

Aplicação de algoritmos classificadores para previsão de vitória em uma partida de *League of Legends*

Rafael Todeschini de Souza – UFRGS – Engenharia de Produção

Prof. Dr. Marcelo Nogueira Cortimiglia – UFRGS – Engenharia de Produção

Resumo: Jogos de computador são grandes negócios, o que reflete no interesse crescente em jogos competitivos, os chamados eSports. Dentre esses jogos competitivos, o LoL é um dos mais bem-sucedidos, tanto em termos comerciais quanto de público. Um assunto recorrente sobre o jogo é a imprevisibilidade dos rumos das partidas, levando ao surgimento de diversas pesquisas que buscam entender melhor o jogo e fornecer conhecimento científico que possa auxiliar as equipes nas tomadas de decisão. Neste artigo se investiga a aplicação dos algoritmos classificadores *Random Forest* e Regressão Logística de forma a gerar um modelo inteligente de previsão de vitória para uma partida do LoL. Os resultados apontam uma precisão média de 75% nas previsões, com o indicador do Ouro como sendo o mais confiável na geração dos modelos das previsões.

Palavras-chave: *League of Legends* (LoL); *machine learning*; previsão de vitória; algoritmos classificadores.

1. INTRODUÇÃO

Ao longo das últimas décadas, os jogos eletrônicos surgiram como uma das principais fontes de receita da indústria de entretenimento mundial. No cenário atual, a indústria dos *games* já movimenta mais dinheiro que a indústria de cinema, por exemplo (FERREIRA, 2015). Este mercado está em plena expansão, podendo atingir até 2017 o patamar de U\$ 100 bilhões ao ano (SINCLAIR, 2014), puxado pelo crescimento do negócio de dispositivos móveis. Projeção parecida foi feita pela Newzoo, a qual estimou um mercado global de U\$ 108,9 bilhões em 2017 envolvendo cerca de 2,2 bilhões de jogadores (NEWZOO, 2017). Quanto ao Brasil, estimou-se que serão gastos U\$ 1,3 bilhões por 66,3 milhões de jogadores brasileiros em 2017 (NEWZOO, 2017).

Essa expansão de mercado e público interessado em jogos eletrônicos, aliado ao fenômeno das mídias sociais fez com que surgissem competições organizadas e transmitidas para este nicho. Dessa forma, refere-se a *eSports* em analogia aos esportes tradicionais, com

seu aspecto competitivo, seus fãs, entusiastas e atletas. Outrora restrito a países asiáticos, como a Coreia do Sul, que já teve no *StarCraft* seu esporte nacional não-oficial (RAY, 2011), hoje podemos encontrar transmissões frequentes de *games* em canais esportivos da TV fechada no Brasil (SporTV, Esporte Interativo) despertando interesse principalmente na parcela mais jovem da população. Neste contexto, o *e-game* foco deste artigo é o *League of Legends*, um MOBA (*Multiplayer Online Battle Arena* ou Arena de Batalha de Multijogadores) desenvolvido e publicado pela empresa Riot Games. No momento, este jogo é disponibilizado no formato *free-to-play* ou seja, gratuito. O jogo possui três mapas para serem jogados e cada mapa possui suas características, são eles: *Summoner's Rift*, *Twisted Treeline* e *Howling Abyss*. A partida consiste em uma disputa entre duas equipes de cinco personagens. Cada equipe começa perto do *Nexus* a base que deve ser protegida. Sua destruição é o objetivo primário do jogo, quem a obtiver automaticamente ganha a vitória. Cinco personagens são escolhidos dentre os 131 possíveis e são chamados de "campeões". O objetivo dos jogadores é avançar nas linhas e destruir o *Nexus* oposto. Para cumprir este objetivo as equipes devem destruir pelo menos cinco das onze torres inimigas e um dos três inibidores. O jogo é muito dinâmico, ocorrendo a intercalação das equipes comandando os ataques ao longo da partida, o que torna comum reviravoltas e dificulta aos times projetar chances de vitória. Para uma descrição mais detalhada do jogo pode se consultar Ferrari (2013).

Entendendo-se o funcionamento básico do jogo, parte-se para as suas relações complexas, as quais ainda não são bem exploradas e a academia tem voltado seus estudos de forma a entendê-las melhor, como as relações que impactam na vitória de uma equipe (BARCELLOS, 2017; ONG, DEOLALIKAR e PENG, 2015; KINKADE e LIM, 2015). Dessa forma, deseja-se explorar métodos que possam projetar as chances de vitória de uma equipe ao longo da partida. É importante ressaltar que a identificação de fatores relevantes para a vitória em uma partida de LoL possui valor no sentido de auxiliar as equipes nas composições dos times, adaptando a estratégia conforme as demandas de resultados aparecerem, adaptando o que se deseja obter na partida (vantagem nas diversas fases *early*, *mid*, ou *late game* do jogo). Egliston (2016) resalta em seu estudo o potencial que técnicas de *Data Mining* têm em extrair informações de grandes bancos de dados para jogos de multijogadores.

Assim, este trabalho tem como objetivo determinar indicadores de previsão do resultado final de uma partida de alto nível de LoL, baseado em poucos indicadores obtidos a partir de um modelo classificador. Com base nessa análise, serão identificados os fatores que mais influenciam na formação do modelo de predição, e conseqüentemente, na vitória, assim apontando indicadores que possam ser mensuráveis ao longo da partida, e que sejam impactantes no resultado de vitória para a equipe.

O artigo possui a seguinte estrutura: a primeira seção contextualiza o tema, esclarece o problema e expõe os objetivos, a segunda seção apresentará o referencial teórico sobre aplicações de priorização e otimização em eSports, principalmente em MOBAs; a terceira seção aborda a metodologia utilizada na aplicação da mineração de dados, enumerando as etapas que envolveram a coleta e manipulação dos dados; a quarta consistirá na análise e detalhamento dos resultados obtidos; e, por fim, a quinta seção exibirá as conclusões obtidas a partir do desenvolvimento do estudo, bem como possibilidades de novos estudos e as limitações do estudo atual.

2. REFERENCIAL TEÓRICO

A literatura referente a eSports ainda não é tão vasta por se tratar de um tema relativamente novo, ainda que esteja crescendo nos últimos anos (EGGERT et al., 2015; ONG, DEOLALIKAR e PENG, 2015). Pesquisadores ao redor do mundo estão se conscientizando sobre o tema e seu potencial de estudo (EGLISTON, 2016), de forma a contribuir com pesquisas para a academia. Esta seção é destinada ao levantamento da literatura sobre trabalhos já realizados sobre aplicações de otimização e priorização em eSports e será estruturada em três subseções.

Primeiramente, são abordadas a contextualização e aspectos do jogo competitivo em eSports, particularmente do LoL, de forma a justificar a importância do estudo. Em um segundo momento, a revisão aprofunda-se nas publicações que estudam fatores que influenciem no desempenho em equipe nos MOBAs. Por fim, no terceiro tópico, são abordadas aplicações de otimização e priorização de forma a entender melhor aspectos implícitos do jogo.

