

Dota 2 Statistic Analysis

Antonio Ambrosio

Ranieri Cristian

Indice

1	Introduzione	2
1.1	Gameplay	2
1.2	Research Questions	2
1.3	Specifica del dataset	3
1.4	Pulizia preliminare del dataset e selezione dei dati	3
1.5	Generazione del dataset sintetico	5
2	Dataset originale	6
2.1	RQ1:Il winrate varia notevolmente tra gli Hero?	6
2.2	RQ2:winrate e pickrate variano tra ranked e unranked?	7
2.3	RQ3:I winrate dei Radiant e dei Dire sono differenti?	16
2.4	RQ4:Esiste una correlazione tra pickrate e winrate?	18
2.5	RQ5:Esistono degli attributi più incisivi all'interno del gioco?	22
2.6	RQ6:Esistono configurazioni dei team più convenienti?	28
2.7	RQ7:Qual'è il rank delle istanze del dataset?	31
3	Dataset sintetico	35
3.1	RQ1:Il winrate varia notevolmente tra gli Hero?	35
3.2	RQ2:winrate e pickrate variano tra ranked e unranked?	37
3.3	RQ3:I winrate dei Radiant e dei Dire sono differenti?	43
3.4	RQ4:Esiste una correlazione tra pickrate e winrate?	43
3.5	RQ5:Esistono degli attributi più incisivi all'interno del gioco?	45
3.6	RQ6:Esistono configurazioni dei team più convenienti?	51
3.7	RQ7:Qual'è il rank delle istanze del dataset?	52
3.8	Conclusioni	53

1 Introduzione

Quella che segue è un'analisi statistica sul gioco multiplayer Dota2, un videogioco uscito nel 2013 sviluppato da Valve. Durante questa analisi si è cercato di indagare lo stato del gioco attraverso le varie research questions. Le varie research questions ci permetteranno di fare chiarezza per quanto riguarda il bilanciamento del gioco e di determinare se il nostro studio è rappresentativo o meno rispetto alla popolazione.

1.1 Gameplay

Di seguito sono introdotti una serie di aspetti del gioco che saranno utili per dare un contesto e comprendere meglio il resto dell'analisi statistica.

In ciascuna partita di Dota2 due team (Radiant e Dire) formati da 5 giocatori si scontrano all'interno della mappa di gioco. La mappa in questione è suddivisa in due parti, ciascuna assegnata e difesa da uno dei due team.

Ciascun team avrà l'obiettivo di distruggere l'Ancient nemico, ovvero una struttura localizzata al centro della base del nemico. Per riuscire nell'intento è però necessario riuscire a ottenere un vantaggio tattico e statistico sul team avversario. Per riuscire a guadagnare il vantaggio statistico è necessario raccogliere esperienza e migliorare il proprio personaggio attraverso degli oggetti. Tali oggetti sono acquistabili attraverso una moneta di gioco che può essere acquisita sconfiggendo i giocatori del team avversario durante la partita oppure conquistando degli obiettivi.

In Dota2 esistono poi varie modalità di gioco, in cui alcune caratteristiche del gioco cambiano. Per ciascuna modalità di gioco si può scegliere poi di fare una partita ranked oppure unranked. La differenza è che in una partita ranked è possibile guadagnare o perdere punti rango rispettivamente se si vince o si perde, mentre nelle unranked questo non avviene.

Ciascun giocatore di ciascuno dei team può scegliere l'hero (personaggio giocabile con abilità e caratteristiche uniche) da giocare. Tale scelta viene effettuata durante una fase preliminare alla partita, dove possono essere effettuate le prime scelte tattiche e dove ci si può organizzare con i membri del proprio team.

1.2 Research Questions

Le research questions che sono state affrontate durante questo studio statistico sono le seguenti:

- RQ1: Il winrate varia notevolmente tra gli Hero?
- RQ2: Il winrate e pickrate variano tra ranked e unranked?
- RQ3: I winrate dei Radiant e dei Dire sono differenti?

- RQ4: Esiste una correlazione tra pickrate e winrate?
- RQ5: Esistono degli attributi più incisivi all'interno del gioco?
- RQ6: Esistono configurazioni dei team più convenienti?
- RQ7: Qual'è il rank delle istanze del dataset?

1.3 Specifica del dataset

Il dataset che è stato utilizzato per effettuare il seguente studio è un insieme di partite al gioco in cui sono presenti i seguenti campi:

- gamemode: Viene riportato l'identificativo numerico della modalità di gioco della partita.
- gametype: Viene riportato se la partita è stata effettuata in modalità ranked o in modalità unranked attraverso un identificativo numerico.
- win: Viene riportato 1 o -1 a seconda di quale delle due squadre ha vinto la partita.
- una colonna per ciascuno degli Hero presenti nel gioco(il nome di ciascuna colonna è hero con concatenato l'identificativo numerico del Hero): nella colonna è presente 0 se l'Hero non è stato giocato da nessuno dei due team, è presente 1 o -1 invece se è stato giocato rispettivamente da uno dei due team.

Durante le fasi successive dello studio, il seguente dataset è stato arricchito dalle informazioni informazioni degli Hero che sono state recuperate direttamente dal sito ufficiale di Dota2 [1] attraverso gli identificativi degli eroi lasciati all'interno di un file json fornito insieme al dataset.

1.4 Pulizia preliminare del dataset e selezione dei dati

Il dataset originale necessitava per poter essere analizzato correttamente una fase preliminare di pulizia. Durante questa fase iniziale, ci siamo resi conto che la maggior parte delle partite riguardavano in realtà una sola gamemode e due sole altre gametype, come mostrato dai seguenti grafici 1-2. Si è deciso quindi di considerare solamente la gamemode 2 e solamente le gametype 2 e 3. Per quanto concerne tutte le altre gamemode, si è ritenuto necessario eliminare quei dati dal dataset, in quanto non avevamo abbastanza istanze per determinare qualcosa rispetto a quelle gamemode. Per quanto riguarda la gametype 1, c'era un numero molto basso di istanze per queste partite, e sono state quindi eliminate.

Il nostro studio si è ristretto solamente a una modalità di gioco e alle gametype 2 e 3 che

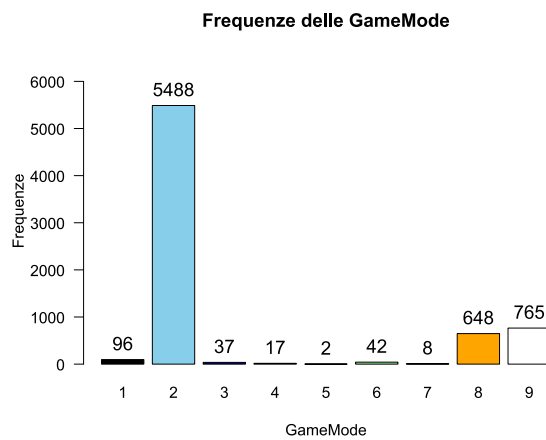


Figura 1: Nella seguente figura è riportata la frequenza delle gamemode

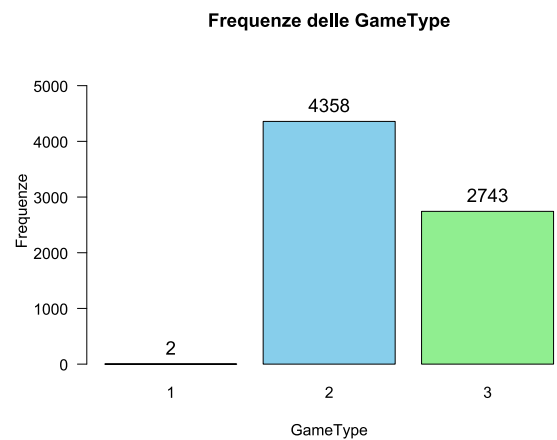


Figura 2: Nella seguente figura è riportata la frequenza dei gametype

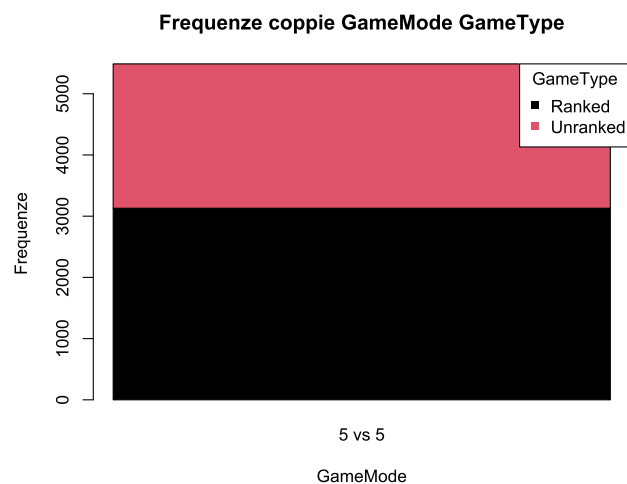


Figura 3

abbiamo poi successivamente mappato a partite classificate e non classificate attraverso le informazioni recuperate sul sito di Dota2.

1.5 Generazione del dataset sintetico

Di seguito è riportata la tecnica utilizzata per generare il dataset. Il Large Language Model che è stato utilizzato è Chat-Gpt 4.0. L'approccio utilizzato inizialmente utilizzato era quello di dare in input al LLM una porzione del dataset per far sì che si allenasse su di esso per poi fargli generare le istanze sintetiche. L'elevato numero di colonne(147 dopo aver integrato i nuovi dati) e di istanze del dataset(5500) portavano LLM a rifiutare il dataset.

Pur utilizzando altri LLM come Gemini e/o diminuendo il numero di istanze fino a 500 il risultato non è cambiato. Di conseguenza si è deciso di utilizzare Chat-Gpt con un approccio differente. L'approccio utilizzato è basato sulla Data Generation con Feedback, che si basa su 4 passaggi:

- Fornire il contesto al LLM dei dati che devono essere generati.
- Fornire degli esempi di input.
- Richiedere la generazione di nuovi dati.
- Fornire informazioni di feedback riguardanti i dati generati per il miglioramento nella generazione dei dati.

Nel tentativo di generare le istanze del dataset nel modo migliore possibile il valore di vittoria non è stato fatto generare da Chat-Gpt. Chat-Gpt non sarebbe stato infatti in grado di generare questo valore, si è deciso quindi di tentare l'utilizzo di modelli di LLM e reti neurali sulle istanze del dataset originale per effettuare successivamente la predizione del valore di vittoria.

Il risultato migliore tra tutti che è stato ottenuto attraverso un algoritmo di classificazione random forest. Il valore di accuratezza raggiunto è 0.55. Si è ritenuto questo valore accettabile dato che all'interno del dataset originale mancano le informazioni che avrebbero permesso di predire in maniera più efficace il valore della colonna win(come ad esempio le informazioni che riguardano l'andamento dei giocatori).

2 Dataset originale

2.1 RQ1: Il winrate varia notevolmente tra gli Hero?

La prima research question riguarda un'analisi della variabilità del winrate di ciascun Hero. Quando ci riferiamo al winrate, intendiamo il rapporto tra le partite vinte e le partite giocate. Nel caso specifico, del winrate di un Hero, si intende il rapporto tra il numero di partite vinte con un Hero e il numero di partite perse con quel determinato Hero.

Avere contezza della variabilità del winrate è una informazione importante per determinare lo stato del gioco. Avere una certa variabilità nel winrate dei campioni, potrebbe indicare che alcuni campioni rispetto ad altri consentono di arrivare più facilmente alla vittoria.

Nel contesto dei giochi MOBA come Dota, dove regna la competizione, non aver bilanciato

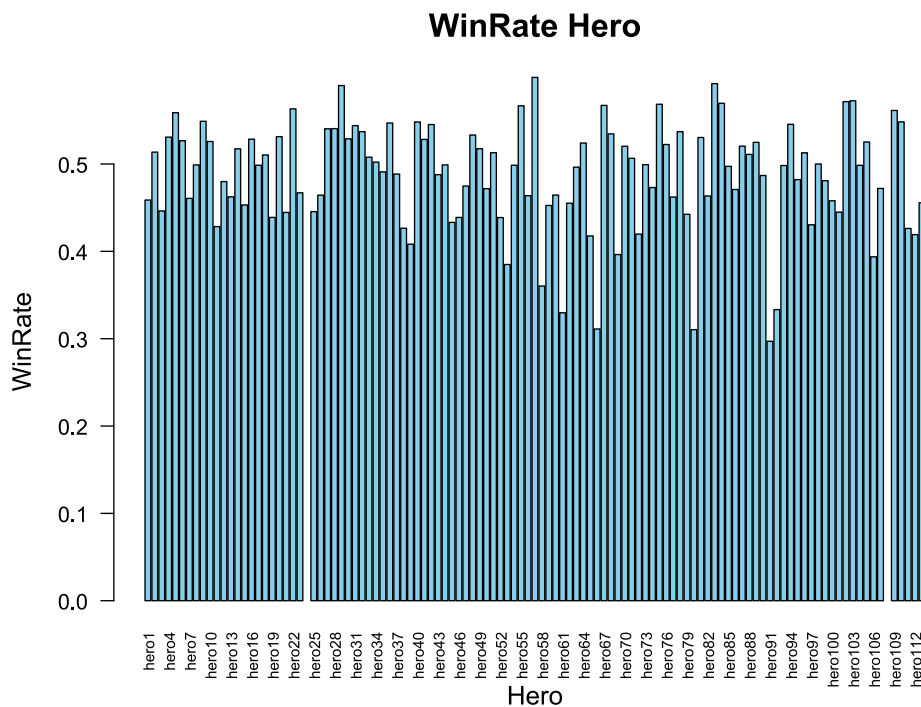


Figura 4: Barplot che mostra il winrate per ciascuno degli Hero

tutti quanti gli Hero potrebbe costringere i giocatori a giocare solamente gli Hero che garantiscono un vantaggio netto sugli altri.

Il primo passo dell'analisi è stato quello di calcolare la varianza del winrate degli Hero, il cui valore è 0.003772695. Questo valore non indica che c'è una forte variabilità tra il

winrate dei vari Hero. Si è deciso quindi di rappresentare attraverso un barplot il winrate di ciascuno degli Hero all'interno della seguente figura 4. All'interno del grafico è possibile notare come siano presenti degli Hero che hanno un winrate pari a zero. In effetti, gli hero 24 e 108 non compaiono all'interno di nessuna delle partite, non essendo giocati, hanno un winrate non determinabile. E' possibile però anche notare che è presente una certa variabilità nel winrate dei vari Hero. In particolare alcuni Hero scendono fino ad un winrate dello 0.3 mentre altri superano addirittura lo 0.5. Successivamente si è deciso di analizzare il winrate degli Hero distinguendo le partite ranked da quelle unranked. La varianza del winrate degli hero sono 0.004792672 e 0.004651123 rispettivamente per le partite ranked e unranked. Non c'è quindi una particolare variazione rispetto alla varianza nel caso in cui le abbiamo considerate tutte insieme. Andando ad analizzare i grafici nelle figure 5 e 6, che sono rispettivamente quelli delle partite ranked e unranked, non si notano particolari differenze rispetto al grafico dove abbiamo considerato tutte quante le partite insieme.

Andando ad analizzare successivamente la distribuzione dei winrate per le partite ranked e quelle unranked(figura 7), ci siamo resi conto che in entrambi i grafici, ci sono un grande numero di Hero che ricadono nel range di winrate di 0.45 e 0.55. In entrambi i grafici è chiaro però che ci sono un grande numero di campioni che hanno un winrate molto più alto rispetto agli altri e altri con un winrate molto più basso(35%).

Riassumendo, gli indici di sintesi delle variabilità, non suggeriscono una forte varianza all'interno dei due campioni. In questo caso però è necessario sottolineare che anche variazioni minime nel winrate, che non possono essere "catturate" da un indice come la varianza, possono determinare uno squilibrio all'interno del gioco. Un buon numero di Hero risiedono all'interno delle code a sinistra e destra dei grafici delle distribuzioni in figura 7. Questo ci permette di determinare che un buon numero di Hero non si trovano in un winrate che varia tra 0.45 e 0.55. Rispetto a questo risultato ottenuto si può concludere che il gioco non è bilanciato, in quanto ci sono un grande numero di Hero per cui c'è una varianza troppo forte del winrate rispetto agli altri.

2.2 RQ2:winrate e pickrate variano tra ranked e unranked?

Durante le fasi preliminari di pulizia del dataset, si è compreso che il dataset era composto da un mix di partite ranked e unranked.

All'interno della seguente research question si cercherà di comprendere se esistono delle differenze notevoli tra il pickrate e il winrate tra le partite ranked e quelle unranked. L'interesse nel rispondere a questa research question è dato dal fatto che ci potrebbe permettere di verificare se sussistono delle differenze nel modo in cui i giocatori approcciano le partite classificate e quelle non classificate.

