

APRENENTATGE AUTOMÀTIC
CURS 2024/25

La victòria a League of Legends

Pol Ballart Álvarez - 1638390

Miguel Ibáñez Molina - 1571691
Juan José Rodríguez Potosí - 1606379

23 de Gener

Universitat Autònoma de Barcelona

1 Introducció

En aquest treball hem decidit centrar el nostre estudi en dades de partides del videojoc i e-sport (esport electrònic) League of Legends.

League of Legends és un MOBA (multiplayer online battle arena), és a dir, un videojoc al qual s'enfronten dos equips de 5 jugadors cadascun amb l'objectiu de destruir el nexe enemic, que és l'estructura principal i més protegida de cada equip.

Des de cada nexe són invocats de manera periòdica els anomenats súbdits, unes unitats que no són controlades per cap jugador i que van en direcció a la seva línia del mapa assignada. El seu objectiu és protegir als jugadors del propi equip i donar experiència i or als jugadors de l'equipo contrari.

A cada base també hi ha 3 inhibidors, estructures que quan són destruïdes permeten a l'equip enemic generar súbdits més poderosos durant 5 minuts en aquell carril determinat.

Al mapa n'hi ha 3 línies, conegeudes per Top lane, Mid(dle) lane i Bot(tom) lane, que es traduirien com a línia superior, del mig i inferior. A aquestes línies acostumen a anar, respectivament, un jugador (anomenat top planer), un altre jugador (anomenat midlaner) i dos jugadors (anomenats ADC i support), on ADC significa attack damage carry (la font principal de mal físic a l'equip).

El terreny que hi ha entre les línies s'anomena jungla, on hi ha els coneguts campaments de la jungla, en els que hi ha unes altres unitats que no controla cap jugador, que no marxen del seu campament i que serveixen per donar or i experiència als jugadors (un a cada equip) amb el rol de jungla.

A cadascuna de les línies hi ha 3 torres que fan la funció d'estructura defensiva per a crear una zona protegida per als jugadors de cada equip.

El mapa queda dividit diagonalment pel riu, que es troba en una regió equidistant a la base d'amb-dos equips. Al riu n'hi ha dues foses, espais on apareixen monstres neutrals molt poderosos donen beneficis a l'equip que els derrota.

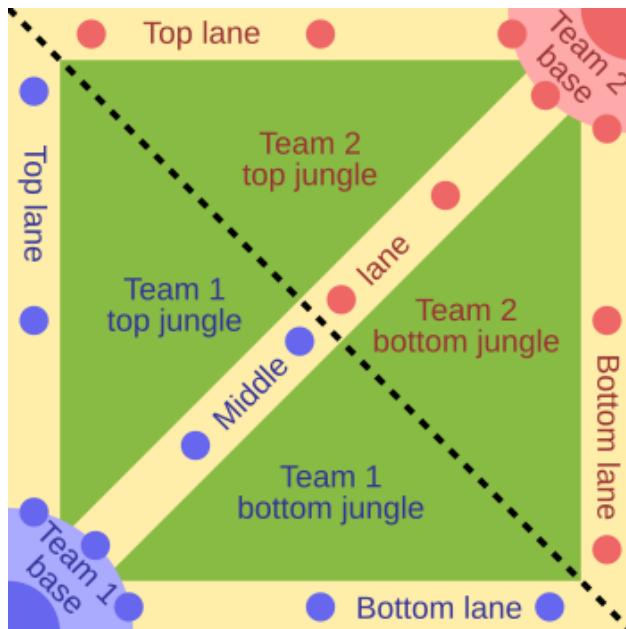


Figura 1: Mapa genèric d'un MOBA.

Hem escollit com a dataset el registre oficial de Riot Games, l'empresa creadora del videojoc. Concretament hem triat un conjunt de dades sobre partides classificatòries a la temporada 7, és a dir, que es van jugar durant l'any 2017.

Les partides classificatòries són aquelles on els participants donen el millor de si per a guanyar, és a dir, no són partides ràpides per passar l'estona. Això és així per què hi ha classificació per nivells dins del joc mesurada per diferents lligues competitives, i l'objectiu de cada jugador és arribar a la màxima lliga possible cada temporada, cosa que s'aconsegueix guanyant punts (League Points) que sumen amb victòries i resten amb derrotes.

Seria la comparació a una lliga de futbol regional, en la que tot i alomillor no tenir el millor nivell, els jugadors s'esforcen al màxim i competeixen. Una partida ràpida seria equivalent a fer un partit de futbol entre amics.

Per això creiem que aquest registre abasteix les dades necessàries i més adients per fer un estudi, ja que recull varis nivells de joc (baix, intermig, i alt) en el que igualment els jugadors competeixen de manera seria.

Els nostres objectius d'estudi són:

- Predir, segons les estadístiques individuals i grupals d'una partida, si un jugador pertany a l'equip guanyador o no.
- Estudiar si hi ha personatges que influeixen més en la victòria d'un equip (sigui escollit o vetat), és a dir, veure la influència de jugar amb els personatges més forts.
- Estudiar la rellevància dels diferents objectius neutrals de la partida i comparar-los.

2 Descripció del conjunt de dades

El conjunt de dades ve donat pel registre oficial de partides de Riot Games l'any 2017, com hem explicat anteriorment. La font de les dades és la pàgina web Kaggle. Al ser el registre de partides oficials donem per fet que no hi ha cap error al mostreig i tota la informació donada és verídica.

