



Universidade de Brasília - UnB
Instituto de Ciências Exatas - IE
Departamento de Estatística - EST

Inferência Bayesiana Aplicada à Previsão de Resultados do Campeonato Brasileiro de Futebol

Bruno Fernandes de Matos

Orientador: Professor Eduardo Yoshio Nakano

Brasília

2017

Bruno Fernandes de Matos

Inferência Bayesiana Aplicada à Previsão de Resultados do Campeonato Brasileiro de Futebol

Relatório final apresentado à disciplina de Trabalho de Conclusão de Curso de graduação em Estatística, Instituto de Exatas, Universidade de Brasília, como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Orientador: Professor Eduardo Yoshio Nakano

Brasília

2017

Bruno Fernandes de Matos

Inferência Bayesiana Aplicada à Previsão de Resultados do Campeonato Brasileiro de Futebol/ Bruno Fernandes de Matos. – Brasília, 2017-
72 p. : il. (algumas color.) ; 30 cm.

Orientador: Professor Eduardo Yoshio Nakano

Relatório Final – Universidade de Brasília

Instituto de Ciências Exatas

Departamento de Estatística

Trabalho de Conclusão de Curso de Graduação, 2017.

1. Bayesiana. 2. Modelagem. 3. Futebol. 4. Previsão. 5. Jogos. 6. Dirichlet.

Bruno Fernandes de Matos

Inferência Bayesiana Aplicada à Previsão de Resultados do Campeonato Brasileiro de Futebol

Relatório final apresentado à disciplina de Trabalho de Conclusão de Curso de graduação em Estatística, Instituto de Exatas, Universidade de Brasília, como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Professor Eduardo Yoshio Nakano
Orientador

Antônio Eduardo Gomes
Membro da Banca

Bernardo Borba de Andrade
Membro da Banca

Brasília
2017

Agradecimentos

Nessa caminhada até aqui tenho que agradecer àqueles que estiveram ao meu lado me dando suporte e apoio nesse etapa da vida.

Primeiramente agradeço aos meus pais, ambos deram a sangue e sonhos para que eu pudesse realizar os meus sonhos, assim espero retribuir no futuro sendo um ser humano com os valores que foram me ensinados.

Agradeço a minha mãe por ser meu exemplo de coragem, independência e amor, por não temer seguir o caminho de seus sonhos e me fazer não temer seguir o caminho dos meus.

Agradeço ao meu pai por ter estado presente sempre em minha vida e mesmo sem condições algumas vezes, fez questão de evitar dizer "não" para seus filhos, nem que isso se tornasse trabalho dobrado.

Agradeço a minha eterna parceira e irmã, a menina com quem convivi mais que meus próprios pais e cada dia ao seu lado foi e é uma lição pra vida ainda maior. A sua determinação me inspira a mesmo em momentos difíceis não desistir dos meus sonhos.

Agradeço a minha família que esteve ao meu lado durante todo esse tempo. Agradeço a minha vó, Clarimunda, que como sem dúvidas, queria que pudesse saber o quanto me inspirou na vida, tendo sempre seu instinto protetor e sua força, como sua forma de demonstrar amor. As minhas tias e tios que me motivaram e cuidaram de mim diversas vezes. Aos meus primos, especialmente Arthur, David e Lays, que me mostraram disposição para lutar pelo seu espaço no mundo e me fizeram tomar parte com vocês.

Agradeço ao meu padrinho, tio Nonô, por ser a pessoa que me apoia a seguir meus sonhos porque vê eles em meus olhos e no meu coração.

Agradeço a minha amiga Isabella Vieira, que sempre esteve ao meu lado desde meu primeiro dia na graduação, me apoiando em minhas dificuldades, ajudando em meus dilemas e aproveitando os momentos de felicidade. Agradeço a meus amigos do FAALE e que conheci em Patos de Minas, são poucas as amizades que continuam as mesmas quando existem quilômetros nos separando e poder reencontrá-los em Patos me fez feliz todo ano. Também agradeço as meninas, Ana Paula Berçot e Thayanne Sales, que me ajudaram nessa caminhada e estiveram sempre ao meu lado.

Agradeço aos meus amigos, Gabriel Ravi, Lucas Lima, Winderson Júnior, Gabriela Vasconcelos, Matheus Maroneze, Alfredo Rossi e todos que fizeram parte da minha caminhada nessa universidade. Durante os momentos felizes, sorriram e riram comigo e durante os momentos difíceis me ajudaram a enfrentar os problemas.

Agradeço ao meu orientador, Eduardo Nakano, por ter aceito me orientar com esse tema e me abrir a cabeça em diversos momentos.

Agradeço à ESTAT, ao OBMigra e ao IPEA por me possibilitarem aplicar a estatística na prática, me fazerem lidar com diversas situações na qual me fez aprender e crescer profissionalmente e pessoalmente como não imaginei. Agradeço aos professores, que me deram a formação teórica para enfrentar essas diversas situações pelas quais passei.

Agradeço à todos os professores com quem tive a honra de aprender durante minha graduação, vocês me possibilitam entender a estatística como a ferramenta que é e que está sempre se desenvolvendo. Também agradeço aos servidores do EST por me ajudarem quando precisei e serem um sinônimo de eficiência.

Agradeço a todos que passaram pela minha vida durante esse período e antes da universidade, o caminho que trilhei foi sem dúvidas influenciado por cada pessoa que passou por ele.

Resumo

A cada ano o interesse em prever resultados de futebol aumenta e junto ao aumento das apostas esportivas, cresce a motivação para se obter diferentes modelos que possam aumentar as chances de acertar os resultados reais. Assim, esse trabalho apresenta uma aplicação de Modelos Bayesianos para a previsão de resultados de partidas de futebol. Nesse trabalho, as técnicas foram aplicadas para o Campeonato Brasileiro de Futebol de 2017, porém, podem ser aplicadas para estimar probabilidades de confrontos de futebol em qualquer natureza e o método de simulação permite obter estimativas sobre o campeonato para qualquer campeonato no modelo de pontos corridos. Foram usados modelos que se baseiam na distribuição de Dirichlet para obter as probabilidades de cada um dos possíveis resultados e foram realizadas simulações via Monte Carlo para obter probabilidades relacionadas ao resultado final do campeonato.