Si potrebbe infatti verificare che dato che all'interno delle partite ranked c'è la possibilità di guadagnare o perdere dei punti in una classifica i giocatori compiano delle scelte più

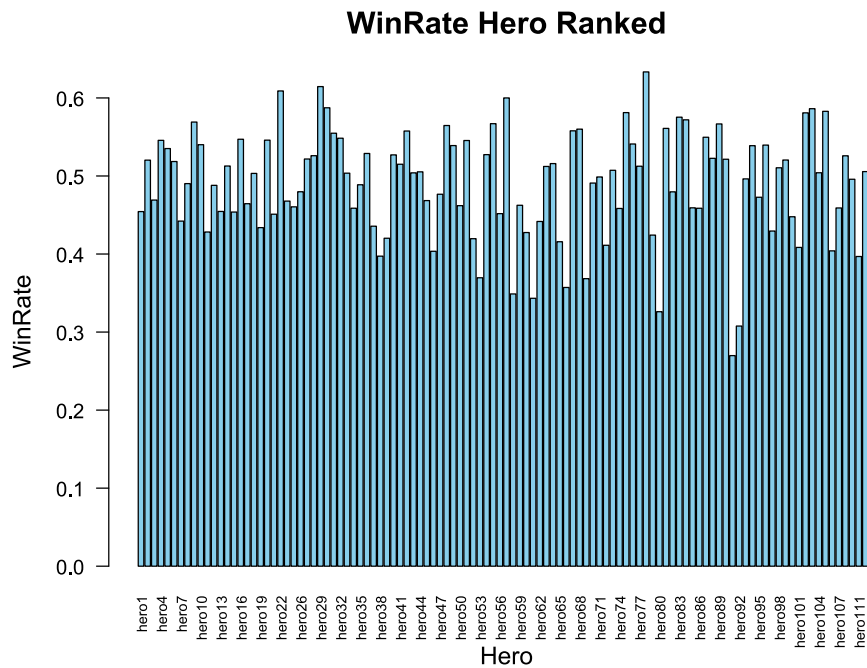


Figura 5: Barplot che mostra il winrate per ciascuno degli Hero nelle partite ranked

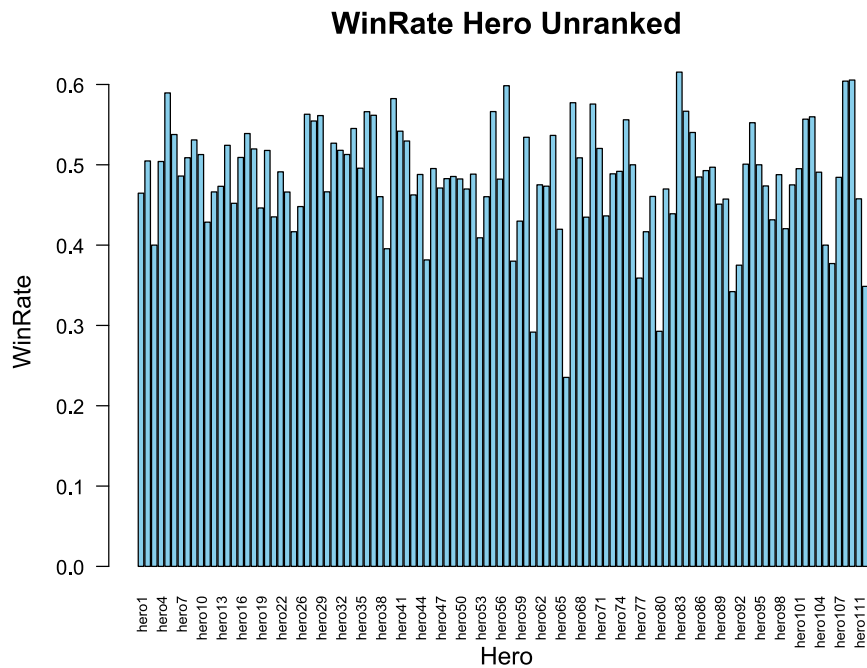


Figura 6: Barplot che mostra il winrate per ciascuno degli Hero nelle partite unranked

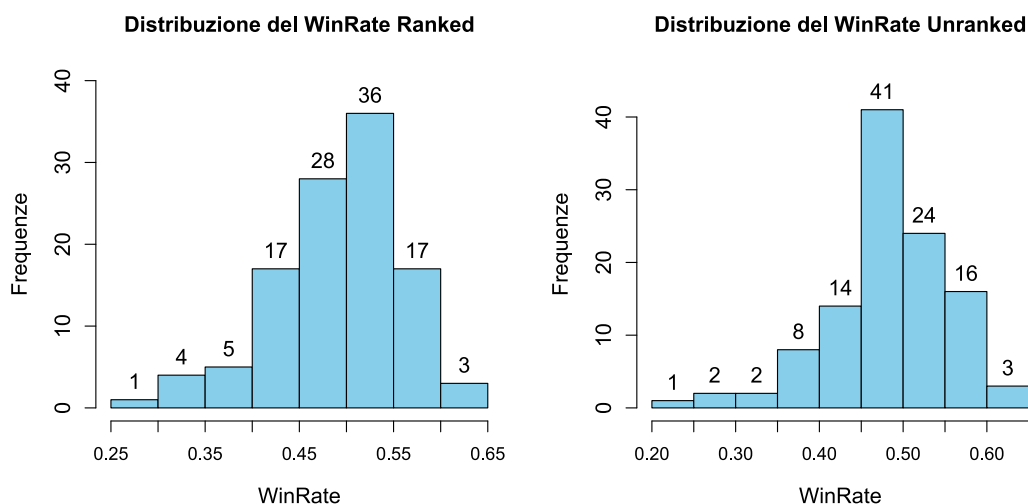


Figura 7: Figura che mostra i barplot per la distribuzione del winrate nelle partite ranked e in quelle unranked

”sicure”. A differenza delle partite non classificate che gli potrebbero dare la possibilità di esplorare nuove possibilità.

Questo si rifletterebbe sul fatto che la scelta dei campioni varia particolarmente tra le partite classificate e non classificate. Per verificare se questo avviene, si è deciso di effettuare un’analisi del pickrate per confrontare come varia tra le due modalità.

All’interno della research question 1 abbiamo già notato alcune differenze tra il winrate all’interno delle partite ranked e unranked, confrontare le differenze tra questi due è interessante per verificare se effettivamente alcuni campioni funzionano meglio all’interno di un contesto di classificate o meno.

La definizione precisa di pickrate è il rapporto tra le partite in cui viene scelto un determinato Hero rispetto al numero totale di partite. All’interno della figura 8 sono rappresentati i pickrate degli Hero nelle partite ranked e unranked. Per rappresentare al meglio la differenza tra il pickrate nelle due modalità, si è deciso di rappresentare tale differenza attraverso il seguente barplot 9. All’interno del grafico è stata riportata una linea rossa che rappresenta il margine secondo il quale non consideriamo sostanziale le differenze dei pickrate dei campioni (2.5%). Fissato questo margine, sembrerebbe che non ci sono grandi differenze tra il pickrate degli Hero nelle due modalità, se non per qualcuno in particolare per cui si nota addirittura una differenza del 10%. Nella seguente figura 10 sono riportati i boxplot ad intaglio per il pickrate degli hero nelle partite ranked e unranked. Da questo grafico, oltre a notare la presenza di alcuni outlier, è possibile notare che le tacche dei due boxplot ad intaglio sembrano sovrapporsi tra di loro.

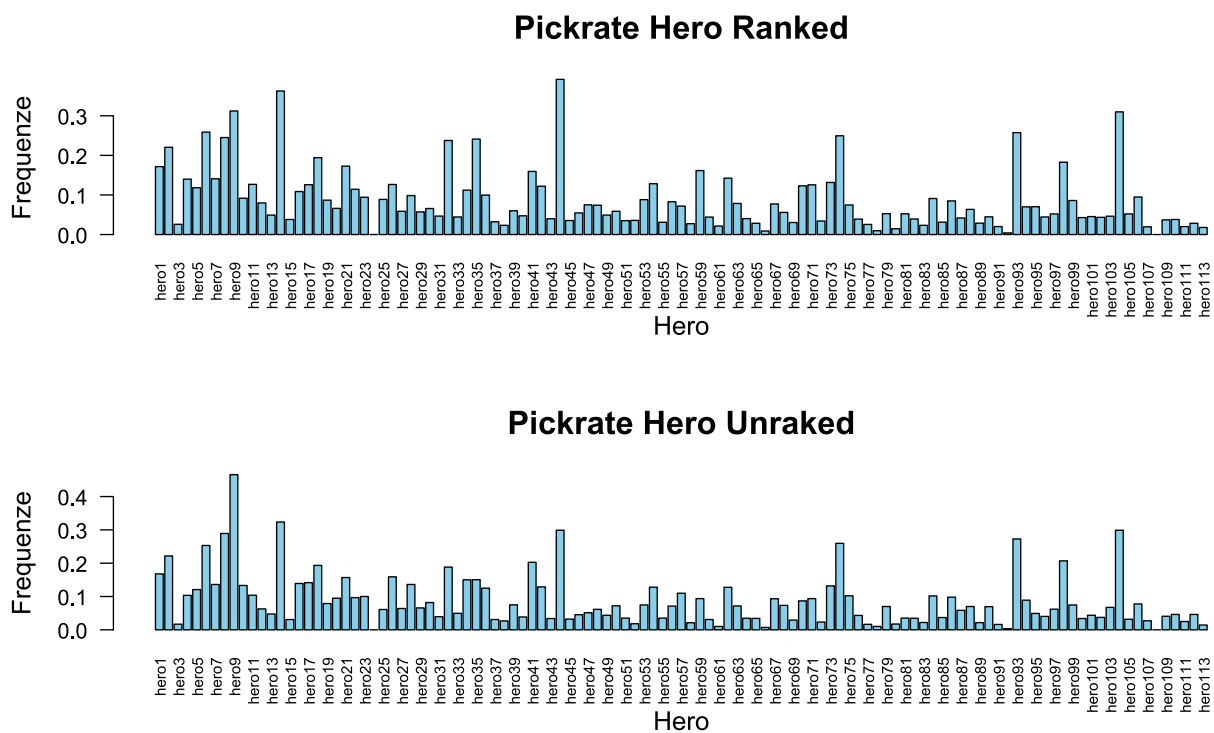


Figura 8: Nella seguente figura sono mostrati i barplot per il pickrate degli hero nelle partite ranked e unranked

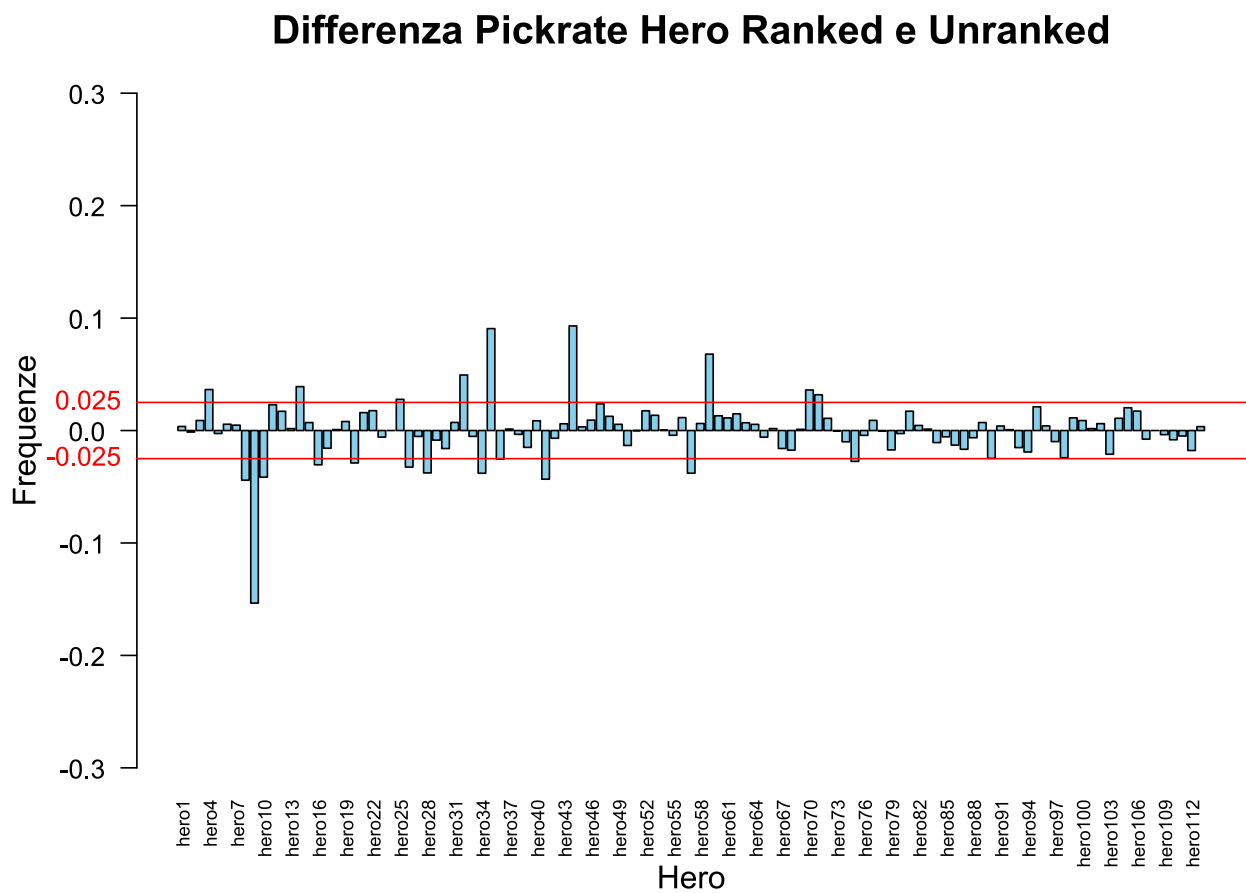


Figura 9: Nella seguente figura è mostrato il barplot per la differenza del pickrate degli hero nelle partite ranked e unranked

Calcolando l'intervallo di confidenza, questa prima impressione viene confermata $(-0.03192, 0.01719)$. Si può quindi concludere, che effettivamente come suggerito dai boxplot ad intaglio non c'è una differenza statistica evidente tra le mediane dei due campioni. Per

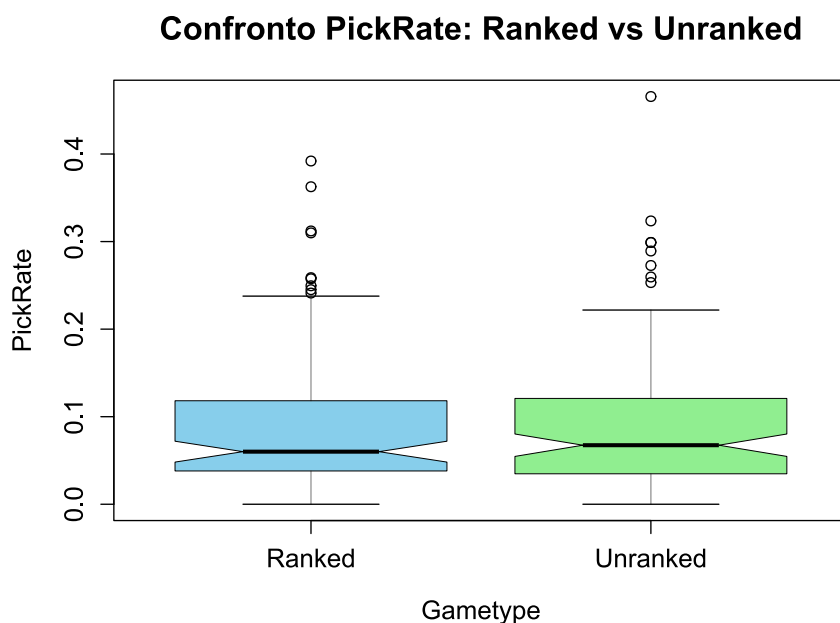


Figura 10: Nella seguente figura sono mostrati i boxplot ad intaglio per il pickrate nelle partite ranked e unranked

confrontare, ulteriormente il pickrate nelle modalità classificata e non classificata, sono stati confrontati gli indici di sintesi che sono riportati all'interno della seguente figura 1. Si può notare che la media è identica, la mediana è molto simile, confermando le informazioni che avevamo ricavato dal boxplot ad intaglio e anche deviazione standard e varianza, risultano essere molto simili. Come si può inoltre notare a partire dai grafici nella figura 11 anche la forma della distribuzione del pickrate, non varia particolarmente tra le partite ranked e quelle unranked. A conferma di questa impressione restituita dai grafici sono stati riportati i valori di skewness e curtosi all'interno della seguente tabella 2. Per quanto riguarda la skewness si può notare che hanno entrambe un valore maggiore di 0 e questo implica che c'è un'asimmetria positiva in entrambe le distribuzioni. I valori di skewness, oltre che essere entrambi maggiori di 0, sono anche molto simili tra di loro. Per quanto riguarda la curtosi è presente una certa variazione nei due casi. Nonostante questa variazione, le distribuzioni sono entrambe leptocurtiche dato che hanno una curtosi maggiore di 3. Riassumendo quindi le informazioni che sono state mostrate riguardo l'analisi del pickrate nelle due modalità, si può concludere che non sembrano sussistere

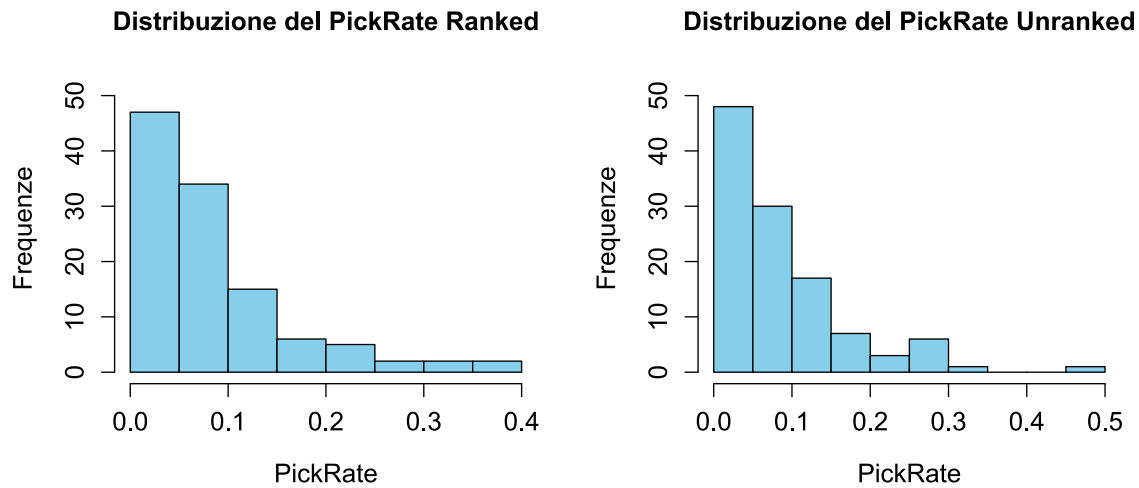


Figura 11: Nella seguente figura sono mostrati gli istogrammi del pickrate per le partite ranked e quelle unranked

Tabella 1: Tabella riassuntiva degli indici di sintesi per il pickrate per le partite ranked e unranked

	ranked	unranked
media	0.08850	0.08850
mediana	0.06006	0.06743
deviazione standard	0.07793	0.07987
varianza	0.00607	0.00638

Tabella 2: Tabella riassuntiva degli indici simmetrici per pickrate nelle partite ranked e unranked

	ranked	unranked
skewness	1.73976	1.89565
curtosi	5.95936	7.39879

delle differenze sostanziali riguardanti il pickrate.

