**Prediciendo partidas en Dota 2 para equipos profesionales**Anteproyecto  
Línea de investigación del grupo FICB-PG: Línea de Investigación en educación y tecnología **Rene Felipe Cardozo 0321040262**Asesor: Javier Niño

*Noviembre 2018*

**\_\_\_\_\_\_\_\_\_\_\_\_**

**Resumen**

La investigación que se realizará en esta tesis consistirá en explorar los datos de las partidas del juego Dota 2, donde se encontrarán las partidas singulares de los jugadores pertenecientes a una liga profesional, y en las partidas realizadas en ligas mayores y premier. Por un lado, analizaremos las partidas singulares de un jugador profesional, entrenando una serie de algoritmos de aprendizaje supervisado y obtener un modelo predictivo. Por otro lado, se implementará un árbol de decisión que será a su vez alimentado por el anterior modelo aplicando este árbol en el campo de las ligas mayores y premier, para predecir un equipo ganador.

**Palabras clave:** Machine learning, modelo predictivo, minería de datos, aprendizaje supervisado.

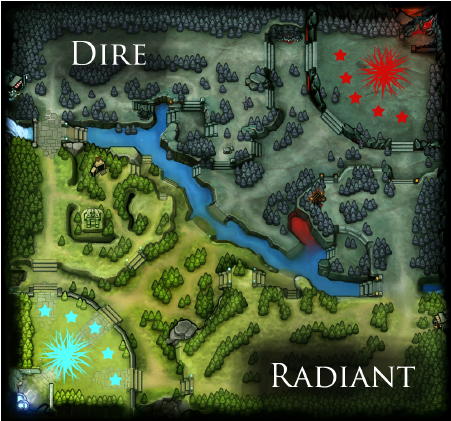
**Abstract**

The research that will be carried out in this thesis will consist in exploring the data of the games of the game Dota 2, where the unique games of the players belonging to a professional league will be found, and in the games made in the major and premier leagues. On the one hand, we will analyze the unique games of a professional player, training a series of supervised learning algorithms and obtain a predictive model. On the other hand, a decision tree will be implemented that will be fed by the previous model, applying this tree in the field of the major and premier leagues, to predict a winning team.

**Key words:** Machine learning, predictive model, data mining, supervised learning.

**INTRODUCCIÓN**

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 limite 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.



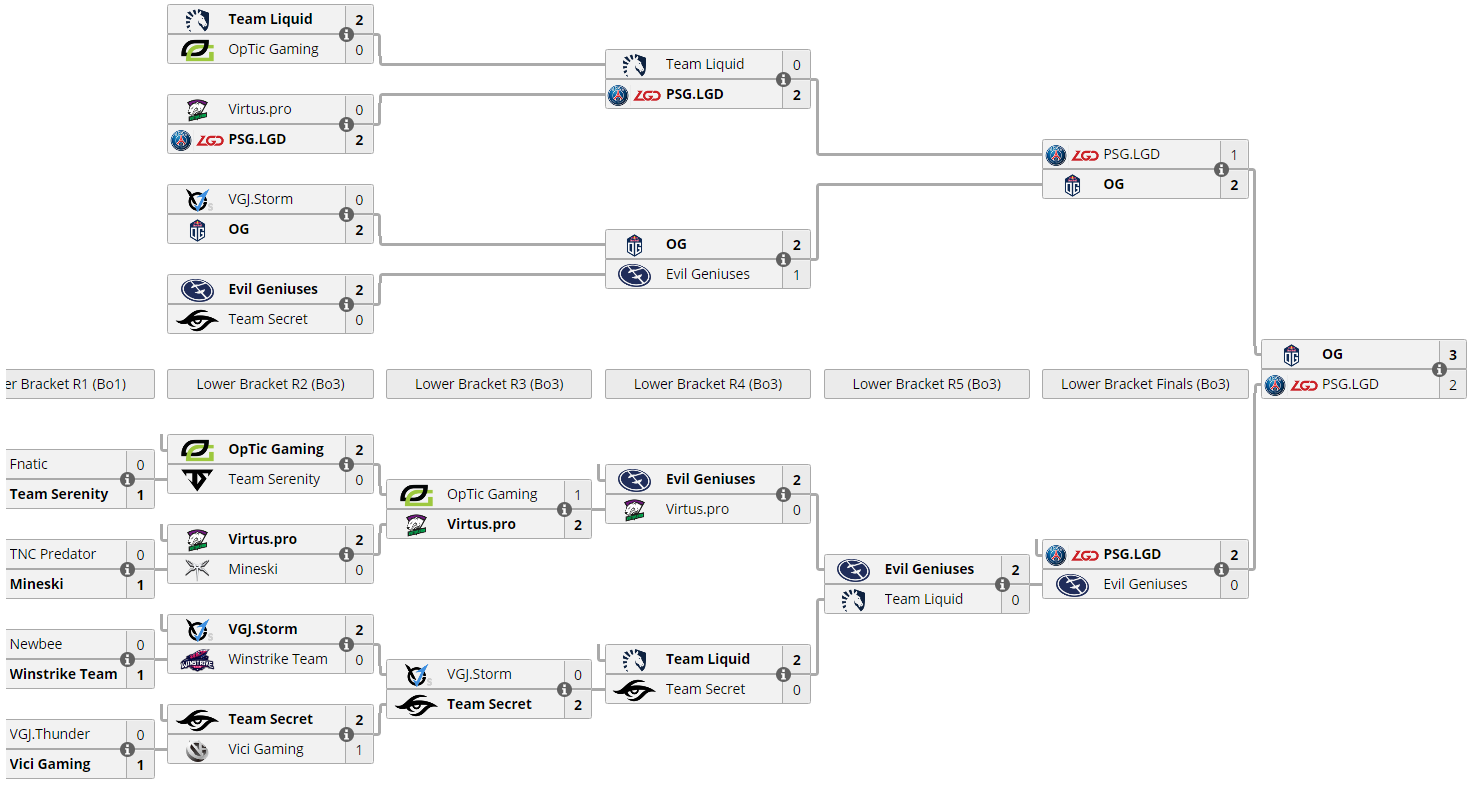
[1]