2.1 Aspectos do jogo competitivo em eSports

Em um dos primeiros estudos realizados com eSports no aspecto competitivo, Rambusch, Jakobsson e Pargman (2007) exploraram um estudo de caso feito sobre o jogo eletrônico de tiro em primeira pessoa *Counter-Strike*. Tratando-se o jogo como um eSport, por apresentar jogabilidade competitiva que recorda esportes tradicionais, foram realizadas entrevistas semiestruturadas com 34 membros de clã entre 19 a 25 anos e obtidas conclusões referentes às ações dos jogadores durante o jogo e interações dentro e entre as equipes. Sobre o primeiro aspecto foi ressaltada a necessidade do jogador apresentar habilidades diretamente relacionados com a destreza mecânica, como reflexos rápidos, destreza manual e excelente coordenação entre olhos e mãos para precisa execução dos movimentos, de forma a poder evoluir de um nível amador para o cenário competitivo. No segundo aspecto foi destacada a importância do entendimento do contexto do jogo pelos jogadores, de maneira que, pelo aprendizado, principalmente devido ao contato com jogadores mais experientes, eles possam

desenvolver corretamente o processo de tomada de decisões e desempenhar seus papéis de acordo com o ambiente cultural do jogo.

Em meados da década de 2010 (GAUDIOSI, 2012), o destaque no cenário competitivo foi gradualmente mudando para os MOBAs. Em um estudo aplicado em *League of Legends*, Donaldson (2015) apresentou dois tipos de perícia de jogo: o *meta* e o *mechanics game*. O *mechanics* se trata da habilidade mecânica e motora do jogador, bem como sua interação com os elementos dentro do jogo tais como navegação pela interface e controle e movimentação do personagem. O *meta*, por sua vez, se trata da consciência e capacidade de entender o jogo e seus aspectos táticos implícitos, tais como a formulação de novas estratégias após modificações em habilidades de certos personagens ou o uso de técnicas matemáticas para determinar a eficácia de certo elemento de jogo ou encontrar a combinação ótima de habilidades entre os jogadores. No estudo, o autor destacou a importância do aspecto *metagame*, realçando que um jogador de similar nível *mechanic* mas com superior entendimento do *meta* levará vantagem ao lidar com um jogador de entendimento inferior. Assim, foi realçada a importância do estudo do aspecto *meta* no cenário de jogo competitivo, especialmente quando se atinge a níveis mais altos do jogo, como o profissional. Finalmente, o autor indicou diversos assuntos relevantes ao *metagame* que renderiam estudos e discussões interessantes; entre eles um estudo entre a relação do *metagame* e a comunicação dentre os jogadores, pois as escolhas de campeões e táticas a serem seguidas nas situações em que os jogadores tem a oportunidade de se comunicar por voz com os colegas de time são diferentes dos cenários em que apenas a comunicação por texto e por *pings*, sinais que são fáceis de ativar e provêm avisos audiovisuais para os colegas de equipe (LEAVITT, KEEGAN E CLARK, 2016), são disponíveis.

2.2 Fatores que impactam no desempenho de equipes em MOBAs

Em seu estudo Kim et al. (2015) trataram sobre o processo de composição de equipe nos MOBAs o qual se dá de forma a equalizar um dilema de proficiência-congruência, entre selecionar os papéis que melhor representam seu estilo de jogo com aqueles que melhor complementam os papéis já existentes na equipe. O estudo foi feito utilizando-se de dados observacionais de 1.9 milhão de partidas do *League of Legends* por meio do uso de métodos quantitativos, como o uso de algoritmos classificadores tal como o de regressão logística, aliado ao uso de técnicas qualitativas, como as entrevistas com jogadores para corroborar com as descobertas. Dessa forma, foram provadas diversas hipóteses, como a de que o desempenho da equipe é significativa e poderosamente influenciado pela semelhança entre o papel do jogador e aquele com o qual ele possui maior familiaridade. Adicionalmente, o grau em que a equipe é composta de papéis que se complementam apresentou contribuição significativa, porém moderada no efeito direto de desempenho da equipe (Tabela 1).

Variáveis	Precisão
Todas variáveis	58,4%
Individuais (proficiência)	57,0%
Equipe (congruência)	52,6%

Tabela 1: Precisão dos algoritmos classificadores utilizados para previsão de vitória conforme o tipo de variável.

Fonte: Adaptado de Kim et al. (2015)

Por fim, foi concluído que usuários especialistas são mais capazes de negociar o dilema de proficiência-congruência do que usuários novatos. Eles buscam maximizar este conjunto tanto quanto possível, ainda que sacrificando, às vezes, sua proficiência individual pela congruência da equipe.

Leavitt, Keegan e Clark (2016) apresentaram um estudo que propõe entender a relação entre comunicação-não-verbal, os chamados *pings*, e o desempenho dos jogadores da equipe no LoL. Foi demonstrado que a quantidade de *pings* depende diretamente do papel do jogador e de seu nível de atividade em jogo. Adicionalmente, demonstrou-se que os *pings* detêm uma relação positiva mas côncava com o desempenho do jogador. Ações diretas de ataque como abates e assistências apresentaram uma relação positiva com o uso de *pings*. Sobre a relação com o número de mortes, se esperava uma correlação negativa pela hipótese de que os *pings* são utilizados também defensivamente, como na forma de avisos dados para evitar ser morto por inimigos. Entretanto, pelos resultados do estudo, as mortes apresentaram uma relação positiva com os *pings*, com os autores explicando este ser um resultado que sofreria pelo efeito secundário de outras variáveis impactando no número de mortes, mas ressaltando a necessidade de mais pesquisa para desmembrar essa possível relação e comprovar os fatos. Por fim, o estudo mostrou que os *pings* apresentaram correlação negativa com o Ouro/min, recurso importante do jogo, adquirido principalmente ao se realizar o ataque final a unidades controladas pelo inimigo, e o qual é utilizado de forma a desenvolver as capacidades do personagem. Este resultado pode ser devido a distração e seu impacto na taxa com a qual os jogadores recebem as recompensas, ainda que novamente seja destacada a dificuldade de avaliar essa possível correlação com o modelo utilizado naquele estudo. Observa-se os resultados na Figura 1 a seguir, na qual foi representado o desempenho da variável Abates na cor laranja, a variável Assistências em cor verde, Mortes em cor azul e Ouro/min na cor roxa.

