode GLM**. Els resultats obtinguts són molt similars als del punt anterior.
- Li apliquem la mateixa **funció** que la que hem fet anteriorment per a restringir les **posicions** dels jugadors.
- **VISUALITZACIÓ:** 4.2.RESULTATS MODEL PREDINT LES DADES D'AQUEST ANY



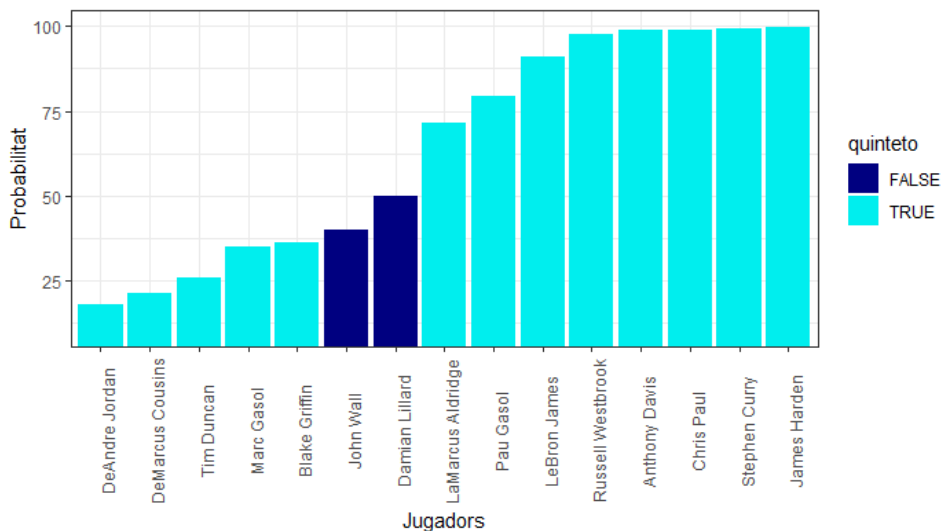
## 4. RESULTATS I CONCLUSIONS

- En aquest punt, analitzarem els resultats obtinguts i traurem les conclusions dels mateixos.

### 4.1. RESULTATS MODEL ON CONEGUEM LES DADES I CONCLUSIONS

#### ➤ Predicció de l'any 2015:

Predicció dels millors quintets  
ANY 2015



-Comprovant el model l'any 2015, apreciem que els resultats obtinguts ens semblen molt precisos. Considerant que tenim una base de dades amb molts jugadors cada temporada, en aquest cas

Sabent que les votacions són subjectives depenent del joc del jugador, i no de les seues estadístiques, observem que el nostre model explica aquestes votacions amb una probabilitat d'acert molt elevada.

En la tabla de sustituciones, aquests son els jugadors que deuriem estar al quintet (Kyrie Irving i Klay Thompson) substituint els que no ha conseguit predir correctament el nostre model (Jhon Wall i Damian Lillard). També mostrem les probabilitats que el nostre model atorga a aquests jugadors. D'aquesta forma podem saber un poc més sobre els nostres errors.

Millors jugadors del 2015

Jugador	Edad	Posicio	Probabilitat	Esta en el quintet?
James Harden	25	SG	99.82453	1
Stephen Curry	26	PG	99.46583	1
Chris Paul	29	PG	98.94948	1
Anthony Davis	21	PF	98.68507	1
Russell Westbrook	26	PG	97.67883	1
LeBron James	30	SF	91.07932	1
Pau Gasol	34	PF	79.18571	1
LaMarcus Aldridge	29	PF	71.48565	1
Damian Lillard	24	PG	49.70695	0
John Wall	24	PG	39.83640	0
Blake Griffin	25	PF	36.01446	1
Marc Gasol	30	C	34.85991	1
Tim Duncan	38	C	25.74312	1
DeMarcus Cousins	24	C	21.20210	1
DeAndre Jordan	26	C	17.97916	1