E' importante sottolineare però che se gli indici di sintesi hanno valori simili e le distribuzioni hanno anch'esse una forma simile, sussistono alcune notevoli differenze per alcuni Hero che è possibile notare all'interno della figura 9. Si procede quindi all'analisi del winrate per verificare se ci sono delle differenze interessanti tra i due campioni analizzati. Si osservano anche questa volta attraverso i barplot il winrate delle partite unranked e delle partite ranked 5-6.

Durante la research question 1, si è determinato che lo squilibrio valeva sia per le partite ranked che per quelle unranked. L'analisi che viene riportata di seguito serve ora per determinare se ci sono Hero per cui è presente una differenza significativa del winrate tra le due modalità.

All'interno della figura 12 sono riportati i boxplot ad intaglio del winrate. A partire da questo grafico è possibile notare un basso numero di outlier rispetto al caso del pickrate. Questo è un risultato prevedibile, dato il fatto che una minima variazione nel winrate ha un forte impatto sull'andamento del gioco, è ovvio che i valori del winrate non possano allontanarsi singolarmente troppo rispetto alla media. Inoltre, si può notare come le due mediane non siano statisticamente così differenti tra di loro. Questo è confermato dall'intervallo di confidenza $(-0.009455, 0.040147)$, che ci conferma che le due mediane non sono statisticamente così differenti.

Anche gli indici di sintesi riportati all'interno della seguente tabella 3 suggeriscono che non ci sono differenze sostanziali tra le due modalità di gioco. E' possibile osservare all'interno della seguente figura 7 i grafici che mostrano la distribuzione del winrate all'interno delle due modalità. Anche in questo caso le due distribuzioni non mostrano sensibili differenze. Questo è confermato dalla curtosi e dalla skewness delle distribuzioni che sono riassunte all'interno della seguente tabella 4. La skewness ci mostra che in entrambi i casi abbiamo un'asimmetria destra, difatti i due valori entrambi negativi. La curtosi invece ci dice che sono entrambe leptocurtiche dato che i due valori sono entrambi maggiori di 3. In questo caso però la distribuzione nel caso ranked è al limite tra essere leptocurtica e normocurtica.

Analizzando solamente gli indici e osservando le distribuzioni, sembrerebbero non sussistere delle grandi differenze tra i campioni utilizzati. E' da sottolineare però che gli strumenti utilizzati fino ad ora, ci avrebbero permesso solamente di osservare dei veri e

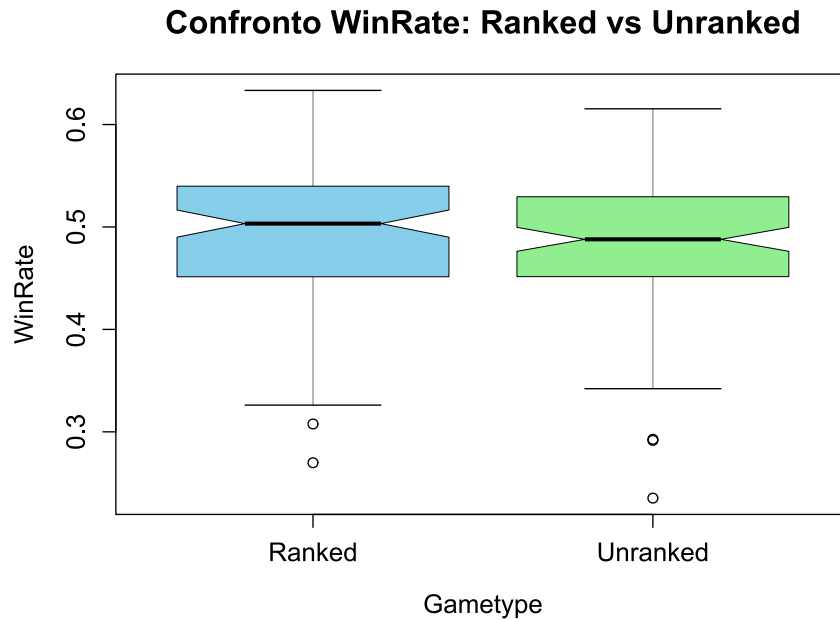


Figura 12: Nella seguente figura sono mostrati i boxplot ad intaglio per il winrate ranked e quelle unranked

Tabella 3: Tabella riassuntiva degli indici di sintesi per il winrate per le partite ranked e unranked

	ranked	unranked
media	0.48950	0.48340
mediana	0.50330	0.48790
deviazione standard	0.06922	0.06819
varianza	0.00479	0.00465

Tabella 4: Tabella riassuntiva degli indici simmetrici per winrate nelle partite ranked e unranked

	ranked	unranked
skewness	-0.57398	-0.78510
curtosi	3.259576	4.24417

propri fenomeni di massa che sussistevano in una modalità di gioco piuttosto che un'altra. Si ritiene che per quanto riguarda lo studio del winrate, non siano interessanti solamente delle differenze macroscopiche tra i due campioni ma anche delle differenze microscopiche e che interessano quindi i singoli Hero.

Si è deciso quindi di riportare in maniera sintetica le differenze che sussistono nei winrate dei singoli Hero attraverso il grafico in figura 13. Anche in questo caso è stato fissato un valore a partire dal quale la differenza tra i winrate comincia a diventare significativa (5%). E' possibile notare come per un grande numero di Hero (37.17%) la differenza superi il valore fissato.

Riassumendo quindi, rispetto ai risultati ottenuti, non è stato possibile individuare delle differenze significative nei pickrate per le due modalità. Per quanto riguarda il winrate invece, seppure non sono emerse delle differenze significative nelle distribuzioni del winrate, sussistono invece delle differenze che possono essere ritenuti significative per quanto riguarda i singoli campioni. Si può quindi concludere che ci sono degli Hero che sembrano molto più funzionanti rispetto ad altri in ranked rispetto che in unranked.

2.3 RQ3:I winrate dei Radiant e dei Dire sono differenti?

Come spiegato all'inizio in Dota2 i giocatori vengono assegnati a uno dei due team Radiant e Dire e l'obiettivo durante una partita è quello di difendere il proprio lato della mappa ed espugnare la base avversaria.

La mappa di gioco è quindi divisa in due parti che non risultano però essere totalmente simmetriche. Le asimmetrie sulla mappa, hanno suggerito spesso ai giocatori quelli che sono dei vantaggi tattici che è possibile sfruttare. Alcune di queste asimmetrie, rendono la navigazione della mappa più sicura, altre rendono alcuni obiettivi più facili da raggiungere per un team piuttosto che un altro.

Una domanda che quindi può sorgere in merito a Dota2 è se esiste effettivamente un lato della mappa che risulta essere vantaggioso rispetto ad un altro, e soprattutto se questo vantaggio si traduce in un winrate maggiore per uno dei due lati.

Questa domanda riguarda in realtà non solo Dota2, ma l'intero genere dei MOBA. All'interno del seguente articolo [2] è riportata una discussione tecnica e tattica su quelli che sono i vantaggi nel trovarsi in un lato della mappa piuttosto che un altro in League Of

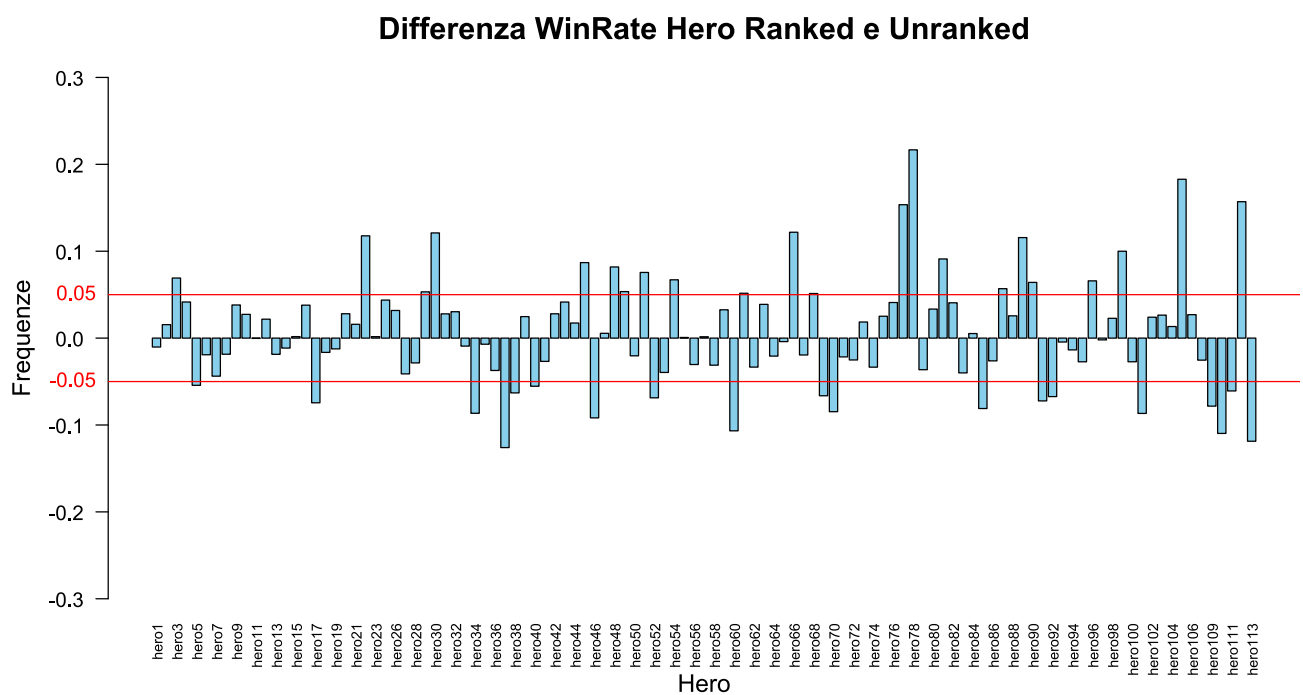


Figura 13: Nella seguente figura è riportato il grafico dove è rappresentata la differenza del winrate tra le partite ranked e quelle unranked

Tabella 5: Tabella riassuntiva del winrate di ciascuno dei due team per le partite ranked e unranked

	ranked	unranked
team 1	47.61%	48.43%
team 2	52.39%	51.56%

Legends, un altro gioco MOBA molto famoso. In tale articolo, è presente una citazione alla frase del coach del team competitivo T1 a seguito di una sconfitta in un torneo competitivo mondiale. Il coach dei T1 afferma quanto segue:” *The reason we lost was, as I just said, because we weren’t able to play on the blue side three times*”. Il fatto che un coach di un team competitivo mondiale attribuisca il motivo della sconfitta del suo team al lato della mappa assegnato suggerisce che anche questo genere di fattori può notevolmente influire su quello che è lo stato del gioco, rendendo il gioco meno bilanciato. Nella seguente research question si cerca di analizzare il dataset per determinare se è possibile che ci sia un vantaggio rispetto al team in cui si gioca e quindi al lato della mappa che viene assegnato. Nel tentativo di determinare se esiste un vantaggio simile, si è calcolato il winrate di ciascuno dei due team per le classificate e per le non classificate, riportandolo all’interno della seguente tabella 5. All’interno della tabella sono stati riportati semplicemente team 1 e team 2, questo perché non è possibile risalire a quale dei due team il dataset si riferisse tra Radian e Dire. Tuttavia a noi interessa semplicemente misurare la differenza di winrate tra i due team nelle partite ranked e nelle partite unranked.

Il team 2 sembra essere in vantaggio in maniera particolare all’interno delle partite classificate. E’ importante notare che anche una variazione minima del winrate in questo caso è un elemento critico per il bilanciamento del gioco. Sono presenti articoli [3] che cercano di analizzare questo aspetto del gioco dove il winrate del 51.6% dei Radian viene considerato un vantaggio netto per questo team.

Si può quindi affermare che rispetto ai due campioni osservati sussiste una differenza di winrate che suggerisce che uno dei due team sia più avvantaggiato rispetto ad un altro.

2.4 RQ4:Esiste una correlazione tra pickrate e winrate?

Come è stato già spiegato nelle research question precedenti, il winrate degli Hero può essere determinante per capire se un gioco risulta essere bilanciato o meno.

Nel genere dei giochi MOBA solitamente minime variazioni nel winrate dei campioni possono condurre i giocatori a giocare un determinato Hero piuttosto che un altro. All’interno di questa research question si cercherà di individuare qualche evidenza di una correlazione tra pickrate e winrate degli Hero. Individuare una simile correlazione è

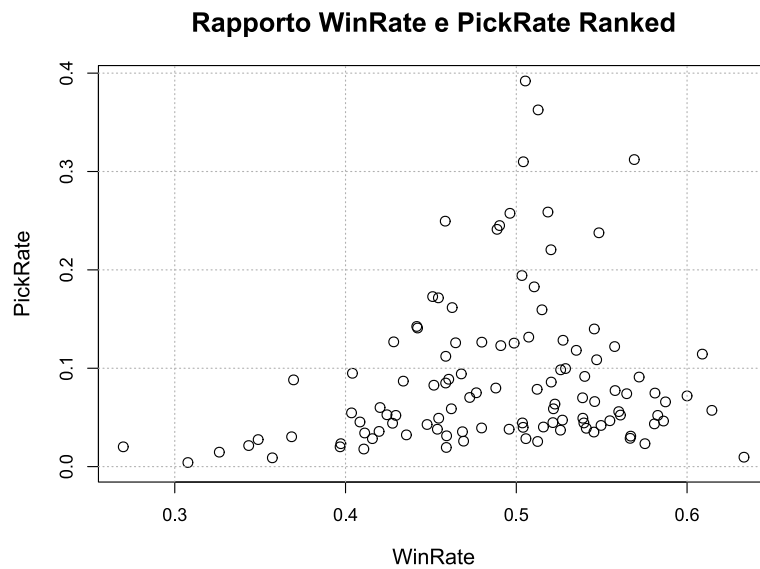


Figura 14: Diagramma di dispersione del winrate e del pickrate degli Hero nelle partite classificate

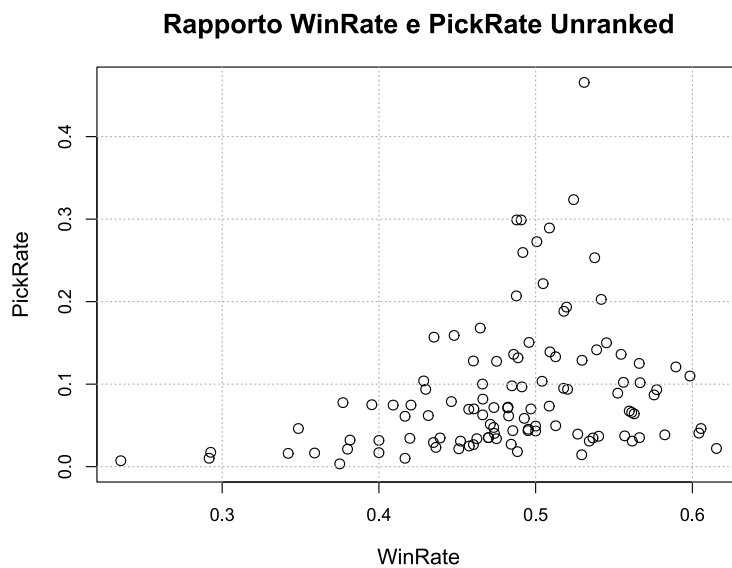


Figura 15: Diagramma di dispersione del winrate e del pickrate degli Hero nelle partite non classificate

interessante perché suggerirebbe che il winrate potrebbe influenzare il pickrate degli Hero. Per cercare di individuare dei palesi pattern all'interno dei dati, abbiamo realizzato dei diagrammi di dispersione (figura 14 15) per le partite ranked e unranked. Entrambi i diagrammi di dispersione non mostrano dei pattern ovvi. Per verificare però, se c'è una correlazione lineare si è deciso di utilizzare le tecniche di regressione e valutare direttamente i coefficienti. All'interno delle seguenti figure (16 17) sono riportati i valori dei coefficienti di correlazione. Si può notare che i coefficienti indicano una correlazione

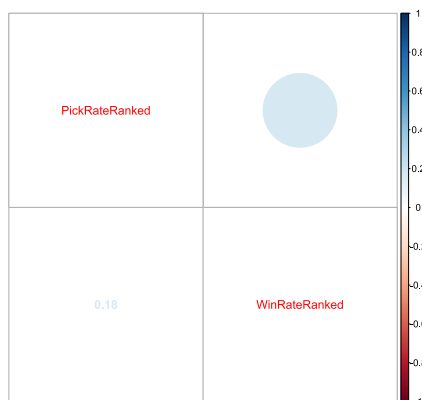


Figura 16: Nella seguente figura è riportato il coefficiente di correlazione winrate e pickrate per le partite ranked

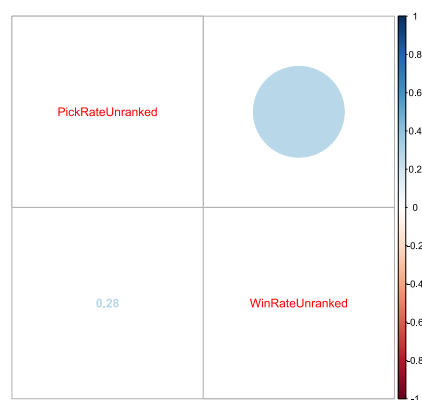


Figura 17: Nella seguente figura è riportato il coefficiente di correlazione winrate e pickrate per le partite unranked

bassa, confermando la sensazione restituita dai diagrammi di dispersione secondo cui non sembrava esistere nessuna correlazione di tipo lineare.