Palavras-chave: Futebol, Bayesiana, Modelagem, Dirichlet, Previsão, Simulação.

Lista de ilustrações

Figura 1 – Diagrama comparativo dos modelos de previsão dos resultados das partidas.	29
Figura 2 – Diagrama da Estrutura de Simulação do Campeonato.	30
Figura 3 – Comparação dos Modelos de Previsão de Título para Corinthians e Grêmio. Campeonato Brasileiro de Futebol 2017.	42
Figura 4 – Comparação dos Modelos de Previsão de Libertadores para Palmeiras e Santos. Campeonato Brasileiro de Futebol 2017.	43
Figura 5 – Comparação dos Modelos de Previsão de Libertadores para Cruzeiro e Flamengo. Campeonato Brasileiro de Futebol 2017.	44
Figura 6 – Comparação dos Modelos de Previsão de Rebaixamento para Atlético-GO e Avaí. Campeonato Brasileiro de Futebol 2017.	45
Figura 7 – Comparação dos Modelos de Previsão de Rebaixamento para São Paulo e Vitória. Campeonato Brasileiro de Futebol 2017.	45
Figura 8 – Comparativo das Distribuições da Proporção de Acertos de Previsão. Jogos do campeonato brasileiro de futebol de 2017. Até a rodada 26.	47
Figura 9 – Comparativo das Distribuições da Proporção de Acertos de Previsão. Jogos do campeonato brasileiro de futebol de 2017. Até a rodada 26.	49
Figura 10 – Comparativo das Distribuições da Proporção de Acertos de Previsão em relação à rodada. Jogos do campeonato brasileiro de futebol de 2017. Até a rodada 26.	50

Lista de tabelas

Tabela 1 – Probabilidades de vitória, empate e derrota dos times 1 e 2.	24
Tabela 2 – Chances de conquistar o título ao longo do Campeonato Brasileiro de Futebol 2017 - Modelo 1.	32
Tabela 3 – Chances de conquistar vaga à Taça Libertadores ao longo do Campeonato Brasileiro de Futebol 2017 - Modelo 1.	34
Tabela 4 – Chances de ser rebaixado ao longo do Campeonato Brasileiro de Futebol 2017 - Modelo 1.	35
Tabela 5 – Chances de conquistar o título ao longo do Campeonato Brasileiro de Futebol 2017 - Modelo 2.	37
Tabela 6 – Chances de conquistar vaga à Taça Libertadores ao longo do Campeonato Brasileiro de Futebol 2017 - Modelo 2.	38
Tabela 7 – Chances de ser rebaixado ao longo do Campeonato Brasileiro de Futebol 2017 - Modelo 2.	40
Tabela 8 – Medidas descritivas do Percentual de Acerto dos Modelos. Jogos do campeonato brasileiro de futebol de 2017. Até a rodada 26.	48
Tabela 9 – Percentuais de Confrontos que tiveram Acerto de Previsão superior à 1/3. Jogos do campeonato brasileiro de futebol de 2017. Até a rodada 26.	48
Tabela 10 – Confrontos com Maiores e Menores Percentuais de Acerto de Previsão para cada Modelo. Jogos do campeonato brasileiro de futebol de 2017. Até a rodada 26.	50

Sumário

	Introdução	15
1	REVISÃO BIBLIOGRÁFICA	17
2	METODOLOGIA	21
2.1	A Estrutura do Campeonato	21
2.2	Modelos por Confrontos Diretos	21
2.2.1	Dados Utilizados	21
2.2.2	Estrutura do Modelo 1	22
2.3	Modelo por Últimas Partidas Disputadas	23
2.3.1	Dados Utilizados	23
2.3.2	Estrutura do Modelo 2	24
2.4	Diagnóstico	26
2.5	Estrutura de Simulação do Campeonato	26
2.5.1	Simulação do Campeonato dado Modelo de Confrontos Diretos	27
2.5.2	Simulação do Campeonato dado Modelo de Últimas Partidas Disputadas	28
2.5.3	Sobre os dados	30
3	RESULTADOS	31
3.1	Modelo 1 - Confrontos Diretos	32
3.2	Modelo 2 - Últimas Partidas	37
3.3	Comparação dos Modelos	42
3.4	Diagnóstico dos Modelos	47
4	CONCLUSÃO	53
	REFERÊNCIAS	55
5	APÊNDICE	57
5.1	Programação - Extração dos Dados	57
5.2	Programação - Modelo Confrontos Diretos	64
5.3	Programação - Modelo Últimas Partidas	68

Introdução

O futebol, como diversos outros esportes, é um dos grandes espetáculos do mundo contemporâneo. Hoje, é o esporte com maior número de espectadores do mundo e a cada dia vem se tornando um esporte com uma circulação monetária ainda maior, criando assim clubes cada vez mais poderosos financeiramente. Assim como diversos esporte, o futebol sempre esteve alinhado com apostas sobre resultados.

Hoje, com a facilidade implementada pela internet, as apostas estão cada vez mais presentes no mundo online e funcionam de forma cada vez mais dinâmica. Os tipos de apostas mais populares são as sobre o resultado de jogos específicos e sobre o resultado final de campeonatos. Assim, no mundo incerto das apostas e do futebol, nasce o interesse em analisar dados e com base neles obter indícios que possam aumentar as chances de prever os resultados reais.

O presente trabalho visa determinar modelos que possam indicar chances de clubes participantes do Campeonato Brasileiro de Futebol - Série A alcançar alguns marcos relativos ao regulamento (Título, vaga na Taça Libertadores da América e Rebaixamento para Série B). Assim, para isso serão desenvolvidos métodos que simulam os resultados de cada uma das partidas, por meio de probabilidades determinadas.