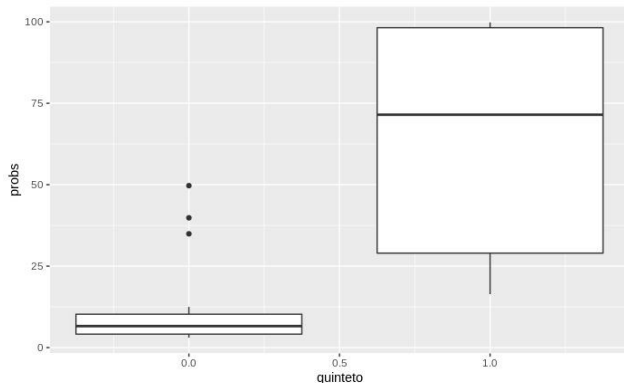
Sustituciones

Jugador	Edad	Posicio	Probabilitat	Esta en el quintet?
Damian Lillard	24	PG	49.70695	FALSE
John Wall	24	PG	39.83640	FALSE
Kyrie Irving	22	PG	32.24694	TRUE
Klay Thompson	24	SG	16.39658	TRUE

Data: 24/05/2020

Assignatura: Models Lineals – L2 (Ciència de Dades)

Autors: Guillem Miralles | Joan Llavata | Miguel Payá

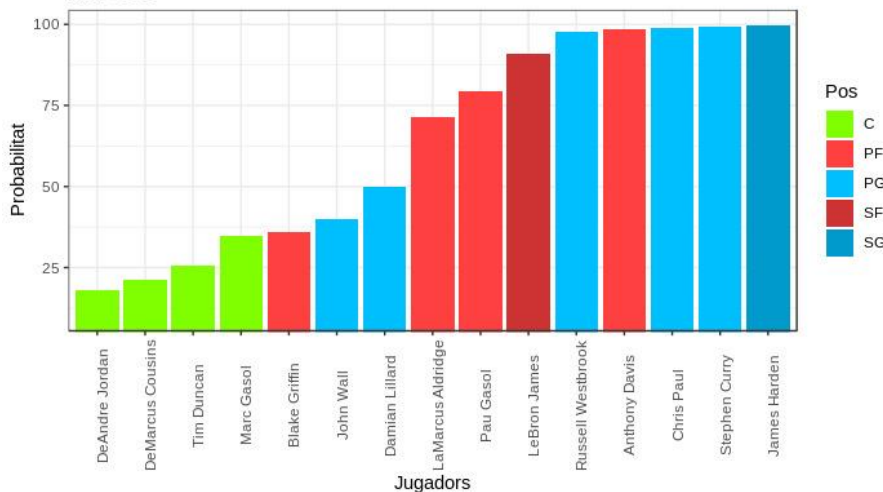


-Per a realitzar aquest boxplot ho fem amb una base de dades dels 30 jugadors que més probabilitat tenen d'estar al quintet.

A la dreta podem veure com la mediana dels jugadors que si que van a estar al quintet aquell any li donem una probabilitat del 71.49%, mentre que als altres jugadors els donem una mediana de 6.59%.

Observem que hi ha 3 outliers que són els jugadors que el nostre model prediu que estaran. Un d'ells no entra degut a la restricció de posicions.

PREDICCIÓ MILLOR QUINTET PER POSICIÓ  
ANY 2015



En aquest gràfic podem observar la divisió de posicions que ens explica un poc més els errors del model, ja que John Wall i Damian Lillard (els dos errors) entren encara que tinguen una probabilitat més elevada que altres jugadors del model, entrarien a la seua posició en els últims dos llocs.

