

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL**

**Métodos de Machine Learning para seleção de variáveis com  
aplicações ao Rugby Sevens Feminino**

Roberta Baptista de Azevedo  
Orientador: Guilherme Pumi

Porto Alegre  
2019

## RESUMO

O artigo busca relacionar as áreas gestão de risco, seleção de variáveis e performance esportiva. Para tanto, são trazidas perspectivas teóricas de diferentes áreas do conhecimento, as quais justificam a relação desses temas na gestão de desenvolvimento esportivo. Além disso, discute-se a importância dos atuários desenvolverem conhecimentos de seleção de variáveis através de métodos de *machine learning*. A metodologia desenvolvida nesta pesquisa focou-se em seleção de variáveis para identificar características de equipes vencedoras em *Rugby Sevens* Feminino. Foram analisados, através de vídeo, 29 jogos escolhidos aleatoriamente das temporadas 2014/2015; 2015/2016; 2017/2018 do Circuito Mundial de Rugby Sevens Feminino. Foram coletadas 112 variáveis para análise e seleção. Após tratamento dos dados, foram selecionadas 52 variáveis para definição das mais importantes. Três métodos de Seleção de Variáveis foram comparados e FoBa foi o mais eficiente na predição dos resultados. Desta forma, utilizando o FoBa, selecionou-se 27 variáveis com importância acima de 0.2 em escala de 0 a 1. Dentre essas variáveis, foi observado que as variáveis de Resultado, *Try* e Conversão, estão entre as 5 mais importantes, sendo *Try* a mais importante com 0.98 de importância. As outras três variáveis do Top 5 de importância, estão relacionadas à Entrada nas 22 de ataque. Ao contrário do imaginado, não tentar tacklear, mesmo que errando, ao perseguir ou marcar dentro das 22 de defesa é uma das variáveis do Top 5.

Palavras-chave: *rugby*, *machine learning*, *support vector machine*, seleção de variáveis, Ciências Atuariais.

## ABSTRACT

This work seeks to relate the areas of risk management, feature selection and sports performance. For that, theoretical perspectives of different areas of knowledge are brought, which justify the relation of these themes in the management of sports development. In addition, it is discussed the importance of actuaries to develop knowledge of feature selection via machine learning methods. The methods developed in this research focused on feature selection to identify characteristics of winning teams in the Women's Rugby Sevens. Twenty-nine randomly chosen games from the 2014/2015; 2015/2016; 2017/2018 seasons of the World Rugby Women Sevens Series were analyzed through video. A total of 112 features were collected for analysis and selection. After data processing, 52 features were selected to define the most important ones. Three methods of Feature Selection were compared and FoBa was the most efficient in predicting the results. Thus, using FoBa, 27 features with importance above 0.2 on a scale of 0 to 1 were selected. Among these variables, it was observed that the Scoring features, Try and Conversion, are among the 5 most important ones, with Try being the most important with 0.98 of importance. The other three features of the Top 5 of importance are related to entering into the opposition's 22 meters zone. Unlike what was expected, not attempting to tackle, even if erroneous, while chasing or defending within the defense 22 meters zone is one of the Top 5 features.

**Keywords:** Rugby, machine learning, support vector machine, feature selection, actuarial science.

## 1 INTRODUÇÃO

Com o avanço das tecnologias, grandes volumes de dados estão disponíveis para a utilização na construção dos modelos de gestão atuarial. Porém, mais relevante do que a quantidade de dados disponíveis é o que pode ser feito com eles (LACERDA, 2017). Mano (2018) aponta que, apesar dos benefícios do Big Data para o mercado atuarial, é necessário ficar atento ao acúmulo e à utilização de informação inútil. Por isso, é importante realizar uma seleção de variáveis específica para o problema a ser resolvido.

Dentro deste contexto, Lacerda (2017) destaca que Big Data tem como aplicações nas Ciências Atuariais a identificação de perfis de clientes; a prevenção e a detecção de fraudes; o monitoramento do comportamento de clientes através de mídias sociais; a identificação de pacientes com propensão a doenças crônicas; as análises de fontes externas para ajuste do cálculo de solvência de instituições. Enquanto que Genovez (2018) destaca que, em relação à Inteligência Artificial (IA), muitas das aplicações atuariais estão alicerçadas em técnicas que servem de base para IA, como técnicas de *Generalized Linear Models* (GLM) que são comumente utilizadas em estimativas e precificação de seguros massificados. De forma a entender melhor como a seleção de variáveis está inserida nas Ciências Atuariais, apresenta-se a seguir revisão de quatro estudos.

Ramos (2011) propõe um método automático para analisar um maior número de Comunicado de Ocorrência de Perda (COP) em seguros agrícolas. Este método é baseado na construção de modelos com técnicas de IA para inferir se há evidências de que pode ocorrer COP a partir dos dados dos laudos obtidos pelo Ministério da Agricultura para as safras de agricultura familiar dos anos de 2006 a 2010. No artigo, o autor utiliza um algoritmo chamado Apriori e diversos métodos de classificação e seleção de variáveis, incluindo *support vector machines* (SVM) e *bagging*.

Em Duma *et. al* (2011), os autores utilizam métodos de seleção de variáveis para melhorar o algoritmo Ripper quando usado em classificação de risco para seguros. Os autores utilizam análise de componentes principais (PCA), redes neurais bayesianas e determinação automática de importância combinados ao algoritmo Ripper. Em 2018, Hall, diretor de pesquisa da Society of Actuaries, publicou relatório sobre análise de risco atuarial utilizando técnicas de análise preditiva; de segmentação; e decomposição. No relatório são discutidas a evolução das tábuas para modelos de mortalidade e a importância de Big Data para as Ciências Atuariais.

Já Oliveira (2016) busca desenvolver um classificador baseado em regras de associação e regressão logística para mineração de dados com intuito de obter um conjunto de regras de detecção de fraudes em cartões. O autor utiliza diversos classificadores para tal, como Árvores de decisão e *Naive Bayes*; os algoritmos Apriori e FP-Growth; e regressão logística com método *stepwise*.

“A criação pelo governo das figuras de auditor atuarial e de gestor de risco ampliam o leque de responsabilidades que a profissão alcançou” (Chaves 2017). Apesar da Circular da Susep nº 517/2015 não explicitar a necessidade de ser um atuário para o papel de gestor de risco, Chaves (2017) defende que a atividade é inerente à formação do atuário.

O risco no mercado esportivo está atrelado tanto à saúde financeira das instituições quanto a questões técnicas, táticas e físicas das equipes. Segundo Kearney (2003), citado por De Souza *et. al* (2005), poucas empresas quantificam o retorno sobre investimento em esportes. A grande maioria delas depende do patrocinador para conhecer o retorno, e 61% não recebe este feedback. Jensen (2014) defende que existe relação entre performance da equipe esportiva e retorno do investimento (ROI).

Analisando-se os relatórios financeiros da Confederação Brasileira de Rugby (CBRu), percebe-se que há maior probabilidade de ocorrer patrocínio privado nas seleções quando os resultados internacionais são positivos, independentemente do nível de rendimento técnico da competição. Desta forma, para uma equipe esportiva ter maior chance de se capitalizar através de patrocinador privado, é necessário obter melhor performance entre os concorrentes.

Métodos e indicadores de análises de performance em *Rugby Sevens* começaram a ser estudados principalmente com a entrada do *Rugby* nas Olimpíadas (ROSS, GIL, *et al.*, 2016). Pesquisas prévias sobre *Rugby Sevens* feminino que tratem especificamente de indicadores de performance técnicos e/ou táticos são praticamente nulas. Em fevereiro e em junho, realizou-se pesquisa utilizando-se os termos “*Rugby*”, “*Sevens*”, “*Women*”, “*indicators*”, “*performance*”, “*KPI*” e combinações dos mesmos no Portal de Periódicos da Capes (CAPES/MEC, 2019). Na pesquisa realizada utilizando-se os termos “*rugby sevens*” AND “*women*”, 147 trabalhos foram apresentados e analisados um a um pelo título, palavras-chave e resumo. Apenas 2 eram diretamente relacionados a indicadores de performance técnico e/ou tático, mas um deles era especificamente sobre *Ruck*.

Utilizando dados oficiais da IRB, Higham, Hopkins, *et al.* publicaram 3 artigos sobre indicadores de performance em *Rugby Sevens* masculino em 2014. Nestes artigos, os autores identificam e caracterizam indicadores de performance relacionados ao desenvolvimento

tático e à probabilidade de vitória<sup>1</sup>; à marcação de pontos e à probabilidade de vencer<sup>2</sup>; e ao posicionamento no ranking mundial<sup>3</sup>. Também relacionando vitórias à marcação de pontos no *Rugby Sevens* masculino, Ross, Gil, *et al.* publicam artigo em 2016. Porém, os dados foram obtidos através de análise de vídeo utilizando o software SportsCode e utilizou outra temporada como amostra.

Apenas em 2016, Barkell, O'Connor e Cotton publicaram pesquisa analisando *Rugby Sevens* feminino. O objetivo do estudo foi analisar as variáveis do jogo relacionadas à vitória nas fases de confrontos diretos. Foram analisados, por meio de vídeo, jogos internacionais masculinos e femininos da mesma temporada, modelados em cada gênero e, ao final, comparando as duas categorias.

No Brasil, foram publicadas pesquisas que definiram e caracterizaram ações de jogo de equipes vencedoras em *Rugby Sevens*. Sansigolo Andrello (2013) analisou passes e *Tackles* de jogos nacionais masculinos. Enquanto Silveira (2018) analisou as ações de jogo da equipe campeã do Super Sevens. Ambas as coletas de dados foram realizadas através de análise de vídeo.

Neste cenário, o presente artigo tem como objetivo revisar algumas técnicas de *machine learning* para a seleção de variáveis úteis para as mais diversas áreas de aplicação, em particular nas ciências atuariais. Estas técnicas são então aplicadas à seleção variáveis de performance técnica e tática em Rugby Sevens Feminino de uma amostra piloto do Circuito Mundial de Rugby Sevens Feminino. Tendo em vista que a maioria das pesquisas em Rugby Sevens utilizam os dados oficiais da World Rugby e não apresentam as definições das variáveis, este artigo apresenta uma nova perspectiva para ser utilizada em trabalhos futuros.

### 3 METODOLOGIA

Nesta seção vamos apresentar fundamentos e problemas de técnicas utilizadas neste trabalho, como análise de vídeo, *machine learning*, seleção de variáveis, *forward* e *backward greedy selection* e o *Adaptive Forward-Backward Greedy Algorithm* (FoBa).

#### 3.1 Análise de vídeo

A análise observacional é utilizada no esporte desde os primórdios. Em 1986, Franks e Goodman publicaram artigo apresentando uma abordagem sistematizada para análises de

---

<sup>1</sup> Identificado como Higham, Hopkins, *et al.* (2014c)

<sup>2</sup> Identificado como Higham, Hopkins, *et al.* (2014b)

<sup>3</sup> Identificado como Higham, Hopkins, *et al.* (2014a)

performances esportivas. Apesar das tecnologias disponíveis na época, Franks e Goodman (1986) já destacavam a necessidade de programação específica para cada esporte. Além disso, destacam a necessidade de adaptabilidade dos softwares, pois, para uma efetiva utilização da abordagem proposta, é necessário o acompanhamento frequente e atualização das informações e necessidades de análises.

A evolução das tecnologias disponíveis tornaram possíveis as realizações de análises de vídeo após o jogo ou mesmo on-line. Análises após o jogo podem ocorrer através de imagens filmadas já com esta finalidade ou através de filmagens oficiais dos jogos (MARQUES, 2018). Estas análises podem ser manuais, semiautomatizadas ou automatizadas.

Existem diversos métodos para análises automatizadas de vídeo, como através de *Semantic Shots*, palavras-chaves, movimento da câmera, trajetória de objetos e detecção de eventos (XU, CHENG, *et al.*, 2009). Porém, ainda hoje, há diversas limitações para a automatização de alguns destes métodos. Um dos pontos de maior dificuldade de automatização está no acompanhamento de variáveis de movimento que contenham nuances não especificáveis na programação dos softwares, como a qualidade dos passes (bom, ok e ruim). Além disso, Xu, Cheng, *et al.* (2019) destaca que a utilização da técnica de palavras-chaves pode enfrentar problemas de programação de reconhecimento de palavras e de entonação ou mesmo problemas técnicos no áudio, que podem ocorrer na gravação, na transmissão ou ainda no tratamento dos dados.

Enquanto isso, Barris e Button (2008) avaliam que análises de vídeo observacionais manuais são mais propensas ao erro, pois a qualificação da situação e a sinalização da mesma são realizadas manualmente por uma pessoa. Para minimizar o erro, sugerem teste de validação da análise, que podem ocorrer de duas formas: análise de uma amostra dos jogos por outros especialistas ou pelo mesmo analista, mas com período de tempo entre as análises. Também é possível tomar alguns cuidados durante a coleta, como utilizar método que não seja necessário trocar os documentos na tela e que haja possibilidade de pausar e retornar situações de forma rápida (BARKELL, O'CONNOR e COTTON, 2016).

### **3.2 *Machine learning* e Seleção de variáveis**

*Machine learning* é um nome genérico dado a um conjunto de métodos para análise de dados desenvolvidos com o intuito de fazer previsão e classificação (DIANA, GRIFFIN, *et al.*, 2019) (HASTIE, TIBSHIRANI e FRIEDMAN, 2016). Embora a área de *machine learning* tenha ganhado força nos últimos anos com o desenvolvimento das tecnologias

relacionadas ao processamento e armazenamento de dados, as técnicas utilizadas aparecem na literatura desde os anos 50. Os computadores modernos tornaram a utilização de métodos de *machine learning* no dia a dia o que impulsionou não apenas a pesquisa e desenvolvimento de metodologias, mas também a utilização de tais técnicas nas mais diversas áreas de aplicação, em particular, para aplicações em diversos tipos de mercados (HALL, PHAN e WHITSON, 2016).