Les variables que tenim són:

Variable	Descripció	Tipus
matchid	Identificador únic de la partida.	Categòrica, independent
platformid	Regió en què s'ha jugat la partida.	Categòrica, independent
seasonid	Temporada de joc a la qual pertany la partida.	Categòrica, independent
duration	Duració de la partida (en segons).	Regressiva, dependent
version	Versió del joc amb què s'ha jugat la partida.	Categòrica, independent
id	Identificador únic del jugador dins la partida.	Categòrica, independent
player	1-5 jugadors de l'equip 1, 6-10 de l'equip 2.	Categòrica, independent
championid	Identificador del campió utilitzat pel jugador.	Categòrica, independent
ss1, ss2	Primer i segon summoner spell utilitzats pel jugador.	Categòrica, independent
role, position	Rol (jungla, suport, etc.) dins l'equip.	Categòrica, independent
win	Si el jugador ha guanyat o no la partida.	Categòrica, dependent
item1-item6, trinket	Objectes comprats pel jugador durant la partida.	Categòrica, independent
kills, deaths, assists	Estadístiques bàsiques del jugador.	Regressiva, dependent
largestkillingspree	Major seqüència d'assassinats sense morir.	Regressiva, dependent
largestmultikill	Major nombre d'assassinats consecutius en poc temps.	Regressiva, dependent
killingsprees	Nombr total de seqüències d'assassinats.	Regressiva, dependent
longesttimespentliving	Període més llarg de temps sense morir.	Regressiva, dependent

Variable	Descripció	Tipus
doublekills, triplekills, quadrakills, pentakills	Nombre d'assassinats dobles, triples, quàdruples i pentakills aconseguits.	Regressiva, dependent
legendarykills	Nombre d'assassinats llegendaris (ratxa de més de 7 assassinats sense morir).	Regressiva, dependent
totdmgdealt, magicdmgdealt, physicaldmgdealt, truedmgdealt	Dany total, màgic, físic i real infligit.	Regressiva, dependent
largestcrit	Major cop crític realitzat pel jugador.	Regressiva, dependent
totdmgtochamp, magicdmgtochamp, physdmgtochamp, truedmgtochamp	Dany infligit a campions (total, màgic, físic i real).	Regressiva, dependent
totheal	Quantitat total de curació realitzada.	Regressiva, dependent
totunitshealed	Nombre total d'unitats curades.	Regressiva, dependent
dmgselfmit	Dany mitigat pel jugador.	Regressiva, dependent
dmgtoobj, dmgtoturrets	Dany infligit a objectius i torres enemigues.	Regressiva, dependent
visionscore	Puntuació de visió (ús de sentinelles i detecció d'enemics).	Regressiva, dependent
totdmgtaken, magicdmgtaken, physdmgtaken, truedmgtaken	Dany rebut pel jugador (total, màgic, físic i real).	Regressiva, dependent
goldearned, goldspent	Or obtingut i gastat pel jugador.	Regressiva, dependent
turretkills, inhibkills	Torres i inhibidors destruïts pel jugador.	Regressiva, dependent
totminionskilled	Nombre total de minions assassinats.	Regressiva, dependent
neutralminionskilled, ownjunglekills, enemyjunglekills	Minions neutrals assassinats, incloent els de la pròpia jungla i la de l'enemic.	Regressiva, dependent
totcctimedea	Temps total de Crowd Control exercit als enemics.	Regressiva, dependent
champlvl	Nivell del campió del jugador al final de la partida.	Regressiva, dependent

Variable	Descripció	Tipus
pinksbought, wardsbought, wardsplaced, wardskilled	Sentinelles de control comprades, col·locades i destruïdes pel jugador.	Regressiva, dependent
firstblood	Si el jugador ha aconseguit el primer assassinat.	Categòrica, dependent
teamid	Identificador de l'equip del jugador.	Categòrica, independent
firsttower, firstinhib, firstbaron, firstdragon, firstharry	Si l'equip ha destruït la primera torre, inhibidor, Baró, drac o herald de la partida.	Categòrica, dependent
towerkills, inhibkills, baronkills, dragonkills, harrykills	Torres, inhibidores, Barons, dracs i heralds destruïts per l'equip.	Regressiva, dependent

El nostre dataset és un conjunt de dades rellevant perquè registra vàries temporades i amb una quantitat fiable d'informació. En el nostre cas ens limitem a la temporada 7 en concret ja que és de la que més dades disposem, i al centrar-nos en una sola temporada el conjunt de dades que ens queda és més fiable.

Les partides entre diferents temporades varien molt ja que es modifiquen característiques molt importants del joc que després es mantenen intactes la resta de la temporada.

Al tenir tanta informació sobre cada partida i jugador podem fer prediccions o estudis sobre ítems molt diversos, com per exemple quin equip guanyarà, si s'han rendit...

Aquest conjunt de dades és idílic ja que disposa de molta informació variada i en molta quantitat, i ens proporciona de sobres les variables que necessitem.

Tenim la intuïció que dins del gran conjunt de dades que tenim, hi haurà un subconjunt de dades que no seran gaire rellevants pels nostres objectius d'estudi i del qual podrem prescindir.