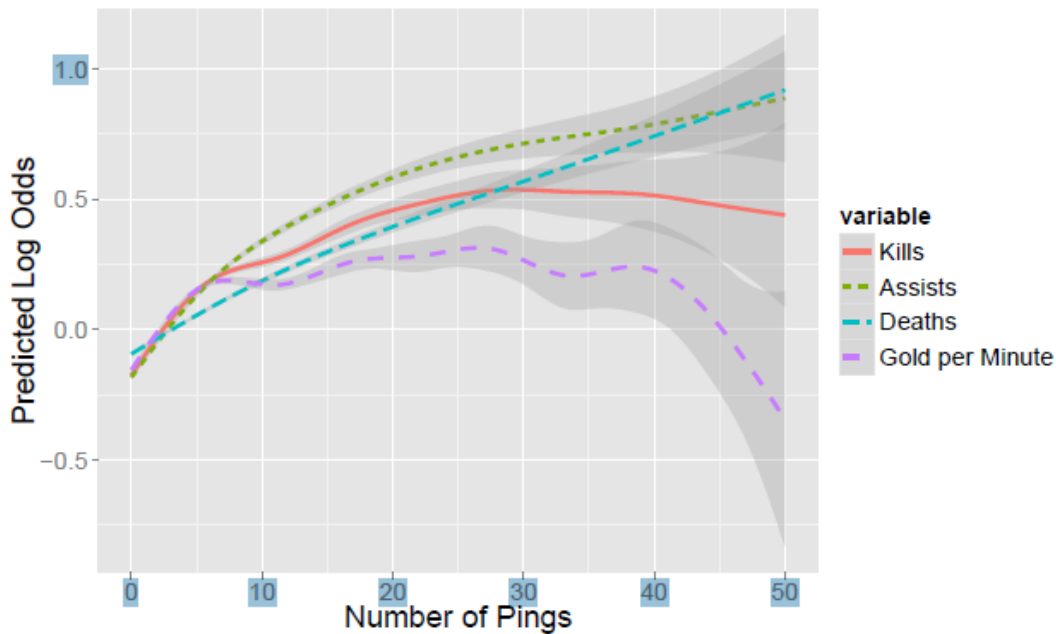


Figura 1. Relação do Número de *Pings* com diversos indicadores do jogo

Fonte: Adaptado de Leavitt, Keegan e Clark (2016)

2.3 Aplicações de Otimização e Priorização em MOBAs

Ong, Deolalikar e Peng (2015) introduziram uma estrutura algorítmica para previsão de resultado da partida (vitória ou derrota). Primeiramente, foram utilizados algoritmos de aprendizagem não supervisionada para achar *clusters* de comportamento, conjunto de objetos que possuem alguma similaridade que os diferencia dos demais em outros *clusters* (BIJURAJ, 2013), conforme as ações e desempenho em jogo dos *players*. Esta informação *clusterizada* foi então, utilizada em algoritmos de classificação de forma a ensinar o algoritmo a prever o resultado das partidas. Dessa forma, o algoritmo desenvolvido pelos autores foi capaz de determinar vitórias ou derrotas com uma precisão acima de 70%. É importante destacar que a informação *clusterizada* proporcionou uma maior acurácia ao algoritmo preditor em comparação à informação bruta extraída diretamente do aplicativo fornecido pela Riot Games, sobre a qual o algoritmo preditor apresentou uma acurácia média de 55%. Outro fator interessante descoberto no estudo foi que o estudo de *clusterização* encontrou papéis diferentes dos personagens em relação a aqueles pré-definidos pela empresa do jogo. O que acontece é que na prática, os jogadores podem utilizar os heróis de forma diferente daquela pressuposta pela empresa em busca do melhor desempenho. Entende-se que os *roles*, papéis personificados pelos campeões ao longo do jogo, os quais podem dinamicamente mudar e ficam sujeitos às decisões de customização do campeão durante a partida, determinados pela *clusterização* representam melhor a realidade do que de fato acontece em jogo, visto que o algoritmo preditor utilizando-se destes dados apresentou nível de acerto superior ao preditor com os papéis tradicionais.

Eggert et al. (2015) realizaram um estudo similar ao citado anteriormente de Ong, Deolalikar e Peng (2015), com a aplicação agora de algoritmos de classificação aliados a estatísticas de heróis de forma a identificar seus respectivos papéis (classe) no jogo Defense of The Ancients 2, o DotA 2, também um MOBA. Pelo método de aprendizado de máquina, foram utilizados diversos algoritmos classificadores no estudo, tendo como o que apresentou os resultados mais estáveis o de regressão logística. A precisão desenvolvida pelo preditor foi de 75%, similar ao estudo realizado no LoL, citado anteriormente. Uma particularidade interessante encontrada foi que reduzindo o número de classes da análise, a precisão do preditor aumentou para 96%. Isso foi justificado como efeito da falta de divisão dos dados de desempenho dos heróis por tempo, pois tendo-se apenas um valor final absoluto se negligencia o fato de que os papéis dos heróis podem sofrer alterações com o desenrolar das partidas. Um resumo dos resultados apresentados pelos classificadores, validados pelo método *10-fold cross-validation* explicitado em Refaeilzadeh, Tang e Liu (2009), está apresentado na Tabela 2.

Classificador	Precisão	Erro médio absoluto
Conjunto completo de classes		
<i>Random Forest</i>	76.27%	0.0905
Regressão Logística	75.85%	0.0826
Redes Bayesianas	72.03%	0.0801
Conjunto reduzido de classes		
<i>Random Forest</i>	91.17%	0.1162
Regressão Logística	96.15%	0.0381
Redes Bayesianas	96.58%	0.0322

Tabela 2: Resumo das precisões do *10-fold cross-validation* e erros médios absolutos com ambos conjuntos de classes

Fonte: Adaptado de Eggert et al. (2015.)

Kinkade e Lim (2015) apresentaram dois preditores de vitória para o DotA 2. O primeiro utiliza-se inteiramente dos dados coletados após a partida, enquanto que o segundo utiliza apenas dados relativos à seleção do herói. Por meio de um conjunto de dados de 62.000 partidas coletadas por meio da Interface de Programação de Aplicação (API) da Steam, e utilizando-se de variáveis como Ouro/min, Experiência/min e Abates/min, os algoritmos da Regressão Logística e *Random Forest* obtiveram acerto acima de 99% em todos os cenários de uso das variáveis, com exceção do cenário em que o preditor foi alimentado apenas da variável Abates/min, obtendo dessa forma um acerto na casa dos 97%. Devido a essa alta taxa de acerto do preditor, os autores sugeriram que a realização de estudos com um preditor que se utilize de informação coletada no início ou no meio das partidas, dessa forma surgindo a possibilidade de encontrar resultados mais interessantes. O preditor que utilizou apenas dados relativos a seleção

dos heróis no início da partida obteve um acerto de 73% nas previsões por meio do algoritmo da Regressão Logística, sobre o qual se concluiu ser mais eficiente do que o *Random Forest* nesse segundo cenário, pois estava apresentando problemas de *overfitting*, percentual de acerto na porção de treino consideravelmente maior do que na porção de teste, conforme esclarecido em Domingos (2012), o que nos indica que o classificador só está funcionando para aquele conjunto de dados. Sobre o segundo preditor, foi ressaltado pelo autor que sua precisão não foi perfeita pois uma partida de DotA não é decidida no estágio de seleção de heróis, mas que os acontecimentos que se desenrolam ao longo do jogo afetam influentemente a saída. Dessa forma, utilizando-se de informações sobre o que acontece durante o jogo iria aumentar de forma considerável a performance do preditor.

Um dos exemplos frequentemente citado como sugestões para trabalhos futuros é finalmente abordado em Barcellos (2017). Nesse estudo, o autor propôs uma sistemática para avaliar quais os principais indicadores de desempenho quanto à contribuição à vitória em partidas de jogadores de alto nível de LoL. A partir da coleta de dados de 724.817 partidas do jogo, o autor aplicou o método GEE (*Generalized Estimating Equations*) com o objetivo de identificar os fatores críticos para o triunfo em uma partida. Ao dividir as partidas em quatro períodos (0-10min, 10-20min, 20-30min e 30min-final) e incorporar os resultados de cada período nas análises seguintes, de forma também a verificar o impacto dos desempenhos iniciais ao longo da partida, concluiu-se que o indicador do Ouro apresentou mais relevância do que os demais, os quais eram dano, tropas abatidas e experiência acumulada. Em todos os períodos, o coeficiente do Ouro nas equações estimadas foi maior do que o das outras variáveis. Por fim, em relação aos resultados dos impactos dos desempenhos ao longo da partida, foi encontrada uma importância cumulativa entre períodos imediatamente consecutivos, entretanto foi observado que os resultados iniciais iam perdendo o valor conforme o jogo se prolongava.