A ideia do machine learning segue o seguinte princípio: divide-se os dados em duas partes, uma chamada de amostra de treino e a outra de amostra de validação. A partir da amostra de treino, um modelo, ou classe de modelos, previamente determinado é “treinado” (isto é, ajustado aos dados) e então sua capacidade de previsão é testada na amostra de validação. O modelo final dentre os modelos da classe é geralmente aquele que melhor prediz, de acordo com alguma métrica, os dados contidos na amostra de validação. A definição dos tamanhos das amostras de treino e validação e também o método de divisão da amostra são problemas importantes neste contexto. Se a amostra de treinamento for muito pequena, devido ao pequeno tamanho amostral poderemos ter poucas informações disponíveis na amostra sobre a sua estrutura, podendo render um modelo pouco fidedigno às características populacionais. Se por outro lado a amostra de validação é pequena, como teremos poucos pontos para verificar a acurácia do modelo treinado, podemos escolher um modelo que possui uma variância de previsão muito grande.

É possível categorizar os métodos de *machine learning* em duas categorias: aprendizagem supervisionada e não supervisionada. A primeira ocorre quando há um grupo de exemplos de entrada-saída, ou seja, para cada observação da preditora  $x_1, \dots, x_p$  há uma resposta associada  $y_i$ . Muitos métodos clássicos de aprendizagem, como regressão linear e regressão logística, assim como abordagens modernas, como *support vector machines* (SVM), estão inclusos nesta categoria (JAMES, WITTEN, *et al.*, 2015).

Aprendizagem não supervisionada descreve situações mais desafiadoras, pois para cada observação  $i = 1, \dots, p$  observa-se um vetor  $x_i$  que não associado à uma resposta  $y_i$ . Portanto, não é possível ajustar um modelo no sentido tradicional da palavra (um modelo de regressão por exemplo), pois não há variável resposta para que se possa realizar a predição. Nestes métodos, o objetivo é geralmente o agrupamento das variáveis em cluster de pontos que apresentam algum tipo de característica semelhante. O desafio se torna então a correta classificação de novos pontos observados nestes grupos predeterminados. Deste modo é possível buscar o entendimento da relação entre as variáveis ou entre as observações, via análise de *cluster* (JAMES, WITTEN, *et al.*, 2015).

Tendo em vista que *machine learning* busca abordar tarefas grandes e complexas, o problema de se focar nas informações mais importantes em uma quantidade imensa de dados é de grande relevância. (BLUM e LANGLEY, 1997). Teoricamente, em aprendizagem supervisionada, quanto maior o número de variáveis (que denotaremos genericamente de  $p$ ), maior o poder de discriminação do modelo. Porém, na prática, em situações onde há uma quantidade excessiva de variáveis e uma quantidade limitada de dados para treino, ocorrem diversos problemas. O principal é o *overfitting*. O *overfitting* ocorre quando o poder de previsão do modelo dentro da amostra de treino fica muito próximo a 100%, o que geralmente acarreta um poder preditivo muito limitado para novas observações. Isto ocorre principalmente devida a inclusão no modelo de variáveis não importantes e/ou variáveis apresentando multicolinearidade (YU e LIU, 2004). Este problema é típico quando temos o tamanho amostra  $n$  muito menor que o número de covariáveis disponíveis ( $p$ ).

Buscando minimizar esses problemas, a seleção de variáveis é um conjunto de ferramentas que visa excluir variáveis que não possuam importância (de acordo com alguma métrica) na predição através de um modelo previamente ajustado (JAMES, WITTEN, *et al.*, 2015). Segundo Yu e Liu (2004), quando bem realizada, a seleção de variáveis aumenta a eficiência do *machine learning* e a acurácia da predição enquanto reduz a complexibilidade dos modelos resultantes. Em um contexto de planejamento de experimento, muitas vezes temos um número grande de variáveis que poderiam ser importantes para o experimento, mas cuja coleta pode ser cara, inviável, ou demorada demais. Nestes casos a seleção de variáveis é fundamental na delimitação de quais variáveis serão levados em conta na hora da realização do experimento. Para isso, tipicamente coletamos uma amostra piloto e a partir desta amostra selecionamos as variáveis de interesse.

Em função da seleção de variáveis depender diretamente do algoritmo escolhido, cada um dos modelos testados apresentará o seu próprio ranqueamento de importância das variáveis, que depende ainda da métrica a ser utilizada para determinar o conceito de importância em si. Segundo Blum e Langley (1997), há diversas definições de importância da variável no modelo dentre a literatura de *machine learning*. Isso ocorre porque diferentes definições podem ser mais apropriadas em relação aos objetivos da modelagem. Blum e Langley (1997) apresenta cinco definições: importante para o objetivo; fortemente importante para amostra/distribuição; fracamente importante para amostra/distribuição; importância como medida de complexidade; utilidade incremental. Sendo esta última, especialmente natural para algoritmos baseados em seleção *stepwise*, por exemplo em *forward* e *backward selection* (BLUM e LANGLEY, 1997).

### 3.3 Estatística aplicada à seleção de variáveis

Para realizar a seleção de variáveis, existem diversos modelos que podem ser utilizados. Antes de definir quais modelos serão considerados, é necessário conhecer os dados, os problemas que os mesmos podem apresentar bem como o objetivo da modelagem. Desta forma, será possível escolher alguns modelos que atendam essas necessidades e realizar as modelagens para classificação da importância das variáveis estudadas (POZO, 2018).

Em aplicações modernas de *machine learning*, tipicamente se está interessado em situações onde há muito mais preditoras  $p$  que o tamanho da amostra  $n$ , ou seja,  $p \gg n$ . Porém, segundo Zhang (2011), quando isso ocorre em modelos de predição linear que utilizam minimização do risco empírico, a solução do problema de minimização pode não ser única. Em uma série de vetores  $\mathbf{x}_1, \dots, \mathbf{x}_n$  pertencentes a um espaço  $R^p$ , com saídas correspondentes  $y_1, \dots, y_n$ , o papel das tarefas de aprendizado supervisionado é estimar uma relação do tipo  $y \sim f(\mathbf{x})$  entre as entradas  $\mathbf{x}$  e as variáveis de saída  $y$ . A qualidade da predição pode ser medida através de uma função de perda  $\varphi(f(\mathbf{x}), y)$  baseada na performance do modelo na amostra de treinamento. O contexto mais utilizado na prática é o modelo de regressão linear  $f(\mathbf{x}) = \beta^T \mathbf{x}$  com função de perda quadrática. Sendo assim, o problema de minimização se torna (ZHANG, 2011) obter

$$\hat{\beta} = \arg \min_{\beta \in R^p} \left\{ \sum_{i=1}^n (\beta^T \mathbf{x}_i - y_i)^2 \right\}. \quad (1)$$

Em geral, são utilizadas condições de regularização de  $\beta$  para solucionar o problema de overfit, o que resulta na restrição do espaço de soluções do problema (ZHANG, 2011). O método de multiplicadores de Lagrange pode ser útil neste contexto. Neste caso, a solução do problema se torna uma minimização com restrição:

$$\hat{\beta} = \arg \min_{\beta \in R^p} \left\{ \sum_{i=1}^n (\beta^T \mathbf{x}_i - y_i)^2 + \lambda g(\beta) \right\}, \quad (2)$$

para alguma condição de regularização  $g(\beta)$ , onde  $\lambda > 0$  é um parâmetro de ajuste.

Entre os métodos de seleção de variáveis, estão os métodos de aplicação de algoritmos *greedy* (algumas vezes traduzidos como algoritmos gulosos ou míopes em português, mas utilizaremos o termo em inglês aqui) *forward* e *backward*. Algoritmos do tipo *forward greedy*, inicializa-se o procedimento com o modelo sem nenhuma preditora. A cada passo o algoritmo determina uma preditoras que é adicionada ao modelo, até que algum critério de parada seja satisfeito. Para a adição das preditoras, o algoritmo analisa e compara-as utilizando algum critério específico, como, por exemplo, o erro quadrático de previsão ou *Akaike Information Criterion* (AIC), até que todas as variáveis presentes no modelo sejam significativas (POZO, 2018). A vantagem deste método é a possibilidade de ser utilizado em situações onde há mais preditoras  $p$  que o tamanho da amostra  $n$  (JAMES, WITTEN, *et al.*, 2015). Porém, Hastie, Tibshirani e Friedman, (2016) destacam que este método possui uma grande limitação: a impossibilidade de correção das preditoras do modelo após terem sido adicionadas pelo algoritmo.

Outro algoritmo *greedy* utilizado é baseado na ideia de *backward selection*, que, ao contrário do método exposto acima, inicializa o procedimento de seleção com o modelo completo, isto é, com todas as preditoras. A cada iteração vai-se removendo uma a uma as preditoras, conforme critério específico (AIC ou erro quadrático por exemplo) até que algum critério de parada seja satisfeito (HASTIE, TIBSHIRANI e FRIEDMAN, 2016). Há duas grandes desvantagens neste método. A primeira é o alto custo computacional pois começamos já com o modelo mais complexos e vamos reduzindo a complexidade a cada passo. A segunda é que não é possível utilizar quando  $p > n$ . Porém este método permite corrigir qualquer erro na determinação das variáveis preditoras que estão no modelo, em claro contraste com a *forward greedy selection* (HASTIE, TIBSHIRANI e FRIEDMAN, 2016).

Uma representação gráfica de um modelo com quatro variáveis destes métodos pode ser vista na Figura 1, com o método *greedy forward* partindo da ponta esquerda e o *greedy backward* partindo da direita. As lacunas em branco representam um espaço vazio de variáveis e os círculos pretos um espaço preenchido, com cada nível à direita contendo um espaço preenchido a mais que o nível respectivamente à esquerda. Os traços representam o caminho que os métodos percorrem até encontrar as variáveis que melhor definem, conforme os critérios especificados, o modelo preditivo buscado (BLUM e LANGLEY, 1997).

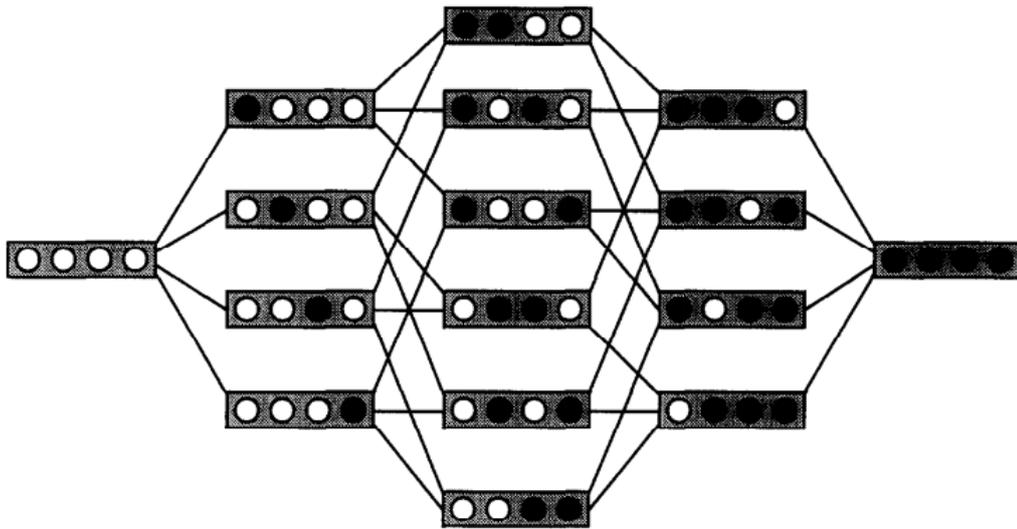


Figura 1 - Representação gráfica dos métodos *greedy forward* e *backward* (BLUM e LANGLEY, 1997)

As maiores vantagens de se utilizar um modelo *forward* são o fato desta técnica sempre funcionar para soluções esparsas (ZHANG, 2011), ou seja, aquelas onde poucas variáveis possuem importância relevante, e em situações onde  $p \gg n$  de maneira eficiente do ponto de vista de custo computacional. Porém, sua maior desvantagem é a de não poder corrigir erros cometidos pelo algoritmo em passos anteriores. Modelos *greedy backward* podem corrigir estes erros, mas deve-se partir de um bom modelo para que não overfit os dados, ou seja, situações onde  $p < n$  (POZO, 2018).

Para solucionar esses problemas, Zhang (2011) propôs a metodologia chamada *Adaptive Forward-Backward Greedy Algorithm* (FoBa), um algoritmo *greedy* que combina as forças das técnicas de *Forward* e *Backward* minimizando as suas limitações. FoBa é um algoritmo *greedy* que utiliza a técnica de seleção *forward* para desenvolver a modelagem e a técnica de eliminação *backward* para corrigir os erros que a seleção *forward* não corrige. O algoritmo apresentado por Zhang (2011) tem como entradas as covariáveis e o vetor resposta bem como  $\epsilon > 0$  que determina o limite da diferença entre o erro quadrático entre o passo atual e o anterior. A saída é o modelo ajustado aos dados de entrada.

O primeiro passo realizado pelo algoritmo é selecionar, entre todas as variáveis, a que minimiza o erro quadrático do modelo preditivo buscado utilizando apenas uma variável. A partir disso, o algoritmo continua a desenvolver a técnica *forward*, sempre selecionando a variável que minimiza o erro quadrático, até que o aumento do erro quadrático de previsão, não seja maior que metade da redução do erro quadrático do passo *forward* anterior. Isso faz com que apenas sejam utilizados passos *backward* que corrijam de forma construtiva os erros gerados pelos passos *forward*. Desta maneira, o número total de passos *backward* nunca será

maior que o número de passos *forward*, tornando o processo computacionalmente eficiente (ZHANG, 2011). O algoritmo detalhado pode ser encontrado em Zhang (2011). Uma implementação deste algoritmo em R pode ser encontrada no pacote *foba*.

