

# Diseño de un modelo predictivo para jugadores profesionales en Dota 2

Felipe Cardozo, codigo 0321040262

**Abstract**—The eSport have been growing since its first version in the World Cyber Games held in Korea in 2000, given this growing is difficult to ignore for its variety of games, competitions and disciplines found in the different aspects that can be covered in this type sports. The research on Dota 2, a completely booming game due to its multiple world championships is becoming increasingly important in the areas of machine learning, a key aspect is the relevance of teams that face in the major leagues. Studies are currently being carried out in different aspects such as Open AI, a company that studies the movements of each player to generate Robots that can compete against said professional players and be able to take the victory, data mining, on the other hand, has been thoroughly studied. and it would not be the exception in this game. Applying supervised learning techniques to predict who wins a game is the objective of this article, trying to show a methodology which leads us to understand a little better the knowledge of the games and what the characteristics might be after processing the data obtained from Open Dota, the recent developments in this video game industry show us that data and information will be knowledge to help us understand which characteristics may be most relevant when winning a game.

**Index Terms**—Machine learning, supervised learning, confusion matrix, accuracy, decision tree



## 1 INTRODUCCION

Dota 2 es un juego gratuito Multiplayer Online battle arena (MOBA) desarrollado por la corporación Valve. Consta de dos equipos llamados Dire y Radiant, cada equipo se conforma por 5 héroes diferentes, gana el equipo que destruya el edificio “ancient” del otro equipo. Cada héroe tiene 25 niveles que se ganan a medida que evoluciona el tiempo de la partida bajo ciertas condiciones de pelea en el mapa, y todos sin excepción inician en el primer nivel. No existe un límite de tiempo para lograr destruir el edificio objetivo “ancient” y su mapa al igual que muchos otros MOBA como League of Legends o Heroes of the Storm, se basa en tres líneas protegidas con torres y para destruir el ultimo edificio es necesario primero destruir estas torres en las diferentes líneas, existen cerca de 215 héroes elegibles, cada uno con diferentes poderes y habilidades. *an article or it will be returned for reformatting.*

Existen ligas de diferentes tipos, existen las más grandes que son la premier y las ligas mayores, que son donde los equipos élite compiten para llevarse un gran premio, su liga de mayor representación es “The International” donde el premio a obtener como primer lugar es de 11’190.158 USD [2]. Existen otras ligas donde fácilmente el primer lugar puede ganar 350.000 USD [3].

Los equipos profesionales, en la actualidad para la liga Premier The International tiene un modo llamado el compendio o “compendium” que se abren una temporada específica, por lo general 1 vez al año donde aquellos usuarios que compran un pase de batalla en cierta temporada del año tienen derecho a esta característica, esta característica tiene la posibilidad de llenar predicciones, en un conjunto de partidas, y predecir quien va a ganar, después de esto a medida que avanza el compendio y el torneo, si el usuario acierta en

su predicción ganará algunos puntos que al final serán redimidos por premios dentro del juego [4].

El juego en este punto es muy versátil, cada liga tiene unos brackets muy particulares, y casi que son únicos, es decir, los equipos al igual que un mundial de Fútbol, tiene clasificados los mejores equipos, y su apertura inicia con una nueva combinación de equipos para cada liga.

Se realizan diferentes tipos de ligas en Dota 2: ligas premier, ligas mayores y ligas menores y algunas líneas amateurs que son online. Al año, se juegan alrededor de 10 ligas premier [6], 17 ligas mayores [7], cada una de estas partidas se puede asistir presencialmente o ser espectador por algún canal de streaming como twitch.tv para ver cada partida.

A Dota 2 se conectan mensualmente cerca de 658.000 personas [8], una de sus ligas más importantes a nivel mundial es “The International” el cual puede llegar a tener un total de 14’960.473 espectadores de todo el mundo y cada partida de una liga premier puede llegar a tener esos mismos 14 millones de espectadores simultáneamente [9]. Su comunidad es bastante grande, no solamente desde el punto de vista de espectadores sino además de jugadores profesionales, amateurs, o jugadores que buscan pasar un rato divertido.

