

Dota 2 Prediction: Match Win Chance Prediction

Estela Miranda Batista¹, João Marcos Alves Modesto Ramos¹, Fabrício Aguiar Silva¹,
Gláucia Braga e Silva¹, Thais Regina de Moura Braga Silva¹

¹Instituto de Ciências Exatas e Tecnológicas
Universidade Federal de Viçosa (UFV)
Florestal - MG - Brasil

{estela.batista, joao.m.ramos, fabricio.asilva, glaucia, thais.braga}@ufv.br

Abstract. Games of the Multiplayer Online Battle Arena (MOBA) genre has stood out in recent years due to their popularity and the creation of professional championships (eSports), and in conjunction with this, studies on match analysis. The objective of this work is the analysis of data from professional matches of the game Defense of The Ancients 2 (Dota 2) to predict the chance of winning at a given time using the smallest possible number of attributes, varying in 4 different configurations of the features. As a result, the best classification model developed, using the random forest algorithm, obtained an accuracy of 76%, using as features the character's identification number and the characters' victory rate. In addition, a web service was developed for use by gamers.

Resumo. Jogos do gênero Multiplayer Online Battle Arena (MOBA) tem se destacado nos últimos anos devido a sua popularidade e criação de campeonatos profissionais (eSports), e em conjunto a isso estudos sobre análise das partidas. O objetivo deste trabalho é a análise de dados de partidas profissionais do jogo Defense of The Ancients 2 (Dota 2) para predição da chance de vitória de um determinado time utilizando do menor número de atributos possível, variando em 4 diferentes configurações dos atributos. Como resultados o melhor modelo de classificação desenvolvido, usando do algoritmo random forest, obteve uma precisão de 76%, usando como atributos o número identificador do personagem e a taxa de vitória dos personagens. Além disso, foi feito o desenvolvimento de um web service para utilização das pessoas jogadoras.

1. Introdução

A indústria dos video games se tornou nos últimos anos um dos setores mais importantes da economia, em que segundo a pesquisa anual newzoo¹ no ano de 2022 os rendimentos chegaram a mais de 174 bilhões de dólares. Dentre os gêneros de jogos, o gênero *Multiplayer Online Battle Arena* (MOBA) tem se destacado pela grande popularidade, em que foram desenvolvidos campeonatos profissionais, fazendo com que os mesmos se tornassem uma categoria de esportes, sendo chamado de esportes digitais ou *eSports*. Tais campeonatos tem disposto de altos valores de premiação, como no campeonato *The International* do jogo *Defense of The Ancients 2* (Dota 2) que teve sua premiação em 2022 de cerca de 8 milhões de dólares.

¹<https://newzoo.com/resources/blog/global-games-market-to-generate-175-8-billion-in-2021-despite-a-slight-decline-the-market-is-on-track-to-surpass-200-billion-in-2023>

Jogos MOBA envolvem ação e estratégia, sendo executado por times formados por 5 pessoas jogadoras cada. As pessoas jogadoras de cada time lutam entre si em um mapa simétrico, como apresentado na Figura 1², usando de personagens que contam com diversas habilidades e aptidões diferentes, tendo como principal objetivo avançar até a base inimiga e destruir o núcleo central, chamado de *ancient*. Levando em consideração especificamente o jogo Dota 2, desenvolvido pela Valve Software³, o mapa conta com 3 torres em cada uma das 3 rotas presentes no mapa, como apresentado na Figura 1, além de 2 barracas em cada rota. Além disso, atualmente cada um dos jogadores pode escolher entre 124 personagens disponíveis, que possuem no mínimo 2 habilidades e no máximo 13. Outro ponto importante sobre o jogo é o conceito das partidas ranqueadas, em que o jogador é avaliado com base em suas habilidades e a vitória ou derrota, em que o jogador pode ser classificado em 8 medalhas: Arauto, Guardião, Cruzado, Arconte, Lenda, Ancestral, Divino e, Imortal. Em cada uma dessas medalhas o jogador pode possuir entre 1 e 5 estrelas, além de poder movimentar entre as medalhas à medida que vai ganhando ou perdendo as partidas a cada temporada.



Figura 1. Mapa do Jogo Dota 2

Observado as principais características do jogo Dota 2 é importante ressaltar que uma das variáveis mais importantes para se definir uma vitória é a escolha dos personagens. Como os personagens possuem diversas habilidades e formas de jogar podem ser criadas diversas estratégias. Dessa forma, entender as chances de vitória que um determinado time possui logo no início da partida se torna de extrema importância para determinar as melhores estratégias a serem usadas pelas pessoas jogadoras para cada um dos personagens.