Data: 24/05/2020

Assignatura: Models Lineals – L2 (Ciència de Dades)

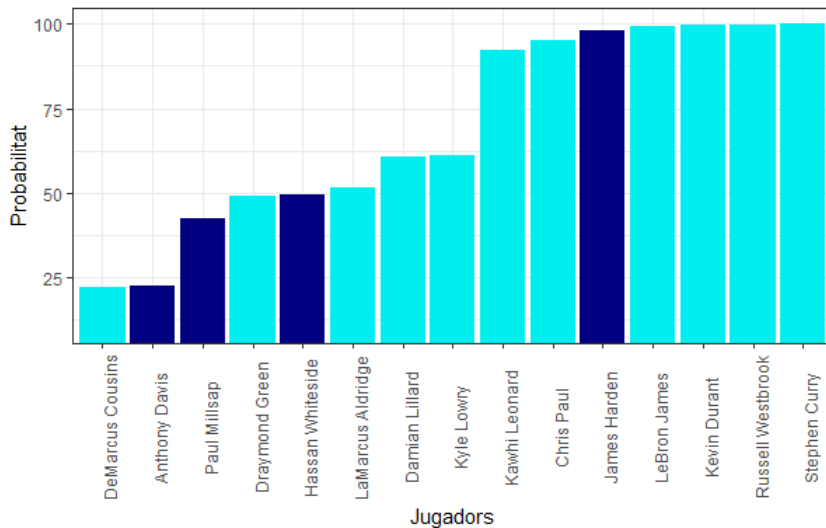
Autors: Guillem Miralles | Joan Llavata | Miguel Payá



### ➤ Predicció de l'any 2016:

#### Predicció dels millors quintets

ANY 2016



-Aquest any és l'any en el qual més errors hi trobem, sobretot ens centrem a l'error de James Hardem qui li donem una probabilitat del 97.8% de pertànyer als quintets. Investigant un poc sobre el jugador ens adonem que pertany als quintets des de l'any 2013 fins al 2019 (amb l'excepció d'aquest any) i des de l'any 2014 apareix sempre al primer quintet.

Observem que aquest any és l'any en el qual el jugador va aconseguir menys victòries (una diferència de 14 en relació en els altres anys), l'error és a causa que el nostre model no les contempla. Pensem que aquesta falta de victòries va influir a les votacions. Encara que les seues estadístiques individuals eren molt destacades. Aquest jugador haguera entrat al quintet segons les votacions d'aquell any a l'NBA, però com ja sabem, existeix la restricció de posicions, fet que va causar que no entrara.

A la taula podem encontrar quins jugadors tenen que estar al quintet substituint les errades del model.

#### Millors jugadors del 2016

Jugador	Edad	Posicio	Probabilitat	Esta en el quintet?
Stephen Curry	27	PG	99.95947	1
Russell Westbrook	27	PG	99.66814	1
Kevin Durant	27	SF	99.59944	1
LeBron James	31	SF	99.27701	1
James Harden	26	SG	97.84001	0
Chris Paul	30	PG	94.96401	1
Kawhi Leonard	24	SF	92.42842	1
Kyle Lowry	29	PG	61.28511	1
Damian Lillard	25	PG	60.52303	1
LaMarcus Aldridge	30	PF	51.54793	1
Hassan Whiteside	26	C	49.35669	0
Draymond Green	25	PF	48.86015	1
Paul Millsap	30	PF	42.52445	0
Anthony Davis	22	C	22.50300	0
DeMarcus Cousins	25	C	22.24952	1

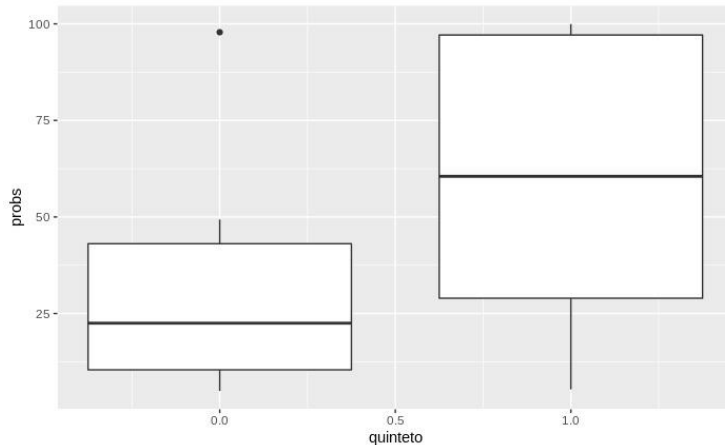
#### Sustitucions

Jugador	Edad	Posicio	Probabilitat	Esta en el quintet?
James Harden	26	SG	97.840012	FALSE
Hassan Whiteside	26	C	49.356690	FALSE
Paul Millsap	30	PF	42.524450	FALSE
Anthony Davis	22	C	22.502996	FALSE
Paul George	25	SF	35.686584	TRUE
DeAndre Jordan	27	C	16.846014	TRUE
Andre Drummond	22	C	6.070247	TRUE
Klay Thompson	25	SG	5.361564	TRUE