## 4 APLICAÇÃO AO RUGBY

*Rugby Union* é um esporte coletivo com bola oval praticado mundialmente e gerido pela World Rugby. Destaca-se por seus valores: Integridade, Disciplina, Respeito, Paixão e Solidariedade. Esses valores devem ser aplicados dentro e fora do campo em todas as instâncias: do escolar ao profissional (WORLD RUGBY, 2014a). O objetivo do jogo é apoiar a bola na zona de *in-goal* de ataque. Para alcançá-lo, utiliza-se estratégias de invasão e evasão.

As duas principais modalidades são XV e Sevens, sendo a última o alvo deste artigo. No XV são 15 jogadores titulares em cada equipe. Já, no Sevens, são 7 titulares em cada equipe. Outra grande mudança entre as modalidades é o tempo de jogo: no XV são 2 tempos de 40min e, no Sevens, é 7min cada tempo. Há diversas outras alterações de Lei de jogo e estratégicas que podem ser encontradas no Apêndice 1, que também apresenta contextualização histórica do esporte, Leis e arbitragem.

### 4.1 Dados

Para a coleta da amostra piloto foram sorteados 29 jogos dos Circuitos Mundial de *Rugby Sevens* Feminino 2014/2015; 2015/2016 e 2017/2018. A temporada pós-olímpica, 2016/2017, não está disponível online. A transmissão foi realizada pelo site da World Rugby, mas os links não são mais válidos.

Os 29 jogos foram sorteados através do software R: um (1) jogo para cada uma das fases da etapa, grupos<sup>4</sup> x confrontos diretos<sup>5</sup>. A Tabela 3 do Apêndice 2 apresenta os jogos analisados discriminando a temporada, a etapa, a fase e o resultado. Para a Seleção de Variáveis, apenas o resultado foi levado em consideração.

Nas temporadas 2014/2015 e 2015/2016, as Finais foram excluídas do circuito por terem duração de tempo muito superior aos demais: cada tempo tinha 10min ao invés de 7min. Segundo Barkell, O'Connor e Cotton (2016), incluir estas Finais aumentaria potencialmente os valores coletados. Após estas temporadas, as Finais passaram a ter 7min

---

<sup>4</sup> 18 jogos

<sup>5</sup> 15 jogos nas temporadas 2014/2015 e 2015/2016 e 16 jogos na temporada 2017/2018

cada tempo e passaram a estar disponíveis nos sorteios. Outros jogos foram excluídos por falhas técnicas de transmissão e imagem dos vídeos conforme a Tabela 4 do Apêndice 2.

Para a definição das variáveis a serem inicialmente coletadas e também das que seriam analisadas, realizou-se uma revisão de artigos sobre indicadores de performance no *Rugby*. Para isto, efetuou-se pesquisas utilizando-se os termos “*Rugby*”, “*Sevens*”, “*Women*”, “*indicators*”, “*performance*”, “*KPI*” e combinações dos mesmos no site da Capes em 25 de fevereiro de 2019. Após a pesquisa, analisou-se quais artigos estavam relacionados apenas ao desempenho técnico e/ou tático. Como o retorno foi baixo<sup>6</sup>, buscou-se referências dentro de trabalhos previamente conhecidos pela autora, como o de Silveira (2018); o de Vaz (2012) ; o de Ortega, Villarejo e Palao (2009). Destes, selecionou-se 13 artigos e trabalhos acadêmicos para a revisão: 6 de *Rugby XV* masculino, 5 de *Rugby Sevens* masculino, 1 de *Rugby Sevens* masculino e feminino e 1 de *Rugby Sevens* feminino. Os artigos de *Rugby XV* são os mais antigos, com exceção de Watson, Durbach, *et al.* (2017), que é o artigo mais recente. Analisou-se os indicadores ou variáveis pesquisadas e definiu-se pelo desmembramento dos principais indicadores em variáveis que possibilitam diversas análises de indicadores diferentes.

## 4.2 Coleta

A coleta foi realizada durante o período de março a maio de 2019 seguindo os preceitos apresentados por Marques (2018) em relação a análises de vídeo. A metodologia utilizada foi notacional, descritiva, quantitativa e qualitativa. A coleta das variáveis para a Seleção de Variáveis ocorreu de forma manual em formulário construído através de variáveis quantitativas e qualitativas. Após, as variáveis quantitativas foram transcritas para uma base de dados. Em situações de jogo aberto, a coleta é subjetiva e foi realizada utilizando-se as definições apresentadas na Tabela 5 e na Tabela 6 do Apêndice 3 e nas experiências da autora como treinadora, analista de vídeo e árbitra de *Rugby Sevens* feminino. As coletas referentes às situações de vantagens e de penalidades foram anotadas conforme a sinalização da arbitragem. Os tempos foram coletados através do programa de análise de vídeos *Rugby Assistant*.

## 4.3 Variáveis

---

<sup>6</sup> Apenas dois dos 22 que apareceram na pesquisa resultante de “*Rugby*” and “*Sevens*” and “*Women*” and “*Performance*”. São os mesmos dois que foram destacados na pesquisa apresentada na **Erro! Fonte de referência não encontrada.**

Segundo Higham, Hopkins, *et al.* (2014b), o ambiente de jogo dinâmico pode tornar difícil para treinadores e comissão técnica definirem quais elementos físicos, técnicos e táticos devem ser o alvo do desenvolvimento para aumentar as probabilidades de vencer o jogo. Os conceitos e quais variáveis devem ser estudadas ainda é campo aberto dentro da pesquisa em *Rugby*. Porém, Vaz (2012); Ortega, Villarejo e Palao (2009) defendem que é necessário analisar o jogo como um todo: jogo aberto e suas ações; formações fixas e seus resultados; situações que resultam em pontuação; disciplina; variáveis de resultado. Para Watson, Durbach, *et al.* (2017), um dos problemas dos estudos de indicadores de performance em *Rugby* é que utilizam uma quantidade pequena de variáveis e/ou indicadores sem uma análise mais profunda sobre a seleção destas variáveis e/ou indicadores. Desta forma, analisou-se 112 variáveis, que estão definidas na Tabela 5 do Apêndice 3. Situações de *Maul* foram excluídas, pois, como destaca Silveira (2018), não é uma ação muito utilizada em *Rugby Sevens*.

Para Ortega, Villarejo e Palao (2009) as equipes vencedoras em *Rugby XV* tem mais efetividade em marcar *Try* ao entrarem nas 22 de ataque. Silveira (2018); Higham, Hopkins, *et al.* (2014a) apontam que esta hipótese também é válida para o *Sevens* além de destacarem as Conversões acertadas. Portanto, determinou-se a coleta das variáveis a cada situação de entrada nesta zona, incluindo-se, de forma descritiva, quais situações originaram cada entrada. Podendo-se, assim, estudar posteriormente se há situações específicas de entrada nas 22 de Ataque que resultam em *Try* e Conversão entre outras situações desta zona.

#### 4.4 Seleção de Variáveis

O primeiro passo no tratamento dos dados após sua coleta foi a remoção de variáveis com variância zero ou muito próxima de zero. Variáveis que não apresentam variabilidade são obviamente inúteis do ponto de vista de modelagem estatística. Além disso é bem conhecido que variáveis com variabilidade próxima a zero também tornam a seleção de modelos problemáticas de forma que é boa prática excluí-las do banco (KUNH, 2008). Um detalhe fundamental é como definiremos se uma variável que assume poucos valores únicos deve ser excluída do modelo. Tipicamente, uma variável é excluída do modelo por ter baixa variabilidade se apresenta as seguintes características:

- a) A razão entre frequência do valor que mais se repete e a frequência do segundo mais frequente estar acima de algum ponto de corte, que utilizamos 95/5;
- b) O percentual de valores únicos estar abaixo de 10 (KUNH, WING, *et al.*, 2019).

Estes valores são valores de referência e podem ser alterados de acordo com a aplicação (KUHN, WING, *et al.*, 2019).

Para o diagnóstico de variáveis com variância zero ou quase-zero utilizamos a função **nearZeroVar** do R. Desta forma, o número de variáveis foi reduzido de 112 para 52. Após, foi realizada a padronização (subtração da média e divisão pelo desvio padrão da variável) dos dados. Depois de padronizados, subtraiu-se os resultados e as covariáveis dos vencedores e dos perdedores, jogo a jogo. Este passo é resultado das limitações características da análise deste tipo de torneio e de mais de uma temporada do mesmo torneio:

- a) Há diversos jogos das mesmas equipes individualmente e também os mesmos cruzamentos<sup>7</sup>. Isto ocorre porque a cada temporada são jogadas de 4 a 6 etapas com as mesmas 12 equipes com pouca renovação entre temporadas.
- b) Para cada vencedor haverá um perdedor e vice-versa. Portanto, é necessário parear os dados.

Após esta etapa, dividimos a amostra em treino e validação na proporção aproximada de 75/25, sendo que a amostra de treino ficou com 25 jogos e a de validação com 4 jogos. A divisão dos jogos foi feita por amostragem aleatória simples. Para garantir a reprodutibilidade do trabalho, utilizamos uma semente aleatória predefinida (utilizando o **set.seed** do R). A divisão em amostra de treino e validação foi feita utilizando a função **createDataPartition** do pacote *caret* do R.

Após a divisão dos jogos da amostra em treino e em validação, passamos para a etapa de treino dos modelos na amostra de treino. Para comparar resultados, utilizamos 3 modelos diferentes para a seleção de variáveis. O primeiro foi o método FoBa explicado na seção 3.3. Utilizou-se ainda os métodos *Bayesian Generalized Linear Model* (Bayes GLM) e SVM (máquinas de vetores de suporte, na sigla em inglês).

O método Bayes GLM é uma técnica baseada em estatística bayesiana (ROBERT, 2007) para modelos GLM. Uma apresentação detalhada do modelo foge dos objetivos do trabalho, mas podem ser encontrada em Gelman, Jakulin *et al.* (2008). Ao leitor interessado em estatística bayesiana, sugerimos o livro Robert (2007) enquanto o livro McCullagh e Nelder (1997) é recomendado a quem possui interesse em modelos GLM. O método Bayes GLM é implementado no pacote **arm** do R.

O SVM é um método de aprendizado supervisionado inicialmente desenvolvido para resolver problemas de classificação a partir de restrições lineares (hiperplanos) em seus

---

<sup>7</sup> Tanto na fase de grupos quanto no confronto direto

espaços de origem. Isto é, dado um conjunto de pontos em um espaço, digamos que cada ponto é marcado como pertencente a uma dentre duas categorias. O SVM é um classificador que, a partir dos dados, encontra um remapeamento do espaço original de forma que os pontos representados neste espaço remapeado podem ser separados da melhor maneira possível utilizando hiperplanos (representado por vetores de suporte) (HASTIE, TIBSHIRANI e FRIEDMAN, 2016).

É importante lembrar que um hiperplano num espaço vetorial de dimensão  $p$  é qualquer subespaço de dimensão  $p - 1$  deste espaço. Em particular um hiperplano em  $R^2$  é uma reta enquanto um hiperplano em  $R^3$  é um plano. Intuitivamente, dado um novo ponto no espaço, este ponto é classificado em um grupo ou outro baseado no remapeamento deste nos vetores de suporte e sua posição em relação ao hiperplano de separação obtido a partir da amostra de treinamento (HASTIE, TIBSHIRANI e FRIEDMAN, 2016). O método pode ser adaptado para lidar com problemas de regressão, mas o detalhamento da técnica foge ao escopo deste trabalho. Ao leitor interessado, sugerimos Drucker *et al.* (1997). Problemas de classificação muitas vezes podem falhar quando os grupos não são separáveis via hiperplanos. Neste caso o problema de classificação é não-linear e a utilização da ideia de SVM vai depender da utilização de um núcleo, que é uma função responsável por linearizar o problema. A linearização de um problema de classificação utilizando-se núcleos é chamado de “truque do núcleo” (*kernel trick*, veja CLARKE, FOKOUÉ e ZHANG, 2009). O núcleo radial, também conhecido como núcleo Gaussiano, é o mais utilizado para este fim e é dado por

$$K(x_i, x_j) = \exp\left(-\frac{|x_i - x_j|^2}{2\sigma^2}\right),$$

para  $x_i$  e  $x_j \in R^p$ .

Utilizando os três modelos mencionados (FoBa, bayesGLM e SVM-Radial) precisamos ainda determinar uma métrica de comparação entre eles. Neste caso definiu-se como modelo mais adequado aquele que tinha tanto o erro quadrático médio quanto o erro absoluto médio de previsão menor que os demais. Lembrando que o ajuste do modelo é feito na amostra de treino enquanto o poder preditivo do modelo ajustado é testado na amostra de validação. Para o ajuste dos modelos utilizamos a função genérica **train** do pacote *caret* do R. Para a predição combinamos a função genérica **predict** do R e **extractPrediction** do pacote *caret*.

Dentre os modelos ajustados, o modelo escolhido para a seleção das variáveis foi o FoBa por apresentar menor erro quadrático médio e menor erro absoluto médio conforme apresentado no Gráfico 1 e no Tabela 1.