Finalmente, Barcellos (2017) buscou adaptar um conceito oriundo da literatura de finanças, a fronteira eficiente de Markowitz, visando sua utilização em um aspecto crítico do jogo: a seleção de campeões. Novamente utilizando-se da coleta de dados de jogadores de alto nível de LoL, o autor propôs três fronteiras eficientes, sobre as quais são sugeridas 19 composições ótimas de equipes. Ainda que nenhuma dessas equipes foram selecionadas nas partidas sobre as quais o estudo foi feito, o autor realça que os resultados indicaram uma convergência significativa das escolhas de jogadores para uma quantidade pequena de campeões. Por fim, equipes compostas por no mínimo quatro dos cinco campeões sugeridos nas composições de equipes ótimas obtiveram uma taxa de vitória média de 62,5%.

3. PROCEDIMENTOS METODOLÓGICOS

Nesta seção do trabalho é explicitado o cenário dos eSports em termos de mercado e participação do LoL. Por fim, é apresentada a classificação do método de pesquisa e a caracterização do método de trabalho, descrevendo as etapas da realização do trabalho.

3.1 Descrição do cenário

Os eSports atingiram o patamar de US\$ 325 milhões de receita gerada no final de 2015 segundo o Relatório anual de eSports da Newzoo (2017). No estudo são considerados valores ganhos com direitos de transmissão, venda de ingressos para eventos de divulgação e partidas, publicidade online, parcerias com patrocinador e investimento adicional das publicadoras dos jogos. Ainda de acordo com este relatório, os eSports seguirão com um aumento constante de receita gerada até o fim da década (35,6% de aumento anual estimado entre 2016 a 2020), conforme o Gráfico 1.

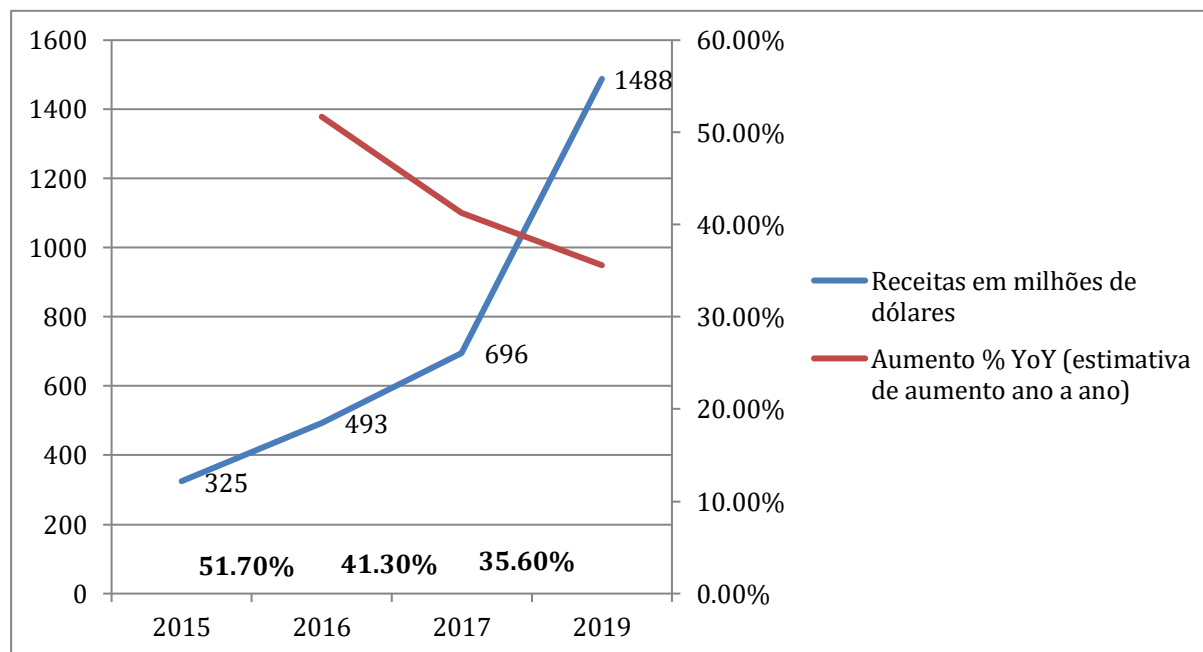


Gráfico 1: Evolução das receitas de eSports

Fonte: Adaptado de Newzoo (2017)

Destaca-se o também o mercado asiático de eSports, sobre o qual se estima que em 2017, China e Coréia do Sul juntas irão gerar US\$ 153 milhões, equivalente a 22% das receitas globais totais dos eSports, como o mostrado no Gráfico 2.

Receitas Globais - 2017

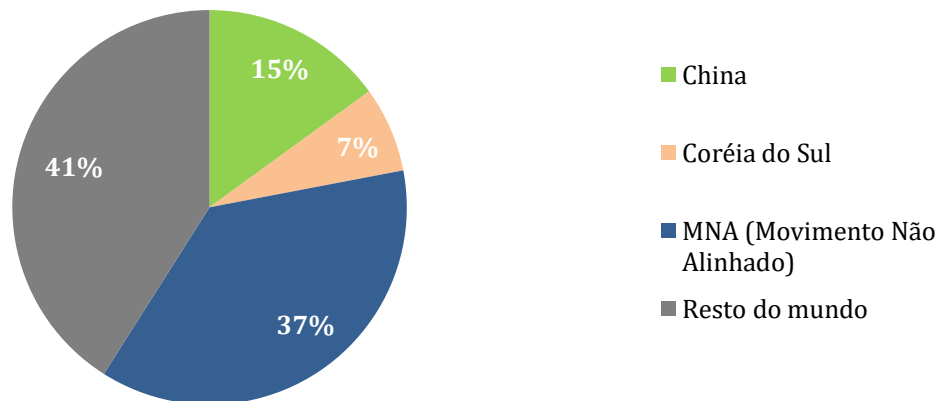


Gráfico 2: Divisão das receitas globais de eSports

Fonte: Adaptado de Newzoo (2017)

Neste crescente segmento de mercado destaca-se o produto desenvolvido pela Riot Games, o *League of Legends*. Trata-se da mais relevante plataforma do eSport atual com mais de 100 milhões de jogadores ativos mensalmente segundo Kollar (2016). Em termos de horas jogadas, ele também supera seus rivais conforme demonstrado na Tabela 3.