3 Exploració de les dades

Taula 1: Temps

Estadística	Mitjana	Desv. Est.	Mínim	Q1	Mediana	Màxim
Duració de partida (s)	1824.08	509.89	190	1536	1830	4991
Temps viu més llarg (s)	626.71	309.60	0	431	587	2982

Taula 2: Estadístiques d'assassinats y morts

Estadística	Mitjana	Desv. Est.	Mínim	Q1	Mediana	Màxim
Kills	4.94	3.94	0	2	4	67
Deaths	5.79	3.87	0	3	5	44
Assists	7.96	5.12	0	4	7	53
Firstblood	0.099	0.299	0	0	0	1
DoubleKills	0.43	0.69	0	0	0	10
TripleKills	0.10	0.32	0	0	0	5
QuadraKills	0.02	0.15	0	0	0	3
PentaKills	0.004	0.06	0	0	0	2
Killingsprees	1.34	1.26	0	0	1	13
Legendarykills	0.000001	0.001	0	0	0	1
Largestkillingspreee	2.66	2.53	0	0	2	35

Taula 3: Estadístiques de dany

Estadística	Mitjana	Desv. Est.	Mínim	Q1	Mediana	Màxim
TotDmgDealt	113353	1632762	-2147484000	59350	108976	1064761
MagicDmgDealt	38636	48798	0	4421	17459	789034
PhysicalDmgDealt	69574	70604	0	11428	42822	917986
TrueDmgDealt	5142	1631145	-2147484000	462	2026	713886
LargestCrit	242	420	0	0	0	4876
TotDmgToChamp	17654	11473	0	9378	1,508	152607
MagicDmgToChamp	7697	9242	0	1307	4106	142469
PhysDmgToChamp	8948	9946	0	1448	4600	123871
TrueDmgToChamp	1008	1642	0	26	460	36143

Taula 4: Estadístiques de curació y mitigació

Estadística	Mitjana	Desv. Est.	Mínim	Q1	Mediana	Màxim
TotHeal	5439	5525	0	1765	3850	99164
TotUnitsHealed	2.25	2.33	0	1	1	49
DmgSelfMit	11456	14842	0	177	7184	311225
Totdmgtaken	23138.46	1186.306	0	15186	21473	165152
Magicdmgtaken	8084.95	5136.359	0	4495	7205	8458
Physdmgtaken	13994.38	7735.014	0	8592	12769	90,430
Truedmgtaken	1058.920	1285.725	0	275	660	31648

Taula 5: Estadístiques d'objetius i or

Estadística	Mitjana	Desv. Est.	Mínim	Q1	Mediana	Màxim
DmgToObj	4365	6771	0	0	1091	84057
DmgToTurrets	2271	2991	0	0	1170	62488
Turretkills	0.91	1.21	0	0	0	10
Inhibkills	0.18	0.46	0	0	0	7
VisionScore	14.64	17.56	0	0	11	179
Pinksbought	0.9889597	1.407279	0	0	1	54
Wardsplaced	11.48005	7.476475	0	7	1	69
Wardskilled	1.769911	17.56	0	1	3	86
GoldEarned	11395	4019	643	8852	11336	40982
GoldSpent	10348	3879	0	7925	10275	70255

Taula 6: Estadístiques de minions y control

Estadística	Mitjana	Desv. Est.	Mínim	Q1	Mediana	Màxim
TotMinionsKilled	120.66	83.91	0	39	125	673
NeutralMinionsKilled	19.91	31.74	0	0	4	299
OwnJungleKills	11.95	19.91	0	0	2	174
EnemyJungleKills	7.96	13.25	0	0	1	181
ChampLvl	14.18	3.14	1	13	15	18
TotCcTimeDealt	426.98	659.15	0	112	251	25204

Taula 7: Estadístiques d'equip

Estadística	Mitjana	Desv. Est.	Mínim	Q1	Mediana	Màxim
Towerkills	5.7504	3.879	0	2	6	11
Inhibkills	1.0373	1.2597	0	0	1	13
Baronkills	0.4135	0.6069	0	0	0	5
Dragonkills	1.4274	1.2317	0	0	1	7