Tabela 1 - Comparativo entre os modelos testados

	FoBa	bayesglm	svmradial
MSE	0.018166	0.084495	0.239456
MAE	0.103239	0.283766	0.398869

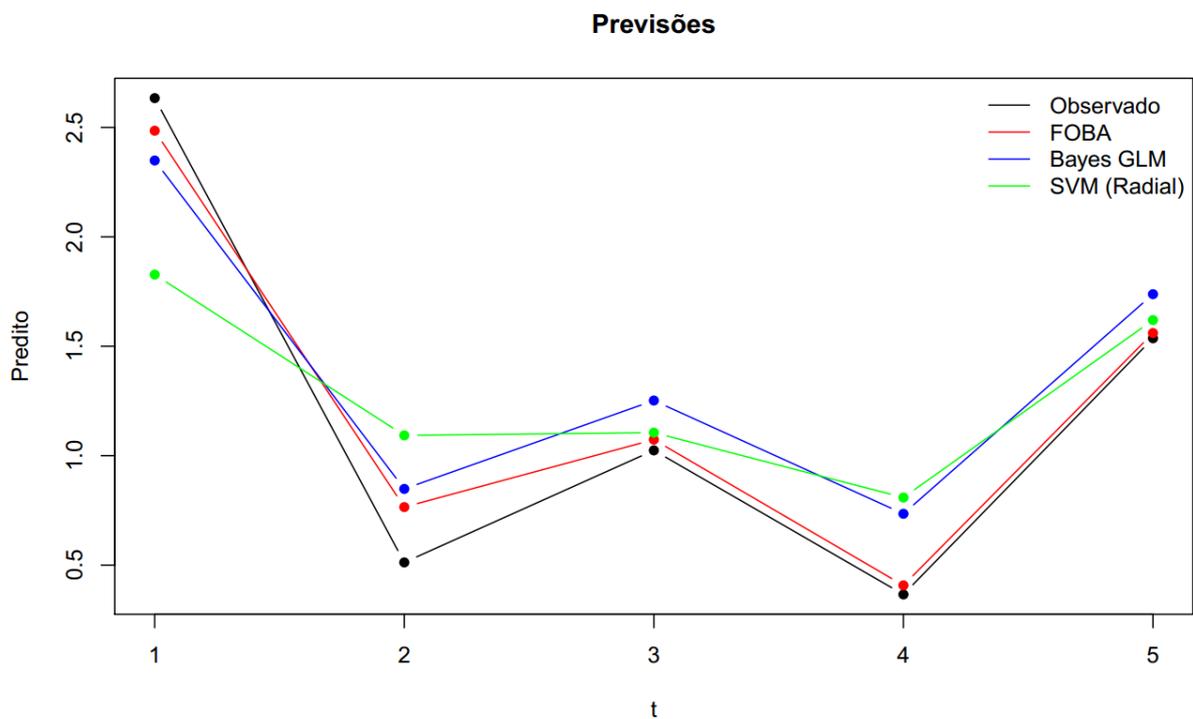


Gráfico 1 - Comparação entre a validação dos modelos

Após a definição do modelo, passamos a etapa de determinação da importância de cada uma das variáveis no modelo. Para isso utilizou-se a função **varImp**, que é um método genérico para calcular a importância das variáveis produzidas nos treinos. Para o caso específico do FoBa, a função **varImp** o valor absoluto da estatística  $t$  de cada parâmetro é utilizada para a determinação da importância de cada parâmetro.

## 5 DISCUSSÃO DOS RESULTADOS

Para essas análises, foram utilizadas as 52 variáveis selecionadas após o tratamento dos dados conforme explicado na Aplicação ao Rugby. A Tabela 2 apresenta a listagem das variáveis conforme seu grau de importância modelado pelo método FoBa.

**Tabela 2 - Variáveis analisadas divididas por grau de importância e apresentadas de forma decrescente**

<b>Variáveis Importantes</b>	<b>Importância</b>	<b>Demais variáveis</b>	<b>Importância</b>
Try	0.98026	Tackle.Eft.Individual	0.14645
Ent.Pos	0.96674	Tackle.NEft.Duplo	0.13906
Convers	0.81540	Tackle.Perdido.Duplo	0.13818
T.Ent	0.80367	Tackle.SP.ind	0.13223
ATQ22_Defesa0	0.62559	LIN_22D.Ganhos	0.08577
Passe.bom	0.39835	ID.Ganhos.1	0.07163
Ruck.e.passe	0.38580	LIN_IA.Ganhos	0.06670
Tackle.NEft.Individual	0.35789	SCR_ID.Inf.Sof	0.05435
Passe.ok	0.35752	Tackle.Eft.Triplo	0.03706
Tackle.SP.duplo	0.34003	SCR_22D.Ganhos	0.03052
ATQ22_Posse	0.33528	SCR_22A.Ganhos	0.02017
ATQ22_Ação.Pos	0.33250	LIN_22A.Ganhos	0.01845
Tackle.Eft.Duplo	0.32874	SCR_5D.Ganhos	0.01734
ATQ22_Defesa.Pos	0.32227	LIN_22D.Perdidos	0.01245
Ruck.e.Drive	0.28360	Ruck.Pick...go	0.01236
Passe.para.trás	0.27388	Ruck.perdido	0.01224
ATQ22_Defesa.Neg	0.27380	ID.Ganhos	0.01082
Tackle.Perdido.Individual	0.27295	TO.passe	0.00391
Knock.on	0.25675	SCR_I.Ganhos	0.00371
Passe.erro..alvo	0.25644	Ruck.erro.Posse	0.00147
Passe.ruim	0.24687	LIN_I.Ganhos	0.00125
Tackle.Penal.Individual	0.24654	LIN_ID.Perdidos	0.00010
Ent.Neg	0.24436	Passe.para.frente	0.00000
ATQ22_Ação.Neg	0.24140	Ruck.e.TO	0.00000
SCR_IA.Ganhos	0.24033	Ent.neutras	0.00000
ATQ22_Erros	0.20357		
SCR_5A.Ganhos	0.20353		

Utilizando como corte 0.2 de importância, obteve-se 27 variáveis, conforme Tabela 2 e Gráfico 2. O corte em 0.2 é arbitrário. Confirmando estudos como Silveira (2018); Barkell, O'Connor e Cotton (2016), *Try* foi a variável de maior importância para as equipes vencerem os jogos. Conversões, como apontado por Higham, Hopkins, *et al.* (2014a); Silveira (2018) também tem grande importância na definição dos vencedores. Isto ocorre principalmente em jogos equilibrados, pois acabam sendo o que diferencia o perdedor do vencedor. Além disso,

esta variável pode estar sendo destacada, pois equipes de ponta, como Austrália e Nova Zelândia, tendem a acertar a maioria das conversões (WORLD RUGBY, 2019g).

Silveira (2018); Barkell, O'Connor e Cotton (2016) destacam que, em jogos de *Sevens* feminino, não há muitas formações de lateral e de *Scrum*. Apesar disso, as variáveis *Scrum* na Intermediária de ataque e nas 5 de ataque são variáveis importantes quando ganhos. Ambas estão entre as últimas antes do corte, mas apontam que, nestas situações, podem ocorrer *Try* em função da proximidade do *In-goal* adversário e em função das estratégias adotadas.

Como discutido por Ortega, Villarejo e Palao (2009); Silveira (2018); Higham, Hopkins, *et al.* (2014a), as equipes que mais entram nas 22 são as que tem mais chances de vencer. A análise realizada aponta esta mesma realidade tendo como segunda variável mais importante as “Entradas positivas nas 22”, ou seja, entradas que resultaram em *Try*. A “Total de entradas nas 22” foi a quarta em importância, ficando em aproximadamente 0.8 na escala de 0 a 1.

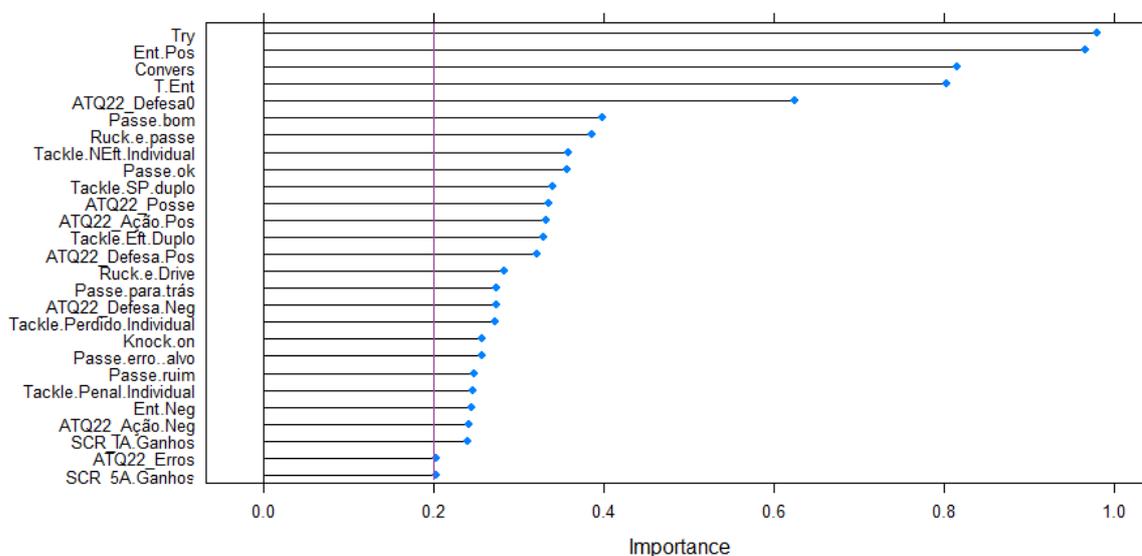


Gráfico 2 - Principais Variáveis por Importância (FoBa)

O alto grau de importância destas duas variáveis relacionadas à entrada das 22 de ataque além do aparecimento positivo<sup>8</sup> da variável Entradas Negativas, ou seja, que ocorreu perda da posse de bola, apontam que é importante o desenvolvimento de táticas de jogo que levam a realizar o maior número de entradas possíveis, buscando não perder a posse de bola. Da mesma forma, ações de ataque<sup>9</sup> e ações da defesa dentro das 22 foram destacadas entre as

<sup>8</sup> com mais de 0.2 de importância

<sup>9</sup> Positiva ou Negativa, mantendo ou não a posse

27 variáveis mais importantes. As ações de defesa neutra e de defesa negativa dentro das 22 são variáveis de importância negativa, ou seja, quanto maior menor a probabilidade de vitória. Já ações de defesa positiva dentro das 22 também aparecem positivamente entre as 27 variáveis importantes, mas seu grau de importância é a metade do da defesa neutra, que é negativa.

Das 27 variáveis selecionadas, 10 estão diretamente relacionadas à zona das 22 de ataque. Destas, 3 estão no Top 5. Estas informações indicam que, assim como no masculino, identificado por Higham, Hopkins, *et al.* (2014a), as ações realizadas dentro da zona das 22 de ataque são fundamentais para a determinação do vencedor.

Em relação a ações de ataque, os passes bom e ok são os que mais importam positivamente. Passes com erro de alvo e ruim também estão entre as variáveis positivas, mas com menor grau de importância, ficando entre 0.2 e 0.3. Já “*Knock-on*” e “*passo para trás*” são variáveis com importância negativa, também com importância entre 0.2 e 0.3. Os dados relacionados a passes demonstram a atratividade de aperfeiçoar essa ação de forma coletiva, pois as variáveis de passe (bom, ok e ruim) são variáveis de análise coletiva e levam em consideração a interação entre passador e receptor.

Os pontos-chave técnicos do movimento passe são analisados e ensinados separadamente dos do movimento recepção de passe, mas em jogo ocorrem sequencialmente e são diretamente relacionados. Por isso, é importante entender como as equipes evoluem coletivamente nestes movimentos. Além disso, os dados relacionados a passe indicam que, taticamente, o importante é a manutenção da posse de bola.

Já nas saídas de *Ruck* ganhos em ataque, as ações relevantes são a saída com passe, acelerando o jogo, e a saída com drive, buscando um *Line break*. Ambas tem uma importância baixa entre as 27, ficando a primeira abaixo de 0.4 e a segunda aproximadamente em 0.3. Com as mudanças de regra para esta temporada, a tendência é diminuir a quantidade de *Rucks* nos jogos.

Em defesa, *Tackles* duplos efetivos, *Tackles* especiais duplos e *Tackles* não-efetivos individuais são as variáveis com mais relevância positiva. Enquanto *Tackles* perdidos individuais e Penais causado em *Tackles* individuais são variáveis de relevância negativa. Porém, com baixa importância, ficando ambas pouco acima de 0.2. *Tackles* duplos especiais são *Tackles* que iniciaram individuais e terminaram como um *Tackle* duplo efetivo. Isso demonstra que, no Sevens, a melhor tática é utilizar *Tackle* duplo apenas quando realmente necessário, como em casos de a “*tackleadora*” inicial não conseguir parar a bola.

Algumas variáveis apresentam multicolinearidade, como Try e Entrada positiva nas 22, pois é a definição de entrada positiva nas 22 é toda a entrada nas 22 que resultou em try e só é possível marcar try entrando nas 22. Porém, isto não é problema para esta etapa da pesquisa, visto que o objetivo do trabalho é a seleção das variáveis mais importantes para o estudo do problema proposto.

## 6 CONCLUSÃO E TRABALHOS FUTUROS

A ciência da informação está em constante evolução, assim como as ciências atuariais. Num mundo onde a informação disponível é abundante, é fundamental para o profissional das atuariais saber discernir informação útil de informação inútil. Neste trabalho apresentamos alguns métodos de *machine learning* para a seleção de variáveis em problemas de regressão, cuja aplicabilidade transcende a área de ciências atuariais. Mostramos a conexão das técnicas com os mais diversos problemas de interesse das ciências atuariais e apresentamos uma aplicação na seleção de variáveis em Rugby Sevens.

O presente trabalho apresentou resultados satisfatórios e alcançou o objetivo de selecionar variáveis em Rugby Sevens Feminino para análises futuras. Foi possível selecionar 27 variáveis importantes utilizando como corte 0.2 de importância e compreender a relação delas com a construção das vitórias e derrotas de forma geral pelas equipes. Dessas, as relacionadas a *Scrum* e a de *Tackle Duplo Especial* foram as grandes surpresas dentro das importantes, pois *Scrum* não costuma ocorrer com frequência e as variáveis de *Tackle Especial* foram criadas pela autora, não tendo sido pesquisadas anteriormente. Além disso, não tentar tacklear, mesmo que errando, ao perseguir ou marcar dentro das 22 de defesa também não era esperado estar no Top 5.

Para a continuação da pesquisa, serão realizadas análises de vídeo das 52 variáveis apresentadas no Apêndice 3 em amostra das temporadas 2014/2015; 2015/2016; 2017/2018 e 2018/2019. Os jogos da última temporada já estão disponíveis no youtube da World Rugby. Porém, as 27 variáveis selecionadas aparentam ser realmente as de destaque, podendo já ser utilizadas em treinamentos e novas pesquisas. Após, será realizado novamente esse procedimento de seleção para verificar se ocorrerá mudanças nas variáveis de maior importância, mantendo-se o corte em 0.2.