Título	Total de horas jogadas na Europa/EUA de Julho de 2011 a Junho de 2012
League of Legends	1,292,502,456
World of Warcraft	622,378,909
Minecraft	371,635,651
Heroes of Newerth	184,520,156
Diablo III	172,907,605
Battlefield 3	171,852,550
MapleStory	165,503,651
StarCraft II	163,980,293
World Of Tanks	145,702,931
Call of Duty: Modern Warfare 3	126,754,082

Tabela 3: Top 10 horas totais jogadas Jul-2011 a Jun-2012

(Fonte: Forbes. Elaboração: Gaudiosi, 2012)

3.2 Classificação da pesquisa

O estudo realizado neste trabalho pode ser classificado como de natureza aplicada, pois foca na busca de soluções de problemas reais. A abordagem da pesquisa é quantitativa, pois seu objetivo é o de testar algoritmos classificadores que possam acuradamente estimar a

possibilidade da equipe vencer ou perder no LoL. A classificação da pesquisa quanto aos objetivos é exploratória, considerando que a pesquisa visa proporcionar maior familiaridade com um fenômeno real e complexo, de modo a torná-lo explícito e desenvolver conhecimento sobre ele, e explicativa, visto que busca identificar fatores que determinam ou contribuem para o fenômeno em estudo. Quanto aos procedimentos, trata-se de pesquisa experimental, pois são manipuladas diretamente as variáveis relacionadas com o objeto de estudo.

3.3 Caracterização do método de trabalho

Este trabalho foi dividido em cinco etapas para a sua realização, sendo elas: (i) coleta de dados, envolvendo a captura de estatísticas das partidas realizadas pelos jogadores melhores qualificados no ranking do LoL em 11 servidores no período de 14 janeiro de 2016 até o dia 1º de setembro de 2016; (ii) revisão sobre estudos já realizados e análises relativas a algoritmos de classificação; (iii) tratamento e adequação dos dados coletados para utilização nos algoritmos propostos; (iv) aplicação dos algoritmos classificadores *Random Forest* e Regressão Logística nos estratos de tempo, para entender qual classificador consegue uma previsão melhor de vitórias, além de identificar qual indicador de desempenho utilizado apresenta uma maior importância sobre a geração dos classificadores; (v) análise dos resultados nas porções teste e treinamento, por meio de *10-fold cross-validation*.

Antes de mais nada, se realça que foi utilizado o mesmo banco de dados que Barcellos (2017). A etapa do trabalho referente à coleta de dados se deu com a captura dos dados dos jogos através da plataforma da API disponibilizada pela desenvolvedora do jogo. Foram listados apenas os jogadores do nível Desafiante, os quais representam os 200 melhores jogadores de cada servidor e incluem grande parte dos jogadores profissionais e de aspirantes a jogar profissionalmente. Essa lista foi coletada pela primeira vez no dia 1º de julho, e depois foi atualizada no dia 1º de setembro de 2016. Como são 11 servidores, cada coleta resultou em uma relação de 2200 jogadores. Foram então relacionadas todas as partidas ranqueadas destes jogadores no mapa Vale dos Invocadores, desde 14 de janeiro de 2016 (início da 6ª temporada) até dia 1º de setembro (ponto de corte da coleta). Foram coletados dados de 724.817 partidas, em um total de 7.248.170 registros individuais. A API da desenvolvedora do jogo fornece uma grande diversidade de informações, porém a maior parte delas refere-se somente aos valores no final da partida. Todos os indicadores que possibilitavam a análise comparativa das equipes ao longo das faixas de tempo foram incluídos na análise, resultando nas seguintes variáveis: monstros abatidos por minuto (“Tropas abatidas”), dano causado por minuto (“Dano”), ouro ganho por minuto (“Ouro”) e Experiência acumulada por minuto (“Experiência”). As variáveis a serem analisadas foram divididas em quatro momentos: A) de 0 a 10 minutos; B) de 10 a 20 minutos; C) de 20 a 30 minutos; e D) de 30 minutos até o final.

A coleta de dados teve que ser feita para cada jogador devido à estrutura de armazenamento de dados da API Riot. Assim, o resultado de cada jogador teve de ser somado ao de seus companheiros, visando a comparação entre times e não indivíduos. A comparação dos valores absolutos de times em partidas diferentes pode levar a conclusões enviesadas, pois mesmo um valor absoluto alto em determinada variável pode significar um desempenho inferior, caso a equipe adversária consiga um resultado ainda mais alto. Dessa forma, os valores totais coletados de cada variável por equipe foram subtraídos, resultando na diferença entre as duas equipes em cada partida, acompanhado do resultado de qual equipe saiu vencedora.

A segunda etapa do trabalho consistiu na revisão e estudo de artigos que indicariam que abordagem deveria ser seguida para, a partir dos dados coletados, realizar análises de forma a extrair informações implícitas destes.

A partir dos dados coletados, foi feito o cálculo do Coeficiente de correlação de Pearson, como exposto em Weston et al. (2001), o qual nos indica o grau de correlação entre diversas variáveis. Pelo cálculo de r , encontramos valores em torno de 0,9 para as variáveis Ouro e Experiência, o que nos indica um alto grau de correlação. Ainda que possa não impactar diretamente na precisão de um classificador, Strobl et al. (2008) demonstra que ao se utilizar certos métodos com dados altamente correlacionadas pode se acabar sofrendo com uma equivocada interpretação dos coeficientes de formação do modelo. Adicionalmente, conforme exposto por Tolosi e Lengauer (2011), modelos que possuem alto número de variáveis em comparação com o número de amostras, podem sofrer com viés de confirmação. Ainda que não seja o caso dessa análise, pois foram utilizadas no máximo quatro variáveis de entrada, deve-se ater a essas recomendações em pesquisas futuras que possam vir a utilizar um maior número de variáveis de entrada.

Sendo assim, visando encontrar classificadores estáveis e de alta precisão de forma prever o resultado de uma partida do LoL, utilizou-se o algoritmo de Regressão Logística. Conforme exposto por Sperandei (2014), a Regressão Logística funciona de forma parecida com a regressão linear, mas com uma variável de resposta binomial, o que é ideal no problema em que a saída é ou vitória ou derrota. A este algoritmo foi incorporada o método “*LASSO*” proposto por Tibshirani (1996), o qual suaviza atributos que sejam relacionadas uns aos outros, reduzindo a complexidade do modelo e o erro, bem como minimizando qualquer possibilidade de *overfitting*. Esse método é também conhecido como regularização L1. Conforme foi citado no Referencial Teórico, observou-se que diversos autores, em aplicações similares, utilizaram o algoritmo classificador do *Random Forest* para previsão da vitória. Esse algoritmo pode, também, cumprir o objetivo de descobrir qual dos indicadores analisados impacta mais em um resultado de vitória.

Para se obter classificações precisas de dados, desenvolveram-se métodos, como o chamado *bagging*, no qual a precisão da classificação vinha da criação de um grupo de árvores de decisão as quais votavam na classe de dados mais popular (BREIMAN, 1996). Com o desenvolvimento destes métodos, chegou-se a um dos métodos que serão testados nesse estudo, o algoritmo *Random Forest*, que consiste em um classificador consistente de uma grande coleção de classificadores de árvores de decisão correlacionadas formado por vetores aleatórios formados a partir dos dados iniciais, e no qual cada árvore de decisão contribui com um único voto para decidir a classe mais popular (BREIMAN, 2001). Essa combinação de modelos de aprendizagem aumenta a precisão da classificação ao se reduzir o ruído e criar um modelo com pequena variância. Outros benefícios do uso do *Random Forest*, conforme Breiman (2001), são sua transparência e simplicidade, as quais nos permitem visualizar as variáveis de entrada que mais estão impactando na formação do algoritmo classificador, por meio da propriedade *feature importance*. Dessa forma, utilizando-se de dados não correlacionados e obtendo-se uma alta precisão no classificador, pode-se utilizar essa propriedade para compreender a relevância das variáveis de entrada no resultado de saída com certa confiança. Assim sendo, também foi utilizado o algoritmo classificador do *Random Forest* para se observar o comportamento das previsões.