Desde 2003 a Série A do Campeonato Brasileiro segue a estrutura de pontos corridos com turno e returno. Essa estrutura se dá com meio de dois confrontos entre todos os times (20 times ao total), totalizando 38 partidas (rodadas) para cada time. À cada jogo, atribui-se 3 pontos ao time vencedor, e zero ao time derrotado. Em caso de empate ambos times ganham 1 ponto. Ao final do campeonato, a classificação geral se dará pela soma dos pontos obtidos por cada time nos 38 confrontos do campeonato. Ao final do campeonato, o time que liderar a pontuação é declarado campeão. Os times que estiverem nas seis primeiras posições têm acesso à Taça Libertadores da América. Já os últimos quatro colocados são rebaixados ao Campeonato Brasileiro Série B.

Essa estrutura de campeonato torna a inferência estatística clássica de difícil tratamento. No entanto, essa mesma inferência pode ser realizada com relativa facilidade com o uso da inferência bayesiana combinada com os métodos de simulação de Monte Carlo. Inferências utilizando técnicas similares aos utilizados nesse trabalho podem ser encontrados em Brunello e Nakano (2015), que combinam a inferência bayesiana com os métodos de simulação de Monte Carlo para a predição dos resultados de eleições proporcionais.

Neste contexto, o objetivo deste trabalho foi apresentar modelos que possam indicar chances de clubes participantes do Campeonato Brasileiro de Futebol - Série A alcançar

alguns marcos relativos ao regulamento (Título, vaga na Taça Libertadores da América e Rebaixamento para Série B).

Este trabalho apresenta dois modelos: Modelo 1, denominado modelo por confrontos diretos e; o Modelo 2, denominado modelo por últimas partidas disputadas. Ambos modelos têm como base estimar o desfecho de um jogo de futebol considerando três resultados possíveis: Vitória, empate ou derrota. Outros estudos consideram modelos que estimam o placar final (saldo de gols) de um jogo. Veja, por exemplo, o trabalho de Arruda (2000), que modela o saldo de gols por meio de uma distribuição Poisson Bivariada. O trabalho de Suzuki (2015) também aplica uma metodologia semelhante a anterior para dados de campeonatos mais recentes. Já Macedo (2014) usa em seu trabalho, regressão logística visando usar algumas co-variáveis para explicar o resultado do jogo.

As inferências foram realizadas dentro de um cenário bayesiano fazendo-se o uso de técnicas de simulação de Monte Carlo. Todos os cálculos e simulações foram realizados por meio do software livre R Team (2017). Os modelos desenvolvidos foram utilizados para realizar previsões do Campeonato Brasileiro de Futebol de 2017.

1 Revisão Bibliográfica

Neste projeto será utilizada a técnica de Inferência Bayesiana para a modelagem dos resultados dos jogos do Campeonato Brasileiro de Futebol e a obtenção das chances de título, vaga para a Copa Libertadores da América e Rebaixamento dos clubes que disputam este torneio.

Dessa forma, pretende-se modelar os jogos por meio das distribuições de probabilidade Multinomial e Dirichlet, considerando os possíveis resultados (Vitória, derrota ou empate). Assim, no modelo mais simples, as probabilidades de cada resultado serão obtidas pelas proporções da ocorrência dos resultados nos últimos confrontos entre os times no Campeonato Brasileiro desde 2015, usando uma priori não informativa para refletir o acaso do resultado do confronto em estudo.

As inferências serão realizadas dentro do cenário Bayesiano fazendo-se o uso das técnicas de simulação de Monte Carlo. Para tanto, este projeto também prevê uma revisão de inferência Bayesiana e o estudo dos métodos de simulação de Monte Carlo. Os cálculos e simulações do trabalho serão realizados através do software livre R (R Core Team, 2017).

Inferência Bayesiana

Dado Ehlers (2013) e Paulino (2003), a Inferência Bayesiana é um método de inferir sobre os parâmetros tendo como base a junção de resultados de conhecimento prévio dos parâmetros (Distribuição à Priori) e os dados observados (Verossimilhança). Dessa forma o método se baseia no Teorema de Bayes apresentado a seguir:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \quad (1.1)$$

Com base nesse resultado e no Teorema de DeFinetti tem-se que a partir de uma distribuição de probabilidade *à priori* e a distribuição do dados, podemos obter a distribuição *à posteriori*

$$\pi(\theta|X) \sim \pi(\theta) * L(\theta; X) \quad (1.2)$$

Onde:

- $\pi(\theta|X)$ é a distribuição à posteriori;
- $\pi(\theta)$ é a distribuição à priori;
- $L(\theta; X)$ é a função de verossimilhança.

Assim, toda a estimação (pontual e intervalar) feita será obtida sobre a distribuição *à posteriori* dos dados. Isso garante um bom resultado sendo esse o fato de que a estimação se trata de probabilidades por estar baseada em uma distribuição, o que leva a ter resultados mais viáveis para as previsões dos resultados.

Para a modelagem inicial desses dados, as distribuições que modelarão serão a Multinomial, como a função de verossimilhança, e a Dirichlet, como distribuição *à priori*. A escolha da Dirichlet vem em razão das duas distribuições fazerem parte da família conjugada Dirichlet-Multinomial (extensão da família Beta-Binomial), o que facilita na obtenção da distribuição *à posteriori* por já existir resultados teóricos.

Distribuição Multinomial

A distribuição multinomial é uma extensão da distribuição binomial quando temos k ($k \geq 2$) categorias. Em particular, o resultado de um jogo de futebol pode ser classificado em $k = 3$ resultados possíveis. Considerando como referência um time 1 qualquer, esses resultados podem ser definidos como: 1 - vitória do time 1; 2 - empate e; 3 - derrota do time 1. Neste caso, a distribuição multinomial é definida por:

$$P(x_1, x_2, x_3 | \theta_1, \theta_2, \theta_3) = \frac{n!}{x_1! x_2! x_3!} \prod_{i=1}^3 \theta_i^{x_i} \quad (1.3)$$

Onde:

- θ_1 é o parâmetro correspondente à probabilidade de vitória do time 1;
- θ_2 é o parâmetro correspondente à probabilidade de empate;
- θ_3 é o parâmetro correspondente à probabilidade de derrota do time 1;
- x_1 é o número de vitórias do time 1;
- x_2 é o número de empates;
- x_3 é o número de derrotas do time 1;
- n é o número de jogos do time 1 ($x_1 + x_2 + x_3$).