**Equipos profesionales**

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 abreen 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 equipo al igual que un mundial de Futbol, tiene clasificados los mejores equipos, y su apertura inicia con una nueva combinación de equipos para cada liga.



[5]

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.

**TRABAJO RELACIONADO (ANTECENDENTES) ¿**

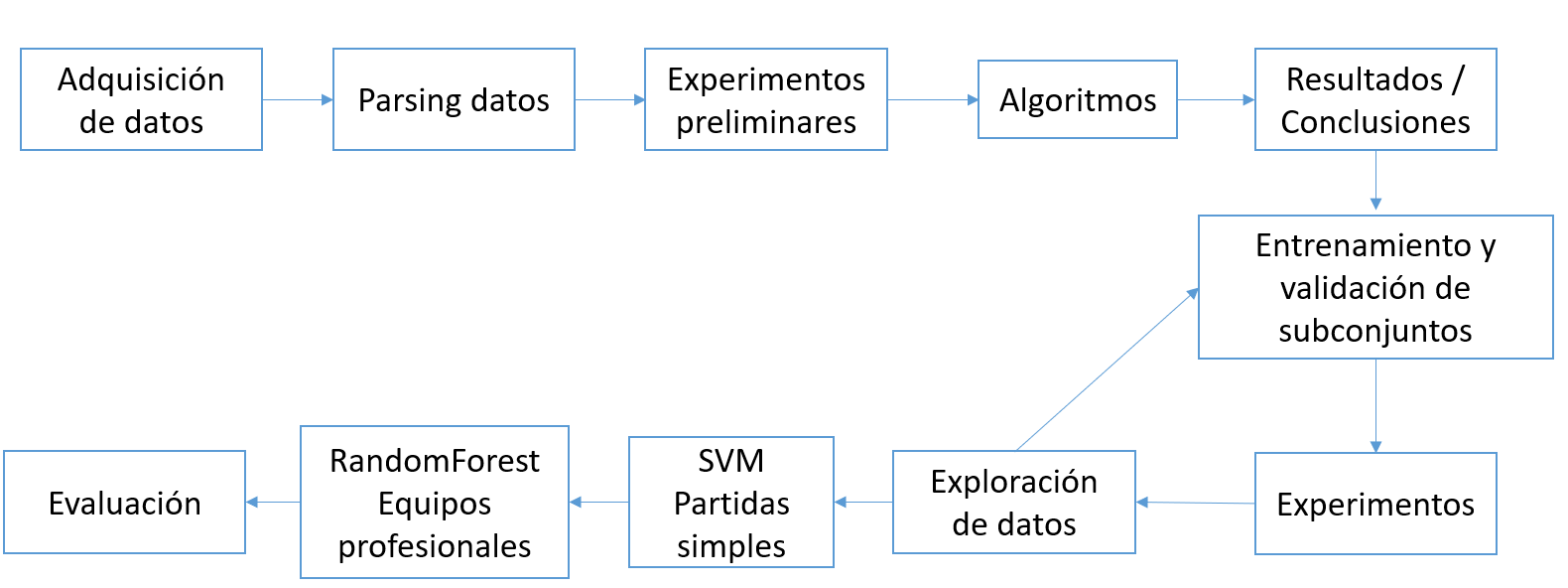
La predicción de juegos se ha realizado en deportes como el futbol [10]. Sin embargo, esas predicciones fueron basadas en partidos de futbol anteriores y circunstancias previas a cada partido: como el clima y el número de lesiones de un jugador. La mayoría de investigación orientada a big data, machine learning y minería de datos que está aplicada a los e-sports (deportes electrónicos) ha sido orientado únicamente al modelamiento de inteligencia artificial para jugadores maquinas o Bots [11]. También se han creado modelo predictivo para predecir que equipo en una partida regular resulta ganador dado un conjunto de datos implementando Random Forest [12].