### 1.1 Machine Learning

El juego en este punto es muy versátil, cada liga tiene unos brackets muy particulares, y casi que son únicos, es decir, los equipos al igual que un mundial de Fútbol, tiene clasificados los mejores equipos, y su apertura inicia con una nueva combinación de equipos para cada liga.

## 2 PREPROCESAMIENTO

### 2.1 Open Dota

Open Dota es un API que ofrece el tipo de informacion

necesaria para llevar a cabo esta investigación, al extraer un set de datos considerablemente grande, en esta etapa, se ha creado un microservicio usando Spring Boot para hacer la petición al API y guardar las partidas guardadas en la base de datos MongoDB. En la actualidad se ha llegado a obtener un poco mas de 22.000 registros de partidas de jugadores profesionales.

## 2.2 Parding Data

Algunos campos como los héroes o "heroId", serán necesarios trabajarlos de manera numérica.

## 2.3 Características

Se presenta la siguiente tabla que puede servir como guía de los campos relevantes para esta investigación.

Campo	Descripción
accountId	Representa el ID del jugador profesional
Assists	La cantidad de asistencias para obtener la muerte de un jugador o héroe opo- nente
HeroId	De 115 heroes disponibles en el juego, el ID representa ese héroe
Lane	Existen una serie de líneas que cada jugador puede elegir
neutralKills	Representa la cantidad de creeps que fueron obtenidos por el héroe
totalXP	La cantidad de experiencia ganada en dicha partida
totalGold	Total, oro ganado en dicha partida
Win	Representa la variable a predecir

## 2.4 Experimentos

Dado el conjunto de datos obtenidos de open dota, se ha corrido con diferentes algoritmos para darnos una primera perspectiva de las características seleccionadas y el conjunto de datos preseleccionado, teniendo en cuenta que win fue configurado como label.

### Naive Bayes

La ejecución de este método nos llevo a un accuracy de 71.12% analicemos la siguiente matriz de confusión.

	Gana	No gana
Gana	777	252
No Gana	293	565

Matriz de Confusión

En la anterior tabla analizamos a través de Bayes, que logra de aquellas partidas que gana un jugador 772 efectivamente son partida que logrará ganar, sin embargo 252 son aquella que prediciría ganar, pero en realidad no ganó.

La Sensibilidad que nos da este método para este data set es de 0.72616 y Especificidad de 0.69155

### ANN

La ejecución de este método aumento considerablemente el accuracy, teniendo un total de 77.36%, observemos a continuación la siguiente tabla.

	Gana	No gana
Gana	3020	877
No Gana	547	1847

En efecto, acertó en 3020 respecto a su opuesto de no ganar en 1847, cuando el algoritmo asumió que no ganaría el resultado fue 877 donde no gana el jugador la partida.

La especificidad fue de 0.84664 y la sensibilidad de 0.67804

### KNN

Ejecutando este método particularmente nos da un accuracy de 58.85% con K=1, es el menor de rendimiento si lo comparamos con el set de algoritmos probados.

	Gana	No gana
Gana	2252	1274
No Gana	1315	1450

Casi se puede concluir que puede acertar la mitad de las partidas de un jugador profesional.

La especificidad fue de 0.63134 y la sensibilidad de 0.53230

Ahora aumentando el K=7 miramos una mejora considerable con un accuracy de 64.08% con la siguiente matriz de confusión.

	Gana	No gana
Gana	2537	1230
No Gana	1030	1494

Vemos notablemente como aumento su exactitud y eficiencia con una distancia mayor en sus vecinos.

La especificidad fue de 0.71124 y la sensibilidad de 0.54845

### SVM

La ejecución de maquina de soporte vectorial nos retorno un accuracy de 62.52% obteniendo la siguiente tabla de matriz de confusion

	Gana	No gana
Gana	3142	1933
No Gana	425	791

La especificidad fue de 0.88085 y la sensibilidad de 0.29038

### Decision Tree

Para este en particular los atributos de arriba abajo dando presendencia según el algoritmo.