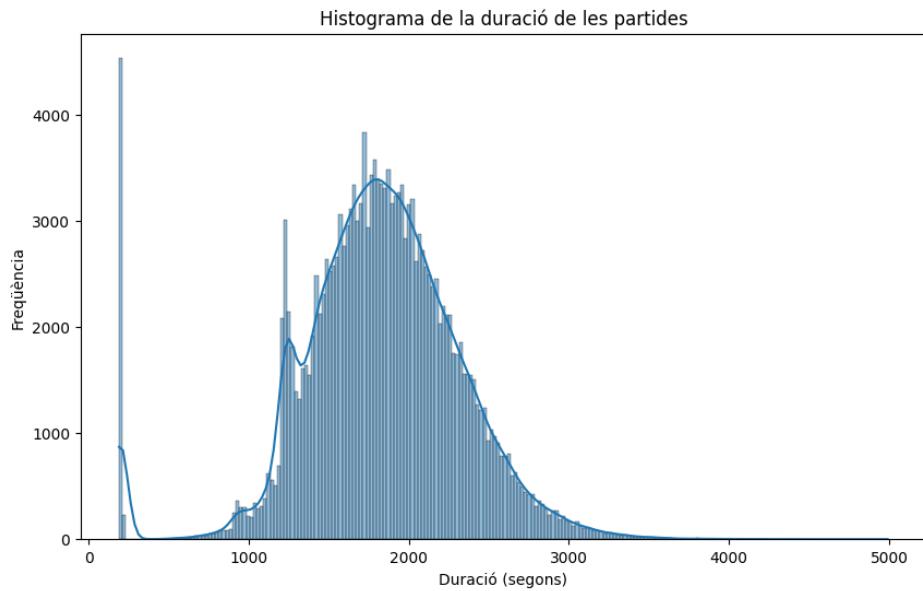


Figura 2: Histograma duració partides

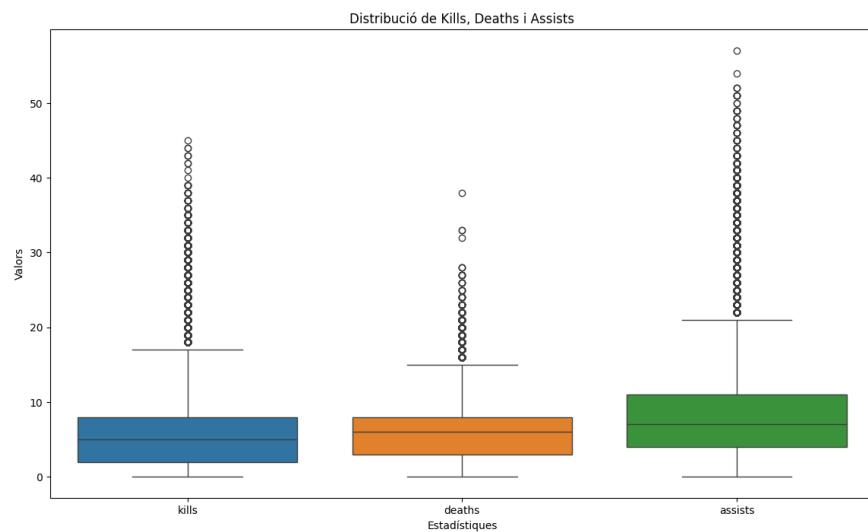


Figura 3: Boxplot de K/D/A

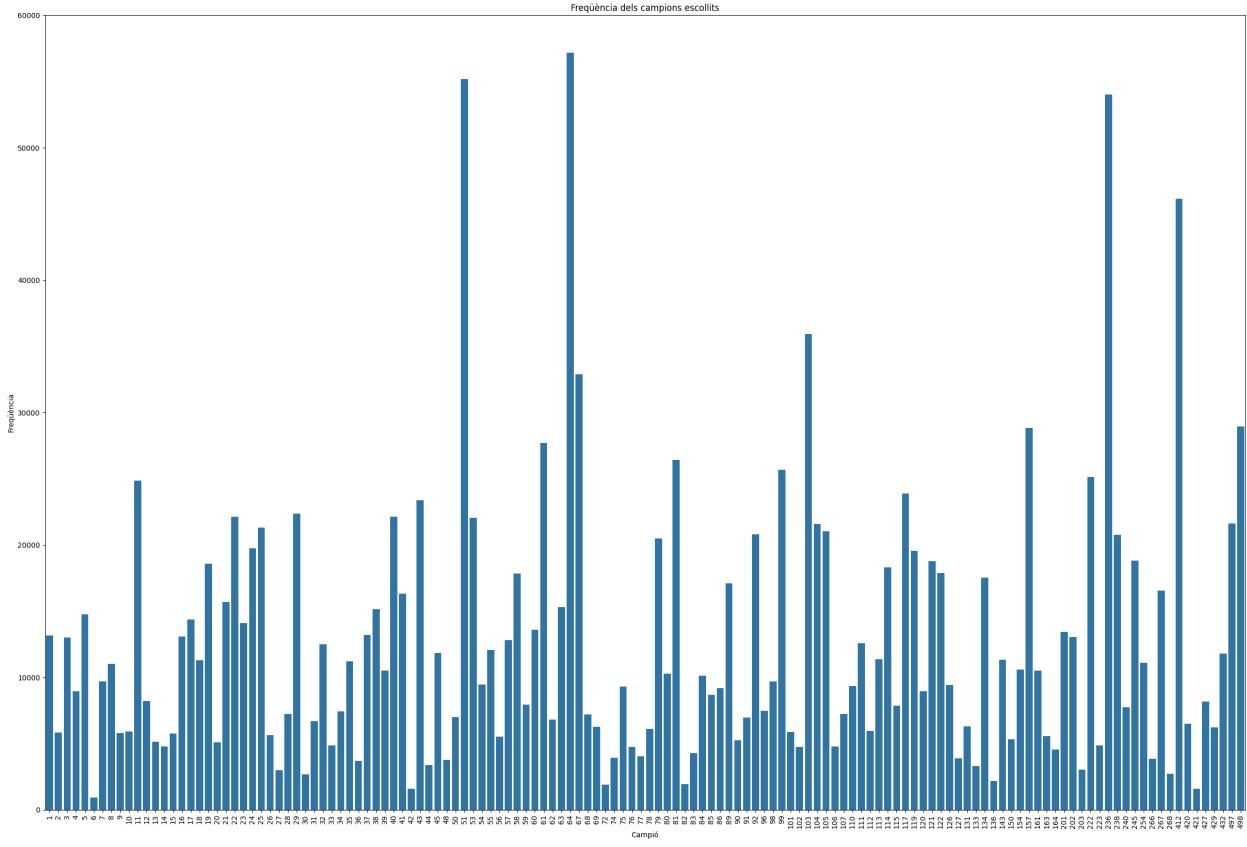


Figura 4: Freqüència dels personatges

Hi ha diferents factors que poden indicar que un individu del dataset és un valor atípic. Entre ells trobem:

- Equips sense el SS (Summoner spell) Adequat

- SMITE: És "obligatori" que n'hi hagi un i només un per equip (la persona que ocupa el rol de JUNGLA).
 - Una combinació de SS poc convencional sol ser indicador de que el jugador vol perdre a propòsit.
 - FLASH: la gran majoria de personatges es porten amb FLASH com a un SS, és a dir, es pot pendre com a indicador. Tot i que hi ha varis exemples en els que no és estrictament necessari.