## Bibliografia

BARKELL, J. F.; O'CONNOR, D.; COTTON, W. G. Characteristics of winning men's and women's sevens rugby teams throughout the knockout Cup stages of international tournaments. **International Journal of Performance Analysis in Sport**, Sydney, 2016.

BARRIS, S.; BUTTON, C. A Review of Vision-Based Motion Analysis in Sport. **Sports Med**, p. 1025-43, 2008.

BLUM, A. L.; LANGLEY, P. Selection of relevant features and learning. **Artificial Intelligence**, v. 97, n. 1-2, p. 245-271, Dezembro 1997. ISSN 0004-3702.

CAPES/MEC. Busca. **Portal de Periódicos CAPES/MEC**, 2019. Disponível em: <<https://www.periodicos.capes.gov.br/>>. Acesso em: 25 fevereiro 2019.

CODECADEMY. Training Set vs Validation Set vs Test Set. **Codecademy**. Disponível em: <<https://www.codecademy.com/articles/training-set-vs-validation-set-vs-test-set>>. Acesso em: jul. 2019.

COMITÊ OLÍMPICO INTERNACIONAL. Olympic History. **Olympic**, 2019. Disponível em: <<https://www.olympic.org/rugby>>. Acesso em: 10 março 2019.

CONFEDERAÇÃO BRASILEIRA DE RUGBY. Brasil classificado! Seleção Feminina de Rugby Sevens garante vaga nos Jogos Olímpicos de Tóquio ao vencer qualificatório. **Brasil Rugby**, 2019a. Disponível em: <<https://ww2.brasilrugby.com.br/blogs/news/brasil-classificado-selecao-feminina-de-rugby-sevens-garante-vaga-nos-jogos-olimpicos-de-toquio-ao-vencer-qualificatorio>>. Acesso em: 13 junho 2019.

CONFEDERAÇÃO BRASILEIRA DE RUGBY. Calendário Oficial. **Portal de Governança e Transparência**, 2019b. Disponível em: <[https://cdn.shopify.com/s/files/1/2151/7049/files/190130\\_calendario\\_de\\_jogos.pdf?15896577562769383336](https://cdn.shopify.com/s/files/1/2151/7049/files/190130_calendario_de_jogos.pdf?15896577562769383336)>. Acesso em: 13 junho 2019.

CONFEDERAÇÃO BRASILEIRA DE RUGBY. Relatório Anual. **Portal de Governança da CBRU**. Disponível em: <<https://brasilrugbygovernanca.com.br/pages/relatorio-anual>>. Acesso em: 25 Fevereiro 2019.

CORTES, C.; VAPNIK, V. Support-Vector Networks. **Machine Learning**, v. 20, p. 273-297, 1995.

DE SOUZA, P. D. B.; DE MATTOS, L. L.; DE SOUZA, M. A. B. Marketing Esportivo e sua Relação com Clubes e Instituições Ligadas ao Esporte, 2005.

DIANA, A. et al. **Machine-Learning Methods**. [S.l.]: Society of Actuaries (SOA), 2019.

DRUCKER, H. et al. Support Vector Regression Machines. **Advances in Neural Information Processing Systems 9**, p. 151-161, 1996.

DUMA, M. et al. Improving the Performance of the Ripper in Insurance Risk Classification: A Comparative Study Using Feature Selection. **arXiv preprint arXiv:1108.4551**, 2011.

DUNN, P. K.; SMYTH, G. K. **Generalized Linear Models with Examples in R**. [S.l.]: Springer, 2018.

FRANKS, I. M.; GOODMAN, D. A systematic approach to analysing sports performance. **Journal of Sports Sciences**, Londres, v. 4, p. 49-59, janeiro 1986.

GELMAN, A. et al. A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models. **The Annals of Applied Statistics**, v. 2, n. 4, p. 1360-83, 2008.

GENOVEZ, G. Inovações, Mudanças e Seguros Gerais. **Revista Brasileira de Atuária**, Rio de Janeiro, v. 2, n. 2, p. 12-15, 2018.

HALL, P.; PHAN, W.; WHITSON, K. **The Evolution of Analytics**. 1. ed. Sebastopol: O'Reilly Media, 2016.

HALL, R. D. **Actuarial Risk Analysis using Predictive Analytics, Segmentation and Decomposition Techniques**. Society of Actuaries. [S.l.]. 2018.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning - Data Mining, Inference and Prediction**. 2. ed. New York: Springer, 2016.

HIGHAM, D. G. et al. Patterns of play associated with success in international rugby sevens. **International Journal of Performance Analysis in Sport** , 2014.

HIGHAM, D. G. et al. Performance Indicators Related to Points Scoring and Winning in International Rugby Sevens. **Journal of Sports Science and Medicine**, p. 358-364, maio 2014.

HIGHAM, D. G. et al. Relationships between rugby sevens performance indicators and international tournament outcomes. **De Gruyter**, v. 1, n. 10, p. 81-87, 2014.

INSTITUTO BRASILEIRO DE ATUÁRIA. O Atuário. **IBA - Instituto Brasileiro de Atuária**. Disponível em: <<https://www.atuarios.org.br/o-atuario>>. Acesso em: 2019.

INTERNATIONAL RUGBY BOARD. **Planejamento Rugby Sevens 2011-2020 IRB**. Dublin. 2010.

INTERNATIONAL RUGBY BOARD. **Rugby Sevens Plan 2011-2020**. International Rugby Board. Dublin, p. 51. 2010.

INTERNATIONAL RUGBY BOARD. **Curso de Árbitros de Rugby Nível I**. Tradução de João Nogueira; Maurício Migliano e Xavier Vouga. [S.l.]: [s.n.], 2012. 56 p.

JAMES, G. et al. **An Introduction to Statistical learning with Applications in R**. 6. ed. New York: Springer, 2015.

JENSEN, J.; COBBS, J. Analyzing Return-of-investment in Sponsorship: Modeling Brand Exposure, Price and ROI in Formula One Racing. **Journal of Advertising Research**, 2014.

JONES, N. M. P.; MELLALIEU, S. D.; JAMES, N. Team performance indicators as a function of winning and losing in rugby union. **International Journal of Performance Analysis in Sport**, v. 4, p. 61-71, Julho 2004.

KUHN, M. et al. Classification and Regression Training - caret Package - Version 6.0-84. **The Comprehensive R Archive Network**, 2019. Disponível em: <<https://cran.r-project.org/web/packages/caret/caret.pdf>>. Acesso em: 15 Maio 2019.

KUNH, M. Building Predictive Models in R Using the caret Package. **Journal of Statistical Software**, v. 8, n. 5, novembro 2008.

LACERDA, W. R. A Nova Revolução com o Big Data. **Revista Brasileira de Atuária**, Rio de Janeiro, v. 1, n. 1, p. 36-40, 2017.

MANO, C. Atuários como Cientistas de Dados ou Cientistas de Dados como Atuários? **Revista Brasileira de Atuária**, Rio de Janeiro, v. 2, n. 2, p. 23-27, 2018.

MARQUES, G. **Curso Análise de Vídeo em Rugby - FGR**. Novo Hamburgo: Federação Gaúcha de Rugby, 2018.

MCCULLAGH, P.; NELDER, J. A. **Generalized Linear Models**. 1. ed. [S.l.]: Chapman and Hall, 1987.

NAURIGHT, J. Rugby. **Enciclopaedia Britannica**, 2018. Disponível em: <<https://www.britannica.com/sports/rugby>>. Acesso em: 15 junho 2019.

OLIVEIRA, P. H. M. A. **Detecção de Fraudes em Cartões: Um Classificador Baseado em Regras de Associação e Regressão Logística**. São Paulo: USP, 2016.

ORTEGA, E.; VILLAREJO, D.; PALAO, J. M. Differences in game statistics between winning and losing rugby teams in the Six. **Journal of Sports Science and Medicine**, p. 523-27, Dezembro 2009.

POZO, A. T. R. **Curso: Análise e Seleção de Variáveis**. Curitiba: UFP, 2018.

RAMOS, R. M. **Identificação de Comunicado de Ocorrência ed Perdas em Seguro Agrícola Utilizando Algoritmos de Inteligência Artificial**. Brasília: UNB, 2011.

ROBERT, C. **The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation**. 2ª Edição. ed. [S.l.]: Springer Verlag, 2007.

RODRIGUES, C. N. M. et al. Evaluation of Machine Learning and Bag of Visual Words Techniques for Pollen Grains Classification. **IEEE Latin America Transactions**, v. 13, n. 10, p. 3498-3504, 2015.

ROSS, A. et al. Defensive And Attacking Performance Indicators In Rugby Sevens. **International Journal of Performance Analysis in Sport**, v. 16, n. 2, 2016.

SANSIGOLO ANDRELLO, G. **Rugby Sevens Brasileiro: Passe e Tackle**. Campinas: Universidade Estadual de Campinas, 2013. 48 p. Trabalho de Conclusão de Curso (Bacharelado em Educação Física)-Faculdade de Educação Física.

SILVEIRA, R. F. D. Caracterização das Ações de Jogo do Time Campeão Brasileiro do Campeonato Super Sevens de Rugby Feminino: Niterói Rugby. **Repositório Institucional da UFSC**, Florianópolis, Novembro 2018. Disponível em: <<https://repositorio.ufsc.br/bitstream/handle/123456789/192038/TCC%20Renata%20Flor%20222.pdf>>. Acesso em: 28 Fevereiro 2019.

SOCIETY OF ACTUARIES. **SOA - Society of Actuaries**. Disponível em: <<https://www.soa.org/>>. Acesso em: 2019.

VAN ROOYEN, M. K. A Statistical Analysis of Tackling Performance during International Rugby Union Matches from 2011. **International Journal of Performance Analysis in Sport**, v. 12, p. 517-30, 2012.

VAZ, L. et al. The importance of rugby game-related statistics to discriminate winners and losers at the elite level competitions in close and balanced games. **International Journal of Performance Analysis in Sport**, v. 11, p. 130-41, 2011.

VAZ, L. M. T. Ações do jogo e de resultado que melhor discriminam vitórias. **Rev. bras. Educ. Fís. Esporte**, São Paulo, v. v.26, p. p.111-17, Março 2012.

WATSON, N. et al. On the validity of team performance indicators in rugby union. **International Journal of Performance Analysis in Sport**, p. 609-621, 2017.

WORLD RUGBY. **Rugby Ready**. Dublin: World Rugby, 2014a.

WORLD RUGBY. The History of Rugby Sevens. **World Rugby**, 09 Novembro 2014b. Disponível em: <<https://www.world.rugby/sevens/history-of-sevens>>. Acesso em: 05 Junho 2019.

WORLD RUGBY. HSBC World Rugby Sevens Series, 2019a. Disponível em: <<https://www.world.rugby/sevens-series/>>. Acesso em: 10 março 2019.

WORLD RUGBY. World Series Standing - Women's. **HSBC World Rugby Sevens Series**, 2019a. Disponível em: <<https://www.world.rugby/sevens-series/standings/womens>>. Acesso em: 16 junho 2019.

WORLD RUGBY. **Laws of the Rugby Union**. Dublin: World Rugby, 2019b.

WORLD RUGBY. Officiating. **Officiating**, 2019c. Disponível em: <<http://officiating.worldrugby.org>>. Acesso em: 25 fevereiro 2019.

WORLD RUGBY. Technical Zone Programme. **Officiating**, 2019c. Disponível em: <<http://officiating.worldrugby.org/?module=2>>. Acesso em: 25 fevereiro 2019.

WORLD RUGBY. Player Welfare Documents. **Player Welfare**, 2019d. Disponível em: <<http://playerwelfare.worldrugby.org/>>. Acesso em: 25 fevereiro 2019.

WORLD RUGBY. Tactics of Sevens. **Coaching**, 2019e. Disponível em: <<https://coaching.worldrugby.org/index.php?page=153>>. Acesso em: 25 fev. 2019.

WORLD RUGBY. World Rugby Game Analysis Reports. **Player Welfare**, 2019g. Disponível em: <<http://playerwelfare.worldrugby.org/?documentid=156>>. Acesso em: 25 fev. 2019.

XU, C. et al. Sports Video Analysis: Semantics Extraction, Editorial Content Creation and Adaptation. **JOURNAL OF MULTIMEDIA**, v. 4, n. 2, p. 69-79, abril 2009.

YU, L.; LIU, H. Efficient Feature Selection via Analysis of Relevance and Redundancy. **Journal of Machine Learning Research**, v. 5, p. 1205-1224, Outubro 2004.

ZHANG, T. Adaptive Forward-Backward Greedy Algorithm for Learning Sparse Representations. **IEEE TRANSACTIONS ON INFORMATION THEORY**, v. 57, n. 7, p. 4689-4708, julho 2011.

## Apêndice 1 – RUGBY

### O RUGBY

*Rugby Union* é um esporte coletivo com bola oval praticado mundialmente. Sua origem não é clara, gerando a lenda de William Webb Ellis. Em 1823, o atleta estaria jogando um esporte próximo ao futebol quando decidiu pegar a bola com as mãos e atravessar o campo correndo. Os adversários teriam tentando pará-lo de diversas formas, mas não obtiveram êxito. Como o jogo ocorria na Rugby School, Inglaterra, passou-se a chamá-lo assim (NAURIGHT, 2018).

O *Rugby Union* é gerido mundialmente pela World Rugby (WR), entidade que congrega as confederações e as uniões Regionais e Nacionais e promove o esporte através da educação, da realização de eventos e do apoio às entidades Regionais e Nacionais (INTERNATIONAL RUGBY BOARD, 2010). A modalidade mais conhecida e praticada é o *Rugby XV*. Já a modalidade olímpica é o *Rugby Sevens (7s)*. Além destas, há o *Rugby Beach* e as alterações da Lei para menores de 19 anos (WORLD RUGBY, 2019b).