Aqui, $\sum_{i=1}^3 \theta_i = 1$ e $\sum_{i=1}^3 x_i = n$

Notação: $(x_1, x_2, x_3 | \theta_1, \theta_2, \theta_3) \sim \text{Multinomial}(n, \theta_1, \theta_2, \theta_3)$

Dentro de um contexto bayesiano, a distribuição multinomial é conjugada com a família de distribuições Dirichlet. A distribuição Dirichlet é uma extensão da distribuição

Beta e, para o caso de $k = 3$ categorias, sua função densidade de probabilidades conjunta é definida por:

$$f(\theta_1, \theta_2, \theta_3 | a_1, a_2, a_3) = \frac{\Gamma(a_1 + a_2 + a_3)}{\Gamma(a_1)\Gamma(a_2)\Gamma(a_3)} \prod_{i=1}^3 \theta_i^{a_i-1} \quad (1.4)$$

Onde, $0 \leq \theta_i \leq 1$; $\sum_{i=1}^3 \theta_i = 1$ e $\Gamma(\cdot)$ é a função Gama definida por:

$$\Gamma(\alpha) = \int_0^{\infty} u^{\alpha-1} e^{-u} du, \alpha > 0$$

Notação: $(\theta_1, \theta_2, \theta_3 | a_1, a_2, a_3) \sim \text{Dirichlet}(a_1, a_2, a_3)$

A inferência bayesiana sobre o vetor de parâmetros $(\theta_1, \theta_2, \theta_3)$ será desenvolvida com base em uma conjugação, no caso a Dirichlet-Multinomial. Uma conjugação é em termos bayesianos um resultado em que uma distribuição à priori é conjugada em uma verossimilhança, se o resultado entre as duas apresenta uma distribuição conhecida. Nesse caso, se a verossimilhança for Multinomial, sua distribuição conjugada é uma Distribuição Dirichlet, na qual resultará numa distribuição à posteriori Dirichlet com alteração nos parâmetros. Este resultado é melhor apresentado a seguir:

Conjugação Dirichlet-Multinomial: Seja (X_1, X_2, X_3) uma única observação de uma distribuição Multinomial com vetor de parâmetros $(\theta_1, \theta_2, \theta_3)$. Assumindo, a priori que $(\theta_1, \theta_2, \theta_3) \sim \text{Dirichlet}(a_1, a_2, a_3)$, tem-se de (1.3) e (1.4) que a distribuição a posteriori é dada por:

$$\begin{aligned} f(\theta_1, \theta_2, \theta_3 | X_1, X_2, X_3) &= \frac{n!}{x_1!x_2!x_3!} \frac{\Gamma(a_1 + a_2 + a_3)}{\Gamma(a_1)\Gamma(a_2)\Gamma(a_3)} \prod_{i=1}^3 \theta_i^{a_i+x_i-1} \\ &= \frac{(x_1 + x_2 + x_3)!}{x_1!x_2!x_3!} \frac{\Gamma(a_1 + a_2 + a_3)}{\Gamma(a_1)\Gamma(a_2)\Gamma(a_3)} \prod_{i=1}^3 \theta_i^{a_i+x_i-1} \\ &= \frac{\Gamma((x_1 + x_2 + x_3 + a_1 + a_2 + a_3))}{\Gamma(a_1 + x_1)\Gamma(a_2 + x_2)\Gamma(a_3 + x_3)} \prod_{i=1}^3 \theta_i^{a_i+x_i-1} \end{aligned}$$

Ou seja, $(\theta_1, \theta_2, \theta_3 | X_1, X_2, X_3) \sim \text{Dirichlet}(a_1 + x_1, a_2 + x_2, a_3 + x_3)$

2 Metodologia

2.1 A Estrutura do Campeonato

O Campeonato Brasileiro de Futebol de 2017 segue a mesma estrutura desde o ano 2003, quando houve a alteração para o sistema de pontos corridos com turno e retorno. Atualmente, o campeonato conta com 20 times de todo o Brasil em sua primeira divisão (conhecida como Série A), nessa divisão disputam o campeonato os clubes da elite do futebol brasileiro.

A estrutura se dá por meio de confrontos entre todos os times. Cada confronto é ocorrido duas vezes, uma na casa do time A e outra na casa do time B, minimizando o efeito de mando de campo no campeonato em geral. Ao fim, se totalizam 38 partidas para cada clube, sendo que as primeiras 19 partidas se repetiram na mesma sequência com mando de campo invertido.

Em relação à pontuação, as partidas seguem o sistema usual de pontuação dos campeonatos de futebol à pontos corridos. Sendo que a vitória atribui ao clube 3 pontos, o empate, 1, e a derrota não concede nenhum ponto. Ao final do campeonato a classificação geral se dará pela soma dos pontos obtidos por cada time nos 38 confrontos do campeonato.

Ao final do campeonato o time que terminar com a maior quantidade de pontos é declarado campeão. Os times que estiverem nas seis primeiras posições conseguem acesso à Taça Libertadores da América, outro objetivo importante no campeonato. Já os últimos quatro colocados são rebaixados ao Campeonato Brasileiro Série B dando possibilidade do acesso dos quatro melhores clubes na Série B.

2.2 Modelos por Confrontos Diretos

2.2.1 Dados Utilizados

Visando prever os resultados dos confrontos do Campeonato Brasileiro, pensa-se que uma das formas mais intuitiva de obter dados seria utilizar os dados de confrontos diretos. Por isso, para a previsão do resultado do jogo entre o time 1 e o time 2 serão utilizados para a estimação o número de cada resultado possível: Vitória do time 1, empate ou derrota do time 1 (vitória do time 2) em n confrontos diretos entre o time 1 e 2. Este modelo é denominado neste trabalho como Modelo 1 ou modelo de confrontos diretos.