O objetivo deste trabalho é criar um modelo preditivo de vitória, no início das partidas, de um time no jogo MOBA Dota 2 utilizando do menor número de atributos possível visto a mecânica complexa presente no jogo. Para a classificação dos dados serão apresentadas 4 abordagens diferentes, que variam os atributos utilizados:

1. Identificador de cada personagem selecionado por cada um dos times;

² Adaptado de: <https://dota2.fandom.com/pt/wiki/Minimapa>

³ <https://www.valvesoftware.com/pt-br/>

2. Identificador de cada personagem selecionado por cada um dos times e a taxa de vitória do personagem escolhido;
3. Identificador de cada personagem selecionado por cada um dos times e a medalha de cada um dos jogadores;
4. Identificador de cada personagem selecionado por cada um dos times, a taxa de vitória do personagem escolhido e a medalha de cada um dos jogadores.

É importante ressaltar que o melhor modelo gerado baseado nas abordagens descritas geraram o desenvolvimento de um *web service* para que a chance de vitória em uma determinada possa ser utilizada por usuários finais: as pessoas jogadoras.

Para realizar a previsão os dados foi utilizado de dados extraídos da API Open Dota⁴ e com enriquecimento dos dados sobre a taxa de vitória de cada personagem usando da ferramenta DotaBuff⁵. Na API Open Dota são disponibilizados partidas em dois tipos: normais e profissionais, em que neste trabalho foi utilizado de partidas profissionais. As partidas profissionais foram utilizadas a fim de se obter um melhor balanceamento em relação à medalha das pessoas jogadoras. Isso porque em partidas normais as pessoas jogadoras possuem diversos níveis de conhecimento, além de haver a privação dos dados da conta da pessoa jogadora, impossibilitando saber sua medalha e o personagem selecionado na partida.

O restante deste trabalho está organizado contando com a apresentação de trabalhos relacionados na seção 2; os materiais e métodos utilizados durante a pesquisa na seção 3; os resultados obtidos na seção 4; e, por fim, as considerações finais na seção 5.

2. Trabalhos Relacionados

No que compete às pesquisas que abordam a predição de vitórias em partidas de jogos MOBA [Almeida et al. 2017] analisaram 123.326 partidas do jogo Dota 2 utilizando dos algoritmos de classificação (executados na ferramenta WEKA) *Naive Bayes*, KNN e, Árvore de Decisão, alcançando uma precisão de aproximadamente 77% levando em consideração duas abordagens: a seleção dos personagens em cada time, e além da seleção dos personagens considerando também a duração da partida. Apesar de ser uma abordagem simples, a utilização do tempo de partida muitas vezes pode não ser viável para a utilização de usuários finais, visto que eles não tem como prever o tempo que a partida em que estão terá.

[Hodge et al. 2021] analisaram partidas ao vivo e profissionais de Dota 2 utilizando de modelos padrão de aprendizado de máquina (usando do software WEKA treinaram um algoritmo *Logistic Regression* e um algoritmo *Random Forest*, além do algoritmo *Microsoft LightGBM*), engenharia de recursos e otimização, atingindo uma precisão de cerca de no máximo 75% em todos os algoritmos.

[Ani et al. 2019] analisaram partidas de outro conhecido jogo MOBA: League of Legends. Foram analisadas 1500 partidas de pessoas jogadoras profissionais, que possuíam em seu *dataset* 97 atributos, de forma que foram feitas análises e predições de vitória antes da partida e durante a partida. Na análise dos dados foi utilizado dos algoritmos Random Forest, AdaBoost, Gradient Boosting, Extreme Gradient Boosting, que

⁴<https://www.opendota.com/>

⁵<https://pt.dotabuff.com>

atingiram antes da partida uma precisão de 95%, durante a partida uma precisão de 98% e ao combinar ambas as abordagens uma precisão de 99.75%.

Por fim, considerando que jogos MOBA possuem uma mecânica muito complexa [Tyran and Chomatek 2021] analisaram a presença de partidas atípicas em bases de dados, ou seja, partidas em que um determinado time perde embora tenha superado a equipe inimiga durante toda a partida. Um ponto importante tratado no trabalho é que tal situação ocorre em sua maioria em partidas em que os jogadores possuem diversos níveis de conhecimento, como em partidas no modo Turbo⁶, dessa forma, como citado anteriormente foi feita a escolha de apenas partidas profissionais para um melhor treinamento dos dados.

Apesar de possuírem trabalhos na literatura que abordem modelos de predição de vitórias em partidas de jogo MOBA similares ao apresentado neste trabalho é importante ressaltar que os mesmos necessitam de uma quantidade maior de atributos, ou de transformações custosas computacionalmente para realizar suas predições. Dessa forma, buscando investigar formas mais otimizadas, e mais fáceis de serem acessadas por pessoas jogadoras este trabalho irá apresentar um modelo de predição que utiliza apenas dos personagens selecionadas e caso esteja disponível, a medalha de cada pessoa jogadora. Em que o principal diferencial apresentado é a utilização de características das pessoas jogadoras que não variam de acordo com o tempo de partida, ou seja, sua medalha.

3. Materiais e Métodos

A metodologia adotada neste trabalho levou em consideração as etapas apresentadas na Figura 2 que serão detalhadas nas subseções a seguir. É importante ressaltar que a análise dos dados desenvolvida utilizou da linguagem Python na ferramenta Jupyter Notebook⁷.