O *Rugby* destaca-se por seus valores: Integridade, Disciplina, Respeito, Paixão e Solidariedade. Esses valores devem ser aplicados dentro e fora do campo em todas as instâncias: do escolar ao profissional (WORLD RUGBY, 2014a). Apesar de ser um esporte que oferece mais contato que o futebol e outros esportes coletivos com bola, o *Rugby* prima pela Disciplina e Respeito, sendo o árbitro peça fundamental do jogo (WORLD RUGBY, 2019b). O árbitro adequado a cada partida saberá reger o jogo conforme as necessidades técnicas, disciplinares e de conhecimento (INTERNATIONAL RUGBY BOARD, 2012). Desta forma, o *Rugby* trata-se de um esporte focado em desenvolver cidadãos ativos em suas comunidades a replicarem esses valores.

#### **2.1 Rugby Sevens**

O *Rugby Sevens* é praticado desde 1883 quando o Melrose Club, em função de dificuldades financeiras e de comprometimento de atletas, decidiu promover um evento com jogos reduzidos: sete atletas em cada equipe e sete minutos para cada lado. Na final, após empate no tempo regulamentar, o capitão da equipe vencedora marcou um *Try* e saiu comemorando o título. Criando, desta forma, o “Ponto de Ouro”, método utilizado ainda hoje para finalizar disputas com tempo extra (WORLD RUGBY, 2014b).

O *Sevens* passou a ser uma modalidade para ser jogada no final ou início das temporadas, um esporte de verão. Apenas em 1999, a International Rugby Board (IRB), precursora da WR, criou o Circuito Mundial de *Rugby Sevens*. Neste período, o torneio era apenas masculino. Só em 2012, teve início o circuito feminino (WORLD RUGBY, 2014b). Em 2009, O *Rugby Sevens* se tornou esporte Olímpico com estreia na Rio2016 (COMITÊ OLÍMPICO INTERNACIONAL, 2019). Após a confirmação do *Rugby* nas Olimpíadas em 2009, o *Sevens* passou a gerar interesse nas entidades nacionais tanto das equipes de elite como das piores ranqueadas (BARKELL, O'CONNOR e COTTON, 2016).

Os circuitos são realizados pela WR, mas organizados pelas Confederações Nacionais. A cada ano, a WR divulga a quantidade de etapas e as sedes da próxima temporada, que costuma ocorrer entre setembro/outubro a maio/junho (WORLD RUGBY, 2019a). As sedes são definidas de forma estratégica e visam o desenvolvimento do *Rugby* (INTERNATIONAL RUGBY BOARD, 2010). De 2014 a 2016, ou seja, por três anos, o Brasil foi sede de etapa do circuito feminino. Neste período, o Brasil participou do circuito como convidado, inicialmente, e como equipe fixa (um ano) (WORLD RUGBY, 2019g). Após a queda do Brasil no circuito, a Brasil Rugby oficializou a saída do Brasil como sede de etapa. Na temporada pré-olimpíadas, 2019/2020, o Brasil volta ao Circuito como equipe fixa e já está com vaga garantida para as Olimpíadas Tokyo2020 (CONFEDERAÇÃO BRASILEIRA DE RUGBY, 2019a).

## 2.2 *Rugby* Feminino

O *Rugby* Feminino ainda está em desenvolvimento no mundo. Nos países onde o *Rugby* é tradição, como Nova Zelândia, Austrália, Inglaterra e Irlanda, ocorrem competições nacionais de *Rugby XV* e *Sevens*, pois estes países (entre outros) jogam torneios internacionais, como a Copa do Mundo de *Rugby XV* e o Circuito Mundial de *Rugby Sevens*. Nacionalmente, até junho de 2019, há apenas competições oficiais de *Rugby Sevens* feminino (CONFEDERAÇÃO BRASILEIRA DE RUGBY, 2019b). Mas há preparação dos estados para torneio nacional de XV. O Super *Sevens* é o Circuito Nacional de *Rugby Sevens* Feminino e conta com oito equipes fixas. Em 2019, essas equipes estão distribuídas em sete estados: Rio Grande do Sul, Santa Catarina, Paraná, São Paulo<sup>10</sup>, Rio de Janeiro, Piauí e Mato Grosso.

---

<sup>10</sup> SP terá duas equipes fixas

As principais equipes de *Rugby Sevens* Feminino são a Nova Zelândia (Black Ferns) e a Austrália. As Black Ferns ganharam 22 das 37 etapas do circuito desde 2012 e foram 5 vezes campeãs do circuito. Já a Austrália é a primeira e atual campeã Olímpica, tendo vencido o circuito 2 vezes. Nenhuma equipe venceu o circuito além destas duas (WORLD RUGBY, 2019a).

Barkell, O'Connor e Cotton (2016) destacam que, desde o estudo de van Rooyen et al (2008), há um grande *gap* na literatura em análises de performance em *Sevens*, especialmente em relação ao feminino, que é quase inexistente. O estudo de Barkell, O'Connor e Cotton (2016) é um dos primeiros a comparar a performance técnica e/ou tática em *Rugby Sevens* masculino e feminino. Tanto o masculino quanto o feminino tem o *Try* como ação definidora do jogo. Além disso, em ambas as categorias, a equipe vencedora terá também maior percentual de passes efetivos, menor percentual de *Tackles* perdidos e farão mais *Line Breaks* (BARKELL, O'CONNOR e COTTON, 2016). Esta pesquisa confirma que, apesar do citado acima, há diferença entre homens e mulheres em relação às estratégias técnicas e às táticas: mulheres utilizam mais reinícios sem disputa e realizam mais passes; enquanto homens tem mais *Scrums* efetivos e optam em chutar para a lateral em Penais mais frequentemente.

### 2.3 Jogando Rugby

O *Rugby* é um esporte de invasão e evasão que tem contato. O objetivo do jogo é a marcação de pontos através da invasão do território adversário (mantendo ou não a posse) (WORLD RUGBY, 2014a). O *Try* é a pontuação máxima do jogo: 5 pontos. Ocorre quando a equipe de ataque consegue encostar a bola no chão dentro do *In-goal* de ataque, região sinalizada na Figura 2. Para ser validado, é necessário que a bola esteja em contato com o chão e a mão ou o tronco do atleta (WORLD RUGBY, 2019b, p. 46-48). Quando marcado, dá a oportunidade do time chutar para o H e alcançar mais 2 pontos, chute chamado de Conversão (WORLD RUGBY, 2019b, p. 48-49). Ainda é possível marcar pontos chutando ao H após um Penal, falta cometida pelo adversário, ou durante o jogo aberto conforme a Lei 8 do livro de Leis do *Rugby*. Ambas as situações agregam 3 pontos quando realizadas corretamente. No *Sevens*, modalidade pesquisada, dificilmente há marcação de pontos através de chute de Penal ou em jogo aberto, visto que é necessário quicar a bola antes de chutar (WORLD RUGBY, 2014a).

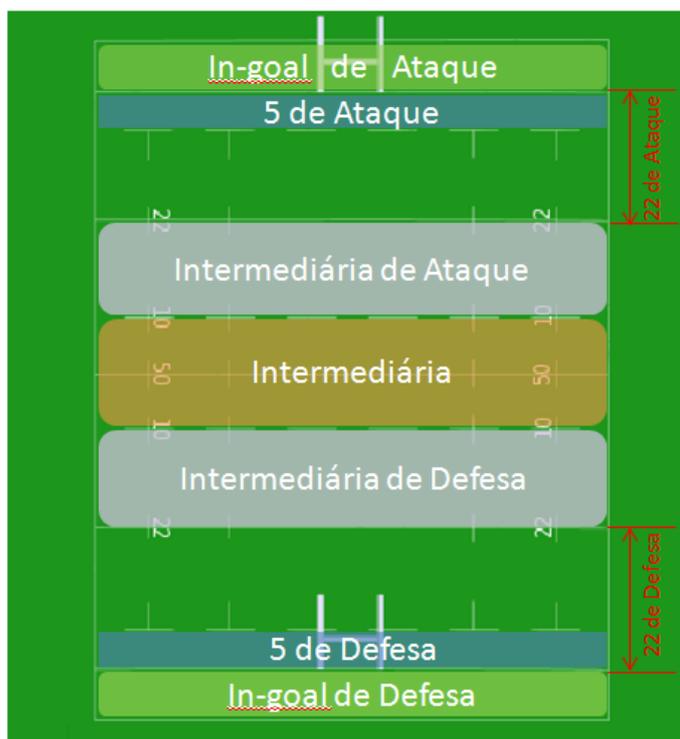


Figura 2 - Campo de Rugby e zonas analisadas

Para o início ou reinícios pós-*Try*, uma das equipes realiza o chute de saída, a bola deve quicar antes de ser chutada e deve passar a linha de 10 metros de ataque (WORLD RUGBY, 2019b, p. 59-61). A atleta que recebe a bola não deve deixá-la cair para frente, pois sempre que a bola cai para frente é sinalizada uma infração chamada *Knock-on*. No *Rugby*, uma das principais formas de avançar e/ou manter a posse de bola é passar a bola com as mãos. Porém, a bola só pode ser passada para trás ou para a mesma linha do passe. Nunca para frente. Quando a bola é passada para frente, é sinalizada a infração *Throw Forward*<sup>11</sup>. Em ambas as situações, pode ser sinalizada a vantagem ou reinicia-se o jogo com um *Scrum* (WORLD RUGBY, 2019b, p. 57-58).

Também é possível avançar correndo com a bola ou chutando-a para frente. Porém, em caso de chute, apenas os jogadores que estiverem atrás do chutador ou que forem alcançados por alguém que saiu de trás do chutador podem correr em direção à bola (WORLD RUGBY, 2019b, p. 55-56). Para tentar evitar o avanço do atleta em posse da bola, a defesa realiza o *Tackle*: movimento que tem como intenção derrubar o oponente, colocando a bola em disputa no chão (WORLD RUGBY, 2019b, p. 64-69). Essa disputa no chão chama-se *Ruck*, que é materializado quando dois oponentes em pé interligam-se ao disputar a bola ao

<sup>11</sup> No Brasil, chama-se de “Passe para frente”

chão (WORLD RUGBY, 2019b, p. 70-72). Geralmente ocorre após um *Tackle*, mas pode ocorrer a qualquer momento que a bola está ao chão (WORLD RUGBY, 2014a).



Figura 3 - *Scrum Sevens* – Foto: Michael Lee – KLC Fotos para World Rugby

Todas essas situações são regidas pelo livro de Leis do *Rugby* que está disponível no Portal da World Rugby. Caso ocorra a materialização de alguma infração, o árbitro sinaliza a vantagem ou a forma que reiniciará o jogo e também sinaliza a infração. O jogo pode reiniciar com *Scrum*, *Free-kick* ou Penal. A forma de reinício está definida no livro de Leis e é relacionada à gravidade da infração. O *Scrum* é aplicado em infrações leves, geralmente técnicas, como *Knock-on*. Para a realização do *Scrum* no *Sevens*, três Atletas de cada equipe alinham-se frente a frente para o encaixe e disputa da bola com os pés. Esta bola é introduzida no túnel formado pelo alinhamento das cabeças, como na Figura 3 (WORLD RUGBY, 2019b, p. 92-101). Em caso de *Free-kick*, as infrações são de gravidade média. A equipe infratora deve recuar 10 metros do local onde ocorreu a infração. A equipe que ganhou o *Free-kick* deve reiniciar o jogo conforme as leis do *Rugby*. Já Penais são situações graves, envolvendo intencionalidade, desrespeito ou insegurança. É possível tentar um chute ao H ou chutar para a lateral e permanecer com a posse da bola. Também pode-se cobrar seguindo as Leis do *Free-kick* ou solicitar um *Scrum* (WORLD RUGBY, 2019b, p. 102-106). Em casos de Penais, a Arbitragem pode dar cartão amarelo, tempo fora do jogo e a equipe permanece temporariamente com um atleta a menos, ou cartão vermelho, expulsão do jogo (WORLD RUGBY, 2019b, p. 51-54).

Quando a bola sai pela lateral, o reinício de jogo pode ser de forma rápida, conforme a Lei 18 do livro de Leis do *Rugby*, ou pela formação de *Lineout*. Para que haja esta formação são necessários dois atletas de cada equipe na região onde a bola saiu entre as linhas de 5 e 15 metros da lateral. A equipe de ataque, a que não saiu com a bola, lança a bola no canal para disputa entre os times (WORLD RUGBY, 2019b, p. 78-91). Quando o *Lineout* é originado na cobrança direta de Penal para fora, a equipe que cobrou o Penal é quem introduz a bola (WORLD RUGBY, 2019b, p. 102-106).

### 2.3.1 Rugby Sevens vs Rugby XV

O *Rugby Sevens* segue o livro de Leis da *World Rugby*, versão 2019, Variações da Lei - *Sevens*. A modalidade segue as mesmas leis básicas, mas possui alterações específicas que mudam as estratégias de jogo (ROSS, GIL, *et al.*, 2016). Por serem sete atletas em cada equipe, ao invés de 15, e as dimensões do campo permanecerem as mesmas, o jogo foca menos no avanço de terreno através do contato, buscando manutenção de posse e avanço por passes ou corridas no espaço aberto (WORLD RUGBY, 2019e). Além disso, o tipo físico dos atletas acaba se tornando mais restrito, pois não existe mais posição fixa em jogo aberto eliminando principalmente os *forwards* que geralmente são mais fortes do que ágeis. Os atletas passam a necessitar mais explosão e menos força. Outra alteração é o tempo de jogo: no *Sevens* são 7 minutos para cada lado enquanto no XV são 40 minutos para cada lado (HIGHAM, HOPKINS, *et al.*, 2014).

Os reinícios de jogo após *Try* são diferentes: enquanto no XV a equipe que sofreu o *Try* é quem reinicia o jogo, no *Sevens*, a equipe marcadora é quem reinicia o jogo, tendo a possibilidade de manutenção da posse da bola (SILVEIRA, 2018). Para chutes ao H, seja em cobrança de Penal ou em Conversões, os atletas do XV podem utilizar um equipamento chamado *Tee* que dá sustento à bola que fica parada para o chute. Já no *Sevens*, é necessário que a bola quique no chão antes do chute, dificultando a execução. Em jogo aberto, sempre é necessário quicar a bola antes do chute ao H além de outras regras específicas conforme a Lei 8 do livro de Leis do *Rugby*.