Destrinchando para a realidade do atual cenário do futebol brasileiro, também é visível que haja uma data limite para os dados que forem coletados, dado que os times

brasileiros mudam constantemente de situação. Assim o ideal é definir uma data a partir da qual que o histórico desses confrontos fossem computados. No caso, parte-se de que o cenário atual se tornou mais estável desde 2015, onde os clubes brasileiros que lideravam são os mesmos, assim como os clubes que eram coadjuvantes.

Outra possível variação desse modelo é a natureza do confronto em relação ao seu campeonato, ou seja, pode ser utilizados somente o histórico de confrontos no Campeonato Brasileiro ou pode utilizar-se de outros campeonatos disputados nesse período. No caso deste trabalho, foram utilizados para esse modelo somente os confrontos do Campeonato Brasileiro Série A.

2.2.2 Estrutura do Modelo 1

O modelo de confrontos diretos estimará as probabilidades dos possíveis resultados de jogo entre um time 1 e um time 2 com base nos n confrontos diretos entre esses dois times nos últimos 2 anos.

Assumindo que θ_1 é a probabilidade de vitória do time 1, θ_2 é a probabilidade de empate e θ_3 é a probabilidade de derrota do time 1 (Vitória do time 2) e que, a priori, $\theta = (\theta_1, \theta_2, \theta_3) \sim \text{Dirichlet}(a_1, a_2, a_3)$ tem-se que a função de verossimilhança do modelo é dada por $X = (x_1, x_2, x_3) \sim \text{Multinomial}(n, \theta_1, \theta_2, \theta_3)$, em que x_1 é o número de vitórias do time 1, x_2 é o número de empates e x_3 é o número de derrotas do time 1 em n jogos.

Assim pela conjugação Dirichlet-Multinomial, tem-se que a distribuição a posteriori de $\theta|X$ é dada por:

$$\theta|X \sim \text{Dirichlet}(a_1 + x_1, a_2 + x_2, a_3 + x_3) \quad (2.1)$$

Note que n pode ser igual a zero, isto é, pode não ter ocorrido no período em questão nenhum confronto direto entre os times 1 e 2 (Resultando em $x_1 = x_2 = x_3 = 0$). Neste caso, a distribuição a posteriori será dada pela própria distribuição a priori.

Será considerado neste trabalho uma priori uniforme para θ , isto é, $\theta \sim \text{Dirichlet}(1, 1, 1)$.

Exemplo: Considere um confronto entre os times A e B. Para estimar a probabilidade dos resultados dos jogos, considere que os resultados dos últimos $n = 20$ confrontos foram: 5 vitórias do time A ($x_1 = 5$), 12 empates ($x_2 = 12$) e 3 derrotas do time A ($x_3 = 3$). Considerando, a priori, que $\theta \sim \text{Dirichlet}(1, 1, 1)$, temos que $\theta|X \sim \text{Dirichlet}(6, 13, 4)$. Assim, a estimativa preditiva de vitória do Time A é dado por:

$$\begin{aligned} P[\text{VitoriaA}] &= P[Y_1 = 1, Y_2 = 0, Y_3 = 0|X_1, X_2, X_3] \\ &= E_{\theta|X}[P[Y_1 = 1, Y_2 = 0, Y_3 = 0|\theta]] \\ &= \int_{\Theta} \theta_1 \frac{\Gamma(23)}{\Gamma(6)\Gamma(13)\Gamma(4)} \theta_1^{6-1} \theta_2^{13-1} \theta_3^{4-1} d\Theta \end{aligned}$$

$$\begin{aligned}
&= \int_{\Theta} \frac{\Gamma(23)}{\Gamma(6)\Gamma(13)\Gamma(4)} \theta_1^{7-1} \theta_2^{13-1} \theta_3^{4-1} d\Theta \\
&= \frac{\Gamma(23)\Gamma(7)}{\Gamma(6)\Gamma(24)} \int_{\Theta} \frac{\Gamma(24)}{\Gamma(7)\Gamma(13)\Gamma(4)} \theta_1^{7-1} \theta_2^{13-1} \theta_3^{4-1} d\Theta \\
\text{Como } \int_{\Theta} \frac{\Gamma(24)}{\Gamma(7)\Gamma(13)\Gamma(4)} \theta_1^{7-1} \theta_2^{13-1} \theta_3^{4-1} d\Theta &= 1 \\
&= \frac{\Gamma(23)\Gamma(7)}{\Gamma(6)\Gamma(24)} \\
&= \frac{\Gamma(23)\Gamma(6)*6}{\Gamma(6)\Gamma(23)*23} \\
&= \frac{6}{23}
\end{aligned}$$

De forma análoga, as probabilidades de empate e derrota do time 1 (vitória do time 2) são dadas respectivamente por $\frac{13}{23}$ e $\frac{4}{23}$.

Note que essas probabilidades preditivas podem ser obtidas numericamente via simulação de Monte Carlo por meio do seguinte algoritmo:

1. Faça $k=1$;
2. Gere $\theta_k = (\theta_{1k}, \theta_{2k}, \theta_{3k}) \sim \text{Dirichlet}(6, 13, 4)$;
3. Gere $(y_{1k}, y_{2k}, y_{3k}) \sim \text{Multinomial}(n = 1, \theta_k)$;
4. Faça $k=k+1$ e repita os passos 2 a 4 m vezes.

Assim, a probabilidade de vitória, empate e derrota do time 1 são dados, respectivamente, por, \bar{Y}_1 , \bar{Y}_2 e \bar{Y}_3 .

2.3 Modelo por Últimas Partidas Disputadas

2.3.1 Dados Utilizados

Para propor um segundo modelo, denominado aqui como Modelo 2 ou modelo por últimas partidas disputadas, a estrutura dos dados utilizado foi mudada. Enquanto para o Modelo de Confrontos Diretos são considerados somente os confrontos em que os times envolvidos na partida disputaram entre si, num período recente, esse segundo modelo busca considerar os jogos mais recentes dos times envolvidos na partida a ser prevista, independentemente de qual foi o oponente.