- Partides que acaben entre els minuts propers a 20:00

Els motius típics per a què passi això són:

- Un dels dos equips ha agafat tanta ventatja en els primers minuts de joc que l’altre equip vota unànimement que es volen rendir.

- Un dels jugadors s'ha desconectat a meitat de la partida.
- El factor mental d'un equip. És a dir, és probable que tot i no haver gran desventatja i poder encara guanyar, un equip es vegi mermat

- **Equips amb jugadors disconnectats:**

Això es pot identificar perquè no han comprat cap objecte en tota la partida.

- **Partides en les que es fa "remake":**

Com que hi ha un dels jugadors disconnectat des de l'inici, la resta de l'equip pot votar per acabar el joc al minut 3:00 sense que compti com a victòria o derrota.

- **Equips guanyadors amb el mínim de torres destruïdes (5):**

Normalment, la quantitat de torres enemigues destruïdes és un clar indicador de les probabilitats que té un equip de guanyar. Si un equip guanyador ha tirat el mínim de torres possibles, això indica que el factor decisiu ha estat una baralla entre els equips a etapes tardanes del joc i han acabat guanyant contra tot pronòstic.

4 Aplicació d'algorismes d'aprenentatge automàtic

Hem escollit els models Random Forest Classifier i Gradient Boosting.

Random Forest Classifier

Un **random forest** és un model d'aprenentatge supervisat que combina múltiples arbres de decisió per realitzar prediccions més precises i robustes. Utilitza un enfocament d'ensamblatge on cada arbre s'entrena amb un subconjunt aleatori de les dades i característiques, i les prediccions finals s'obtenen fent una mitjana (en regressió) o una votació (en classificació) dels resultats dels arbres individuals, cosa que redueix el sobreajustament i millora la generalització. Hem trobat adequat aquest model per què dona bons resultats a l'hora de classificar dades i té en compte varis models entrenats amb subconjunts aleatoris del dataset, el que assegura més fiabilitat per la seva part. A més, recordem que els nostres objectius eren:

- Predir la victòria d'un jugador donat amb informació d'ell i les estadístiques del seu equip.
- Estudiar si hi ha personatges que influeixen més en la victòria que altres.
- Estudiar la rellevància dels diferents objectius neutrals.

Per tant, es tracta al final d'un model de classificació entre victòria i derrota, i veure quina és la importància que té cada variable per arribar a extreure conclusions sobre el que volem. Per això, i pel tipus de dades que tenim, hem creut que un model de Random Forest és adequat. Com que es basa en la combinació de certs arbres de decisió, observem que aquests models en concret segueixen una estructura molt entenedora intuitivament parlant a la hora de fer prediccions, és a dir, creiem que amb aquest tipus de dades (una clara variable objectiu, que és victòria o derrota, i una gran quantitat de variables predictores) és raonable seguir una estructura com la dels arbres de decisió a l'hora de fer prediccions, ja que podem anticipar que hi haurà certes variables que destacaràn molt més per sobre d'altres.

A l'hora de fer el model ens hem trobat amb que la immensa majoria de variables tenien rellevància nul·la, pel que hem fet un filtrat per quedar-nos només amb les més importants, fixant-nos en l'histograma de rellevància i en el nostre coneixement de les dades.

Hem entrenat a continuació un altre model amb aquest filtrat de dades:

A continuació les variables amb les que el model ha sigut entrenat: `championid`, `role`, `win`, `totdmgdealt`, `totheal`, `visionscore`, `totdmgtaken`, `goldearned`, `towerkills`, `inhibkills_y`, `firsttower`, `firstinhib`, `firstbaron`, `firstdragon`, `baronkills`, `dragonkills`, `harrykills`, `firstblood_y`, `totcctimeddealt`, `kda`

Observem que hem afegit una variable nova anomenada `kda`, que és una manera d'ajuntar `kills`, `deaths`, `assists` en $KDA = \frac{kills+deaths}{assists}$.

D'aquest model i l'anterior, i com esperàvem que passaria, concloem que la variable més important és sense dubte el nombre de torres destruïdes per l'equip. "Si eliminem totes les variables que tenen a veure amb l'eliminació d'estructures com de bó seria el nou model?"

La resposta a aquesta pregunta pensàvem que seria negativa, però resulta que el KDA és la variable més rellevant per determinar això i no dona mals resultats.

A continuació els resultats obtinguts a partir dels 3 models:

Esquerra: Métriques de rendiment.

Dreta: Importància relativa.

Accuracy: 0.9470	
Classification Report:	
precision	recall f1-score support
0.0 0.97 0.92 0.95 173458	
1.0 0.93 0.97 0.95 173273	
accuracy	0.95 0.95 0.95 346731
macro avg	0.95 0.95 0.95 346731
weighted avg	0.95 0.95 0.95 346731

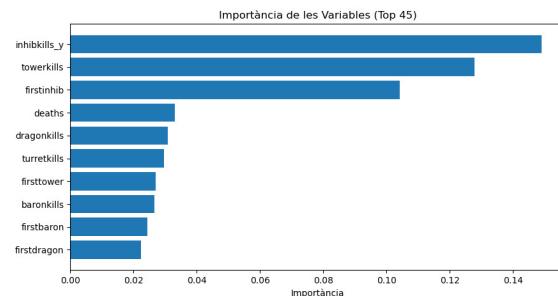


Figura 5: Primer model

Accuracy: 0.9535	
Classification Report:	
precision	recall f1-score support
0.0 0.97 0.94 0.95 173458	
1.0 0.94 0.97 0.95 173273	
accuracy	0.95 0.95 0.95 346731
macro avg	0.95 0.95 0.95 346731
weighted avg	0.95 0.95 0.95 346731