Figura 2. Etapas do Processo de Análise dos Dados

3.1. Coleta dos Dados

Durante a etapa de coleta dos dados, como citado anteriormente, na seção 1, foi utilizado da API Open Dota, em que inicialmente foi feita uma extração para salvar os valores de

⁶Neste modo de jogo é como se o tempo de jogo andasse de uma forma mais rápida, em que o *gold* por minuto é aumentado, tempo de dia/noite mais curtos, entre outras características.

⁷<https://jupyter.org/>

número identificador das partidas profissionais. De posse do valor identificador de cada partida foi realizado a extração dos dados de cada partida, em que além de informar os personagens selecionados também é possível verificar as medalhas em que cada jogador se encontra no momento da extração, dessa forma salvando essas informações em um arquivo CSV.

Na última etapa da coleta de dados visto que a API Open Dota realiza a extração apenas das últimas 100 partidas de cada personagem, podendo não indicar assim a taxa de vitórias real de um determinado personagem na temporada foi utilizado da ferramenta DotaBuff para enriquecer os dados. A ferramenta realiza o cálculo da taxa de vitórias de um determinado personagem de acordo com as medalhas das pessoas jogadoras, visto que personagens mais complexos podem se sair melhor em medalhas mais altas, na Figura 3 é apresentada essa diferenciação. Os dados da taxa de vitória de cada personagem consideraram então as duas últimas medalhas (Divino e Imortal), visto que a maioria das pessoas jogadoras profissionais se encontram em tais medalhas. É importante ressaltar que os dados extraídos não utilizaram de *scripts* automáticos para a extração, sendo salva tal informação manualmente em um dicionário, que unido ao nome e número identificador de cada personagem foram adicionados à base de dados de acordo com os personagens selecionados em cada partida extraída. Ao final da extração dos dados foram extraídas informações de 1387 partidas profissionais de Dota 2.

Herói	Pick %	Win %	Pick %	Win %	Pick %	Win %	Pick %	Win %	Pick %	Win %
Pudge	30.03%	49.88%	29.69%	49.39%	29.92%	49.07%	29.98%	49.26%	23.39%	49.07%
Witch Doctor	22.57%	50.51%	19.13%	49.11%	14.71%	47.69%	11.00%	46.97%	6.24%	46.80%
Legion Commander	21.85%	54.77%	22.08%	55.15%	22.05%	54.95%	22.40%	54.55%	22.48%	54.23%
Axe	19.92%	54.79%	19.13%	53.39%	17.61%	52.66%	16.51%	51.67%	14.19%	50.32%
Slark	19.40%	50.95%	19.47%	50.56%	19.21%	50.70%	19.72%	51.53%	20.73%	52.45%
Crystal Maiden	19.00%	53.21%	17.89%	52.44%	17.72%	51.67%	18.15%	51.33%	19.43%	51.32%

Figura 3. Exemplo de Classificação de Taxa de Vitória de Personagens na Ferramenta Dota Buff

Ao final, a base de dados, que pode variar seus atributos de acordo com cada abordagem, obteve um total de 32 atributos, contendo o número identificador de cada partida, o identificador do time vencedor, podendo assumir os valores *dire* ou *radiant*, um conjunto de características dos personagens, e um conjunto de características das pessoas jogadora. Sendo P o conjunto de personagens, em que cada personagem $p_i \in P$, tal que $i = 1, 2, \dots, 10$, é representado pelas características < número identificador, taxa de vitória >. Além disso, sendo J o conjunto de pessoas jogadoras, em que cada pessoas jogadora $j_i \in J$, tal que $i = 1, 2, \dots, 10$, é representado pelo número identificador da medalha ao qual ela possui.

3.2. Preparação dos Dados

Primeiramente foi realizada uma análise exploratório dos dados coletados para verificar a necessidade de transformações e pré-processamentos nos dados. Foi observado que não haveria a necessidade de realizar o balanceamento dos dados visto que as classes de times dados como vencedores da partida possuem quantidade similares, sendo 676 vitórias de times no lado *radiant* do mapa, e 711 vitórias de times no lado *dire*. Foi verificado também que não haveria a necessidade de transformações e nem remoção de dados em relação ao

número identificador dos personagens selecionados e da medalha das pessoas jogadoras. Em relação ao número identificador dos personagens selecionados, como apresentado na Figura 4(a), sendo representado pelos primeiro personagem selecionado de cada time, eles possuem uma variação entre 1 e 137, e apesar de possuírem 124 personagens alguns valores de número identificador não foram utilizados para identificar os personagens. Em relação as medalhas de cada pessoa jogadora, como apresentado na Figura 4(b), pode-se verificar que a maioria assume valores entre 70 e 80, que como apresentado anteriormente a maioria das pessoas jogadoras profissionais possuem medalhas altas. Porém, outras medalhas, de valores menores foram registradas, em que foi verificado que em partidas de ligas menores profissionais os times realizam a inserção para treinamento de pessoas jogadoras iniciantes, ou que não realizaram o balanceamento da medalha na temporada. Verificou-se então que não era necessário transformações e pre-processamento dos dados durante a criação dos modelos na execução da seleção dos algoritmos.