Dessa forma, os dados aqui são obtidos pelos resultados dos últimos n jogos de cada um dos times. Vale ressaltar que a modelo de previsão definido a seguir permite com que esse número de jogos seja diferente para cada um dos dois times em confronto, isso porque esse modelo será baseado somente nas proporções de resultados de cada um dos times.

O número de jogos considerados deve ser definido dado um período válido, pois é necessário que o modelo ajude a trazer a realidade atual de cada time envolvido no

confronto. Para isso nesse caso serão considerados os jogos da temporada 2017, desde o início do ano, onde cada time passou por uma preparação por meio de campeonatos estaduais e regionais.

2.3.2 Estrutura do Modelo 2

Visando construir um modelo que considere os últimos resultados dos times, independente do adversário, é preciso definir uma série de regras de como é feita a atribuição de probabilidade de vitória (derrota ou empate) de cada clube. O primeiro passo é obter as proporções de vitórias, empates e derrotas de cada um dos times e dessa forma, determinar um parâmetro para cada um dos resultados de cada um dos times.

Tabela 1 – Probabilidades de vitória, empate e derrota dos times 1 e 2.

Resultados Possíveis	Time 1	Time 2
Vitória	θ_{11}	θ_{21}
Empate	θ_{12}	θ_{22}
Derrota	θ_{13}	θ_{23}
Total	1	1

Sejam x_{i1} o número de vitórias do time i nos últimos n_i jogos, x_{i2} o número de empates e x_{i3} o número de derrotas do time i nos últimos n_i jogos ($n_i = x_{i1} + x_{i2} + x_{i3}$, $i=1,2$). Considerando a priori que $\Theta_i = (\theta_{i1}, \theta_{i2}, \theta_{i3}) \sim Dirichlet(a_{i1}, a_{i2}, a_{i3})$, temos pela conjugação Dirichlet-Multinomial, que

$$(\theta_{i1}, \theta_{i2}, \theta_{i3} | x_{i1}, x_{i2}, x_{i3}) \sim Dirichlet(a_{i1} + x_{i1}, a_{i2} + x_{i2}, a_{i3} + x_{i3}).$$

Onde θ_{ij} (Probabilidade do resultado para o time i) pode ser obtido como a razão entre o número de ocorrências do resultado nas últimas n partidas disputadas. Dessa forma, como decorrência de uma propriedade da probabilidade, as soma dessas probabilidades totalizariam 1.

$$\theta_{i1} + \theta_{i2} + \theta_{i3} = 1 \quad (2.2)$$

E assim, cada um desses eventos poderiam ser modelados pela distribuição Dirichlet, sendo definidos como:

$$\Theta_i = \{\theta_{i1}, \theta_{i2}, \theta_{i3}\} \sim Dirichlet(a_{i1}, a_{i2}, a_{i3}) \quad (2.3)$$

Partindo desse ponto para se obter a probabilidade de cada um dos possíveis resultados da partida entre dois times, suponha, time 1 e 2, seria obtida com base nos θ_{ij} definidos na Tabela 1. Assim, o Modelo 2 se baseia no seguinte pensamento: a probabilidade do time 1 ganhar jogando contra o time 2 é estimada como a soma de todas

as probabilidades favoráveis ao time 1, considerando independentes as probabilidades dos resultados dos dois times. De forma similar, a probabilidade do time 1 ganhar jogando contra o time 2 é estimada como a soma de todas as probabilidades desfavoráveis ao time 1. Finalmente, a probabilidade de empate será dada pela soma das probabilidades dos resultados similares aos dois times.

$$\phi_1 = \theta_{11}\theta_{22} + \theta_{11}\theta_{23} + \theta_{12}\theta_{23} \quad (2.4)$$

$$\phi_2 = \theta_{11}\theta_{21} + \theta_{12}\theta_{22} + \theta_{13}\theta_{23} \quad (2.5)$$

$$\phi_3 = \theta_{21}\theta_{12} + \theta_{21}\theta_{13} + \theta_{22}\theta_{13} \quad (2.6)$$

Sendo ϕ_1 a probabilidade de que na partida o resultado seja uma vitória do time 1, ϕ_3 a probabilidade de que o time 2 vença a partida e ϕ_2 a probabilidade de que o resultado da partida se dê por um empate.

Vale ressaltar que para que os resultados acima sejam válidos é suposto que $\Theta_1 = (\theta_{11}, \theta_{12}, \theta_{13})$ seja independente do $\Theta_2 = (\theta_{21}, \theta_{22}, \theta_{23})$. Para isso, se põe a questão que seja necessário excluir os confrontos entre os times 1 e 2 para que possa se garantir a independência e assim obter as probabilidades da forma indicada pela equações acima.

Note ainda que:

$$\begin{aligned} \phi_1 + \phi_2 + \phi_3 &= \\ \theta_{11}\theta_{21} + \theta_{12}\theta_{22} + \theta_{13}\theta_{23} + \theta_{11}\theta_{22} + \theta_{11}\theta_{23} + \theta_{12}\theta_{23} + \theta_{21}\theta_{12} + \theta_{21}\theta_{13} + \theta_{22}\theta_{13} &= \\ \theta_{11}(\theta_{21} + \theta_{22} + \theta_{23}) + \theta_{12}(\theta_{21} + \theta_{22} + \theta_{23}) + \theta_{13}(\theta_{21} + \theta_{22} + \theta_{23}) &= \\ \theta_{11} + \theta_{12} + \theta_{13} &= 1 \end{aligned}$$

Note que a distribuição a posteriori de $\Phi = (\phi_1, \phi_2, \phi_3)$ pode ser obtida numericamente, via simulação de Monte Carlo, por meio da distribuição conjunta de (Θ_1, Θ_2) . Segue abaixo um algoritmo para estimação das probabilidades de vitória, empate e derrota do time 1 ao confrontar o time 2.