Si è deciso quindi di tentare l'utilizzo di tecniche più avanzate di regressione, per cercare una correlazione di tipo non lineare. Nello specifico sono stati utilizzati metodi avanzati basati sulle curve di fitting. Il primo metodo che è stato utilizzato è la spline, il risultato è stato riportato graficamente all'interno delle seguenti figure 18 19. Nel caso di ranked il valore di R-squared è di 0.1205, mentre per le unranked 0.1462. Dato che il valore di R-squared varia tra 0 e 1, si può concludere che questo valore è basso e non rappresentativo di una correlazione effettiva.

Si è deciso quindi di tentare anche un altro metodo, in particolare la Gaussian Process Regression. Il risultato di questa regressione è visibile all'interno dei seguenti grafici 20 21. Il valore di R-squared nel caso delle partite ranked è 0.2412 e invece è 0.2625 nel caso delle partite unranked. In entrambi i casi i valori ottenuti sono troppo bassi, di conseguenza anche queste curve non sono rappresentative. Rispetto a quelli che sono i risultati che sono stati ottenuti, non siamo riusciti ad individuare una correlazione banale tra il pickrate e il winrate che fosse rappresentabile attraverso una linea oppure una curva.

Ragionando però su quello che è il dominio sul quale stiamo conducendo l'analisi, questo

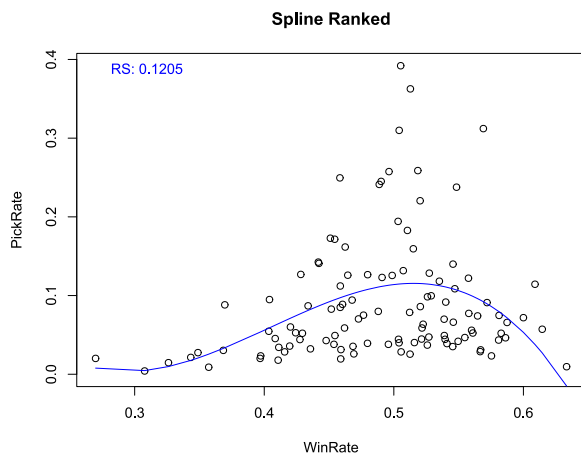


Figura 18: Nella seguente figura è riportato il diagramma di dispersione con la curva computata attraverso la Spline sulle partite ranked

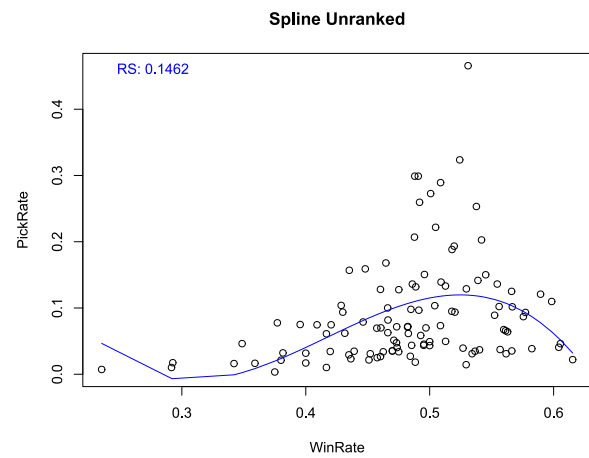


Figura 19: Nella seguente figura è riportato il diagramma di dispersione con la curva computata attraverso la Spline sulle partite unranked

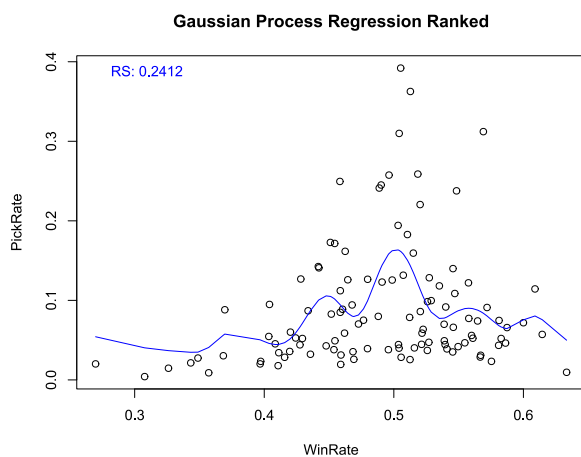


Figura 20: Nella seguente figura è riportato il diagramma di dispersione con la curva computata attraverso la Gaussian Process Regression sulle partite ranked

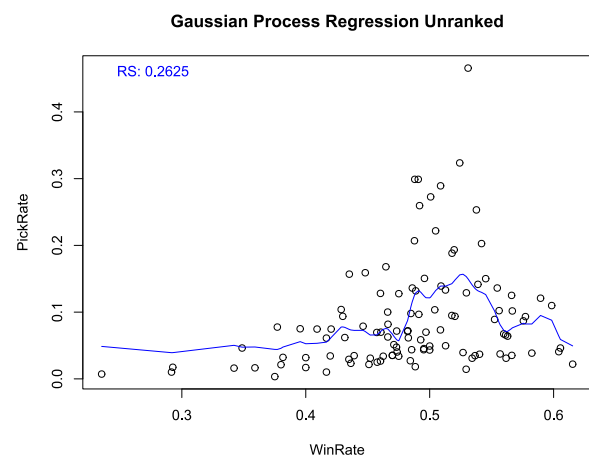


Figura 21: Nella seguente figura è riportato il diagramma di dispersione con la curva computata attraverso la Gaussian Process Regression sulle partite unranked

risultato assume facilmente un significato. Il winrate può indicare uno squilibrio nel gioco e molto spesso nei MOBA pochissimi personaggi giocabili sono quelli utilizzabili per competere. Non è però solamente questo parametro che influenza quali personaggi verranno giocati, a influenzare la scelta del personaggio da giocare all'interno dei MOBA sono anche la complessità dell'Hero, le sue abilità, il suo ruolo all'interno del team e quanto quest'ultimo risulta divertente da giocare.

Se quindi è naturale che non possa esistere una correlazione così banale tra il winrate e il pickrate, è possibile comunque ragionando sul dominio del problema fare determinate osservazioni. Prendendo in considerazione i diagrammi di dispersione riportati nelle figure 22-23 è possibile apprezzare che tali diagrammi possono essere facilmente divisi in 4 aree definendo degli intervalli per il winrate e per il pickrate. E' quindi possibile notare che non esistono Hero con un basso winrate e un alto pickrate. Se il winrate non è il solo elemento che influenza se un Hero viene scelto da un giocatore, contemporaneamente si deve considerare che se il winrate di un Hero è troppo basso è raro che quest'ultimo verrà scelto dai giocatori. Addirittura rispetto agli intervalli che sono stati fissati, questa area appare vuota o quasi in entrambi i diagrammi di dispersione. Esistono invece in un numero esiguo gli Hero che hanno un basso winrate e un basso pickrate.

Per quanto riguarda invece gli Hero che possiedono un alto winrate è possibile notare che la maggior parte di questi ultimi hanno un pickrate basso e solamente un numero esiguo hanno un pickrate alto. Anche questa situazione non stupisce rispetto a quello che di solito è lo stato dei giochi MOBA. Difatti, il winrate non può da solo determinare se un personaggio verrà giocato molto o meno. Spesso i personaggi che hanno un alto winrate possono risultare noiosi da giocare. In altri casi, come anche quello di Dota2, i match sono organizzati per elo(rango del giocatore) e alcuni personaggi, troppo semplici possono facilmente diventare troppo limitati o addirittura inutili quando si gioca con degli avversari maggiormente esperti. Questi sono solo due esempi rispetto ai molti elementi che influenzano notevolmente il pickrate di un determinato personaggio all'interno dei giochi MOBA. Dai risultati ottenuti e dalle informazioni riguardanti il dominio si può concludere che è verosimile che non esista una correlazione (quanto meno banale) tra il pickrate e il winrate.

2.5 RQ5:Esistono degli attributi più incisivi all'interno del gioco?

Ciascuno dei personaggi giocabili all'interno del gioco possiede delle caratteristiche uniche, come ad esempio le sue skill, i suoi ruoli e il suo attributo. Gli attributi in Dota2 indicano quali sono le statistiche principali del personaggio che vanno a determinare come quest'ultimo crescerà ed evolverà durante la partita. Gli attributi sono quindi le caratteristiche degli Hero che li caratterizzano maggiormente.

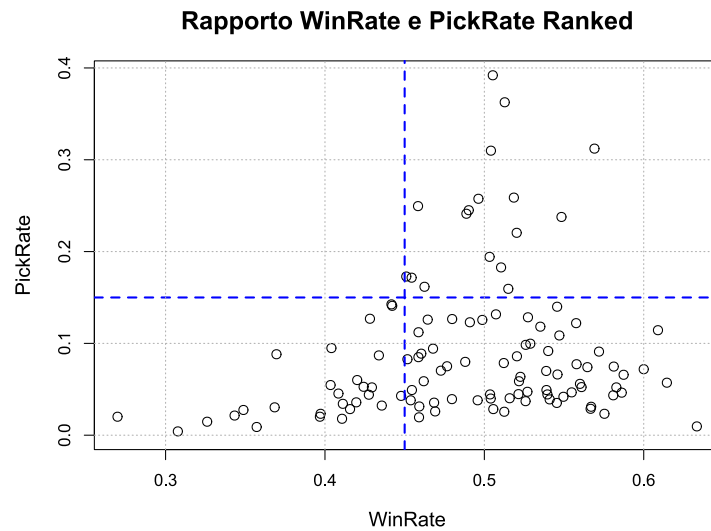


Figura 22: Nella seguente figura è riportato il diagramma di dispersione delle partite ranked diviso in 4 quadranti

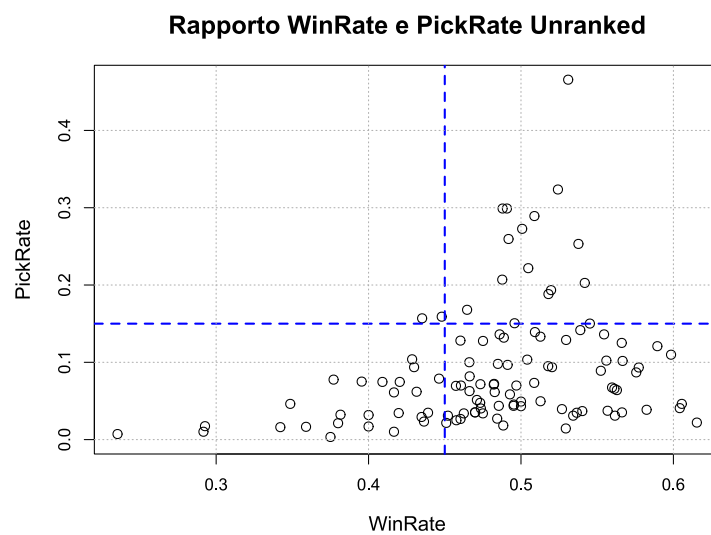


Figura 23: Nella seguente figura è riportato il diagramma di dispersione delle partite unranked diviso in 4 quadranti

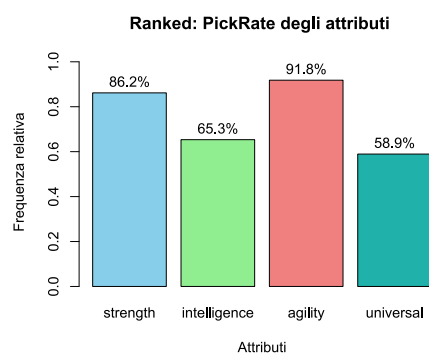


Figura 24: La seguente figura è un barplot del pickrate per ciascuno degli attributi nelle partite ranked

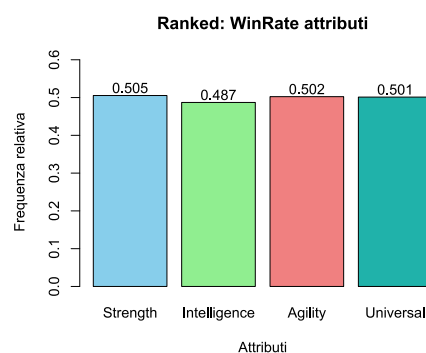


Figura 25: La seguente figura è un barplot del winrate per ciascuno degli attributi nelle partite ranked

In Dota2 sono presenti 4 attributi differenti:

- Strength
- Agility
- Intelligence
- Universal

Ai fini di questo studio non si sono ritenute necessarie le informazioni tecniche che riguardano ciascuno degli attributi.

Gli attributi permettono di determinare degli elementi fondamentali del Hero e hanno quindi una grande influenza su quello che è lo stato del gioco. Nella seguente research question si cercherà di determinare se tra tutti quanti gli attributi ne è presente qualcuno che risulta essere più "conveniente" dal punto di vista statistico rispetto agli altri.

Per rispondere a questa research question, il primo passo è stato quello di analizzare il pickrate e il winrate per gli attributi. Il pickrate degli attributi è stato calcolato come il rapporto tra la frequenza assoluta delle partite in cui è stato giocato almeno un Hero con quell'attributo fratto il numero di partite. Il winrate è stato calcolato come il rapporto tra il numero di partite in cui il team vincente aveva selezionato almeno un Hero con quell'attributo fratto il numero di partite in cui compariva un team con quell'Hero.

I valori del pickrate di ciascun attributo nelle partite ranked è stato riportato nel grafico 24. E' possibile notare che nella maggior parte delle partite viene giocato almeno un Hero con attributo Strength e uno con Agility. All'interno del grafico 25 è stato riportato il valore del winrate per ciascuno degli attributi nelle partite ranked, ed è possibile notare che

tutti quanti gli attributi hanno un winrate che si aggira attorno al 50%. Per come è stato misurato il winrate, anche una minima variazione (più bassa rispetto a quelle considerate nelle research question precedenti) è importante. Avere uno degli attributi con un winrate

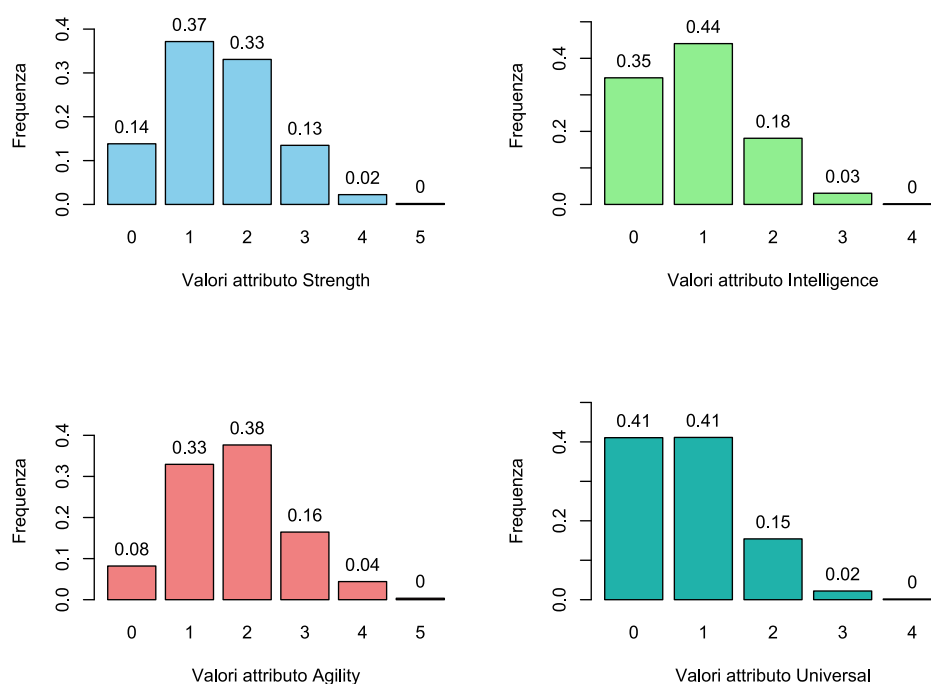


Figura 26: Nella seguente figura sono riportati 4 barplot, uno per ogni attributo del gioco. All'interno di ciascun grafico è stata utilizzata una colonna per riportare la percentuale di squadre in cui compaiono esattamente un certo numero di Hero con un determinato attributo. I seguenti grafici sono riferiti alle partite ranked.

decisamente più basso rispetto agli altri significherebbe che un Hero con quell'attributo compare molto più di rado rispetto agli altri in un team vincente. E' quindi da notare che Intelligence rispetto a tutti gli altri attributi è associata ad un winrate leggermente minore.