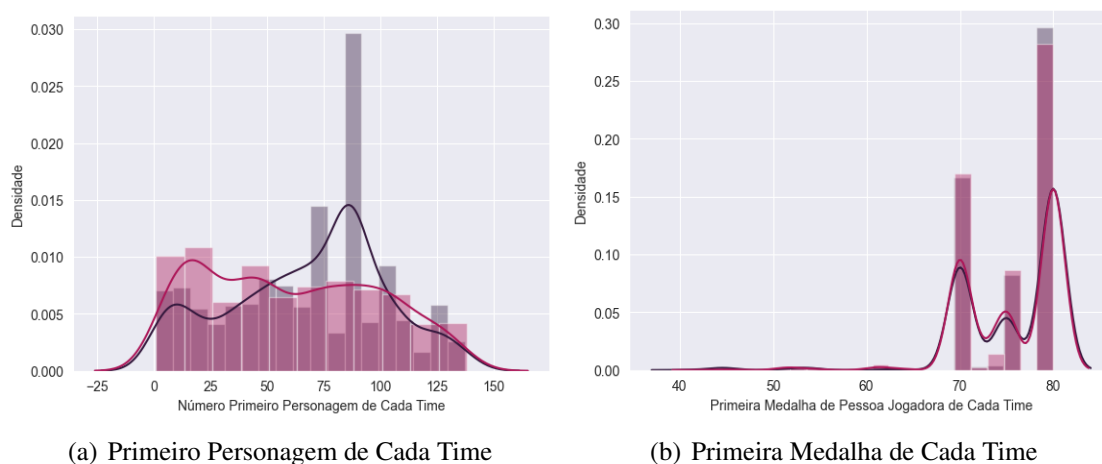


Figura 4. Distribuições dos Dados

Porém, deve-se ressaltar que ao receber dados de um usuário no *web service* desenvolvido, que será apresentado posteriormente, é realizada uma transformação e adição nos dados, de forma que o nome de cada personagem digitado é transformado em seu número identificador, e adicionado ao vetor de predição juntamente com a sua respectiva taxa de vitória. Para realizar a conversão do nome em um número identificador, e verificar a taxa de vitória do mesmo os dados foram registrados em um dicionário, de forma a realizar a conversão em uma forma mais otimizada computacionalmente.

3.3. Escolha do Modelo

Para a escolha do modelo foram selecionados previamente alguns algoritmos apresentados na literatura para a classificação de resultados em partidas de jogos MOBA. Os algoritmos selecionados foram *Logistic Regression*, *Random Forest*, *Decision Tree*, *K-Nearest Neighbors* (KNN) e *Naive Bayes*.

O algoritmo *Logistic Regression* é baseado no recurso de estimar a probabilidade associada à ocorrência de um determinado evento baseado em um conjunto de variáveis exploratórias, ou seja, ele estima a probabilidade de cada atributo ocorrer em uma determinada classe de classificação [Cramer 2003]. O algoritmo *Decision Tree* é baseada na

divisão dos atributos em grupos suficientemente distintos entre si, de forma que ao gerar a árvore os nós filhos atribuídos devem possuir características diferentes [Quinlan 1996], como por exemplo uma sequência 0111, em que os dois primeiros valores são na cor rosa, os dois últimos na cor azul, os rosas passariam por uma segunda divisão, enquanto os azuis apenas verificariam a cor por serem iguais. O algoritmo *Random Forest* por sua vez utiliza de árvores de decisão em sua construção, de forma que são geradas diversas árvores de decisão, que possuem uma baixa correlação entre si, e durante a classificação aquelas que melhor se adequem para os dados de entrada são usadas [Ho 1995]. O algoritmo *K-Nearest Neighbors* (KNN) utiliza de cálculos de distância (euclediana, manhattan, etc.) para verificar a distância de um determinado dado em relação aos demais, de forma que ele utiliza do valor dos atributos para determinar essa distância [Cover and Hart 1967], ou seja, valores de atributos similares tem uma distância menor. Por fim, o algoritmo *Naive Bayes* é um modelo probabilístico que utiliza como base o teorema de Bayes, que determina a probabilidade de um determinado evento A ocorrer em relação a um evento B que ocorreu, fazendo com que os atributos sejam independentes para se obter melhores resultados [Webb 2010].

Após a escolha dos algoritmos, a base de dados foi separada em 80% para treino, e 20% para teste, usado posteriormente para a comparação com os resultados apresentados na literatura. Com os 80% de dados para treino foi realizada uma validação cruzada com 10 partes, em que a mesma foi repetida 30 vezes para diminuir o impacto da aleatoriedade na separação das bases. Para determinar o melhor algoritmo para cada uma das abordagens foi escolhido o uso da acurácia, porque neste problema ambas as classes são igualmente importantes.