Data: 24/05/2020

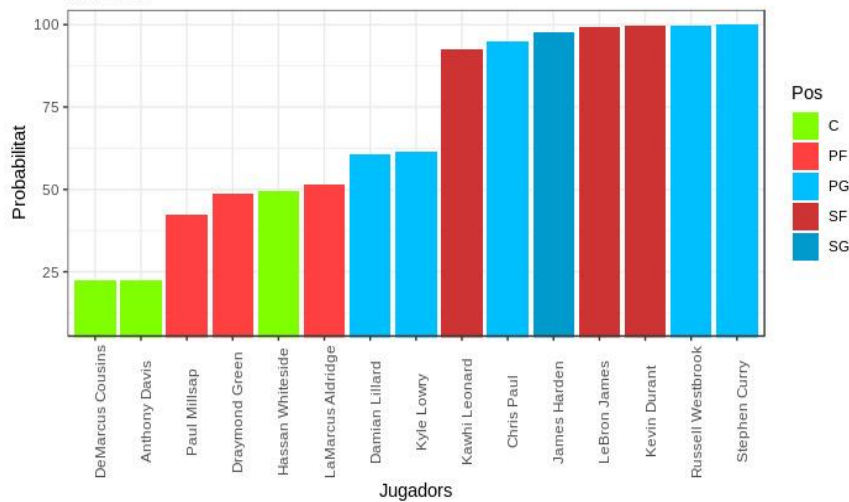
Assignatura: Models Lineals – L2 (Ciència de Dades)

Autors: Guillem Miralles | Joan Llavata | Miguel Payá



Aquest any tenim la mediana de jugadors que no estan al quintet més alta, encara que, podem observar que la mediana de probabilitat dels jugadors que estan al quintet es el doble dels que no hi estan. Per tant podem considerar que realitzem una bona predicció.

### PREDICCIÓ MILLOR QUINTET PER POSICIÓ ANY 2016



Com poder comprovar a quesymodel, ja hi trobem 6 jugadors a la posició de “guard”, aquest fet causa que James Harden no aconsegueixca entrar al quintet d’aquest any.

Data: 24/05/2020

Assignatura: Models Lineals – L2 (Ciència de Dades)

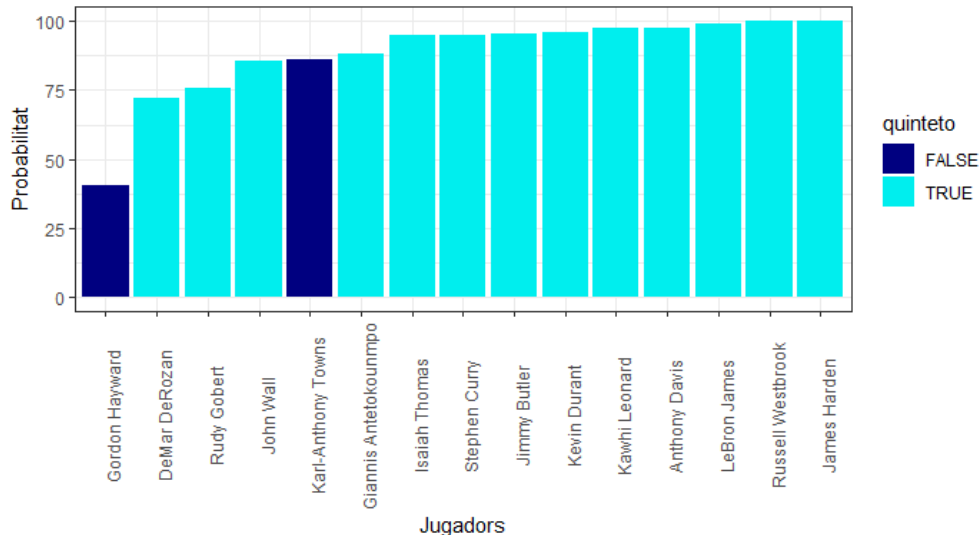
Autors: Guillem Miralles | Joan Llavata | Miguel Payá