Successivamente si è deciso di analizzare in maniera più precisa il winrate di ciascuno degli attributi. All'interno della seguente figura 26 è stato riportato un barplot per ciascuno degli attributi nelle partite ranked. All'interno di ciascun grafico è stata utilizzata una colonna per riportare la percentuale di squadre in cui compaiono esattamente un certo numero di Hero con un determinato attributo. Questi grafici ci sono utili principalmente per comprendere per quali tipologie di team si hanno abbastanza istanze per poterne poi successivamente valutare il winrate. Tutte quante le tipologie di team per cui la frequenza

calcolata è minore del 5% sono state tagliate dal resto dello studio per questa research question.

All'interno della figura 27 è riportato un grafico per ciascuno degli attributi. All'interno di ciascuno dei grafici ciascuna colonna rappresenta il winrate dei team che possiedono un certo numero di Hero con un determinato attributo. E' possibile notare che per quanto riguarda l'attributo Strenght sembra essere statisticamente più conveniente avere un numero di Hero in squadra tra 1 e 3 rispetto a non averne affatto. Per quanto riguarda Agility la scelta conveniente sembra averne un numero tra 1 e 2. Mentre per quanto riguarda Intelligence e Universal, sembra convenire avere al più un solo Hero con quel determinato attributo.

Da questa analisi è emerso che avere in squadra Hero con Universal o Intelligence non fa variare statisticamente il winrate, ed è possibile osservare che i team in cui sono presenti Hero con attributi Strenght e Agility hanno un winrate maggiore rispetto a quelli che non li hanno. Se quindi non è possibile determinare con precisione se un attributo sia nettamente più incisivo degli altri, per quanto riguarda le partite ranked gli attributi Strenght e Agility sembrano risultare più determinanti per ottenere la vittoria. Questo risultato, sembra giustificare il fatto che il pickrate per Strenght e Agility rappresentato nel grafico 24 sia molto più alto per Strenght e Agility rispetto agli altri due attributi.

Ripetiamo l'analisi che è stata effettuata per le partite ranked anche per le partite unranked. Il grafico del pickrate è presente nella figura 28. Quest'ultimo risulta essere molto simile a quello delle partite unranked, si ha però un incremento per quanto riguarda Intelligence e Universal. Per quanto riguarda il winrate degli attributi, che è riportato nel grafico 29 è possibile notare che abbiamo una distribuzione del winrate tra gli attributi praticamente perfetta.

Si procede ora con l'analisi maggiormente specifica per ciascuno degli attributi. All'interno della seguente figura 30 è stato riportato un baplot per ciascuno degli attributi nelle partite unranked. Nei grafici le colonne rappresentano la percentuale di squadre in cui compaiono esattamente un certo numero di Hero con un determinato attributo. Anche in questo caso questi dati sono stati utilizzati per scartare le tipologie di team per le quali si hanno troppe poche istanze per effettuare una valutazione, utilizzando lo stesso principio utilizzato precedentemente.

All'interno della figura 31 è riportato un grafico per ciascuno degli attributi. Nei grafici ciascuna colonna rappresenta il winrate dei team che possiedono un certo numero di Hero con un determinato attributo. Nello specifico in questo caso è possibile notare che nel caso dell'attributo Strength conviene avere tra 1 e 3 Hero con questo attributo. Per agility invece conviene avere tra 1 e 2 Hero. Ma la situazione rispetto a prima cambia notevolmente per Universal e Intelligence. In questo caso, infatti, non avere un Hero con questi attributi ci porta ad avere un winrate notevolmente minore.

Nel caso quindi delle partite unranked, Strenght e Agility sono incisivi per lo meno quando

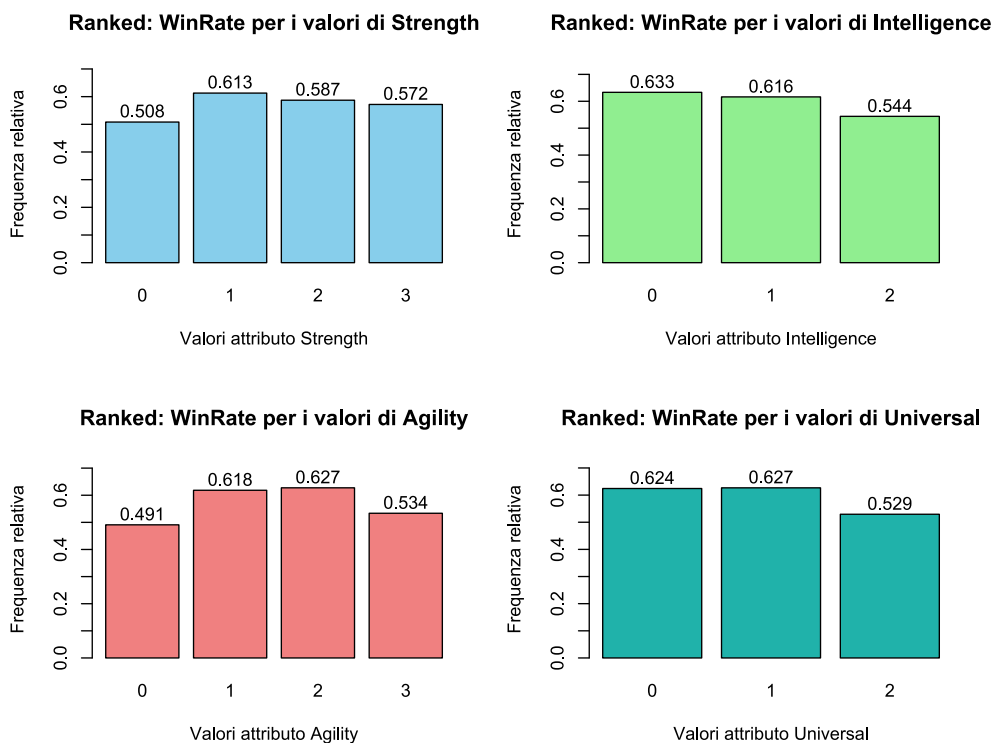


Figura 27: Nella seguente figura sono riportati 4 barplot, uno per ogni attributo del gioco. All'interno di ciascun grafico è riportato il winrate dei team che possiedono un certo numero di Hero con un determinato attributo. I seguenti grafici sono riferiti alle partite ranked.

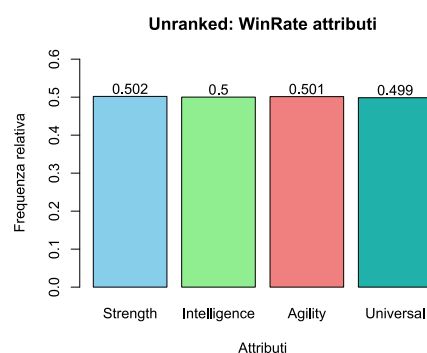
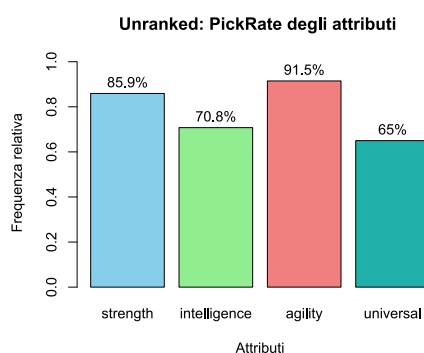


Figura 28: La seguente figura è un barplot del pickrate per ciascuno degli attributi nelle partite unranked

Figura 29: La seguente figura è un barplot del winrate per ciascuno degli attributi nelle partite unranked

si posseggono da 1 a 2 Hero, mentre nel caso di Universal e Intelligence, conviene avere esattamente un Hero con quel determinato attributo.

Riassumendo, sono stati ottenuti risultati differenti per le partite ranked e per le partite

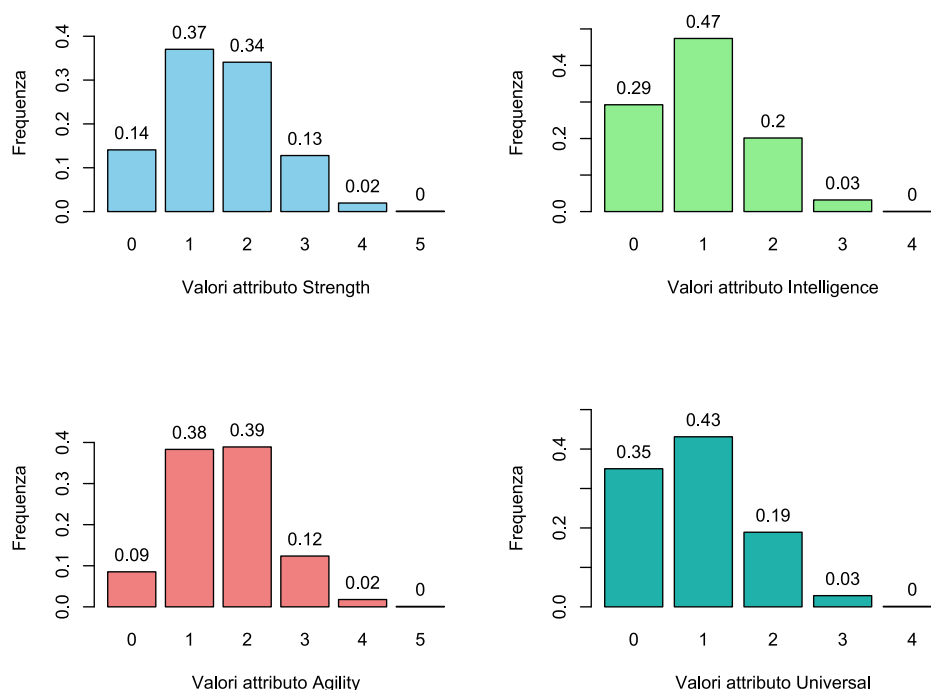


Figura 30: Nella seguente figura sono riportati 4 barplot, uno per ogni attributo del gioco. All'interno di ciascun grafico è stata utilizzata una colonna per riportare la percentuale di squadre in cui compaiono esattamente un certo numero di Hero con un determinato attributo. I seguenti grafici sono riferiti alle partite unranked.

unranked. In particolare, nelle partite ranked si è rivelata fondamentale la presenza di un Hero con attributo Strength o Agility mentre avere più di un Hero con attributo Intelligence o Universal sembra diminuire statisticamente il winrate. Per quanto riguarda invece le partite unranked la situazione di Strength e Agility non è mutata in maniera particolarmente interessante, invece per Universal e Intelligence la scelta migliore è avere esattamente un membro in squadra per ciascuno dei due attributi.

2.6 RQ6:Esistono configurazioni dei team più convenienti?

All'interno di questa analisi statistica si è cercato di definire lo stato del gioco indagando anche se c'erano Hero e attributi che convenivano statisticamente rispetto agli altri.

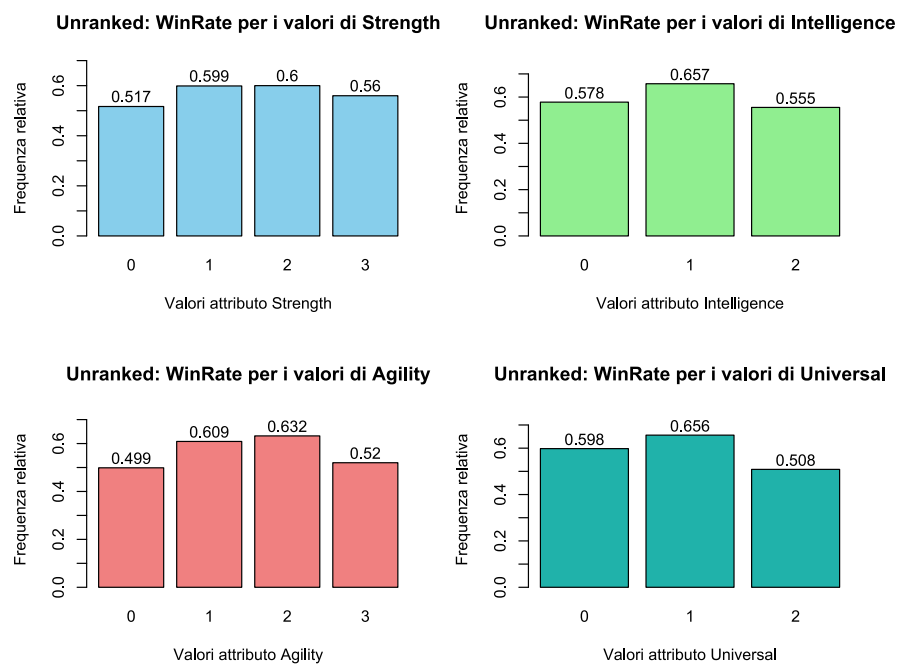


Figura 31: Nella seguente figura sono riportati 4 barplot, uno per ogni attributo del gioco. All'interno di ciascun grafico è riportato il winrate dei team che possiedono un certo numero di Hero con un determinato attributo. I seguenti grafici sono riferiti alle partite unranked

Tabella 6: Nella seguente tabella sono riportati le 5 configurazioni di team che hanno il winrate più alto all'interno delle partite ranked.

Strength	Intelligence	Agility	Universal	WinRate
1	2	2	0	54.9%
3	1	1	0	53.7%
3	0	2	0	52.9%
1	1	2	1	52.1%
2	0	3	0	51.6%

Tuttavia, nel genere dei MOBA spesso la vittoria non è determinata dai singoli ma è piuttosto determinata dall'efficacia del team. Se anche un giocatore compie tutte quante le scelte giuste per quanto gli è possibile, probabilmente non sarà abbastanza se il team non fa altrettanto bene.

Rispetto a questa caratteristica dei MOBA è interessante analizzare il winrate delle varie configurazioni di team, in modo da capire se emerge una qualche configurazione di team che è statisticamente più conveniente rispetto a tutte quante le altre.

Ciascun Hero è caratterizzato da vari elementi e di conseguenza per valutare la configurazione del team si possono utilizzare varie metriche. Ovviamente, non c'è in assoluto nessuna metrica che si possa dire migliore rispetto alle altre. La metrica che ci è sembrato più sensato utilizzare per valutare l'efficacia di un team è quella basata sulla configurazione di attributi. Semplicemente, si andranno a creare delle classi di team rispetto al numero di Hero che possiedono per ciascun attributo.

Non tutte quante le classi di team avranno lo stesso numero di istanze e alcune configurazioni di team non compariranno affatto all'interno dei dati. Questo è plausibile dato che alcune configurazioni risultano semplicemente assurde per poter essere giocate. Tra tutte quante le configurazioni di team, si è deciso di scartare le configurazioni di team che sono presenti in meno del 10% delle partite.

Anche in questo caso l'analisi è stata effettuata in maniera distinta per le partite ranked e per quelle unranked. All'interno della seguente tabella 6 sono state riportate le 5 configurazioni di team che hanno un winrate maggiore rispetto alle altre. Per le partite ranked è possibile notare che il risultato dell'analisi, è in un certo qual modo coerente rispetto ad alcuni risultati ottenuti in precedenza. Difatti, all'interno di 4 team su 5, l'attributo Universal si trova a 0 rispecchiando quelli che erano i risultati già ottenuti in precedenza per quanto riguarda il winrate dell'attributo Universal nelle partite ranked (figura 27). Anche per quanto riguarda Intelligence quest'ultimo compare all'interno dei top team con una media di 0.8, conforme rispetto a quelle che erano le scelte convenienti che erano emerse durante l'analisi degli attributi.

Tabella 7: Nella seguente tabella sono riportati le 5 configurazioni di team che hanno il winrate più alto all'interno delle partite unranked.

Strength	Intelligence	Agility	Universal	WinRate
1	2	2	0	54.6%
2	1	1	1	51.8%
1	2	1	1	50.8%
1	1	2	1	50.7%
2	0	2	1	50.6%

All'interno della seguente tabella 7 sono invece riportati i risultati per le partite unranked. Se confrontiamo le configurazioni dei team che appaiono migliori nelle partite unranked con i dati riguardanti il winrate degli attributi in figura 31, è possibile notare che Universal è utilizzato esattamente una volta nella maggior parte dei team, Strength varia tra 1 e 2 e lo stesso vale per Agility.

E' possibile inoltre notare che ci sono differenze meno marcate rispetto ai top team delle partite ranked, se non per il team più vincente tra tutti. Il team più vincente tra tutti è anche lo stesso tra le partite ranked e quelle unranked.

In conclusione, è importante sottolineare che rispetto al winrate di un determinato Hero questo valore risulta essere molto più significativo. Questo proprio per i motivi che sono stati spiegati precedentemente, la configurazione del team è sicuramente un parametro molto più importante rispetto a quella che è la scelta del singolo giocatore. Si può concludere quindi che c'è una differenza significativa tra i winrate di alcuni team, che ci porta a concludere che effettivamente esistono delle configurazioni di team maggiormente convenienti rispetto ad altre.

2.7 RQ7:Qual'è il rank delle istanze del dataset?

In Dota2 vincendo le partite classificate è possibile accumulare punti rank. Questi punti determinano l'elo(praticamente il rango) del giocatore che viene poi utilizzato durante la creazione dei due team per una determinata partita(matchmaking).