Campo	Humbral
Assits	6500
HeroKills	11500
TotalXP	9500

<b>NeuralKills</b>	161500
<b>totalKills</b>	36640

### Random Forest

Se ejecuta este algoritmo con numero de arboles igual a 100. Si comparamos con decisión tree toma los mismos atributos y en cada árbol que genera aleatoriamente tomando totalGold con un humbral de 2178, en otros arboles toma como raíz heroKills, assits nuevamente como raíz del árbol con diferentes humbrales.

## 7 CONCLUSION

Aunque se puedan apreciar diferentes métodos de aprendizaje supervisado, es importante valorar la validez de los datos para este tipo de investigación. Escencialmente Dota es un juego complejo que comprende muchas características como se puede apreciar en el API, sin embargo, resulta interesante poder analizar los datos frescos dado las investigaciones anteriores. En este punto cabe aclarar que el método mas acertado y efectivo se encuentra con Redes Neuronales Artificiales, el cual nos brinda un accuracy del 77.36%.

## 8 TRABAJO FUTURO

Es importante la predicción de este tipo de partida, porque nos da una significancia y además nos permite ver que los jugadores profesionales pueden tener cierto rendimiento en un campeonato profesional, seria interesante teniendo este set de datos y resultados, poder modelar una predicción acerca de las ligas mayores y predecir que equipo puede ganar una partida dada la probabilidad de sus partidas anteriormente jugadas de manera singular.

## REFERENCES

- [1] V. Corp, "Dota 2 - The International," Valve, 20 08 2018. [Online]. Available: <http://www.dota2.com/international/overview/>. [Accessed 06 11 2018].
- [2] Research Gate, "Discover scientific knowledge, and make your research visible,," researchgate.net, 2018. [Online]. Available: [https://www.researchgate.net/figure/Map-of-Dota-2-from-Dota-2-wiki-7\\_fig1\\_262207918](https://www.researchgate.net/figure/Map-of-Dota-2-from-Dota-2-wiki-7_fig1_262207918). [Accessed 21 11].
- [3] Liquid, "Liquipedia Kuala Lumpur Major," Team Liquid, 2018. [Online]. Available: [https://liquipedia.net/dota2/PGL/Kuala\\_Lumpur\\_Major](https://liquipedia.net/dota2/PGL/Kuala_Lumpur_Major). [Accessed 06 11 2018].
- [4] Dota2, "Battlepass TI," Dota 2, 2018. [Online]. Available: <https://www.dota2.com/international/battlepass/>. [Accessed 08 11 2018].

- [5] Liquipedia, "Liquipedia," Liquipedia, 2018. [Online]. Available: [https://liquipedia.net/dota2/The\\_International/2018](https://liquipedia.net/dota2/The_International/2018). [Accessed 20 11 2018].
- [6] T. Liquid, "Premier tournaments," Liquid, 2018. [Online]. Available: [https://liquipedia.net/dota2/Premier\\_Tournaments](https://liquipedia.net/dota2/Premier_Tournaments). [Accessed 05 11 2018].
- [7] t. Liquid, "Major tournaments," Team Liquid, 2018. [Online]. Available: [https://liquipedia.net/dota2/Major\\_Tournaments](https://liquipedia.net/dota2/Major_Tournaments). [Accessed 05 11 2018].
- [8] , "An ongoing analysis of Steam's concurrent players,," Steamcharts.com, 2018. [Online]. Available: <https://steamcharts.com/app/570#All>. [Accessed 05 11 2018].
- [9] "Researching esports and streaming trends," ESM.one, 2018. [Online]. Available: <https://esc.watch/blog/post/stats-international-2018>. [Accessed 05 11 2018].
- [10] [10] M. M, R. A and T. and A, Foundations of Machine learning, London, England: The MIT Press, 2012.