### ➤ Predicció de l'any 2017:

#### Predicció dels millors quintets

ANY 2017



-Aquest any podem considerar que es presenten molts jugadors amb una probabilitat molt alta de pertànyer al quintet. Observem que hi ha molt pocs errors. Hi ha dues errades, i no estan entre els 10 primers.

Karl-Anthony Towns del equip Minnesota Timberwolves, té un total de 31 victòries i 51 derrota. Sent aquestes el mínim de victòries de tots els jugadors predits. Una cosa que també podem destacar es que aquest jugador va estar el 16é a les posicions dels quintets, amb 4 punts menys que Deandre Jordan que va entrar el 15é.

Gordon Hayward aquell any va fer el any en millors estadístiques. Va ser el seu únic any amb més de 20 punts per partit jugat. Va ser l'únic any en que va ser elegit per al All Star de l'NBA.

#### Millors jugadors del 2017

Jugador	Edad	Posicio	Probabilitat	Esta en el quintet?
James Harden	27	PG	99.90859	1
Russell Westbrook	28	PG	99.90621	1
LeBron James	32	SF	98.82146	1
Anthony Davis	23	C	97.47270	1
Kawhi Leonard	25	SF	97.37003	1
Kevin Durant	28	SF	96.08908	1
Jimmy Butler	27	SF	95.59811	1
Stephen Curry	28	PG	95.06756	1
Isaiah Thomas	27	PG	94.65219	1
Giannis Antetokounmpo	22	SF	88.08650	1
Karl-Anthony Towns	21	C	86.31123	0
John Wall	26	PG	85.78017	1
Rudy Gobert	24	C	75.74142	1
DeMar DeRozan	27	SG	71.95088	1
Gordon Hayward	26	SF	40.50799	0

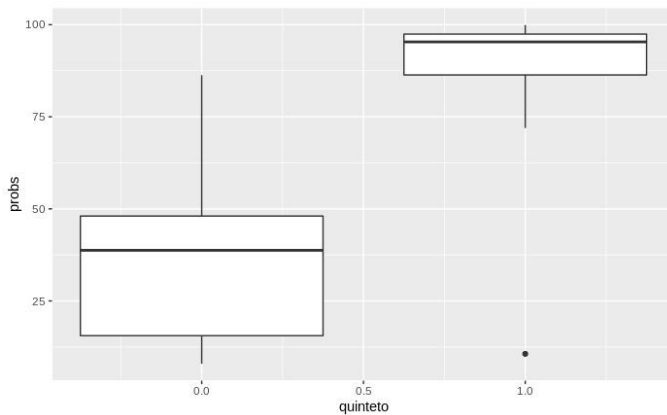
#### Sustitucions

Jugador	Edad	Posicio	Probabilitat	Esta en el quintet?
Karl-Anthony Towns	21	C	86.311228	FALSE
Gordon Hayward	26	SF	40.507995	FALSE
DeAndre Jordan	28	C	10.635179	TRUE
Draymond Green	26	PF	4.138469	TRUE