Quanto ao método utilizado para a seleção das variáveis de entrada dos algoritmos classificadores, observou-se abordagens de otimização das seleções, como o *wrapper approach* proposto em Kohavi e John (1997). Todavia, conforme os autores, deve-se procurar, na prática, por variáveis ótimas de acordo com o tipo específico de algoritmo de aprendizado utilizado, além dos dados de treino disponíveis. Isso, aliado ao fato delimitador da coleta de dados, que restringiu as variáveis coletadas, levou à decisão de formular todas as combinações possíveis entre as quatro entradas disponíveis e avaliar então os resultados das permutações. Se realça essa escolha devido ao número relativamente pequeno de variáveis disponíveis para se formar o modelo (quatro). Foram usados também nas análises iniciais os algoritmos classificadores *k-nearest neighbors*, conforme descrito em Zhang (2016) e observado o método de *decision tree* explicitado em Song e Lu (2015), mas como os resultados preliminares apresentaram menor precisão, descartaram-se em análises posteriores. Particularmente na situação do *decision tree*, a menor precisão relativa aos outros métodos pode ser explicada pois o próprio *Random Forest* se utiliza de uma lógica mais complexa de formação de um conjunto de árvores de decisão, o que aumenta a robustez e precisão do classificador. As aplicações de algoritmo foram inicialmente realizadas em linguagem R, posteriormente migrando para a linguagem *Python*, no software *Spyder*, devido a algumas facilidades de execução encontradas, como no tempo de execução dos algoritmos. Por fim, utilizou-se o método do *10-fold cross-validation* para estimar a variação dos modelos encontrados.

4. RESULTADOS E DISCUSSÕES

Os resultados dos classificadores do Momento A até o Momento D oscilaram entre 66 a 78%. Os cálculos com o algoritmo do *Random Forest* foram feitos utilizando-se 50 a 500 estimadores, variando conforme o número de amostras disponíveis para a análise. Ao algoritmo classificador da Regressão Logística, adicionou-se a regularização L1 (“Lasso”) de forma a reduzir a importância da correlação entre as variáveis. Nos cálculos das variações já estão incorporados três desvios padrões. Após diversas experimentações, constatou-se que a Regressão Logística apresentou as melhores precisões e estabilidade.

Realça-se que as precisões de classificação sempre se referem à porção de teste do conjunto de dados. A diferença na acurácia entre as porções treino e teste das classificações com o algoritmo da Regressão Logística foi desprezível. Entretanto, o acerto na porção de treino para o *Random Forest* foi acima de 90% sempre que houve duas ou mais variáveis na formação do modelo, enquanto que na porção de teste desses modelos obteve-se até mesmo a precisão de 66% em uma permutação no Momento A. Essa diferença considerável de mais de 20% nos indica a ocorrência de *overfitting*, todavia como acertos acima de 66% em algoritmos de classificação são consideráveis, esses resultados não serão omitidos de forma a incentivar a discussão. De qualquer forma, é preferível um classificador que seja 75% preciso em ambos conjuntos de dados do que um que seja 100% preciso no conjunto de treino e 50% no conjunto de teste (DOMINGOS, 2012), reforça o que foi concluído no parágrafo anterior de que o algoritmo da Regressão Logística apresenta a melhor relação precisão/confiabilidade. Os modelos gerados apenas com as variáveis Tropas Abatidas e/ou Dano obtiveram precisões de 50 a no máximo 60% em ambos classificadores, dessa forma não serão expostos seus resultados em cada estrato a seguir.

4.1 Momento A (de 0 a 10 Minutos)

Os classificadores com seu nível de acerto e variações calculadas para o período de zero a dez minutos de jogo estão reportados na Tabela 4.

Classificador	Precisão	Variação
<i>Random Forest</i>		
Ouro + Dano + Tropas Abatidas + Experiência	68.98%	0.33%
Ouro + Dano + Experiência	67.20%	0.33%
Ouro + Experiência	66.18%	0.46%
Ouro	70.66%	0.43%
Experiência	69.46%	0.24%
Regressão Logística		
Ouro + Dano + Tropas Abatidas + Experiência	70.94%	0.38%
Ouro + Dano + Experiência	70.92%	0.31%
Ouro + Experiência	70.91%	0.31%
Ouro	70.68%	0.34%
Experiência	69.60%	0.46%

Tabela 4 – Precisões dos classificadores pelo *10-fold cross-validation* no Momento A

Os classificadores gerados a partir da variável Ouro apresentaram os melhores resultados. Observa-se que o algoritmo da Regressão Logística apresentou maior precisão do que o *Random Forest* em todas comparações.

4.2 Momento B (de 10 a 20 Minutos)

Os classificadores com seu nível de acerto e variações calculadas para o período de dez a vinte minutos de jogo estão reportados na Tabela 5.

Classificador	Precisão	Variação
<i>Random Forest</i>		
Ouro + Dano + Tropas Abatidas + Experiência	77.82%	0.61%
Ouro + Dano + Experiência	76.36%	0.45%
Ouro + Experiência	74.94%	0.47%
Ouro	78.83%	0.41%
Experiência	74.08%	0.41%
Regressão Logística		
Ouro + Dano + Tropas Abatidas + Experiência	78.96%	0.39%
Ouro + Dano + Experiência	78.93%	0.37%
Ouro + Experiência	78.93%	0.32%
Ouro	78.93%	0.38%
Experiência	74.90%	0.46%

Tabela 5 - Precisões dos classificadores pelo *10-fold cross-validation* no Momento B

Houve um aumento considerável nas precisões comparando-se com o Momento anterior, o que evidencia uma importância maior dos indicadores Ouro e Experiência nesta fase.

As classificações com a Regressão Logística apresentam virtualmente a mesma precisão em todos modelos que consideraram a variável Ouro em sua formação.

4.3 Momento C (de 20 a 30 Minutos)

Os classificadores com seu nível de acerto e variações calculadas para o período de vinte a trinta minutos de jogo estão reportados na Tabela 6.

Classificador	Precisão	Variação
<i>Random Forest</i>		
Ouro + Dano + Tropas Abatidas + Experiência	77.10%	0.73%
Ouro + Dano + Experiência	75.41%	0.66%
Ouro + Experiência	74.29%	0.60%
Ouro	77.63%	0.71%
Experiência	71.70%	0.84%
Regressão Logística		
Ouro + Dano + Tropas Abatidas + Experiência	78.15%	0.75%
Ouro + Dano + Experiência	77.93%	0.61%
Ouro + Experiência	77.91%	0.66%
Ouro	77.66%	0.61%
Experiência	71.87%	0.97%

Tabela 6 - Precisões dos classificadores pelo *10-fold cross-validation* no Momento C

Mesmo com a permutação de variáveis nas formações de modelos, observa-se uma certa estabilidade nas precisões encontradas pelos classificadores, as que possuem a variável de entrada Ouro mantém valores similares de precisão. A Regressão Logística vem apresentando melhores resultados, tanto em precisão como em estabilidade.