Unido a validação cruzada foi realizada o ajuste dos parâmetros de cada um dos algoritmos a fim de se determinar o melhor resultado possível para cada um dos algoritmos. O algoritmo *Logistic Regression* teve variações nos parâmetros de penalidade (*penalty* = 11, 12 e *elasticnet*), e tolerância (*tol* = 0.0001, 0.001, 0.01, 0.0004 e 0.004). O algoritmo *Decision Tree* teve como variação os parâmetros o critério (*criterion* = *entropy* e *gini*), profundidade máxima da árvore (*max_depth* = None, 50 e 100), divisão mínima dos nós (*min_samples_split* = 2, 4, 6), e o número máximo de atributos por nó da árvore (*max_features* = auto, log2 e sqrt). O algoritmo *Random Forest* teve variações nos parâmetros similar ao algoritmo *Decision Tree* com o acréscimo dos parâmetros o número de estimadores (*n_estimators* = 50, 100 e 200), e como forma de construção da árvore (*bootstrap* = True, para usar todo o conjunto de dados e False caso contrário). O algoritmo KNN por sua vez teve como variação os parâmetros o número de vizinhos (*n_neighbors* = 5, 10, 15 e 20), o algoritmo de construção (*algorithm* = auto, brute, kd_tree, ball_tree), e a métrica de definição da distância dos vizinhos (*metric* = cosine, euclidean, manhattan, minkowski). Por fim, temos o algoritmo *Naive Bayes* que teve variações nos parâmetros de probabilidades (*priors* = None e $0.1 \times \text{número de atributos de cada abordagem}$), e a maior variância que deve ser adicionada às demais variâncias (*var_smoothing* = $1e-9$, $1e-6$ e $1e-12$).

Na Tabela 1 pode ser verificado os resultados da média de acurácia de cada um dos algoritmos apresentados em cada uma das abordagens após a validação cruzada juntamente com o ajuste de parâmetros, onde foi demarcado em negrito o algoritmo escolhido.

Em ambas as quatro abordagens apresentadas o algoritmo *Random Forest* foi es-

Abordagem	Logistic Regression	Decision Tree	Random Forest	KNN	Naive Bayes
(1) Personagem Selecionado	54.6%	57.8%	60.5%	57.8%	53.8%
(2) Personagem Selecionado e Taxa de Vitória	55.7%	62.5%	74.2%	56.1%	54.9%
(3) Personagem Selecionado e Medalha da Pessoa Jogadora	53.7%	58.5%	61.2%	56.1%	51.3%
(4) Personagem Selecionado, Taxa de Vitória e Medalha da Pessoa Jogadora	54.9%	63.5%	71.8%	55.6%	52.8%

Tabela 1. Comparação de Algoritmos por Abordagem na Validação Cruzada

colhido pois, além de possuir uma maior média de acurácia, como apresentadas nas Figura 5(a), 5(b), 6(a), 6(b), que apresentam o *boxplot* gerado a partir da validação cruzada em cada abordagem, o algoritmo também apresentou uma menor variação nos valores de sua acurácia durante as execuções. É importante ressaltar que os modelos das abordagens (2) e (4) foram escolhidos para serem implementados no *web service* desenvolvido, que será apresentado posteriormente, por possuírem os melhores valores de acurácia.