As formações fixas, *Scrum* e *Lineout*, sofrem alterações na Lei e na estratégia, respectivamente. No XV, o *Scrum* conta com 8 jogadores: Primeira linha (3 atletas), Segunda Linha (2 atletas) e Terceira linha (3 atletas) além do *Scrum-half*, atleta que introduz a bola ou marca quem a está introduzindo. Já no *Sevens*, é apenas a Primeira linha (3 atletas) e o *Scrum-half*. Essa alteração resulta em menos estabilidade da formação e mais agilidade na reposição

da bola. Já o *Lineout* não sofre alteração de Lei na sua formação, mas estrategicamente, as equipes de XV utilizam de 5 a 7 atletas além do lançador/opositor, sendo 7 o máximo permitido por Lei. Enquanto que no *Sevens* se utiliza de 2 a 3 atletas além do lançador/opositor, sendo 2 o mínimo permitido por Lei.

Em situações mais graves ou de desrespeito, pode ser aplicado o cartão amarelo. No *Rugby XV*, significa que o atleta ficará fora do jogo por 10 minutos e a equipe ficará esse mesmo tempo com um atleta a menos. Já no *Sevens*, o tempo é reduzido para 2 minutos (WORLD RUGBY, 2019c). Já em caso de substituições temporárias: sangramento ou protocolo de concussão, os tempos de 15 e 10 minutos, respectivamente, não são alterados por serem tempos de protocolos de segurança médica (WORLD RUGBY, 2019d).

## 2.4 Arbitragem

Segundo Higham, Hopkins, *et al.* (2014b), indicadores de performance podem variar conforme diversos fatores, entre eles a arbitragem. A arbitragem é composta por um árbitro principal e dois auxiliares. Em torneios World Rugby de *Sevens*, são incluídos árbitro de vídeo e dois auxiliares de *In-goal*<sup>12</sup>. Há também a equipe extracampo que é responsável pela súmula, pelas substituições, pela coordenação dos tempos dos cartões amarelos etc. Para eventos internacionais, como os pesquisados, a arbitragem principal é formada por uma equipe global semiprofissional contratada pela WR que trabalha e se desenvolve conjuntamente durante toda a temporada (WORLD RUGBY, 2019c).

Os árbitros auxiliares podem ser tanto da equipe da WR quanto das Confederações Nacionais. A WR define com base em qualidade técnica e objetivos de desenvolvimento para a definição dos auxiliares a cada jogo. A equipe extra-campo é formada por árbitros e agentes ativos no *Rugby* Nacional, coordenados por equipes específicas da WR. Todo evento é entendido como oportunidade de formação e desenvolvimento do esporte e seus agentes locais (INTERNATIONAL RUGBY BOARD, 2010).

A formação dos árbitros e demais agentes do esporte é realizada de forma escalonar, ou seja, a WR é responsável pela formação dos níveis mais altos e dos educadores, já as Confederações e Federações são responsáveis pelos níveis mais baixos. Todos ficam sob a acreditação da WR (WORLD RUGBY, 2019c). O Rio Grande do Sul tem a primeira *Trainer*, Educadora de educadores, em Primeiros Socorros no *Rugby*. Para atuar como árbitro em

---

<sup>12</sup> Um em cada *In-goal*

competição oficial, é necessário no mínimo ter realizado curso de Árbitro nível I, oferecido pelas Federações.

A atuação dos árbitros em qualquer nível de jogo segue protocolos e orientações globais. Ao identificar uma infração, o árbitro deve analisar e definir se há vantagem para a equipe não infratora. Em situação de vantagem, o árbitro fará comunicação verbal da vantagem e da forma de reinício além da comunicação corporal ao estender o braço apontando para quem tem direito à bola. Se não ocorrer a vantagem, o árbitro deve apitar sinalizando com o tipo de silvo qual a forma reiniciará o jogo. Após, deverá comunicar verbal e corporalmente como e quem reiniciará o jogo e qual foi a infração (INTERNATIONAL RUGBY BOARD, 2012). A comunicação verbal da arbitragem a nível internacional é realizada em microfone aberto, sendo possível ouvir o árbitro nos jogos analisados.

## Apêndice 2 – Tabelas de Jogos

Tabela 3 - Jogos Analisados

Temporada	Etapa	Fase	Seleção 1	Pontos S1		Pontos S2	Seleção 2
2017/2018	Dubai	Grupos	Austrália	31	x	12	Rússia
2017/2018	Dubai	Confrontos diretos	NZ	24	x	7	França
2017/2018	Sydney	Grupos	EUA	38	x	0	Japão
2017/2018	Sydney	Confrontos diretos	EUA	10	x	19	Irlanda
2017/2018	Japão	Grupos	Canadá	5	x	19	Rússia
2017/2018	Japão	Confrontos diretos	NZ	17	x	12	Austrália
2017/2018	Canadá	Grupos	Austrália	31	x	19	Irlanda
2017/2018	Canadá	Confrontos diretos	Canadá	35	x	12	Inglaterra
2017/2018	França	Grupos	Inglaterra	45	x	12	Gales
2017/2018	França	Confrontos diretos	EUA	28	x	7	Fiji
2015/2016	Dubai	Grupos	EUA	5	x	12	Irlanda
2015/2016	Dubai	Confrontos diretos	Canadá	19	x	19	NZ
2015/2016	Brasil	Grupos	Austrália	26	x	17	Fiji
2015/2016	Brasil	Confrontos diretos	Fiji	26	x	12	Inglaterra
2015/2016	EUA	Grupos	Rússia	10	x	7	Irlanda
2015/2016	EUA	Confrontos diretos	Fiji	0	x	38	EUA
2015/2016	Canadá	Grupos	NZ	26	x	5	Fiji
2015/2016	Canadá	Confrontos diretos	Espanha	5	x	21	Canadá
2015/2016	França	Grupos	Austrália	40	x	5	Irlanda
2015/2016	França	Confrontos diretos	Fiji	10	x	10	Espanha
2014/2015	Dubai	Grupos	França	24	x	7	África do Sul
2014/2015	Dubai	Confrontos diretos	França	10	x	31	NZ
2014/2015	Brasil	Grupos	NZ	54	x	0	França
2014/2015	Brasil	Confrontos diretos	China	5	x	7	Espanha
2014/2015	EUA	Grupos	Austrália	36	x	0	Fiji
2014/2015	Inglaterra	Grupos	Austrália	31	x	0	China
2014/2015	Inglaterra	Confrontos diretos	NZ	24	x	19	EUA
2014/2015	Holanda	Grupos	Austrália	48	x	0	China
2014/2015	Holanda	Confrontos diretos	China	12	x	38	Fiji

Tabela 4 - Jogos Excluídos

Temporada	Etapa	Fase	Seleção 1		Seleção 2	Motivo
2017/2018	Dubai	Confrontos diretos	Rússia	x	Espanha	Jogo incompleto
2017/2018	Dubai	Confrontos diretos	Japão	x	Fiji	Jogo incompleto
2017/2018	Dubai	Confrontos diretos	Espanha	x	Inglaterra	Jogo incompleto
2017/2018	Sydney	Grupos	EUA	x	Inglaterra	Jogo incompleto
2017/2018	Sydney	Grupos	NZ	x	Japão	Jogo incompleto
2017/2018	Sydney	Confrontos diretos	Canadá	x	França	Jogo incompleto
2015/2016	Canadá	Grupos	Canadá	x	EUA	Jogo incompleto
2014/2015	EUA	Confrontos diretos	Toda a fase			Não há vídeos disponíveis
2014/2015	Canadá	Grupos	Toda a fase			Imagens travando
2014/2015	Canadá	Confrontos diretos	Toda a fase			Imagens travando

Além destes jogos, foram excluídas as finais de 2014/2015 e 2015/2016.

### Apêndice 3 – Tabelas de Definições de Variáveis

**Tabela 5 – Definição das variáveis utilizadas para a seleção de variáveis**

Situação	Variável	Definição (por equipe por jogo)
Resultado	Vencedor/Perdedor	Variável binária que demonstra quem venceu o jogo: Vencedor (1) x Perdedor (0)
Resultado	Pontos marcados	Quantidade de pontos que cada equipe marcou na partida
Ataque - Passe/Recepção	Acerto - Bom	Quantidade de passes que chegam para o receptor na altura equivalente ao seu Tronco, preferencialmente em um ponto futuro da sua trajetória de corrida para que seja possível atacar a trajetória na qual a bola voa, ganhando velocidade e propiciar ao receptor as melhores condições para receber a bola
Ataque - Passe/Recepção	Acerto - Ok	Quantidade de passes que chegam para o receptor na altura equivalente ao seu tronco, porém de forma que a trajetória e velocidade da bola interferem sobre a velocidade e trajetória do receptor em baixo grau ainda permitindo ao receptor boas condições de obter a bola
Ataque - Passe/Recepção	Acerto - Ruim	Quantidade de passes que chegam para o receptor mais alto ou mais baixo do que a altura equivalente ao seu tronco, ou ainda, um passe que chega num ponto atrás do receptor. Nesse passe a trajetória e velocidade da bola influenciam bastante na velocidade e trajetória do jogador devido às más condições para receber a bola.
Ataque - Passe/Recepção	Erro - Alvo	Quantidade de passes que não chegam ao receptor, que cai no chão antes, depois, atrás ou na frente do receptor
Ataque - Passe/Recepção	Erro - Para Trás	Quantidade de passes que chegam ao receptor, mas o mesmo não domina a bola que cai para trás do mesmo, permitindo a continuidade do jogo
Ataque - Passe/Recepção	Erro - Turn Over (TO)	Quantidade de “roubadas” de bola em situações de Passe, como após erro de alvo ou para trás, ou mesmo em passe lento que é interrompido por defensor
Ataque - Passe/Recepção	Erro - Knock-on	Quantidade de passes que chegam ao receptor, mas o mesmo não domina a bola deixando-a cair para a frente: infração leve, vantagem curta ou reinício com <i>Scrum</i> para o outro time
Ataque - Passe/Recepção	Erro - Throw Forward	Quantidade de passes que são executados para frente da linha do passador: infração leve, vantagem curta ou reinício com <i>Scrum</i> para a outra equipe
Ataque - Ruck	Ganho - Drive	Quantidade de Ruck de Ataque ganhos e que teve como sequência um Drive (corrida)
Ataque - Ruck	Ganho - Pick & Go	Quantidade de Ruck de Ataque ganhos e que teve como sequência Pick & Go (contato intencional)
Ataque - Ruck	Ganho - Passe	Quantidade de Ruck de Ataque ganhos e que teve como sequência um Passe
Ataque - Ruck	Ganho - Erro e posse	Quantidade de Ruck de Ataque ganhos e que teve como sequência um erro na saída com manutenção da posse de bola
Ataque - Ruck	Ganho - Erro e TO	Quantidade de Ruck de Ataque ganhos e que teve como sequência um erro na saída e a perda da posse de bola
Ataque - Ruck	Perdido	Quantidade de Ruck de Ataque perdidos
	Tackle	Ação da defesa que tem como objetivo parar o atacante e a bola, colocando-a em disputa, preferencialmente ao chão
Defesa - Tackle	Efetivo - Triplo	Quantidade de Tackles que conseguiram parar o atacante e a bola utilizando 3 defensores
Defesa - Tackle	Efetivo - Duplo	Quantidade de Tackles que conseguiram parar o atacante e a bola utilizando 2 defensores