4.4 Momento D (de 30 até o fim)

Os classificadores com seu nível de acerto e variações calculadas para o período de trinta minutos até o fim do jogo estão reportados na Tabela 7. Houve queda na precisão dos modelos gerados unicamente a partir da variável Experiência na fase final do jogo, em que os campeões provavelmente já alcançaram seu nível máximo de habilidade.

Classificador	Precisão	Variação
<i>Random Forest</i>		
Ouro + Dano + Tropas Abatidas + Experiência	76.88%	1.15%
Ouro + Dano + Experiência	73.57%	1.15%
Ouro + Experiência	71.98%	1.10%
Ouro	74.26%	0.95%
Experiência	66.27%	1.06%

Regressão Logística

Ouro + Dano + Tropas Abatidas + Experiência	77.77%	1.11%
Ouro + Dano + Experiência	76.20%	1.07%
Ouro + Experiência	75.97%	1.06%
Ouro	74.38%	0.81%
Experiência	66.87%	1.16%

Tabela 7 - Precisões dos classificadores pelo *10-fold cross-validation* no Momento D

4.5 Partida inteira (de 0 até o fim)

Pela predominância do indicador Ouro nos algoritmos classificadores de maior precisão acima, foram usados dados acumulados de Ouro do início ao fim da partida. Utilizando 500 estimadores no *Random Forest* obteve-se um acerto na previsão de 81,36%, enquanto que com a Regressão Logística junta da regularização L1, se obteve 81,87%. Os valores de formação desse modelo de *Random Forest* estão na Tabela 8.

Variáveis	Feature Importance
Ouro (0 a 10min)	13.2%
Ouro (10 a 20min)	17.8%
Ouro (20 a 30min)	25.3%
Ouro (30 até o fim)	43.7%

Tabela 8 – *Feature Importances* do indicador Ouro em vários momentos no *Random Forest*

Como os coeficientes de correlação de Pearson obtiveram resultado desprezível para todas as relações entre as quatro variáveis de ouro, aliado a precisão do classificador, concluiu-se que momentos mais tardios da partida apresentam um impacto maior nas chances de vitória dos times. Além disso, foi utilizado um número alto de estimadores (500) de forma que os resultados das *Feature Importances* se mantiveram constantes com diversas simulações (STROBL, 2008), permitindo uma interpretação sensata sobre a importância dos valores.

4.6 Importâncias dos demais indicadores

De forma a entender melhor o impacto das quatro variáveis coletadas (Ouro, Dano, Tropas Abatidas e Experiência) na formação dos modelos de *Random Forest* dos itens 4.1 ao 4.4, foi feito o cálculo das *Feature Importances* e esquematizado na Tabela 9.

<i>Feature Importance</i>	Ouro	Experiência
Momento A	31.9%	29.0%
Momento B	39.4%	28.3%
Momento C	41.6%	26.9%
Momento D	40.6%	23.5%

Tabela 9 - *Feature Importances* nos estratos A, B, C e D para os modelos calculados com as variáveis Ouro, Dano, Tropas Abatidas e Experiência

Como os indicadores de Ouro e Experiência apresentaram alto valor de correlação coeficiente de Pearson, podem-se ter como inconclusivos os valores das *features importances* calculados na Tabela 9. De fato, ainda que a partir do momento B a diferença de importância da variável Ouro fica evidente, os valores calculados no momento A reforçam nossa dúvida de que essa diferença pode não ser significativa para afirmar que o Ouro apresenta o maior impacto na criação de um modelo classificador confiável para a previsão de vitória. Dessa forma, calculou-se o impacto das variáveis nos modelos separadamente, juntando-as aos dois remanescentes indicadores de Dano e Tropas Abatidas, os quais apresentaram baixo grau de correlação. Os resultados para os modelos com a variável Ouro estão na Tabela 10.

<i>Feature Importance</i>	Ouro	Dano	Tropas Abatidas	Precisão
Momento A	47,6%	35,6%	16,8%	66,44%
Momento B	52,6%	29,4%	18,0%	75,96%
Momento C	57,4%	27,2%	15,4%	75,39%
Momento D	53,5%	27,8%	18,7%	75,35%

Tabela 10 - *Feature Importances* nos estratos A, B, C e D para os modelos calculados com as variáveis Ouro, Dano e Tropas Abatidas

Os resultados para os modelos com a variável Experiência estão na Tabela 11.

<i>Feature Importance</i>	Experiência	Dano	Tropas Abatidas	Precisão
Momento A	48,3%	37,6%	14,1%	65,57%
Momento B	48,0%	32,5%	19,5%	71,77%
Momento C	50,8%	32,1%	17,1%	69,74%
Momento D	44,3%	31,9%	23,8%	69,70%

Tabela 11 - *Feature Importances* nos estratos A, B, C e D para os modelos calculados com as variáveis Experiência, Dano e Tropas Abatidas

Observa-se que as variáveis Ouro e Experiência são dominantes na formação de seus modelos, enquanto que as variáveis Dano e Tropas Abatidas são coadjuvantes. Assim sendo,

fica exposto que se atingem maiores precisões com o uso da variável Ouro em comparação à variável Experiência.

5. CONSIDERAÇÕES FINAIS

O intuito de se prever o resultado final de uma partida de alto nível de LoL, baseado em poucos indicadores, foi cumprido. A inclusão da regularização Lasso no modelo de Regressão Logística proporcionou a geração de um modelo de classificação confiável e estável com média de 75% de acerto de previsão apenas com o indicador do Ouro, em todos os momentos da partida com exceção do A, o qual parece não impactar tanto no resultado final conforme os valores gerados pelo classificador, fato reforçado pela conclusão em Barcellos (2017).

A limitação dos dados (apenas quatro categorias e quatro estratos) limita os resultados, uma vez que existem muitas variáveis de interesse que não estão disponíveis via API. A obtenção destes parâmetros manualmente necessitaria de um trabalho desgastante; dessa forma, incentiva-se o desenvolvimento de ferramentas as quais auxiliem no processo de coleta, as quais poderiam enriquecer significativamente o modelo.

O preditor teve um acerto menor na primeira fase e valores estáveis ao longo das três remanescentes, isso indica que os desempenhos obtidos nos primeiros dez minutos de jogo, ainda que importantes, perdem relevância ao longo da partida e impactam menos na determinação da vitória do que em períodos posteriores. Isso pode ser explicado por exemplo, pelo *respawn* mais tardio dos heróis na medida em que o jogo avança. Dessa forma, um desempenho negativo resultante em uma morte de herói no fim de uma partida é muito mais punitivo do que no início, em que o *respawn* dos campeões é mais rápido.

Os algoritmos poderiam ser mais efetivos em prever as chances de vitória ao longo de partidas de alto nível do LoL. Fica a indicação a trabalhos futuros de que a inclusão de mais variáveis significativas possa auxiliar ainda mais na precisão de novos modelos. Mesmo assim, considerando o aspecto dinâmico do jogo e suas diferentes etapas, o Ouro mostrou-se significativamente superior aos outros três indicadores para a geração dos modelos preditores de estimativa de vitória. Além disso, constatou-se que o desempenho nas fases iniciais do jogo vai perdendo importância na medida em que o mesmo se prolonga.

6. REFERÊNCIAS BIBLIOGRÁFICAS

BARCELLOS, R. L. **Suporte à tomada de decisão estratégica no âmbito de esports: o caso do League of Legends**. 2017. 111 f. Dissertação (Mestrado em Engenharia) – Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2017.