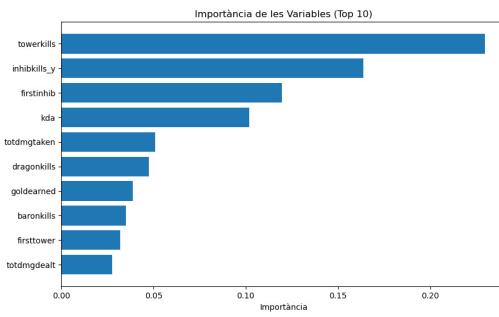


Figura 6: Segon model

Accuracy: 0.8741	
Classification Report:	
precision	recall f1-score support
0.0 0.89 0.86 0.87 173458	
1.0 0.86 0.89 0.88 173273	
accuracy	0.87 0.87 0.87 346731
macro avg	0.87 0.87 0.87 346731
weighted avg	0.87 0.87 0.87 346731

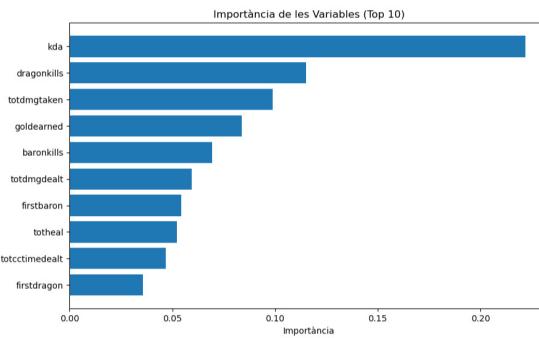


Figura 7: Tercer model

Veiem que la precisió al primer model i al segon és quasi la mateixa, amb una diferència de 0.8%, però quan arribem al tercer model clarament la precisió baixa de un 95% a un 87%, la qual cosa reflecteix un altre cop que les variables més importants són les torretes, com era anticipable.

Gradient Boosting

Els models de tipus **gradient boosting** són robustos i tenen una gran capacitat per captar relacions complexes entre les variables. És un dels models més utilitzats actualment gràcies a la seva forta

capacitat predictiva, i per aquest motiu l'hem seleccionat.

L'objectiu del model és predir, segons les dades d'una partida ja jugada, si un jugador determinat formava part de l'equip guanyador o no. La gran quantitat de dades disponibles i la forta influència que tenen algunes variables en el resultat final permeten fer prediccions amb molta confiança.

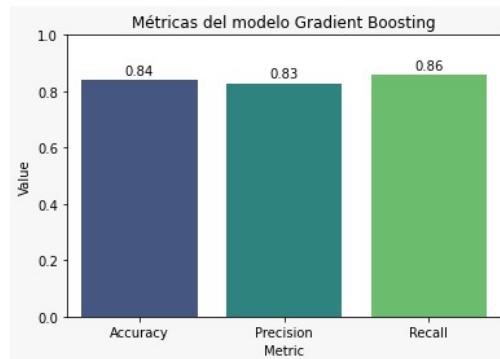
Les mesures estructurals com ara: `towerkills`, `inhibkills_y`, `firsttower` i `firstinhib` són altament determinants en una partida. Fins i tot, gairebé "a simple vista", és possible, seguint criteris senzills i consistents, determinar si el jugador va guanyar amb el seu equip o no. Quan el model disposa d'aquestes variables, ofereix mètriques de precisió (*precision*), exactitud (*accuracy*) i sensibilitat (*recall*) que superen el 95%.

Tanmateix, aquest resultat no és especialment interessant per al nostre estudi, ja que les dades estructurals gairebé garanteixen el resultat. Per aquest motiu, hem entrenat dos models addicionals on hem exclòs aquestes dades estructurals.

- **Primer model:** Incloem les dades relacionades amb el KDA (també molt determinants).
- **Segon model:** Excloem també les dades del KDA per posar a prova encara més el rendiment del model.

Les mètriques i la importància de les característiques (*feature importances*) dels models són les següents:

Esquerra: Mètriques de rendiment.



Dreta: Importància relativa.

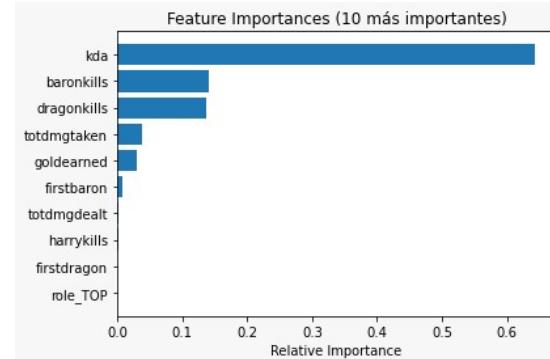


Figura 8: Primer model

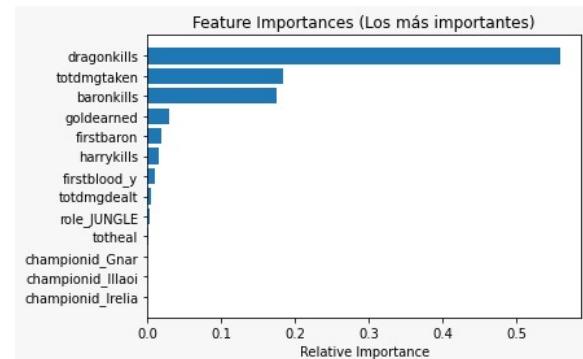
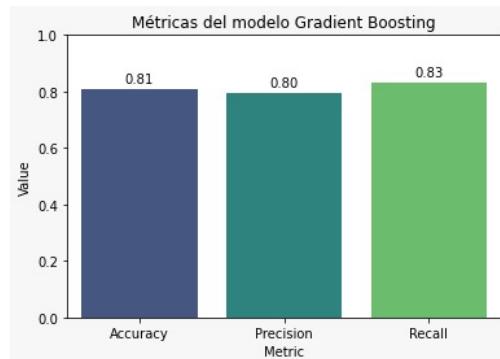