I team per le partite vengono creati con strategie differenti rispetto al fatto che le partite siano ranked o unranked. Per le partite unranked non si utilizza l'elo per determinare da quali giocatori saranno composti i team. Questa informazione viene invece utilizzata per le partite ranked.

Se i rank infatti sono diventati nel tempo un motivo di orgoglio e degli obiettivi da raggiungere per i giocatori, originariamente erano stati soprattutto pensati per essere utilizzati come parametro ai fini di organizzare le partite in maniera che potesse esserci

un agonismo sensato.

L'obiettivo di questa research question è provare a determinare attraverso altri parametri il rank delle partite, verificare se la distribuzione del rank delle partite è nota e se questa distribuzione individuata è riconducibile dal campione analizzato alla popolazione. Anche in questo caso è stato fondamentale analizzare i due campioni(ranked e unranked) in maniera distinta.

All'interno del dataset non è presente come informazione il rank di ciascuno dei giocatori. Di conseguenza, determinare il rank di una partita non può essere fatto analizzando quei valori. Al variare dell'elo dei partecipanti ad una partita possono variare alcuni aspetti, che diventano quindi indici che ci possono suggerire qual'è il rank della partita. In particolare ciascuno degli Hero nel gioco possiede un parametro di complessità che va da 1 a 3, che indica quanto l'Hero è difficile da utilizzare. E' plausibile che l'utilizzo di Hero più complessi sia maggiormente diffuso in partite dove ci sono giocatori con un elo maggiore.

Un altro parametro influenzato dal rank, riguarda il numero di counter-pick che vengono effettuati. Un counter-pick non è altro che una scelta di un personaggio fatta per mettere in difficoltà l'avversario in base alla sua scelta. All'interno di Dota2 per ciascun Hero è possibile individuare almeno un altro Hero che attraverso le sue caratteristiche nullifica le potenzialità dell'altro. E' logico pensare che un numero maggiore di counter-pick saranno presenti all'interno di partite dove ci sono giocatori con un elo maggiore, che conoscono meglio il gioco e sanno quindi giocare meglio di strategia.

Dopo aver considerato l'elo come una variabile aleatoria sono stati utilizzati i valori di complessità e i counter-pick per determinarne il valore. In particolare, la complessità della partita è stata definita come la somma delle complessità degli Hero che compaiono in una partita, mentre il valore di counter-pick è il numero di counter-pick che sono stati effettuati per quella partita. Il valore del rank sarà quindi dato la seguente combinazione lineare dei valori di counter-pick e complessità $500 + ((0.85 \cdot \frac{C-10}{20} + 0.15 \cdot \frac{CP}{25}) \cdot 5000)$. All'interno della seguente formula è possibile osservare due parametri C e CP, che rappresentano rispettivamente la complessità e il counter-pick della partita. Il valore C viene normalizzato sottraendo 10 e dividendo il risultato per 20, in quanto il valore della complessità nella partita può variare tra 10 e 30. Mentre invece il valore di CP viene diviso per 25, in quanto il numero massimo di counter-pick in una partita è 25. Il valore massimo è dato dal fatto che attraverso un pick è possibile al più fare un counter-pick per i cinque personaggi della squadra avversaria. Per quanto riguarda i pesi che sono stati scelti per il parametro C e il parametro CP ci consentono di valutare principalmente il valore della complessità, che riteniamo più incisiva dei counter-pick. Dato che il valore dell'elo nel gioco varia tra 500 e 5500, si è deciso di passare i valori in questa scala.

All'interno delle seguenti figure 32 33 sono presenti dei grafici per la distribuzione dell'elo da noi calcolato rispettivamente per le partite ranked e quelle unranked. Entrambi i grafici

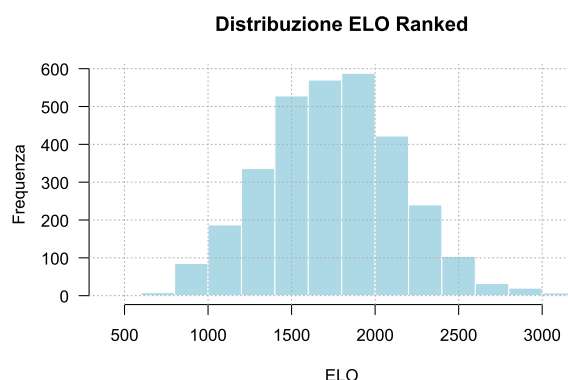


Figura 32: Nella seguente figura è presente un grafico della distribuzione dell'elo calcolato all'interno delle partite ranked

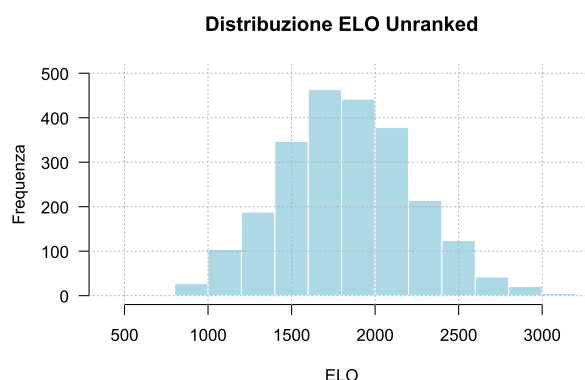


Figura 33: Nella seguente figura è presente un grafico della distribuzione dell'elo calcolato all'interno delle partite unranked

ci suggeriscono che l'elo abbia una distribuzione normale. Successivamente verrà utilizzato il test del chi-quadro per verificare se la distribuzione è effettivamente una normale o meno.

Nel caso delle partite ranked il valore di chi-quadro è 1.597444 e l'intervallo (0.05063562, 7.377759), di conseguenza il test del chi-quadro è passato. Il test del chi-quadro invece nel caso delle partite unranked ha restituito per chi-quadro il valore 13.4207 e come intervallo (0.05063562, 7.377759). Nel caso delle partite unranked il test del chi-quadro è quindi risultato non valido, di conseguenza il resto dello studio verrà portato avanti solamente per le partite ranked.

Sapendo che l'elo nelle partite ranked segue una distribuzione normale, si cerca ora di determinare se le considerazioni effettuate per questo campione possono essere estese alla popolazione. Nello specifico ci interessa farlo per il valore medio dell'elo della popolazione, cioè il valore medio dell'elo delle partite di Dota2. Per effettuare la stima puntuale si è utilizzato il metodo della massima verosimiglianza. In questo caso lo stimatore di massima verosimiglianza è la media campionaria dell'elo. Il valore della media campionaria dell'elo è: 1742.19, ma è scontato in questo caso che tale valore rientra nell'intervallo di confidenza (1727.62760612041, 1756.75258557288). E' stato così ottenuto con un grado di affidabilità del 95% che il valore medio della popolazione rientra all'interno dell'intervallo di confidenza.

Per ottenere una maggiore certezza, si è deciso anche di effettuare la verifica delle ipotesi con un grado di affidabilità del 99%. Dove nello specifico ciò che è stato fatto è un test bilaterale dove l'ipotesi nulla è che il valore medio dell'elo della popolazione è uguale a quello del nostro campione. In particolare si vuole verificare che l'ipotesi sul valore medio è vera nel caso in cui la varianza della popolazione non è nota. Dato che il test ha

dato successo sull'intervallo $(-1.960722, 1.960722)$ con il valore 0 e che il p-value(1) è maggiore di $\alpha(0.01)$ non si può rifiutare l'ipotesi nulla con un grado di significabilità del 99%.

3 Dataset sintetico

3.1 RQ1: Il winrate varia notevolmente tra gli Hero?

Si cerca ora di rispondere alla research question numero 1 anche per il dataset sintetico. Nella figura 34 è riportato il barplot per il winrate di ciascuno degli Hero, anche all'interno

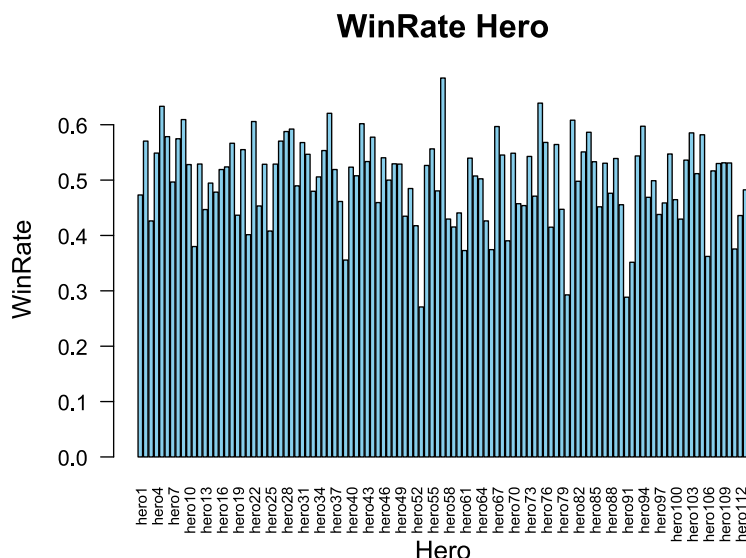


Figura 34: Barplot che mostra il winrate per ciascuno degli Hero

del dataset sintetico è presente una certa variabilità del winrate. In particolare la varianza del winrate è pari a 0.005914222. Seppure non c'è una forte varianza quest'ultima risulta essere maggiore rispetto a quella del dataset non sintetico. Anche per il dataset sintetico è stato deciso di analizzare in maniera distinta le partite ranked da quelle unranked. La varianza degli hero nelle partite ranked è 0.00697383 e quella delle partite unranked è 0.006019371. Non c'è una particolare variazione neanche in questo caso. Non si notano notevoli differenze neanche in questo caso tra ranked e unranked andando a guardare i barplot nelle figure 35 36. Anche per il dataset sintetico si sono analizzate le distribuzioni dei winrate per le partite ranked e quelle unranked (figura 37). In questo caso si può notare un notevole peggioramento nella distribuzione rispetto al winrate degli Hero rispetto al dataset originale. Difatti, si hanno ben il 47% degli Hero per le ranked e il 57% per le unranked che si trovano fuori dal range (0.45, 0.55). Di conseguenza, il gioco risulta avere addirittura una maggiore variabilità per quanto concerne il winrate degli Hero.

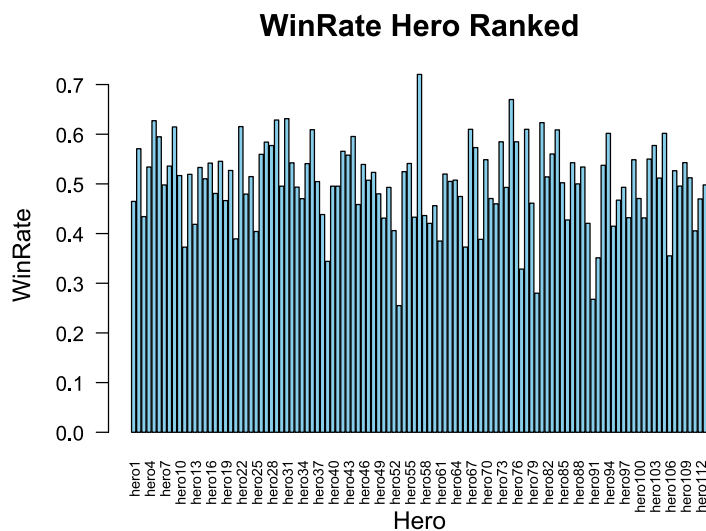


Figura 35: Barplot che mostra il winrate per ciascuno degli Hero nelle partite ranked

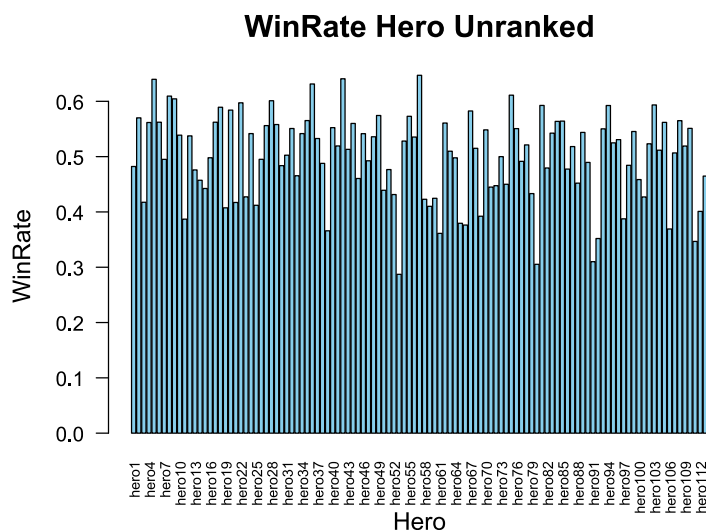


Figura 36: Barplot che mostra il winrate per ciascuno degli Hero nelle partite unranked

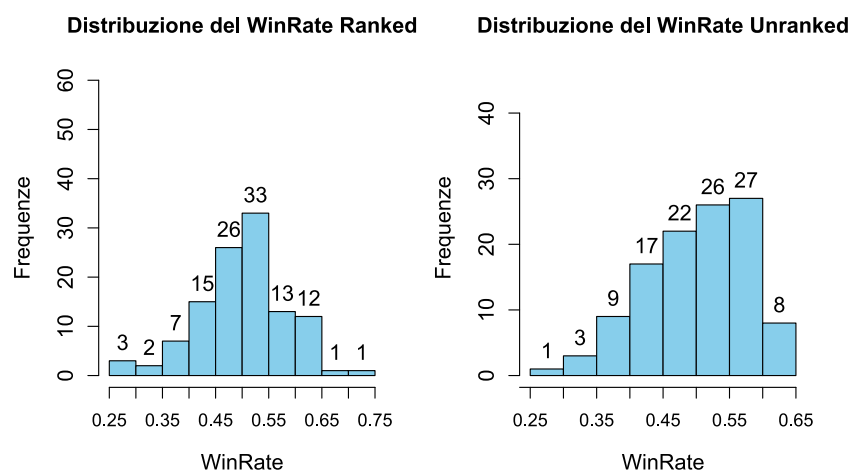


Figura 37: Figura che mostra i barplot per la distribuzione del winrate nelle partite ranked e in quelle unranked

3.2 RQ2:winrate e pickrate variano tra ranked e unranked?

Allo scopo di verificare se all'interno del dataset sintetico è presente una variabilità tra le partite classificate e quelle non classificate, si procede a rispondere alla research question 2 per il dataset sintetico.

Nelle figura 38 sono riportati i pickrate degli Hero nelle partite ranked e unranked. La differenza dei pickrate tra le due modalità è osservabile meglio all'interno del grafico 39. Rispetto al dataset originale, è possibile osservare che le differenze tra le modalità ranked e unranked sono molto minori. In questo caso praticamente non abbiamo mai un Hero che supera la soglia del 2.5%. Sono stati poi realizzati anche in questo caso i boxplot ad intaglio 40 per il pickrate degli Hero per le partite ranked e unranked. Si nota la quasi totale assenza di outlier, che ritorna come informazione rispetto al grafico delle differenze dei pickrate 39. Come nel caso del dataset originale, l'intervallo $(-0.002648102, 0.001843585)$ suggerisce che non ci sono differenze statistiche particolari tra le mediane dei due campioni.

Nella tabella 8 sono stati riportati gli indici di sintesi per il pickrate nelle partite ranked e unranked. Se la media dei campioni è rimasta invariata rispetto al dataset originale, è interessante notare che la mediana è invece cambiata. La varianza e la deviazione standard sono diminuite moltissimo rispetto al campione originale.

I valori del campione sintetico si allontanano di più da quelli dell'originale proprio per skewness e curtosi. Quindi, a differenza del campione originale, nel campione sintetico sussistono delle differenze importanti riguardo la distribuzione del pickrate, nonostante non ci siano delle forti variazioni per quanti riguarda i singoli Hero. Infatti rispettivamente per

Tabella 8: Tabella riassuntiva degli indici di sintesi per il pickrate per le partite ranked e unranked

	ranked	unranked
media	0.08850	0.08850
mediana	0.08820	0.08860
deviazione standard	0.005701508	0.005403042
varianza	3.25072e-05	2.919287e-05

le partite ranked e unranked abbiamo una skewness positiva e negativa, che comportano quindi rispettivamente un'asimmetria a destra e una a sinistra. Si verifica anche per il

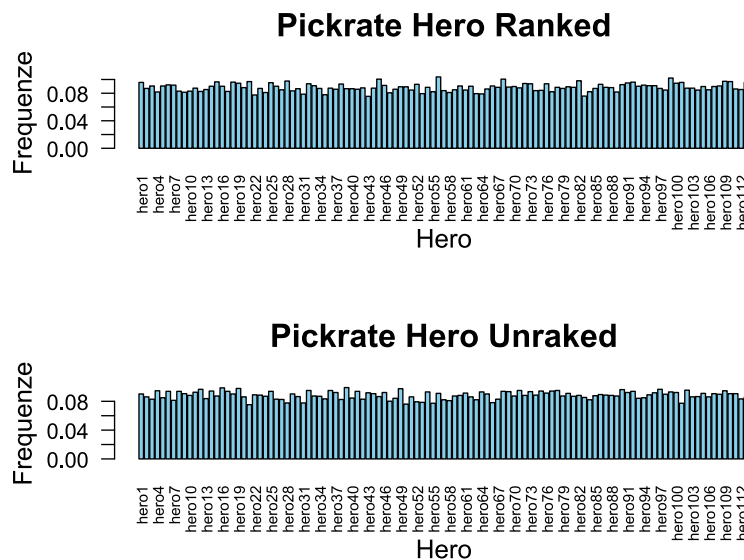


Figura 38: Nella seguente figura sono mostrati i barplot per il pickrate degli hero nelle partite ranked e unranked

winrate se ci sono delle differenze interessanti tra i due campioni analizzati. All'interno della figura 41 sono riportati i boxplot ad intaglio del winrate. Come nel caso del pickrate, anche per il winrate si hanno un basso numero di outlier e anche in questo caso l'intervallo di confidenza $(-0.03541395, 0.02387649)$ indica che le mediane dei due campioni non sono troppo differenti tra di loro.