1. Faça $k=1$;
2. Gere $\Theta_{1k} = (\theta_{11k}, \theta_{12k}, \theta_{13k}) \sim \text{Dirichlet}(a_{11} + x_{11}, a_{12} + x_{12}, a_{13} + x_{13})$ e gere também $\Theta_{2k} = (\theta_{21k}, \theta_{22k}, \theta_{23k}) \sim \text{Dirichlet}(a_{21} + x_{21}, a_{22} + x_{22}, a_{23} + x_{23})$;
3. Calcule $\Phi = (\phi_1, \phi_2, \phi_3)$, segundo as expressões (2.4), (2.5) e (2.6).
4. Gere $(y_{1k}, y_{2k}, y_{3k}) \sim \text{Multinomial}(n = 1, \Phi_k)$;

5. Faça $k=k+1$ e repita os passos 2 a 5 m vezes.

Assim, a probabilidade de vitória, empate e derrota do time 1 são dados, respectivamente, por \bar{Y}_1 , \bar{Y}_2 e \bar{Y}_3 .

2.4 Diagnóstico

Afim de avaliar o desempenho dos modelos propostos, será utilizada uma medida definida como Percentual de Acerto de Previsão (PAPc), uma medida baseada na capacidade do modelo prever o resultado que realmente ocorreu.

1. Cria-se um vetor para abrigar a variável indicadora
2. Realize-se uma simulação de Monte Carlo para a partida;
3. Se o resultado simulado for igual ao resultado real, a variável indicadora recebe 1, caso contrário, receberá 0;
4. Ao final das m simulações, para cada partida será obtido o percentual de partidas em que o resultado simulado acertou, com base na variável indicadora.

Ao considerar o complementar dessa medida, tem-se o Percentual de Erro de Previsão do Confronto (PEPc), isto é, $PEPc = 1 - PAPc$.

Visto que PEPc é uma proporção o seu resultado varia entre 0 e 1, sendo que quanto mais próximo de 0 for seu resultado, melhor é a previsão do modelo. Outro ponto importante é que como só existem 3 resultados possíveis (vitória, empate e derrota), tem-se que se o PEPc (PAPc) estiver próximo de 0.666 (0.333) o indicativo é de que o modelo não consegue agregar informação nas previsões.

2.5 Estrutura de Simulação do Campeonato

Dada a forma como foi modelado a previsão de jogos, com base nelas é preciso definir o processo de simulação de todo o campeonato para obter as probabilidades de metas (Título, vaga à Libertadores ou rebaixamento para a Série B). Nesse processo, é importante tentar definí-lo de forma mais semelhante o possível em relação à formulação do campeonato Brasileiro, para que dessa forma, consiga-se obter probabilidades que indicam da melhor forma o que há de decorrer no campeonato.

De certa forma, parte da simulação do campeonato é fixa, ou seja, dado o resultado previsto de um jogo, o processo para obter os resultados continua o mesmo, independente da forma como foi obtida a previsão. Dado que serão considerados dois tipos de modelagem

dos jogos (Modelo de Confrontos Diretos e de Últimas Partidas Disputadas), haverão adaptações no processo de simulação dos jogos para cada um dos modelos, dado suas características.

2.5.1 Simulação do Campeonato dado Modelo de Confrontos Diretos

No caso dos jogos sendo previstos pelo modelo de confrontos diretos, o processo de simulação se dá por níveis. Primeiro se simula um resultado de dado confronto, num segundo nível se simulando 10 confrontos dos diferentes 20 times, tem-se a simulação de uma rodada. Vale ressaltar que a cada rodada os todos os times possuem apenas um confronto e todos realizam uma única partida.

Após realizar a simulação de uma rodada de jogos é necessário armazenar essas informações para atualizar os dados utilizados para prever a próxima rodada. Como nesse caso, são considerados apenas os confrontos diretos e tendo que os dois times só se confrontam duas vezes. os resultados simulados no primeiro turno (primeira metade do campeonato) serão usados para estimar resultados de partidas somente a partir da metade do campeonato (segundo turno ou retorno).

Um ponto importante é que por falta de consistência dos calendários dos times, não são simulados confrontos diretos entre dois times ocorridos em virtude de outro campeonato, mesmo que esses venham a ocorrer de fato. Isso torna-se uma limitação desse modelo de simulação do campeonato.

Dessa forma, após prever cada jogo é atribuída e armazenada a pontuação para cada time. Nesse sistema, a vitória vale 3 pontos, o empate 1 ponto e a derrota não há atribuição de ponto. Assim cada time vai acumulando os pontos dada a ocorrência dos resultados simulados.

Também é armazenada a quantidade de vitórias de cada time, isso em virtude de se obter um critério de desempate mais semelhante ao do campeonato brasileiro, onde pelo regulamento é o primeiro critério de desempate. Os critérios subsequentes se baseiam em informações que o modelo de simulação de jogos não prevê, nesse caso, saldo de gols e quantidade de cartões amarelos (forma de advertência).

Dessa forma, ao fim da simulação de todo o campeonato, considerando as previsões da primeira metade para prever a segunda metade do campeonato, têm-se os resultados finais. Com base nesses resultados são obtidas as posições de cada time, considerando primeiramente, os pontos obtidos em sequência o número de vitórias e por fim, caso continue tendo empate, o sorteio (último critério de desempate) entre os times empatados.

Esse processo é repetido por m vezes para que haja um erro de distribuição pequeno. Nesse trabalho foi usado o número de 10.000 repetições desse processo. Assim todo o campeonato é simulado 10 mil vezes e são obtidas as posições para cada time em cada uma

dessas simulações. A probabilidade de determinado time conseguir alcançar tal posição é dada pela proporção de vezes nas m simulações de Monte Carlo em que o time esteve na posição de interesse. Assim a probabilidade de ser campeão é obtida pela probabilidade de estar na primeira posição.

Nesse caso a probabilidade de ir para conseguir vaga na Copa Libertadores é obtida pela soma das probabilidades de estar em alguma posição contemplada por esse benefício, nesse caso, inicialmente estar entre os 6 primeiros colocados (pode variar dado algumas situações particulares). Da mesma forma a probabilidade de ser rebaixado para a Série B é obtida pela soma de estar na última 4 posições.