Figura 9: Segon model

5 Resultats i interpretació

Respecte a la comparació entre els models **Random Forest Classifier** i **Gradient Boosting** ens hem de limitar a comparar els resultats presentats per l'últim model de Random Forest i el primer de Gradient Boosting, ja que són ambdós amb el conjunt de variables més interessant. Fixant-nos en les mètriques observem que el Random Forest és el que millors resultats té, però això té una explicació que va més enllà de la diferència de models, i és la diferència del nombre d'estimadors emprats per ambdós models. Per a fer el Random Forest hem usat `n_estimators=100`, `random_state=42`, és a dir, 100 arbres de profunditat il·limitada, a canvi d'un major temps de càrrega computacionalment. Pel Gradient Boosting, en canvi, hem usat (`n_estimators=20`, `learning_rate=0.1`, `max_depth=3`, `random_state=42`), la qual cosa implica un menor temps de càrrega computacional però pitjors resultats. Tot i així amb poca diferència.

Els resultats que hem obtingut no ens sorprenen en general, són resultats esperats. Això succeeix perquè tenim coneixement sobre el tema d'estudi i ja teníem una intuïció prèvia que s'ha acabat complint.

És fàcil d'entendre que en un videojoc que consisteix a destruir la base enemiga, les dades més importants a tenir en compte per predir què equip ha guanyat són les estructures que s'han destruït a la partida.

Un cop obviem les dades referents a les estructures també és comprensible que la importància la tinguin els objectius neutrals, on destaca el baró (objectiu molt important que en moltes partides pot decidir el resultat de les mateixes) i els dracs (que donen estadístiques addicionals a l'equip que li dona el cop de gràcia durant tota la partida).

En aquest aspecte ens sorprèn una mica la diferència negativa en importància de l'objectiu neutral anomenat "herald" (harrykills), ja que aquest objectiu serveix per donar un cop fort a una estructura enemiga i debilitar-la i abans hem vist que les estructures són el més important del joc.

Tot i això, coneixent el context es pot entendre aquest fet, pel fet que aquest objectiu va ser una incorporació al videojoc, és a dir, el videojoc des que existeix té dracs, el baró i les estructures, però l'herald va aparèixer al videojoc més tard (5 anys després del llançament del videojoc, i l'estudi és de partides de 7 anys després del llançament). Això justifica el fet que els jugadors no estiguin tan familiaritzats amb aquest objectiu neutral i no se sàpiga aprofitar tan bé com els altres, tot i el seu potencial.

Un altre tema important és el KDA, i dintre d'aquest concretament les morts (deaths). Per com està fet el joc, és natural que el KDA sigui important, ja que assassinat i assistir en assassinats dona or que es pot usar per comprar millors objectes.

També s'entén que les morts, dintre de les variables que té en compte el KDA sigui la més rellevant, perquè quan un campió mor, aquest desapareix de la partida i no pot tornar a jugar durant uns segons. Aquest temps en què no es pot jugar quan un campió mor va augmentant de manera proporcional al temps de partida, és a dir, si un campió mor al principi de la partida estarà sense jugar 5 segons, però si mor al minut 40 de partida pot estar sense jugar 50 segons. Aquest fet dona molt avantatge a l'equip que ha assassinat al campió, ja que durant aquest espai de temps compta amb avantatge numèric.

Tot i això, sí que hi ha alguns resultats que ens han sorprès, com el fet que els campions escoïllits és un factor que en general influeix poc en el resultat de la partida. Sobretot comparant-ho

amb els altres factors de la partida.

Si només tenim en compte la rellevància dels campions, els resultats un cop més no sorprenden, ja que els que tenen més importància són els més jugats que, alhora, són els més jugats generalment perquè són els més forts. Hi ha altres raons per les quals un campió es juga més, com és el cas de campions que s'han afegit fa poc al videojoc o campions que han rebut noves "skins" (aparences diferents del campió) fa poc.

Dintre d'això és lògic que els rols tinguin més importància que els campions que es juguen en si, ja que un rol engloba molts campions. També l'ordre dintre dels rols és natural, pel fet que el midlaner es troba al mig del mapa i té relativament fàcil el fet d'influir als altres dos carrils o al jugador de la jungla. El mateix es pot dir del jungla, que juga per tot el mapa i pot ajudar a tots els seus companys o fer-se més fort aprofitant possibles assassinats a les diferents línies. També és el jugador amb el rol de jungla l'encarregat d'assegurar els objectius neutrals, ja que tenen un summoner spell i un objecte dedicats a això dins del joc.