Gli indici di sintesi riportati all'interno della seguente tabella 10 non mostrano particolari differenze tra media, mediana e varianza tra le partite ranked e unranked. Rispetto al dataset originale è possibile notare un minimo incremento degli indici di dispersione.

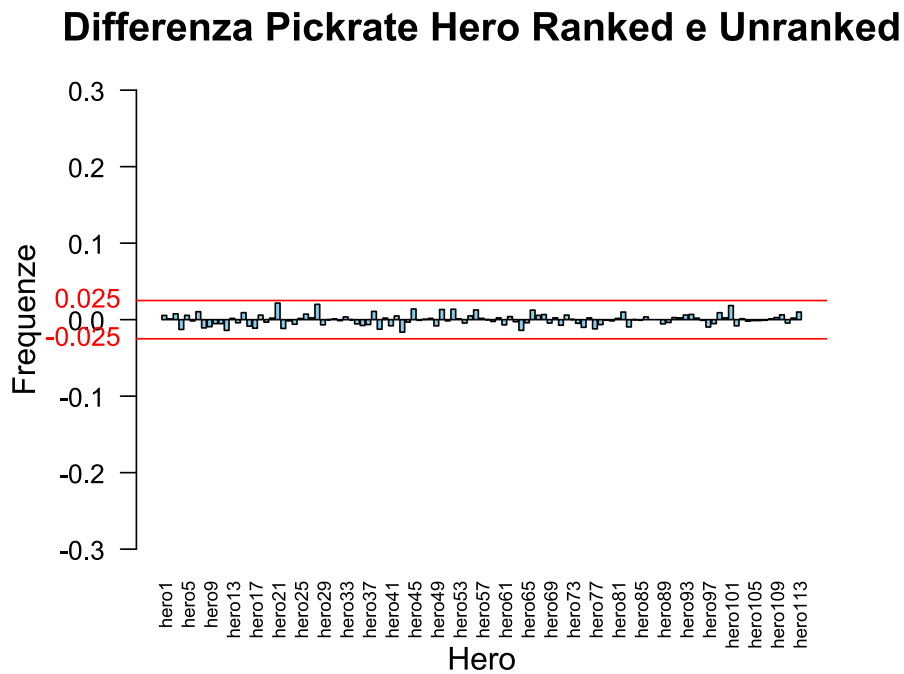


Figura 39: Nella seguente figura è mostrato il barplot per la differenza del pickrate degli hero nelle partite ranked e unranked

Tabella 9: Tabella riassuntiva degli indici simmetrici per pickrate nelle partite ranked e unranked

	ranked	unranked
skewness	0.1577882	-0.3684828
curtosi	2.848794	2.547602

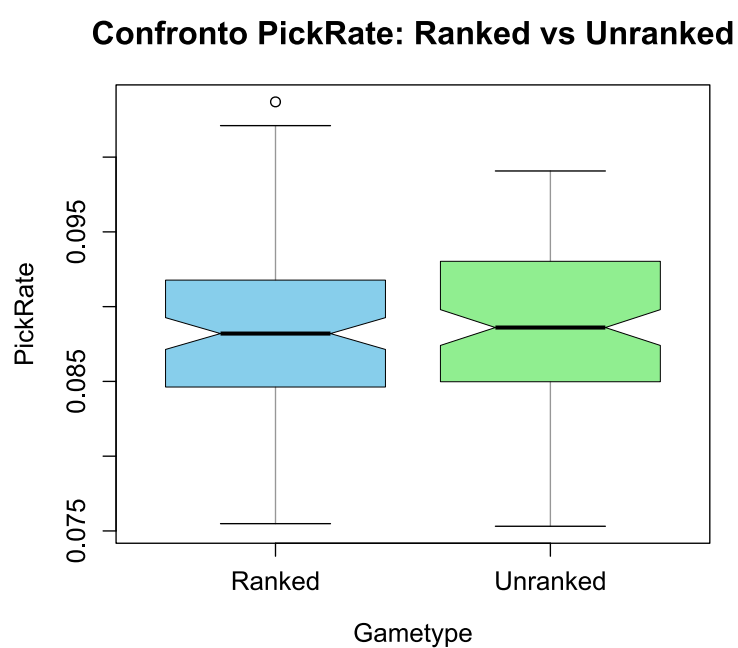


Figura 40: Nella seguente figura sono mostrati i boxplot ad intaglio per il pickrate nelle partite ranked e unranked

Tabella 10: Tabella riassuntiva degli indici di sintesi per il winrate per le partite ranked e unranked

	ranked	unranked
media	0.5008	0.5000
mediana	0.5074	0.5132
deviazione standard	0.08350946	0.07758461
varianza	0.00697383	0.006019371

In questo caso le distribuzioni non mutano sensibilmente tra le partite ranked e quelle unranked, come è possibile vedere nella figura 37, il che è testimoniato anche dagli indici presenti nella tabella 11. Rispetto al dataset originale però, si hanno meno differenze tra gli Hero che superano quella che è la soglia stabilita. In particolare 27.4% contro il 37% del dataset originale.

In conclusione per quanto riguarda il pickrate, è presente una bassissima variabilità

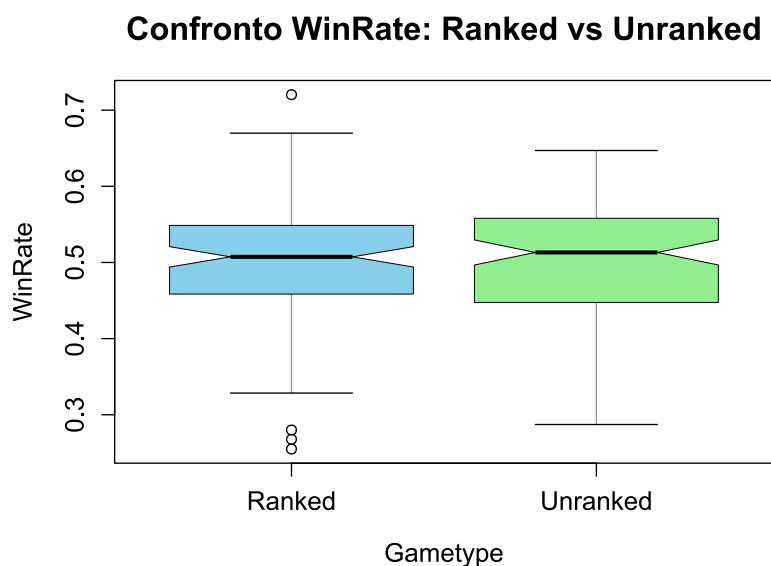


Figura 41: Nella seguente figura sono mostrati i boxplot ad intaglio per il winrate nelle partite ranked e unranked

per il singolo Hero ma una variazione della distribuzione del pickrate. Per il winrate, le distribuzioni risultano essere simili tra di loro mentre per i singoli Hero sussistono comunque delle differenze importanti.

Tabella 11: Tabella riassuntiva degli indici simmetrici per winrate nelle partite ranked e unranked

	ranked	unranked
skewness	-0.4500729	-0.5209041
curtosi	3.495748	2.78405

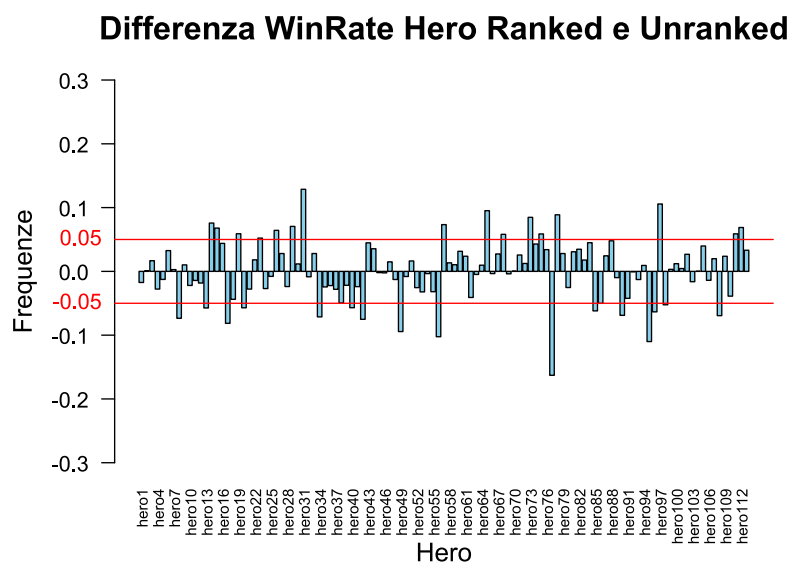


Figura 42: Nella seguente figura è riportata la differenza dei winrate per i singoli Hero nelle partite ranked e unranked

Tabella 12: Tabella riassuntiva del winrate di ciascuno dei due team per le partite ranked e unranked

	ranked	unranked
team 1	57.77%	55.66%
team 2	42.23%	44.34%

3.3 RQ3:I winrate dei Radiant e dei Dire sono differenti?

Si verifica ora se sussistono delle differenze notevoli di winrate tra i due team di Dota2. I winrate per ciascuno dei due team sono riportati all'interno della seguente tabella 12. Rispetto al dataset originale si può notare che la differenza di winrate tra i due team risulta essere molto più marcata. Se si dovesse interpretare l'insieme delle partite generate come un insieme di partite in un determinato stato del gioco, tale stato di gioco sarebbe da considerare estremamente sbilanciato a favore del team 1.

3.4 RQ4:Esiste una correlazione tra pickrate e winrate?

Si procede quindi a verificare se esiste una correlazione tra winrate e pickrate nei dati generati in maniera sintetica. Di seguito si mostrano i diagrammi di dispersione 43 44 per le partite ranked e unranked. A partire da questi ultimi è possibile notare come sia difficile individuare una correlazione tra pickrate e winrate. Infatti, tutte quante le tecniche che sono state utilizzate sul dataset originale hanno restituito dei risultati per r-squared ancora più bassi(spesso quasi approssimabili a zero).

Si conferma che anche per quanto riguarda i dati generati in maniera sintetica non sono presenti delle correlazioni banali tra pickrate e winrate. Per quanto riguarda però i dati sintetici, pur dividendo in quattro quadranti lo scatterplot non si riesce a trarre nessuna conclusione che avrebbe un significato rispetto al dominio di riferimento. Questo perchè i dati generati hanno una variabilità del pickrate praticamente inesistente, tanto da renderli palesemente artificiali.

La relazione tra pickrate e winrate che era difficile da descrivere ma che era presente per quanto riguarda i dati del dataset originale, sembra quindi non esser stata replicata dalla generazione dei dati.

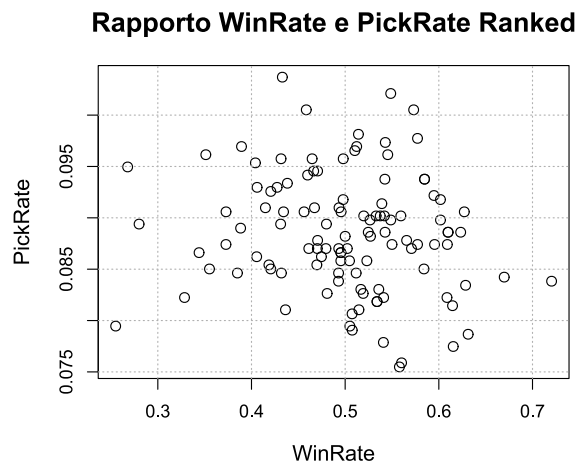


Figura 43: Diagramma di dispersione del winrate e del pickrate degli Hero nelle partite classificate

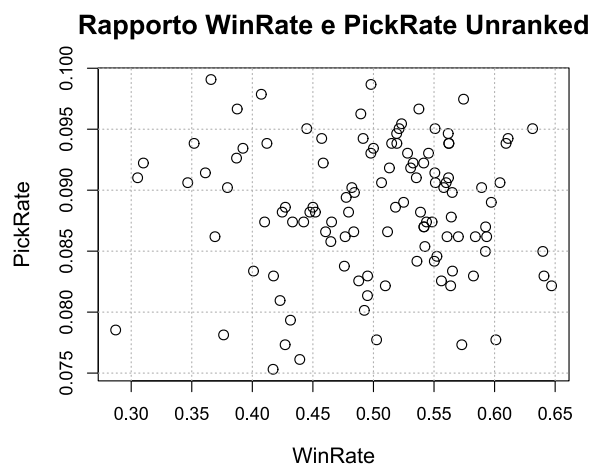


Figura 44: Diagramma di dispersione del winrate e del pickrate degli Hero nelle partite non classificate

3.5 RQ5:Esistono degli attributi più incisivi all'interno del gioco?

Si tenta ora di rispondere alla research question 5 anche per quanto riguarda il dataset sintetico. I concetti di pickrate e winrate, sono i medesimi di quando è stata effettuata l'analisi per il dataset originale. Nel seguente grafico 45 è stato riportato il pickrate per ciascuno degli attributi per le partite ranked. Rispetto al dataset originale, si può notare che i pickrate degli attributi sono molto più simili tra di loro. Per quanto riguarda invece il winrate, si osservano diverse differenze nei valori associati agli attributi Strength e Agility, che risultano nettamente superiori rispetto a Intelligence e Universal. Di conseguenza, emerge una discrepanza rispetto al dataset originale, in cui tutti i valori di winrate sono simili (intorno al 50%) e solo Intelligence risulta effettivamente inferiore rispetto agli altri attributi. All'interno della seguente figura 47 è stato riportato un baplot per ciascuno degli

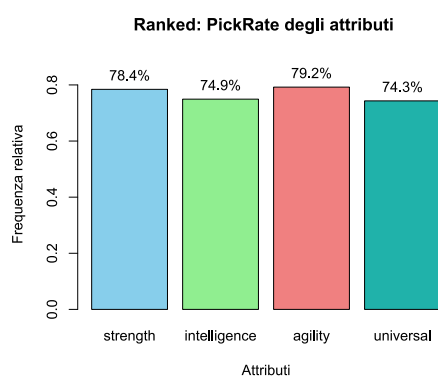


Figura 45: La seguente figura è un barplot del pickrate per ciascuno degli attributi nelle partite ranked del dataset sintetico

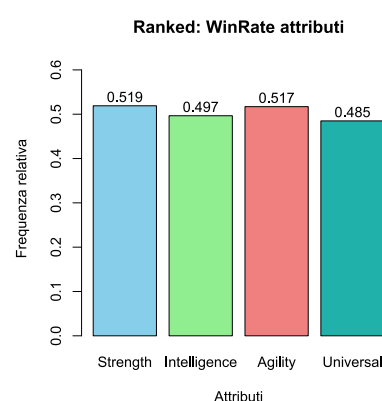


Figura 46: La seguente figura è un barplot del winrate per ciascuno degli attributi nelle partite ranked

attributi nelle partite ranked. All'interno di ciascun grafico è stata utilizzata una colonna per riportare la percentuale di squadre in cui compaiono esattamente un certo numero di Hero con un determinato attributo. Anche per il dataset sintetico, si è deciso di escludere tutte quante le configurazioni di team che hanno una frequenza minore del 5% sono state tagliate dal resto dello studio per questa research question. Le frequenze relative di team in cui compaiono un esatto numero di Hero con un determinato attributo hanno valori molto simili tra di loro. Questa è una notevole differenza rispetto al dataset originale. All'interno della seguente figura 47 invece sono riportati i barplot del winrate per ciascuno degli attributi. In questo caso è possibile notare, che rispetto al caso originale per gli attributi Strength e Agility la situazione sembra essere più o meno invariata, sono invece presenti delle notevoli differenze per quanto riguarda Intelligence e Universal, per i quali la cosa migliore è avere esattamente un Universal o Intelligence in squadra. Osservando i

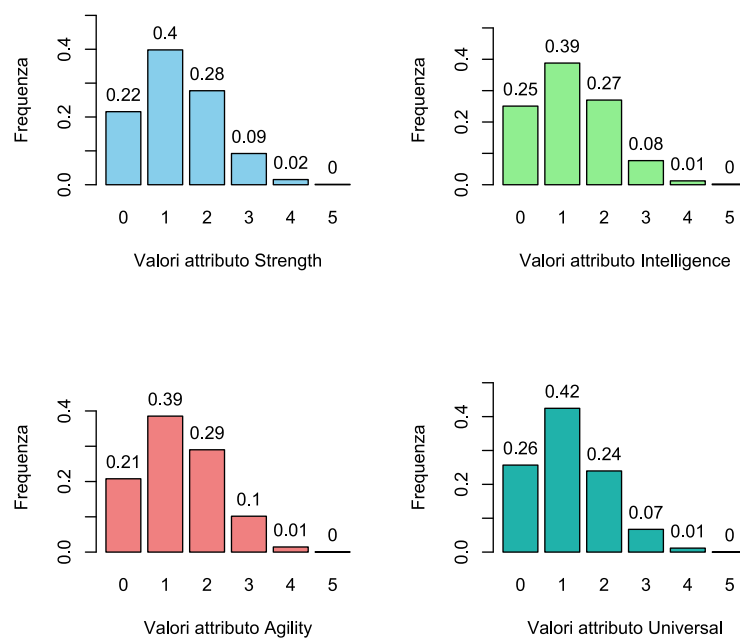


Figura 47: Nella seguente figura sono riportati 4 barplot, uno per ogni attributo del gioco. All'interno di ciascun grafico è stata utilizzata una colonna per riportare la percentuale di squadre in cui compaiono esattamente un certo numero di Hero con un determinato attributo. I seguenti grafici sono riferiti alle partite ranked del dataset sintetico.