Data: 24/05/2020

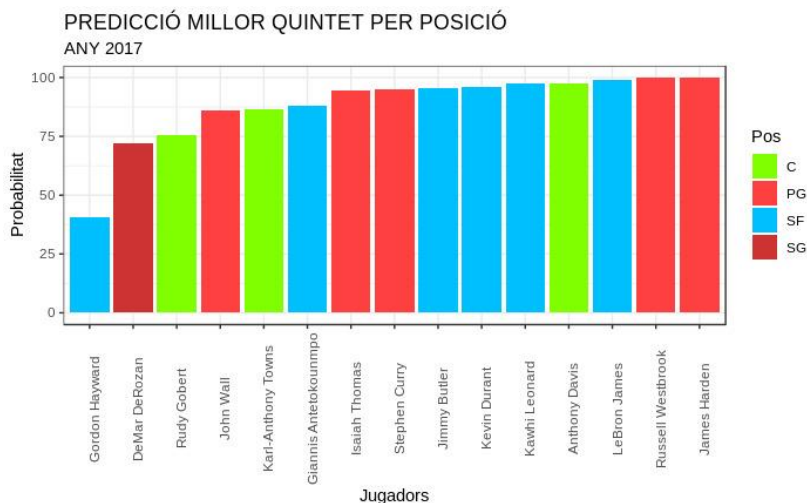
Assignatura: Models Lineals – L2 (Ciència de Dades)

Autors: Guillem Miralles | Joan Llavata | Miguel Payá



Aquest diagrama de caixes mostra una gran diferència entre ambdós grups, una mica inclús més destacable que els altres anys. Amb unes medians de 8.76% respecte a 95.07%

Centrant-nos en l'outlier que trobem, s'adonem que és Deandre Jordan amb una probabilitat de 10.64% i que ocupa la mateixa posició que Karl-Anthony Towns (l'error d'abans). Deandre Jordan és un jugador amb una faceta molt defensiva, per tant no tenia molt bones estístiques, però té molt bona reputació a la lliga. El seu equip aquest any va aconseguir 20 victòries més que l'equip de Karl-Anthony Towns, a més, entrant a playoffs per la part alta del quadre.