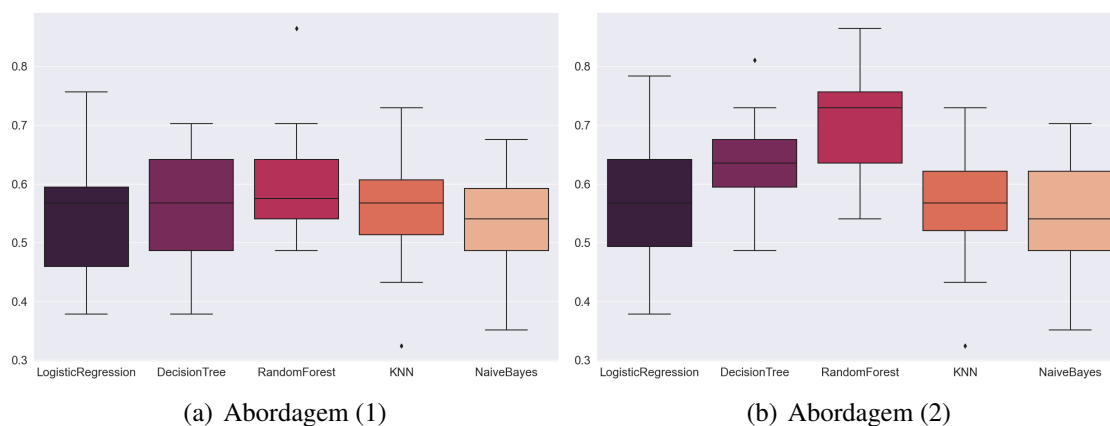


Figura 5. Variação de Acurácia na Validação Cruzada

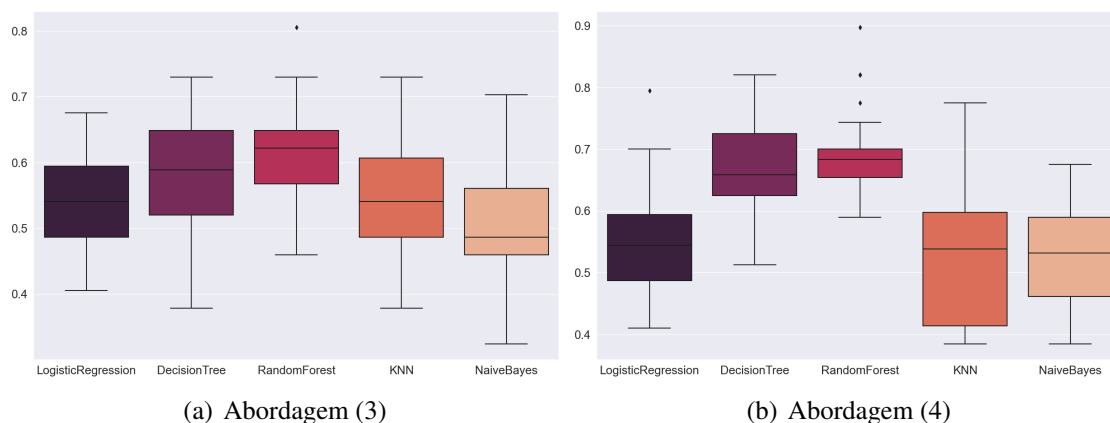


Figura 6. Variação de Acurácia na Validação Cruzada

3.4. Implantação

Após a escolha dos modelos utilizados foi realizada a implantação dos modelos um *web service* para que possam ser realizadas a entrada de personagens selecionados, e caso

haja acesso a informação, da medalha de cada pessoa jogadora para assim apresentar a taxa de vitória e o time predito como vencedor da partida. Como citado anteriormente, a informação sobre a taxa de vitória de cada personagem é adicionada no vetor de predição a partir dos dados coletados anteriormente.

Como todos os algoritmos, e testes apresentados anteriormente foram realizados utilizando da linguagem Python foi optado por utilizar do framework Flask⁸ para realizar a implantação do *back-end* do *web service*. Nesta aplicação foi então configurado a criação de duas páginas: a página inicial onde podem ser digitados os valores, e a apresentação do resultado, além de fazer a chamada da atualização dos dados, ou seja, inserir novos dados a partir de chamadas de requisição à API Open Dota. A segunda página é utilizada quando a atualização dos dados é chamada, informando que os novos dados foram inseridos.

Para a interface, ou seja, o *front-end* do *web service* foi construído utilizando de HTML e CSS, sendo os mesmos usados em sua forma pura, sem a necessidade da utilização de frameworks. E além disso, visando um melhor desenvolvimento da interface, todos os modelos criados foram inicialmente modelos por meio da ferramenta Figma⁹, e posteriormente implementados.

A apresentação dos resultados foi feita utilizando de uma interface *web service* para facilitar o acesso as jogadores leigos, que não possuem um conhecimento prévio sobre as técnicas de desenvolvimento de software. Além disso, devido a facilidade de acesso pelo *web service* cada um dos times pode nos minutos iniciais decidir possíveis estratégias para serem utilizadas durante a partida para reverter, ou confirmar a predição dada. Na Figura 7 é apresentado a principal interface desenvolvida, onde foi realizada a chamada da predição de uma partida. Na Figura 8 é apresentada a interface ao usuário após realizado a chamada de atualização dos dados, para que assim o usuário possa verificar as mudanças feitas na página inicial.

Olá! Bem-Vinde ao
Dota 2 Predictor

Essa plataforma foi desenvolvida para apresentar a chance de vitória de um time em partidas de Dota 2 utilizando de poucos atributos: os personagens selecionados por cada jogador de cada time, ou além disso, caso esteja disponível a medalha de cada jogador, sendo 0 Arauto, 2 Arauto com 2 estrelas, e assim por diante.

Os dados foram treinados com partidas profissionais e usam como atributo extra a taxa de vitória dos personagens selecionados, sendo essa informação adicionada pela própria plataforma na hora da predição.

Precisão do Modelo sem Medalha: 70%
Precisão do Modelo com Medalha: 69%
Número de Partidas na Base de Dados: 1387
[Inserir Novas Partidas!](#)

personagens time radiant
personagens time dire
medalhas time radiant
medalhas time dire
Send

51% de chance do time
DIRE vencer!