**Machine Learning**

Machine learning es un amplio campo de las ciencias de la computación que gobiernan sistemas que pueden aprender de un conjunto de datos en orden de mejorar el rendimiento de una tarea específica [13]. En este campo se encuentra una subsección de problemas llamado aprendizaje supervisado, donde la entrada es usada para crear un modelo el cual produce una salida deseada [13]. Si vemos los métodos de aprendizaje como una caja negra, decimos que la entrada son un conjunto de instancias donde cada instancia tiene un vector de atributos, uno de esos siendo la salida deseada. Los modelos son construidos para emparejar las relacionas desde los atributos al atributo objetivo o deseado encontrado en los datos. Si este atributo es un valor nominal, es llamada la clase y el modelo un clasificador, allí es donde hay bastante variedad de algoritmos de aprendizaje debió a las diferencias en como estos operando para construir modelos que funcionen internamente diferente, es decir cada uno tiene su propia implementación si lo vemos desde el punto de vista estratégico.

**METODOLOGIA**

La metodología que ya se encuentra en progreso en este proyecto de investigación se presenta a continuación, seguida de su explicación por cada etapa.

****

**Adquisición de datos**: En esta etapa se presenta el API y se construye un software capaz de obtener los datos por cada jugador profesional que activamente pertenezca a un equipo de una liga profesional. Este proceso se encuentra en progreso, en la actualidad para la construcción del microservicio se eligió Spring Boot por su facilidad y adaptabilidad a llamados REST. Almacenando todos estos llamados en una Base de datos mongoDB.

**Parsing data**: Esta etapa se trata de garantizar la homogenización de datos encontrados a lo largo de los documentos json, en algunos casos, algunos atributos se encuentran de tipo nominal, y se pretende hacerles un tratamiento para este tipo de características y hacer conversión numérica.

**Experimentos preliminares**: En esta etapa nos centraremos en generar un primer conjunto de entrenamiento obtenido del histórico de partidas jugadas de los jugadores profesionales seleccionados.

**Algoritmos**: En esta etapa, a través del conjunto de entrenamiento se harán pruebas de algoritmos de machine learning, donde obtendremos una perspicacia inicial sobre la exactitud y primeras posibles salidas.

**Resultados-Conclusiones**: Aquí nos centraremos en analizas y concluir las etapas primarias de la metodología, que mejoras podremos hacer en los datos capturados, que algoritmos y métodos podríamos potencialmente elegir.

**Entrenamiento y validación de subconjuntos**: Esta etapa particularmente se harán validación de un conjunto de datos mas grande y completo, teniendo en cuenta lo aprendido en las etapas anteriores, y aplicar los modelos a los algoritmos elegidos.

**Experimentos**: Esta etapa será fundamental para aplicar los modelos y conjuntos de entrenamiento y de prueba elegidos en etapas previas.

**Exploración de datos**: Aquí presentamos nuestras ultimas etapas de prueba y error, dando la posibilidad de ser necesario retroceder un par de etapas para garantizar que tanto el conjunto de datos, como las características y atributos son los ideales para el modelo predictivo.

**SVM para partidas simples**: Aquí aplicaremos al conjunto de prueba nuestro modelo predictivo para partidas singulares de los jugadores profesionales.

**Random Forest para equipos profesionales**: Esta etapa, es importante porque se alimenta consecuentemente de la etapa previa, para poder construir el árbol de entrenamiento y poder aplicar el conjunto de prueba.