Endinsant-nos dintre dels campions trobem els objectes que ha comprat cadascun d'ells. També és natural que els diferents objectes tinguin una rellevància interior i aquesta, de fet, ja es té en compte a la rellevància dels campions perquè el fet que un campió sigui millor es pot donar per diferents raons: les estadístiques pròpies del campió són molt bones, ja sigui que fa molt dany, que aguanta molt el dany que se li fa o que té habilitats complexes que tot i no fer dany o evitar-ho ajuden molt a l'equip o perjudiquen els jugadors de l'equip enemic. D'altra banda, si els objectes que s'acostumen a comprar amb un campió tenen molt bones estadístiques o una molt bona habilitat, el campió que els compra es veu beneficiat.

Per últim, cal tenir en compte que les dades són de partides de tots els nivells, tot i ser partides on la gent s'hi juga el seu rang competitiu. En partides dels jugadors professionals aquests paràmetres que hem analitzat tindrien diferents rellevàncies.

Un que veiem molt clar és el visionscore, que consisteix en la informació del mapa que cada jugador ha obtingut mitjançant uns objectes de visió que hi ha al joc (objectes invisibles que donen visió d'una part del mapa, objectes que permeten veure aquests objectes invisibles, etc.). La visió és molt més important en partides professionals, ja que els jugadors estan comunicant-se verbalment, cosa que no passa en general a les partides no professionals, com a molt dues persones poden jugar juntes en aquestes partides.

6 Conclusions i treball futur

En aquest estudi hem utilitzat els models Gradient Boosting i Random Forest per predir si un jugador formava part de l'equip guanyador d'una partida de joc, basant-nos en dades tant individuals com globals de la partida. La complexitat inherent d'aquestes dades requereix models capaços de captar relacions subtils entre les variables i el resultat (target), les quals no són fàcilment observables mitjançant estudis estadístics simples o la nostra pròpia experiència empírica del joc.

Els models seleccionats no només han demostrat una alta capacitat predictiva, sinó que també ens han permès aprofundir en el grau de determinisme de certes variables. Això ens ha ajudat a entendre millor quins valors són més influents a l'hora de decidir el resultat d'una partida. Aquest coneixement és especialment valuós, ja que ens proporciona una visió més clara de com alguns elements del joc impacten directament en la victòria o derrota.

Aquest enfocament pot tenir aplicacions futures tant en el disseny d'estratègies de joc com en l'anàlisi de rendiment d'equips i jugadors i és un exemple de com un model pot ajudar no només a predir però a aportar entendiment sobre el funcionament profund d'un procés complex. A l'hora de fer aquest treball hem tingut clares limitacions. Algunes d'aquestes limitacions són:

- **Volum de dades:** Com a cada partida participen 10 jugadors i cada jugador té moltes variables associades com ja hem vist al treball, el processament de les dades es fa molt pesat tot i prescindir d'algunes variables.
El gran volum de dades resultava beneficiós per a l'entrenament dels models, però complicava la gestió, així com la tasca de neteja i la selecció de les dades més rellevants.
- **Tenir en compte diferents estats de les partides:** El que hagués estat més interessant de fer en aquest tipus de projectes seria un càcul a temps real de les probabilitats de guanyar de cada equip, que de fet a les partides professionals del mundial anual de League of Legends ja hi és implementada aquesta característica i de tant en tant mostren quina és la probabilitat de victòria de cada equip en aquell moment. Si més no, almenys fer una predicción de victòria amb les dades d'un minut concret intermedi, com per exemple seria el minut 20 o el 15, moments en els quals no hi ha encara possibilitat de rendició, i que faria interessant un enfocament de predicción a futur d'una partida a mig fer.

També tenim propostes del que es podria fer com a continuació d'aquest treball amb més recursos:

- **Aplicar el que aprenem als bots del joc:** És una molt bona oportunitat per a millorar els bots que té el joc. Aquests són molt limitats en termes de com juguen al videojoc. Juguen molt malament i són més una eina per introduir-se al videojoc que per millorar el rendiment un cop saps jugar.

Es podria fer un projecte de cara a afegir diferents nivells de joc dels bots, és a dir, millorar molt la qualitat dels bots fins al punt de crear un prototip de bots que guanyessin el 100% de les partides (com succeeix als escacs). Seria una eina molt útil, per exemple pels jugadors professionals, per poder estudiar el joc i veure com entén el joc una màquina "perfecta". Aquí també sorgeix el debat de si existeix la forma perfecta de jugar al League of Legends, ja que hi ha diferents maneres d'interpretar el joc.

- **Exemple de la vida real:** Recentment, hem trobat per xarxes socials un treball de machine learning semblant al que hem fet nosaltres realitzat per un equip professional de League of Legends espanyol (Giants Gaming).

Hem trobat que han estat treballant en una eina que calcula, segons els personatges que ja s'han escollit a la partida (aliats i enemics), quin és el campió que té una probabilitat més alta de guanyar la partida. Tenint en compte les sinergies amb els campions aliats i els desavantatges davant dels campions enemics.

És un treball més centrat en els campions, el nostre treball és més general i té en compte variables que aquest treball no enfoca amb tanta importància. D'altra banda, el seu treball té un enfocament més comercial.

7 Annexos

El codi utilitzat pel treball està com fitxers separats, a banda del pdf.

Les dades han sigut obtingudes de la pàgina web:

<https://www.kaggle.com/datasets/paololol/league-of-legends-ranked-matches>

On la persona que va penjar els dataframes a la pàgina web comenta que les va obtenir de la base de dades de Riot Games (els propietaris del videojoc). Aquesta base de dades es pot consultar a la pàgina web següent: <https://developer.riotgames.com/docs/lol#data-dragon>