Made with ♥ by Estela

Figura 7. Interface da Página Inicial do Web Service Desenvolvido

⁸<https://flask.palletsprojects.com/en/2.3.x/>

⁹<https://www.figma.com/>



Figura 8. Interface da Página de Atualização dos Dados Web Service Desenvolvido

4. Resultados

Nesta seção será apresentado o resultado de acurácia obtido para cada abordagem por cada um dos algoritmos em relação à validação utilizando da base de teste, apresentada anteriormente na seção 3, comparando os resultados com abordagens apresentadas pela literatura e com ajustes de parâmetros realizados de forma similar ao apresentado para as abordagens apresentadas.

4.1. Comparação de Resultados com a Literatura

Após o ajuste de parâmetros dos algoritmos em cada uma das abordagens, como citado anteriormente foram validados utilizando da base de 20% de teste, sendo os resultados apresentados na Tabela 2. Além disso, foi inserido na tabela a mesma validação utilizando da base de 20% para os algoritmos nas abordagens apresentadas no trabalho de [Almeida et al. 2017], de forma a se utilizar da base de dados apresentada neste trabalho. É importante ressaltar que para realizar a validação das abordagens do trabalho de [Almeida et al. 2017] foi realizado o mesmo processo descrito na seção 3 para as abordagens descritas nesse trabalho, realizando o ajuste de parâmetros para cada algoritmo.

Abordagem	Logistic Regression	Decision Tree	Random Forest	KNN	Naive Bayes
(1) Personagem Selecionado	47.9%	58.3%	55.8%	53.9%	47.5%
(2) Personagem Selecionado e Taxa de Vitória	50.7%	71.6%	72.7%	54%	53.6%
(3) Personagem Selecionado e Medalha da Pessoa Jogadora	47.9%	57.2%	53.2%	50.7%	46.4%
(4) Personagem Selecionado, Taxa de Vitória e Medalha da Pessoa Jogadora	52.7%	63.2%	71.6%	50.7%	50.7%
(5) Vetor Binário de Personagem Selecionado [Almeida et al. 2017]	49%	49.7%	50.3%	50.4%	54.5%
(6) Vetor Binário de Personagem Selecionado e Tempo de Partida [Almeida et al. 2017]	50.7%	53.6%	51.8%	54%	55.2%

Tabela 2. Comparação de Acurácia entre Abordagens e a Literatura

Observando-se então a Tabela 2 pode-se verificar que as abordagens apresentadas neste trabalho que utilizaram de uma abordagem similar as de [Almeida et al. 2017] obtiveram valores similares acurácia, como no caso da abordagem (1), em que no caso da abordagem apresentada neste trabalho foi utilizado de uma menor quantidade de atributos e uma conversão mais otimizada computacionalmente. Apesar da abordagem (3) possuir uma menor acurácia em relação ao apresentado no trabalho de [Almeida et al. 2017] as abordagens (2) e (4) obtiveram melhores resultados, tanto em relação ao trabalho quanto as demais abordagens apresentadas neste trabalho.

4.2. Importância de Atributos

Ao final da execução de cada um dos algoritmos das abordagens foi gerado os atributos mais importantes, trazendo assim explicabilidade dos dados. Na abordagem (1), como podemos observar na Figura 9(a) todos os atributos possuíram importância similar, visto que como se utilizou apenas do número identificador dos personagens selecionados todos devem ser analisados em conjunto para determinar o time vencedor da partida. Na abordagem (2) como pode se observar na Figura 9(b) mesmo com alguns atributos se sobressaindo a outros, todos os atributos possuem valores similares de importância, e como foi utilizado apenas do número identificador de cada personagem e da taxa de vitória dos mesmos essas são informações relevantes para determinar a vitória de um time, porque aliado ao apresentado na abordagem (1) a informação da taxa de vitória se torna relevante visto que personagens com altas taxas de vitória podem indicar habilidades mais poderosas usadas pelos jogadores que se tornam facilitadores.