**Evaluación**: Finalmente, después de los datos y la información interpretada, podremos evaluar todos estos factores y poder llevar todo esto al conocimiento y generar un análisis basado en la experiencia de este ciclo metodológico.

|  |  |
| --- | --- |
| Nombre campo | Descripción |
| Win | Booleano que potencialmente será nuestro atributo por predecir |
| Players | Arreglo, trae información como experiencia ganada en la partida, jugadores que participaron, y los IDs de esos jugadores |
| Teamfights | Arreglo, que trae peleas en equipo con experiencia inicial y final |
| Deaths | Entero, indica cuantas muertes en un teamfight ocurrieron |
| League | Arreglo, trae información si viene de un tipo de liga |
| xp\_start | Entero, muestra la experiencia inicial |
| xp\_end | Entero, muestra la Experiencia final obtenida |
| killed | Arreglo, muestra que héroes pudo asesinar durante la partida |

Todo este procesamiento se debe hacer gradual debido a que algunos atributos en el documento de respuesta deberán ser preprocesados para un mejor uso en los métodos a usar.

**Objetivos**

Objetivo General

Diseñar un modelo predictivo usando máquinas de aprendizaje supervisado que prediga el equipo ganador en un torneo o liga de Dota 2 dado los datos parciales recopilados de ligas anteriores y partidas comunes de cada jugador profesional que participe en estas ligas.

Objetivos Específicos

* Determinar cuál es la relación entre las partidas que juega un jugador profesional fuera de la liga y el rendimiento de ese jugador cuando participa en una liga profesional.
* Proponer un modelo predictivo de que un equipo pueda o no ganar una partida en una liga teniendo en cuenta el historial de ligas.
* Elaborar un modelo en el cual se analicen los resultados singulares de los jugadores profesionales en sus partidas fuera de campeonato.

**Preguntas de investigación**

P1: ¿Cuál es la precisión de nuestro modelo predictivo para determinar si un equipo puede o no ganar una partida en una liga?

P2: ¿Cuál es la relación entre las partidas singulares y el rendimiento de un jugador en medio de una liga?

**TRABAJO REALIZADO – ESTADO DEL ARTE**

**Deep Reinforcement Learning using Capsules in Advanced Game Environments**

En esta tesis, se presenta Reinforcement Learning (RL) como un área de investigación que ha florecido enormemente en los últimos años y ha mostrado un potencial notable para los oponentes basados en inteligencia artificial en los juegos de computadora. Presenta este éxito principalmente como las vastas capacidades de las redes neuronales (ConvNet), que permiten a los algoritmos extraer información útil. Información de ambientes ruidosos. Demuestra redes como una arquitectura para la clasificación de imágenes, con un rendimiento superior para la clasificación del conjunto de datos MNIST [15].

**On using Artificial Neural Network models to predict game outcomes in Dota 2**

Esta tesis investiga si las redes neuronales artificiales se pueden usar para predecir los resultados del juego, basándose únicamente en la selección de personajes realizada en cada juego. Además, el informe explorará si la modificación de los parámetros básicos de los modelos ANN utilizados puede mejorar el rendimiento predictivo. Los modelos considerados en la tesis variarán en número de neuronas y capas ocultas en la red neuronal. Los resultados muestran que los diversos modelos tienen una precisión de predicción promedio que oscila entre 53-59%. Encuentran que usar un número bajo de neuronas con muchas capas mejora la precisión de la predicción. Si bien los resultados de este estudio parecen indicar una correlación entre la selección de personajes y los resultados del juego, recomiendan que se realice un análisis más extenso para reproducir estos resultados y así garantizar la validez externa [16]. **REVISAR 15 referencia es la 19**

**Result Prediction by Mining Replays in Dota 2**

Esta tesis, marca un contexto en los juegos en tiempo real como Dota 2 afirma que carecen del modelo matemático extenso de los juegos basados ​​en turnos que se pueden usar para hacer declaraciones objetivas sobre la mejor manera de jugarlos. Comprender un juego de computadora en tiempo real a través del mismo tipo de modelado que un juego por turnos es prácticamente imposible. Los objetivos En esta tesis se centran en crear un modelo utilizando el aprendizaje automático que puede predecir el equipo ganador de un juego de Dota 2 dado los datos parciales recopilados a medida que avanzaba el juego. Además, Se probaron un par de clasificadores diferentes, de estos bosques aleatorios se eligió para estudiarlos en mayor profundidad. Su metodología consiste en idear un método para recuperar las repeticiones de Dota 2 y analizarlas en un formato que se puede usar para entrenar modelos clasificadores. Se realizó un experimento comparando la precisión de varios algoritmos de aprendizaje automático con el algoritmo (RF) Random Forest para predecir el resultado de los juegos Dota 2. Se llevó a cabo un experimento adicional que comparó la precisión promedio de 25 modelos de bosques aleatorios con diferentes configuraciones para el número de árboles y atributos. En sus Resultados, Random Forest tuvo la mayor precisión de los diferentes algoritmos con el mejor ajuste de parámetros con un promedio de 88.83% de precisión, con un 82.23% de precisión en el punto de cinco minutos [12].