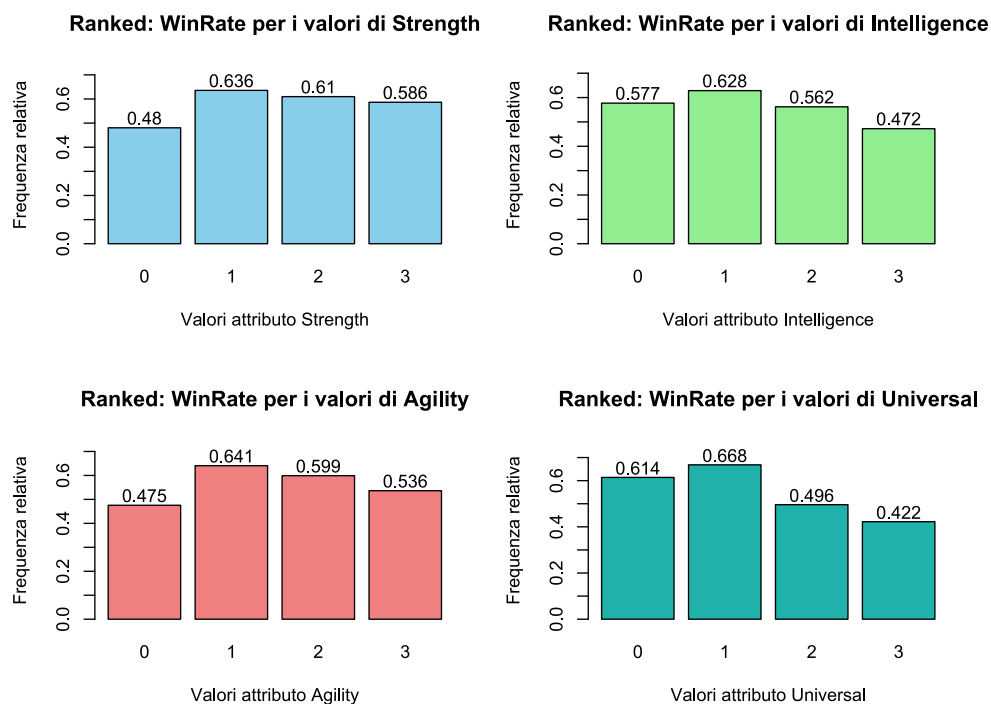


Figura 48: Nella seguente figura sono riportati 4 barplot, uno per ogni attributo del gioco. All'interno di ciascun grafico è riportato il winrate dei team che possiedono un certo numero di Hero con un determinato attributo. I seguenti grafici sono riferiti alle partite ranked dell'attributo sintetico.

grafici è possibile quindi notare che per le partite ranked del dataset sintetico la scelta migliore è quella di avere un Hero per ciascuno degli attributi. Per quanto riguarda le partite unranked, dai seguenti grafici (49 50 51 52) si nota che la situazione rimane sostanzialmente invariata rispetto alle partite ranked. L'unica differenza significativa è che, per l'attributo Agility, la scelta ottimale è avere due giocatori in squadra anziché uno. Pertanto, nel dataset sintetico, non sussiste quindi nessuna particolare differenza su quelli che sono gli attributi più convenienti tra le partite ranked e quelle unranked del dataset sintetico, a differenza del dataset originale.

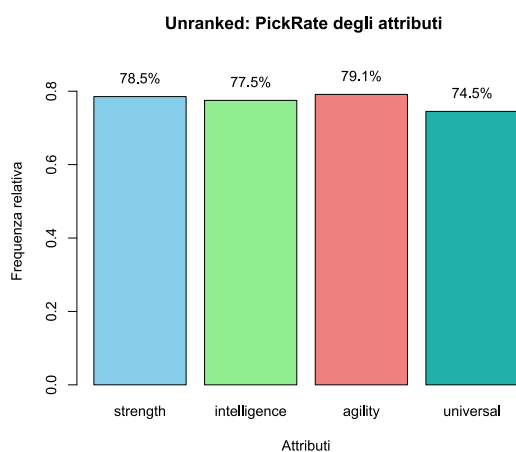


Figura 49: La seguente figura è un barplot del pickrate per ciascuno degli attributi nelle partite unranked del dataset sintetico

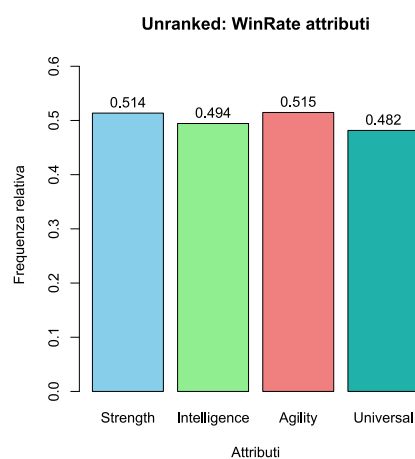


Figura 50: La seguente figura è un barplot del winrate per ciascuno degli attributi nelle partite unranked del dataset sintetico

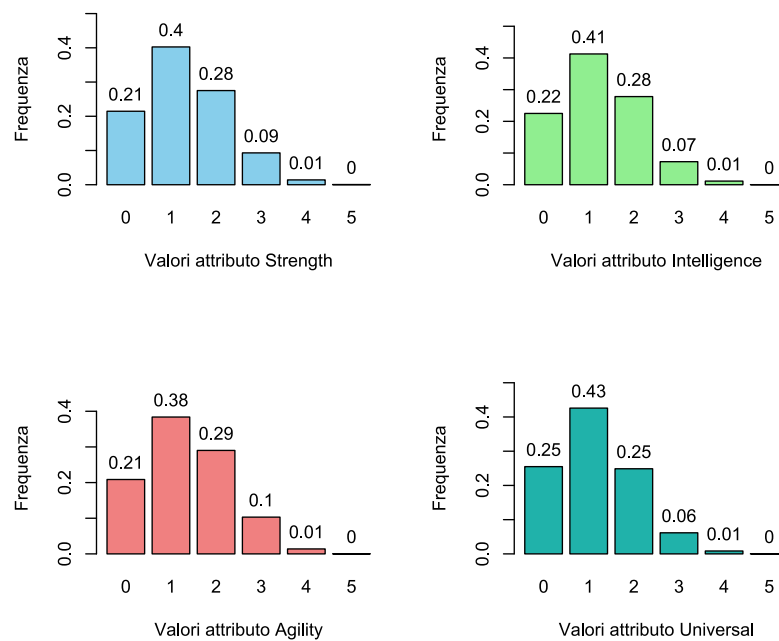


Figura 51: Nella seguente figura sono riportati 4 barplot, uno per ogni attributo del gioco. All'interno di ciascun grafico è stata utilizzata una colonna per riportare la percentuale di squadre in cui compaiono esattamente un certo numero di Hero con un determinato attributo. I seguenti grafici sono riferiti alle partite unranked del dataset sintetico.

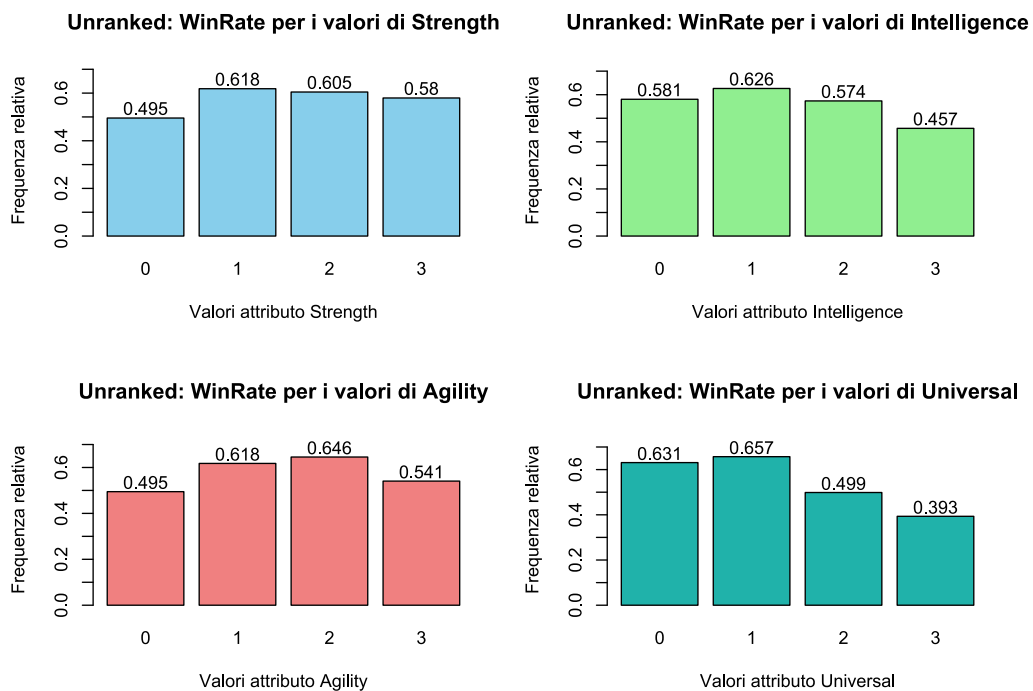


Figura 52: Nella seguente figura sono riportati 4 barplot, uno per ogni attributo del gioco. All'interno di ciascun grafico è riportato il winrate dei team che possiedono un certo numero di Hero con un determinato attributo. I seguenti grafici sono riferiti alle partite unranked dell'attributo sintetico.

3.6 RQ6:Esistono configurazioni dei team più convenienti?

Nella seguente research question 6 si cercherà di definire se esistono dei team che sono più convenienti dal punto di vista statistico per ottenere la vittoria. Anche in questo caso si è deciso di scartare tutte quante le configurazioni di team che sono presenti in meno del 10% delle partite. In questo caso abbiamo un team in meno nella top rispetto alle classifiche del dataset originale.

In questo caso come è possibile osservare dai winrate e dalle configurazioni dei team che sono presenti all'interno delle tabelle 13 14 sia nelle partite ranked che in quelle unranked i team presenti sono gli stessi anche se con delle differenti percentuali di vittoria.

Come nel caso del dataset originale, le configurazioni di team più vincenti sono coerenti rispetto a quelli che sono i risultati ottenuti nelle research question precedente, quindi durante l'analisi del winrate per attributi. Si era infatti concluso che convenisse sia nelle partite ranked che in quelle unranked avere almeno un Hero per ciascun attributo per massimizzare le possibilità di vittoria.

In modo simile al dataset originale, si può concludere che anche in questo caso esistono configurazioni di team più convenienti rispetto ad altre. L'unica differenza significativa è che, nel dataset sintetico, le configurazioni di team più vincenti presentano una percentuale di vittoria decisamente più elevata.

Tabella 13: Nella seguente tabella sono riportati le 4 configurazioni di team che hanno il winrate più alto all'interno delle partite ranked.

Strength	Intelligence	Agility	Universal	WinRate
2	1	1	1	58 %
1	1	2	1	57 %
1	2	1	1	49.5%
1	1	1	2	48.4%

Tabella 14: Nella seguente tabella sono riportati le 4 configurazioni di team che hanno il winrate più alto all'interno delle partite unranked.

Strength	Intelligence	Agility	Universal	WinRate
1	1	2	1	57.1 %
1	2	1	1	52.7 %
2	1	1	1	50.2 %
1	1	1	2	43.5 %

3.7 RQ7:Qual'è il rank delle istanze del dataset?

Infine si procede a rispondere alla research question 7 anche per il dataset sintetico, all'interno della quale si cercherà di determinare se l'elo per le partite ranked e unranked segue o meno una distribuzione nota e se questa distribuzione individuata è riconducibile dal campione alla popolazione. Come per il dataset originale il valore dell'elo di ogni partita del dataset sintetico è stato calcolato attraverso l'utilizzo della medesima formula utilizzata per il dataset originale.

All'interno delle seguenti figure () sono presenti i grafici delle distribuzioni dell'elo, rispettivamente per le partite ranked e unranked. Proprio come nel caso del dataset originale entrambi i grafici ci suggeriscono che l'elo abbia una distribuzione normale. Quindi per determinare ciò abbiamo utilizzato il test del chi-quadro per verificare se la distribuzione è effettivamente una normale o meno. Nel caso delle partite ranked il valore di chi-quadro è 15.16329 mentre per le partite unranked è 9.958115, dato che in entrambi i casi il valore non rientra nell'intervallo di accettazione del test del chi-quadro(0.05063562, 7.377759) si può affermare che l'elo sia per le partite ranked che per quelle unranked non segue una distribuzione normale.

Di conseguenza, anche se la forma della distribuzione non sembra seguire una distribu-

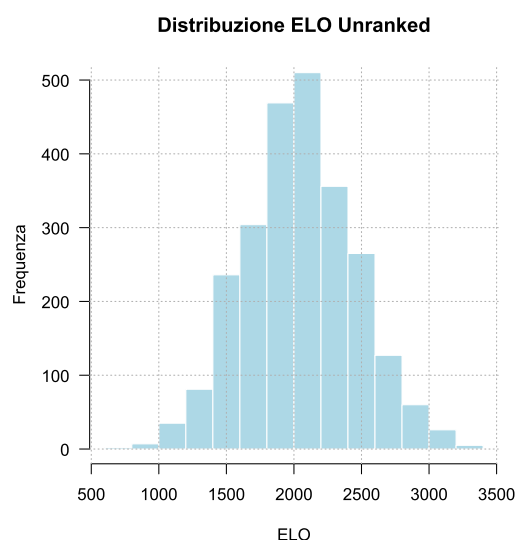
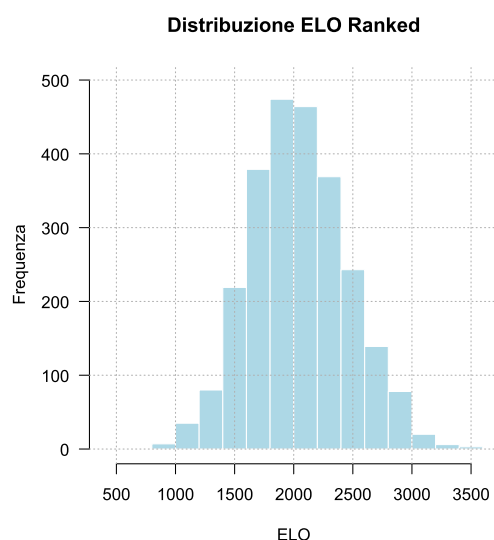


Figura 53: Nella seguente figura è presente un grafico della distribuzione dell'elo calcolato all'interno delle partite ranked

Figura 54: Nella seguente figura è presente un grafico della distribuzione dell'elo calcolato all'interno delle partite unranked

zione di Poisson, si è comunque deciso di applicare il test del chi-quadro per verificarlo. Nel caso delle partite ranked, il valore del chi-quadro è 397.6751, mentre per le partite

unranked è 879.4748. Poiché questi valori non rientrano minimamente nell'intervallo di accettazione del test, si può confermare che le osservazioni fatte sulle distribuzioni dell'Elo sono corrette.

Si conclude quindi che entrambe le distribuzioni non seguono alcuna distribuzione nota, il che impedisce di proseguire lo studio e di determinare il valore medio dell'Elo della popolazione per il dataset sintetico.

3.8 Conclusioni

In conclusione, la generazione di dati sintetici è riuscita a simulare le differenze tra le partite ranked e unranked per quanto riguarda il pickrate e il winrate degli Hero. Per il pickrate però è necessario notare che rispetto al dataset originale i dati generati dal dataset sintetico dimostrano una perfetta ripartizione del pickrate tra i vari Hero. La generazione dei dati non è riuscita a cogliere le complicate relazioni che intercorrono tra i vari elementi che costituiscono le partite. Questo è dimostrato dal fatto che non è stato possibile neanche individuare delle aree all'interno dello scatterplot winrate/pickrate, come invece è stato possibile per il dataset originale. I dati generati variano sensibilmente anche per quanto riguarda il winrate dei team Radian e Dire. Se per quanto riguarda il dataset originale la distribuzione dell'elo per le partite ranked è riconducibile a una distribuzione nota(normale) questo non è vero per il dataset sintetico.

Il dataset sintetico generato in questo caso non si ritiene attendibile per riuscire a migliorare l'analisi statistica oppure per migliorare i risultati della regressione. A nostro parere, le feature del dataset risultano inadeguate per cercare di generare delle nuove istanze di partite che possano rispecchiare le relazioni tra intercorrono tra gli elementi del gioco e quindi risultare utili per migliorare i risultati dell'analisi statistica.

Bibliografia

- [1] *Dota2 Sito Ufficiale*. <https://www.dota2.com>.
- [2] *LoL: How Much Does Side Selection Matter, Really?*
<https://www.itero.gg/articles/side-selection>.
- [3] *Radiant side has a massive advantage in the current Dota 2 meta*.
<https://win.gg/news/radiant-side-has-a-massive-advantage-in-the-current-dota-2-meta/>.