Defesa - <i>Tackle</i>	Efetivo - Individual	Quantidade de <i>Tackles</i> que conseguiram parar o atacante e a bola utilizando 1 defensor
Defesa - <i>Tackle</i>	Não-efetivo - Duplo	Quantidade de <i>Tackles</i> que não conseguiram parar o atacante ou a bola utilizando 2 defensores
Defesa - <i>Tackle</i>	Não-efetivo - Individual	Quantidade de <i>Tackles</i> que não conseguiram parar o atacante ou a bola utilizando 1 defensor
Defesa - <i>Tackle</i>	Perdido - Duplo	Quantidade de <i>Tackles</i> que não conseguiram parar o atacante nem a bola utilizando 2 defensores
Defesa - <i>Tackle</i>	Perdido - Individual	Quantidade de <i>Tackles</i> que não conseguiram parar o atacante nem a bola utilizando 1 defensor
Defesa - <i>Tackle</i>	Duplo Especial	Quantidade de <i>Tackles</i> que iniciaram individuais e terminaram como um <i>Tackle</i> duplo efetivo.
Ataque - Penais	<i>Ruck</i>	Quantidade de penais cometidos pela equipe atacante em situações de <i>Ruck</i> , como selar o <i>Ruck</i> , mão no <i>Ruck</i> , <i>Offside</i> entre outras que infrinjam as Leis do <i>Rugby</i> , especialmente a Lei 15
Ataque - Penais	<i>Scrum</i>	Quantidade de penais cometidos pela equipe atacante em situações de <i>Scrum</i> , como derrubar o <i>Scrum</i> , <i>Offside</i> entre outras que infrinjam as Leis do <i>Rugby</i> , especialmente a Lei 19
Ataque - Penais	<i>Tackle</i>	Quantidade de penais cometidos pela equipe atacante em situações de <i>Tackle</i> , como segurar a bola após o <i>Tackle</i> , <i>Offside</i> entre outras que infrinjam as Leis do <i>Rugby</i> , especialmente a Lei 14
Ataque - Penais	Impedimento de jogo	Quantidade de penais cometidos pela equipe atacante em situações que impedem a continuação do jogo, como jogar a bola com as mãos intencionalmente para fora entre outras que infrinjam as Leis do <i>Rugby</i> , especialmente a Lei 9
Ataque - Penais	Jogo Sujo	Quantidade de penais cometidos pela equipe atacante que enquadram-se como Jogo Sujo, como puxão de cabelo, agressão física entre outras que infrinjam as Leis do <i>Rugby</i> , especialmente a Lei 9
Ataque - Penais	<i>Offside</i>	Quantidade de penais cometidos pela equipe atacante em situações de <i>Offside</i> não contempladas nas demais variáveis, como em situações de <i>Lineout</i> , jogo aberto, pós- <i>Knock-on</i> entre outras que infrinjam as Leis do <i>Rugby</i> , especialmente a Lei 10
Defesa - Penais	<i>Ruck</i>	Quantidade de penais cometidos pela equipe defensora em situações de <i>Ruck</i> , como selar o <i>Ruck</i> , mão no <i>Ruck</i> , <i>Offside</i> entre outras que infrinjam as Leis do <i>Rugby</i> , especialmente a Lei 15
Defesa - Penais	<i>Scrum</i>	Quantidade de penais cometidos pela equipe defensora em situações de <i>Scrum</i> , como derrubar o <i>Scrum</i> , <i>Offside</i> , <i>Tackle</i> sem bola entre outras que infrinjam as Leis do <i>Rugby</i> , especialmente a Lei 19
Defesa - Penais	<i>Tackle</i>	Quantidade de penais cometidos pela equipe defensora em situações de <i>Tackle</i> , como <i>Tackle</i> Alto ( <i>Tacklear</i> acima da linha dos ombros), <i>Tackle</i> com as pernas, <i>Offside</i> entre outras que infrinjam as Leis do <i>Rugby</i> , especialmente a Lei 14
Defesa - Penais	Impedimento de jogo	Quantidade de penais cometidos pela equipe defensora em situações que impedem a continuação do jogo, como jogar a bola longe intencionalmente após apito do árbitro entre outras que infrinjam as Leis do <i>Rugby</i> , especialmente a Lei 9
Defesa - Penais	Jogo Sujo	Quantidade de penais cometidos pela equipe defensora que enquadram-se como Jogo Sujo, como puxão de cabelo, agressão física entre outras que infrinjam as Leis do <i>Rugby</i> , especialmente a Lei 9
Defesa - Penais	<i>Offside</i>	Quantidade de penais cometidos pela equipe defensora em situações de <i>Offside</i> não contempladas nas demais variáveis, como em situações de <i>Lineout</i> , jogo aberto, entre outras que infrinjam as Leis do <i>Rugby</i> , especialmente a Lei 10
Ataque - <i>Scrum</i>	5 m de Defesa - Ganho	Quantidade de <i>Scrum</i> em ataque ganhos na 5 m de Defesa
Ataque - <i>Scrum</i>	5 m de Defesa - Perdido	Quantidade de <i>Scrum</i> em ataque perdidos na 5 m de Defesa
Ataque - <i>Scrum</i>	5 m de Defesa - Infrações	Quantidade de Infrações Cometidas em <i>Scrum</i> em ataque na 5 m de Defesa

	cometidas	
Ataque - <i>Scrum</i>	5 m de Defesa - Infrações sofridas	Quantidade de Infrações Sofridas em <i>Scrum</i> em ataque na 5 m de Defesa
Ataque - <i>Scrum</i>	22 m de Defesa - Ganho	Quantidade de <i>Scrum</i> em ataque ganhos na 22 m de Defesa
Ataque - <i>Scrum</i>	22 m de Defesa - Perdido	Quantidade de <i>Scrum</i> em ataque perdidos na 22 m de Defesa
Ataque - <i>Scrum</i>	22 m de Defesa - Infrações cometidas	Quantidade de Infrações Cometidas em <i>Scrum</i> em ataque na 22 m de Defesa
Ataque - <i>Scrum</i>	22 m de Defesa - Infrações sofridas	Quantidade de Infrações Sofridas em <i>Scrum</i> em ataque na 22 m de Defesa
Ataque - <i>Scrum</i>	Intermediária de Defesa - Ganho	Quantidade de <i>Scrum</i> em ataque ganhos na intermediária de Defesa
Ataque - <i>Scrum</i>	Intermediária de Defesa - Perdido	Quantidade de <i>Scrum</i> em ataque perdidos na intermediária de Defesa
Ataque - <i>Scrum</i>	Intermediária de Defesa - Infrações cometidas	Quantidade de Infrações Cometidas em <i>Scrum</i> em ataque na intermediária de Defesa
Ataque - <i>Scrum</i>	Intermediária de Defesa - Infrações sofridas	Quantidade de Infrações Sofridas em <i>Scrum</i> em ataque na intermediária de Defesa
Ataque - <i>Scrum</i>	Intermediária - Ganho	Quantidade de <i>Scrum</i> em ataque ganhos na intermediária
Ataque - <i>Scrum</i>	Intermediária - Perdido	Quantidade de <i>Scrum</i> em ataque perdidos na intermediária
Ataque - <i>Scrum</i>	Intermediária - Infrações cometidas	Quantidade de Infrações Cometidas em <i>Scrum</i> em ataque na intermediária
Ataque - <i>Scrum</i>	Intermediária - Infrações sofridas	Quantidade de Infrações Sofridas em <i>Scrum</i> em ataque na intermediária
Ataque - <i>Scrum</i>	Intermediária de Ataque - Ganho	Quantidade de <i>Scrum</i> em ataque ganhos na intermediária de Ataque
Ataque - <i>Scrum</i>	Intermediária de Ataque - Perdido	Quantidade de <i>Scrum</i> em ataque perdidos na intermediária de Ataque
Ataque - <i>Scrum</i>	Intermediária de Ataque - Infrações cometidas	Quantidade de Infrações Cometidas em <i>Scrum</i> em ataque na intermediária de Ataque
Ataque - <i>Scrum</i>	Intermediária de Ataque - Infrações sofridas	Quantidade de Infrações Sofridas em <i>Scrum</i> em ataque na intermediária de Ataque
Ataque - <i>Scrum</i>	22 m de Ataque - Ganho	Quantidade de <i>Scrum</i> em ataque ganhos na 22 m de Ataque
Ataque - <i>Scrum</i>	22 m de Ataque - Perdido	Quantidade de <i>Scrum</i> em ataque perdidos na 22 m de Ataque
Ataque - <i>Scrum</i>	22 m de Ataque - Infrações cometidas	Quantidade de Infrações Cometidas em <i>Scrum</i> em ataque na 22 m de Ataque
Ataque - <i>Scrum</i>	22 m de Ataque - Infrações sofridas	Quantidade de Infrações Sofridas em <i>Scrum</i> em ataque na 22 m de Ataque
Ataque - <i>Scrum</i>	5 m de Ataque - Ganho	Quantidade de <i>Scrum</i> em ataque ganhos na 5 m de Ataque
Ataque - <i>Scrum</i>	5 m de Ataque - Perdido	Quantidade de <i>Scrum</i> em ataque perdidos na 5 m de Ataque
Ataque - <i>Scrum</i>	5 m de Ataque - Infrações cometidas	Quantidade de Infrações Cometidas em <i>Scrum</i> em ataque na 5 m de Ataque
Ataque - <i>Scrum</i>	5 m de Ataque - Infrações sofridas	Quantidade de Infrações Sofridas em <i>Scrum</i> em ataque na 5 m de Ataque



	cometidas	m de Ataque
Ataque - <i>Lineout</i>	5 m de Ataque - Infrações sofridas	Quantidade de Infrações Sofridas em <i>Lineout</i> de ataque, cobrança rápida ou não, na 5 m de Ataque
Ataque - 22 m	Entrada - Positiva totais	Quantidade de entradas que resultam em <i>Try</i>
Ataque - 22 m	Entrada - Neutra totais	Quantidade de entradas que não resultam em <i>Try</i> , mas a posse permanece com a equipe que estava em ataque
Ataque - 22 m	Entrada - Negativa totais	Quantidade de entradas que não resultam em <i>Try</i> e ainda resultam em perda da posse da bola - <i>Turn Over (TO)</i>
Ataque - 22 m	Ações positivas de ataque totais	Quantidade de ações positivas do ataque, como passe bom ou ok, <i>Drive</i> , disputa de <i>Ruck</i> , cobrança de penal ou <i>Scrum</i>
Ataque - 22 m	Ações negativas de ataque totais	Quantidade de ações negativas do ataque, como passe ruim, com erro de alvo, situações de passe ou <i>Ruck</i> que permitam <i>Turn Over (TO)</i> , infrações cometidas
Ataque - 22 m	Ações de Posse em ataque totais	Quantidade de ações positivas ou negativas do ataque mas que o ataque mantém a posse da bola
Ataque - 22 m	Ações que causam <i>Turn Over (TO)</i> totais	Quantidade de ações positivas ou negativas do ataque, mas que o ataque perde a posse da bola
Ataque - 22 m	Tempo em ataque	Tempo em segundos que o ataque permanece com a bola dentro da 22 m de ataque
Ataque - 22 m	Ações positivas de defesa totais	Quantidade de ações positivas da defesa, como <i>Tackle</i> efetivo, <i>Turn Over (TO)</i> em <i>Ruck</i> ou passe
Ataque - 22 m	Ações neutras de defesa totais	Quantidade de ações neutras da defesa, como <i>Tackle</i> não-efetivo, marcação na entrada da 22m, disputa de <i>Scrum</i>
Ataque - 22 m	Ações negativas de defesa totais	Quantidade de ações negativas da defesa, como <i>Tackle</i> perdido e infrações cometidas
Ataque - 22 m	<i>Try</i>	Quantidade de <i>Tries</i> marcados
Ataque - 22 m	Conversão	Quantidade de Conversões efetivas

Fonte: Autoria Própria com referências de Andrello (2013) e Silveira (2018)

Tabela 6 - Definições de outras variáveis

Ataque - 22 m	Entrada - Positiva por entrada	Variável binária sinalizando se a entrada resulta em <i>Try</i> (1) ou não (0) por entrada
Ataque - 22 m	Entrada - Neutra por entrada	Variável binária sinalizando se a entrada não resulta em <i>Try</i> , mas a posse permanece com a equipe que estava em ataque (1) OU não (0) por entrada
Ataque - 22 m	Entrada - Negativa por entrada	Variável binária sinalizando se a entrada não resulta em <i>Try</i> e ainda resulta em perda da posse da bola ( <i>Turn Over (TO)</i> - TO) (1) ou não (0) por entrada
Ataque - 22 m	Ações positivas de ataque por entrada	Quantidade de ações positivas do ataque, como passe bom ou ok, <i>Drive</i> , disputa de <i>Ruck</i> , cobrança de penal ou <i>Scrum</i> ; a cada entrada na 22 m de ataque
Ataque - 22 m	Ações negativas de ataque por entrada	Quantidade de ações negativas do ataque, como passe ruim, com erro de alvo, situações de passe ou <i>Ruck</i> que permitam <i>Turn Over (TO)</i> , infrações cometidas; a cada entrada na 22 m de ataque
Ataque - 22 m	Ações de Posse em ataque por entrada	Quantidade de ações positivas ou negativas do ataque, mas que o ataque mantém a posse da bola a cada entrada na 22 m de ataque
Ataque - 22 m	Ações que causam <i>Turn Over (TO)</i> por entrada	Quantidade de ações positivas ou negativas do ataque, mas que o ataque perde a posse da bola a cada entrada na 22 m de ataque
Ataque - 22 m	Tempo em ataque por entrada	Tempo em segundos que o ataque permanece com a bola dentro da 22 m de ataque a cada entrada
Ataque - 22 m	Ações positivas de defesa por entrada	Quantidade de ações positivas da defesa, como <i>Tackle</i> efetivo, <i>Turn Over (TO)</i> em <i>Ruck</i> ou passe, a cada entrada na 22 m de ataque
Ataque - 22 m	Ações neutras de defesa por entrada	Quantidade de ações neutras da defesa, como <i>Tackle</i> não-efetivo, marcação na entrada da 22m, disputa de <i>Scrum</i> ; a cada entrada na 22 m de ataque
Ataque - 22 m	Ações negativas de defesa por entrada	Quantidade de ações negativas da defesa, como <i>Tackle</i> perdido e infrações cometidas, a cada entrada na 22 m de ataque
Ataque - 22 m	<i>Try</i> por entrada	Variável binária afirmando a ocorrência (1) ou não (0) de <i>Try</i> a cada entrada na 22 m de ataque
Ataque - 22 m	Entrada por Line Break por entrada	Variável binária afirmando a ocorrência (1) ou não (0) da entrada nas 22 através de uma situação de Line Break ocorrida em qualquer parte do campo
Ataque - 22 m	<i>Try</i> por Line Break por entrada	Variável binária afirmando a ocorrência (1) ou não (0) <i>Try</i> através de uma situação de Line Break ocorrida em qualquer parte do campo
Ataque - 22 m	Conversão por entrada	Variável binária afirmando a ocorrência (1) ou não (0) de cobrança efetiva de Conversão a cada entrada na 22 m de ataque
Arbitragem	Jogos em WSS	Quantidade de jogos que o árbitro já havia apitado no Circuito no momento da partida
Chute de Saída/Reinícios	1º Tempo	Variável binária sinalizando a equipe chutadora com 1 e a equipe receptora com 0
Chute de Saída/Reinícios	2º Tempo	Variável binária sinalizando a equipe chutadora com 1 e a equipe receptora com 0
Ataque - Penais	Cobrança: H total	Quantidade de Penais cobrados ao H
Ataque - Penais	Cobrança: H	Quantidade de Penais cobrados e convertidos ao H
Ataque - Penais	Cobrança: Lateral	Quantidade de Penais cobrados com chute para lateral
Ataque - Penais	Cobrança: <i>Tap &amp; GO</i>	Quantidade de Penais abertos como <i>Free-kick</i> e a própria atleta avança com a bola
Ataque - Penais	Cobrança: <i>Tap &amp; Passe</i>	Quantidade de Penais abertos como <i>Free-kick</i> e a atleta realiza um passe na sequência
Ataque - Penais	Cobrança: Rápida	Quantidade de Penais cobrados rápidos conforme a Lei 20 do Livro de Lei do <i>Rugby</i>

Tabela de definições de variáveis coletadas para pesquisas futuras e não utilizadas na amostra piloto.

Fonte: Autoria própria