**Outcome Prediction of DOTA2 Based on Naïve Bayes Classifier**

Aunque los supuestos independientes son a menudo inexactos, de hecho, algunas de las propiedades del clasificador Naive Bayes lo hacen sorprendentemente efectivo en la práctica. El autor proporciona una forma de analizar las alineaciones y la probabilidad de ganar en el Dota2 con el clasificador Naive Bayes, presenta la idea básica de cómo analizar el juego del clasificador Naive Bayes y verifica la posibilidad de analizar el juego con datos cuantitativos en el modelo de clasificador Naive Bayes [17].

**Predicting Future States in DotA 2 using Value-split Models of Time Series Attribute Data**

En este trabajo, se introduce un enfoque para pronosticar cambios en la salud del héroe en DotA 2 al dividir los datos en cambios grandes y pequeños. Usando este enfoque de división de valores, se predice ambos tipos de cambios por separado utilizando diferentes modelos estadísticos. Para cambios pequeños, usan un modelo de media móvil autorregresiva (ARMA) y para cambios grandes usan una combinación de métodos. Los cambios grandes ("puntos de salto") se predijeron utilizando una estimación no homogénea del proceso de puntos de Poisson, mientras que la regresión logística y la regresión lineal se usaron para predecir el signo y la magnitud de estos puntos, respectivamente [18].

**Prediction of Dota 2 Game Result**

La parte teórica de esta tesis se centra en aclarar brevemente el árbol de decisiones y la teoría de redes neuronales artificiales y explica los factores básicos que tienen un impacto significativo en el resultado del juego. En la parte práctica, el enfoque se centra en experimentar con los parámetros de la técnica de aprendizaje automático, extender los datos de entrada con información sobre las composiciones de héroes, comparar y evaluar el rendimiento de estas extensiones. Todo esto da como resultado la implementación de un programa experimental que producirá un modelo ANN predictivo. Este modelo se puede usar más adelante para predecir el resultado del juego según el conocimiento de las composiciones del héroe del equipo inicial [19].

# Bibliography

|  |  |
| --- | --- |
| [1] | 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. [Accessed 21 11]. |
| [2] | V. Corp, "Dota 2 - The International," Valve, 20 08 2018. [Online]. Available: http://www.dota2.com/international/overview/. [Accessed 06 11 2018]. |
| [3] | Liquid, "Liquipedia Kuala Lumpur Major," Team Liquid, 2018. [Online]. Available: 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. [Accessed 20 11 2018]. |
| [6] | T. Liquid, "Premier tournaments," Liquid, 2018. [Online]. Available: 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. [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] | A. Joseph, N. E. Fenton and M. Neil, "Predicting football results using bayesian nets and other machine learning techniques," in *Knowledge-Based Systems*, 2006, pp. 544-553. |
| [11] | S. G and B. P, A Bayesian model for plan recognition in RTS games applied to StarCraft, 2011. |
| [12] | J. W. Filip Johansson, Result Prediction by Mining Replays in Dota 2, Karlskrona, 2015. |
| [13] | M. M, R. A and T. and A, Foundations of Machine learning, London, England: The MIT Press, 2012. |
| [14] | "Open Dota," The OpenDota team, 2018. [Online]. Available: https://opendota.com/. [Accessed 05 11 2018]. |
| [15] | P.-A. ANDERSEN, Deep Reinforcement Learning using Capsules in Advanced Game Environments, Grimstad, Norway, 2018. |
| [16] | W. VIKTOR and A. JULIEN, On using Artificial Neural Network models to predict game outcomes in Dota 2, Stockholm, Sweden, 2017. |
| [17] | K. Wang and W. Shang, Outcome Prediction of DOTA2 Based on Naive Bayes Classifier, Beijing, 2017. |
| [18] | Z. Cleghern, O. Ozaltın, S. Lahiri and D. Roberts, Predicting Future States in DotA 2 using Value-split Models of Time Series Attribute Data, North Carolina, 2017. |
| [19] | F. Beskyd, Result, Prediction of Dota 2 Game, Prague, 2018. |
| [20] | P.-A. Andersen, M. Goodwin and O.-C. Granmo, Deep RTS: A Game Environment for Deep Reinforcement Learning in Real-Time Strategy Games, Grimstad, Norway, 2018. |
| [21] | "SullyGnome - detailed Twitch statistics and analysis," SullyGnome, 5 11 2018. [Online]. Available: https://sullygnome.com/game/Dota\_2. [Accessed 5 11 2018]. |