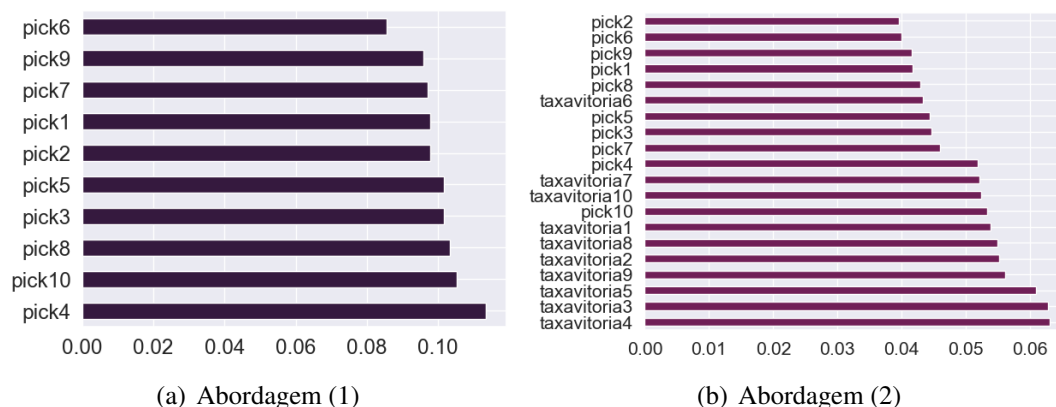


Figura 9. Atributos Importantes

Na abordagem (3) como pode ser observado na Figura 10(a) os atributos que dizem respeito a medalha de cada uma das pessoas jogadoras possuíram uma baixa importância, e ao se analisar os dados é possível verificar que isso possa ter ocorrido neste modelo em específico pelo fator de que os dados estão relacionados a pessoas jogadoras profissionais, e como citado anteriormente possuindo como valor atribuído a sua medalha valores entre 70 e 80. Por fim, na abordagem (4) foi possível observar uma relação com a abordagem (2), em que a informação da taxa de vitória e da escolha do personagem possuem importância similar, enquanto como apresentado na abordagem (3) a informação sobre a medalha da pessoa jogadora não se tornou tão relevante, porém ao olharmos para a precisão gerada é possível que tenha contribuído para um pior resultado no modelo, sendo isso apresentado na Figura 10(b).

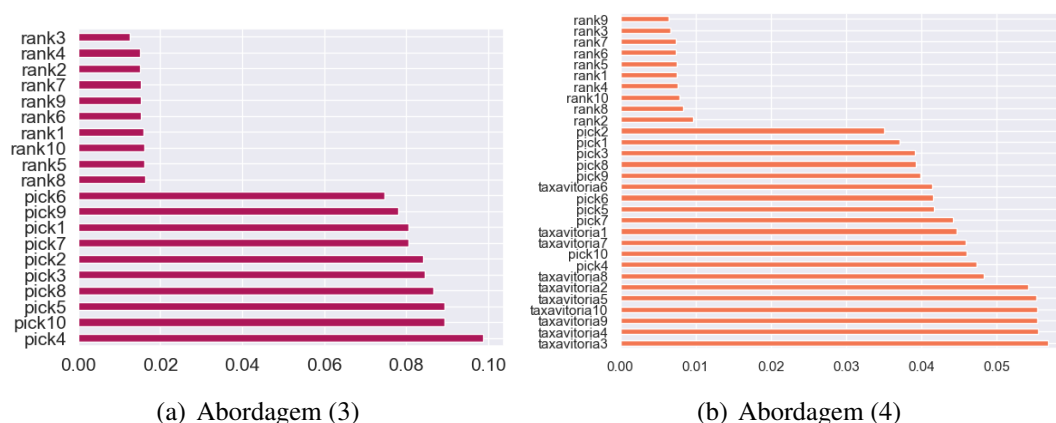


Figura 10. Atributos Importantes

5. Considerações Finais

Com base nas análises realizadas durante o trabalho foi possível verificar uma acurácia máxima de 73% durante a predição do time vencedor em uma partida do jogo MOBA Dota 2, onde foi utilizado apenas como atributos os personagens selecionados e a taxa de vitória dos personagens. Ao utilizar como atributo a medalha que uma determinada pessoa jogadora possui foi possível observar que para o caso de pessoas jogadoras profissionais, abordado neste trabalho, onde o valor das mesmas não possui uma alta variação, este atributo se tornou não muito relevante, fazendo com que até mesmo o valor de acurácia em relação à uma abordagem similar, entre as abordagens (2) e (4), diminuísse.

Como trabalhos futuros sugere-se a ampliação da base de dados para se verificar o comportamento da acurácia dos modelos em base de dados maiores. Outro trabalho futuro importante é a aplicação dos modelos em base de dados que possuam partidas normais, de forma a se variar as medalhas que as pessoas jogadoras possuam e entender se tal atributo continua a não possuir uma importância significativa para o modelo.

Referências

- Almeida, C. E. M., Correia, R. C. M., Eler, D. M., Olivete-Jr, C., Garci, R. E., Scabora, L. C., and Spadon, G. (2017). Prediction of winners in moba games. In *2017 12th Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–6.
- Ani, R., Harikumar, V., Devan, A. K., and Deepa, O. (2019). Victory prediction in league of legends using feature selection and ensemble methods. In *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, pages 74–77.
- Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27.
- Cramer, J. (2003). The Origins of Logistic Regression. *SSRN Electronic Journal*.
- Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, volume 1, pages 278–282 vol.1.
- Hodge, V. J., Devlin, S., Sephton, N., Block, F., Cowling, P. I., and Drachen, A. (2021). Win prediction in multiplayer esports: Live professional match prediction. *IEEE Transactions on Games*, 13(4):368–379.

- Quinlan, J. R. (1996). Learning decision tree classifiers. *ACM Computing Surveys*, 28(1):71–72.
- Tyran, J. and Chomatek, L. (2021). Influence of outliers in moba games winner prediction. *Procedia Computer Science*, 192:1973–1981. Knowledge-Based and Intelligent Information Engineering Systems: Proceedings of the 25th International Conference KES2021.
- Webb, G. I. (2010). *Naïve Bayes*, pages 713–714. Springer US, Boston, MA.