BIJURAJ, L.V.; Clustering and its Applications. **Proceedings of National Conference on New Horizons in IT - NCNHIT 2013**.

- BREIMAN, Leo. Bagging Predictors. **Machine Learning**, Volume 24, Issue 2, pp. 123–140. 1996
- BREIMAN, Leo. Random Forests. **Machine Learning**, Volume 45, Issue 1, pp. 5–32. 2001.
- DOMINGOS, Pedro. A Few Useful Things to Know about Machine Learning. **Communications of the ACM**, Volume 55, Issue 10, pp. 78-87. 2012.
- DONALDSON, Scott. Mechanics and Metagame: Exploring Binary Expertise in League of Legends. **Games and Culture** 1-19. 2015.
- EGGERT, Christoph; HERRLICH, Marc; SMEDDINCK, Jan David; MALAKA, Rainer. Classification of Player Roles in the Team-Based Multi-player Game Dota 2. **Entertainment Computing - ICEC 2015: 14th International Conference, ICEC**. 2015.
- EGLISTON, Ben. Big playerbase, big data: On data analytics methodologies and their applicability to studying multiplayer games and culture. **First Monday**, 2016, Volume 21, Issue 7. Disponível em < <http://journals.uic.edu/ojs/index.php/fm/article/view/6718/5531>>. Acesso em 13 nov. 2016.
- FERRARI, Simon. From Generative to Conventional Play: MOBA and League of Legends. **Proceedings of DiGRA 2013: DeFragging Game Studies**. Georgia Institute of Technology, Atlanta, Georgia, Estados Unidos, 2013.
- FERREIRA, 2015. Indústria de games supera o faturamento de Hollywood. Disponível em <<https://webnoticias.fic.ufg.br/n/68881-industria-de-games-supera-o-faturamento-de-hollywood>>. Acesso em 04 nov. 2016.
- GAUDIOSI, 2012. Riot Games' League Of Legends Officially Becomes Most Played PC Game In The World. Disponível em <<http://www.forbes.com/sites/johngaudiosi/2012/07/11/riot-games-league-of-legends-officially-becomes-most-played-pc-game-in-the-world/#7030a76b6ae2>>. Acesso em 05 nov. 2016.
- KIM, Jooyeon; KEEGAN, Brian C.; PARK, Sungjoon; OH, Alice. 2015. The Proficiency-Congruency Dilemma: Virtual Team Design and Performance in Multiplayer Online Games. Disponível em < <https://arxiv.org/abs/1512.08321>>. Acesso em 12 nov. 2016.
- KINKADE, Nicholas; LIM, Kyung yul Kevin. 2015. DOTA 2 Win Prediction. Disponível em <<https://cseweb.ucsd.edu/~jmcauley/cse255/reports/fa15/018.pdf>>. Acesso em 10 jun. 2017.
- KOHAVI, Ron; JOHN, George H. Wrappers for feature subset selection. **Artificial Intelligence**, Volume 97, Issue 1-2, pp. 273-324. 1997.
- KOLLAR, P. **The Past, Present and Future of League of Legends Studio Riot Games**, 2016. Disponível em <<http://www.polygon.com/2016/9/13/12891656/the-past-present-and-future-of-league-of-legends-studio-riot-games>>. Acesso em: 16 jun. 2017.
- LEAVITT, A; KEEGAN, B. C.; CLARK, J.; 2016. Ping to Win? Non-Verbal Communication and Team Performance in Competitive Online Multiplayer Games. Disponível em <http://alexleavitt.com/papers/2016_CHI_LeavittKeeganClark_PingTeamPerformance.pdf>. Acesso em 12 nov. 2016.

- NEWZOO, 2017. **Esports revenues will reach \$696 million this year and grow to \$1.5 billion by 2020 as brand investment doubles.** Disponível em <<https://newzoo.com/insights/articles/esports-revenues-will-reach-696-million-in-2017/>>. Acesso em 30 jun. 2017.
- NEWZOO, 2017. **Newzoo's 2017 report: insights into the \$108.9 billion global games market.** Disponível em <<https://newzoo.com/insights/articles/newzoo-2017-report-insights-into-the-108-9-billion-global-games-market/>>. Acesso em 30 jun. 2017.
- NEWZOO, 2017. **The Brazilian Gamer,** Disponível em <<https://newzoo.com/insights/countries/brazil/>>. Acesso em 30 jun. 2017.
- ONG, Hao Yi; DEOLALIKAR, Sunil; PENG, Mark. 2015. Player Behavior and Optimal Team Composition in Online Multiplayer Games. Disponível em <<https://arxiv.org/pdf/1503.02230v1.pdf>>. Acesso em 12 nov. 2016.
- RAMBUSCH, Jana; JAKOBSSON, Peter; PARGMAN, Daniel. 2007. Exploring E-sports: A Case Study of Gameplay in Counter-strike. Disponível em <<https://pdfs.semanticscholar.org/3e92/3bacd4cc82b295066a13e47dd42be9de0f64.pdf>>. Acesso em 19 nov. 2016.
- RAY, Michael. 2011. StarCraft. *Britannica Academic*, Encyclopædia Britannica. Disponível em <academic-eb-britannica.ez45.periodicos.capes.gov.br/levels/collegiate/article/544682>. Acesso em 04 nov. 2016.
- REFAEILZADEH, Payam; TANG, Lei; LIU, Huan. Cross Validation. **Encyclopedia of Database Systems**, pp. 532-538. 2009.
- SINCLAIR, B. Mobile could push games to \$100 billion by 2017, Digi-Capital, 2014. Disponível em: <<http://www.gamesindustry.biz/articles/2014-07-09-mobile-could-push-games-to-usd100-billion-by-2017-digi-capital>>. Acesso em: 16 jun. 2017.
- SONG, Yan-yan; LU, Ying. Decision tree methods: applications for classification and prediction. **Shanghai Archives Psychiatry**, Volume 27, Issue 2, pp. 130-135. 2015.
- SPERANDEI, Sandro. Understanding logistic regression analysis. **Biochemia Medica**, Volume 24, Issue 1, pp. 12-18. 2014.
- STROBL, C; BOULESTEIX, A-L; KNEIB, T; AUGUSTIN, T; ZEILEIS, A. Conditional variable importance for random forests. **Bioinformatics**, Volume 9, pp. 307. 2008.
- TIBSHIRANI, Robert. Regression Shrinkage and selection via the lasso. **Journal of the Royal Statistical Society. Series B**, Volume 58, Issue 1, pp. 267-288. 1996.
- TOLOSI, Laura; LENGAUER, Thomas. Classification with correlated features: unreliability of feature ranking and solutions. **Bioinformatics**, Volume 27, Issue 14, pp. 1986-1994. 2011.
- WESTON, Jason; MUKHERJEE, Sayan; CHAPELLE, Olivier; PONTIL, Massimiliano; POGGIO, Tomaso; VAPNIK, Vladimir. Feature selection for SVMs. **Neural Information Processing Systems**, Volume 12, pp. 668-674. 2001.
- ZHANG, Zhongheng. Introduction to machine learning: k-nearest neighbors. **Annals of Translational Medicine**, Volume 4, Issue 11, pp. 218. 2016.