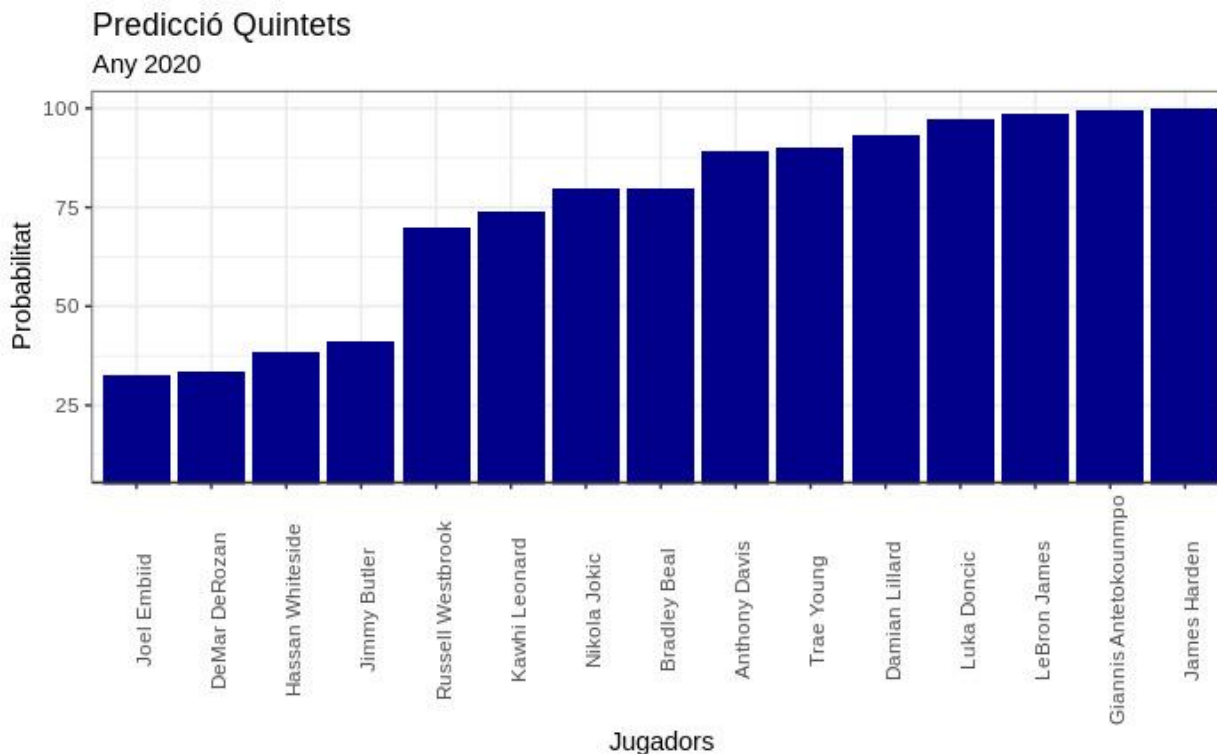


En aquest gràfic ens sembla interessant veure com no hi ha cap jugador en la posició PS que en l'actualitat presenta molts canvis a la competició. Ocupant-la en la mesura per jugadors de SF o C.





## 4.2. RESULTATS MODEL PREDINT LES DADES D'AQUEST ANY I CONCLUSIONS



-Ací presentem la nostra predicció per a aquest any, encara que no podem comprovar pel fet que no s'ha atorgat aquest premi. De forma subjectiva podem valorar com els jugadors obtinguts han realitzat fins al moment una bona temporada. D'aquesta manera podem dir amb una certa base que el model prediu el quintet d'una forma raonable.

Com ja hem dit, no podem comprovar que aquest jugadors seràn elegits per al millor quintet, pero podem comparar-ho amb la votació dels All Stars. Encara que és una votació amb finalitat diferent, ja que no voten periodistes sinó aficionats, però pot servir per orientar-nos. En aquesta 12 dels jugadors predits al model han sigut elegits.

Millors jugadors del 2020			
Jugador	Edad	Posicio	Probabilitat
James Harden	30	G	99.82000
Giannis Antetokounmpo	25	F	99.67420
LeBron James	35	F	98.82920
Luka Doncic	21	G-F	97.26112
Damian Lillard	29	G	93.29723
Trae Young	21	G	90.10713
Anthony Davis	27	F-C	89.39499
Bradley Beal	26	G	79.80287
Nikola Jokic	25	C	79.67576
Kawhi Leonard	28	F	74.03270
Russell Westbrook	31	G	69.90656
Jimmy Butler	30	F	41.15219
Hassan Whiteside	30	C	38.69001
DeMar DeRozan	30	F	33.67154
Joel Embiid	25	C-F	32.54220





Puesto	Jugador	Votos
1	Giannis Antetokounmpo	50
2	LeBron James	39
3	James Harden	23
4	Luka Doncic	15
5	Anthony Davis	9

-Per comparar una mica més els resultats d'aquest any, podem observar que a l'últim article que publica NBA Spain per la carrera de l'MVP. Observem que els 5 jugadors que entren obtenen més d'un 89% de probabilitat d'estar al quintet. El primer quintet que es pot conformar segons la restricció de posicions, el conformaríem amb aquestes 5 posicions.

Per tant sembla les prediccions d'aquest any són raonables.

## 5. CONCLUSIONS FINALS

-En conclusió, el nostre model obté una fiabilitat molt elevada quan parlem sobretot dels 5 jugadors amb més probabilitats d'estar al quintet per any. En jugadors als quals el nostre model els dona probabilitats més baixetes, els resultats finals també solen errar un poc més. De totes maneres, pensem que la fiabilitat que obté el nostre model es molt alta, com em vist als resultats. Realitzant la predicció correcta en un total de 37 jugadors dels 45 (82'22%).

Com visualment podem apreciar, d'aquests 8 errors, sols hi ha dos que fallen dels que otorguem una probabilitat superior al 50%.

En ambdós casos, el total de victòries de l'equip es prou significatiu, ja que no hi són nombroses. Aquest és el punt on més errors pensem que ha trobat el nostre model. Segons la nostra hipòtesi, si tinguérem una base de dades d'on poder traure aquesta possible variable, els nostres resultats millorarien notablement l'estudi.