Um ponto importante é que em meio a essa simulação é realizado o cálculo dos Percentuais de Acerto e Erro de Previsão. Nesse caso é considerado a ocorrência do resultado real do campeonato até determinada rodada, e até essa determinada rodada os resultados simulados servirão somente para obter se o resultado previsto coincidiu com o resultado simulado. O restante do campeonato continuará sendo obtido por meio de simulação.

2.5.2 Simulação do Campeonato dado Modelo de Últimas Partidas Disputadas

Considerando o modelo de últimas partidas tem-se que pouco se diferenciará do processo de simulação do campeonato pelo modelo de confrontos diretos. Nesse caso as principais mudanças se darão na obtenção do resultado simulado dos confrontos.

Nesse caso, para obter o resultado de cada uma das partidas, será inserido na simulação da distribuição de Dirichlet, as respectivas quantidades de cada resultado nas últimas n partidas disputadas para cada time. Assim, cada time terá um vetor com 3 entradas, uma para o número de vitórias, outra para os empates e por fim uma para derrotas. Esses vetores serão determinados pela contagem feita com base na data dos jogos, ou seja, só serão considerados resultados anteriores a data inicial da rodada.

Após isso o cálculo das probabilidades será realizado e por meio de uma observação da distribuição Multinomial será definido o resultado do confronto. Esse resultado será adicionado ao vetor de ambos os times envolvidos para ser considerado na simulação dos próximos resultados. Dessa forma, esse modelo possui a limitação de não simular resultados de confrontos que ocorrem no decorrer do campeonato brasileiro, mas que não são referentes a essa competição.

No caso do modelo de confrontos diretos, a atualização dos dados quando se simulava um resultado, se dava apenas no segundo turno. Nesse caso, a atualização se dá rodada a rodada, ou seja, a cada rodada simulada os vetor de todos os times se alteram para simular a próxima rodada.

Em relação ao sistema de pontuação, a consideração de resultados que já ocorreram

e o cálculo dos Percentuais de Acerto e Erro de Previsão são realizados da mesma forma que no modelo de confrontos diretos. Assim, pode-se notar que em suma, o campeonato é simulado de forma bem semelhante, sendo alterado somente na forma de se obter o resultado para cada confronto e na forma de considerar resultados já simulados para as simulações dos confrontos decorrentes.

Abaixo seguem dois diagramas para que possa ser mais intuitivo a compreensão de como se dá a simulação dos jogos via os dois modelos e da estrutura do campeonato.

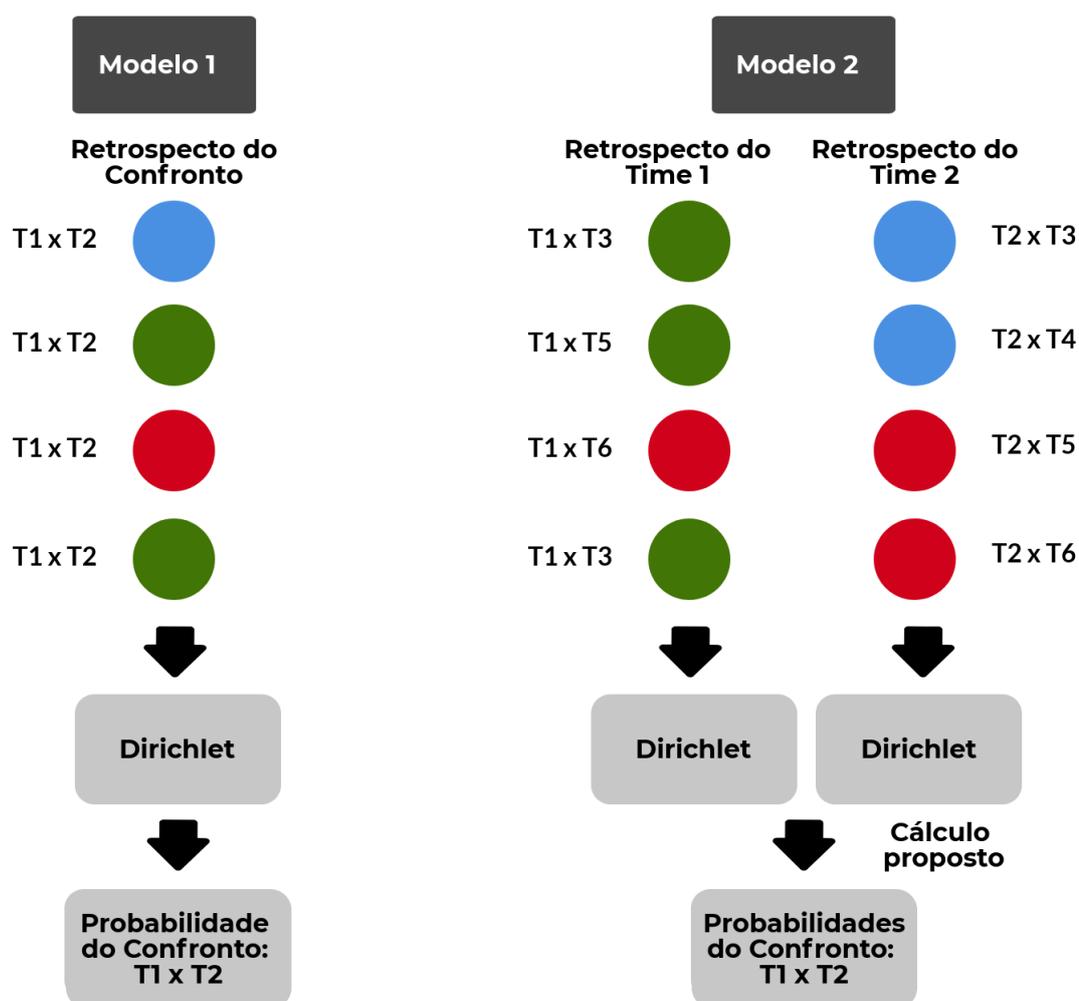


Figura 1 – Diagrama comparativo dos modelos de previsão dos resultados das partidas.

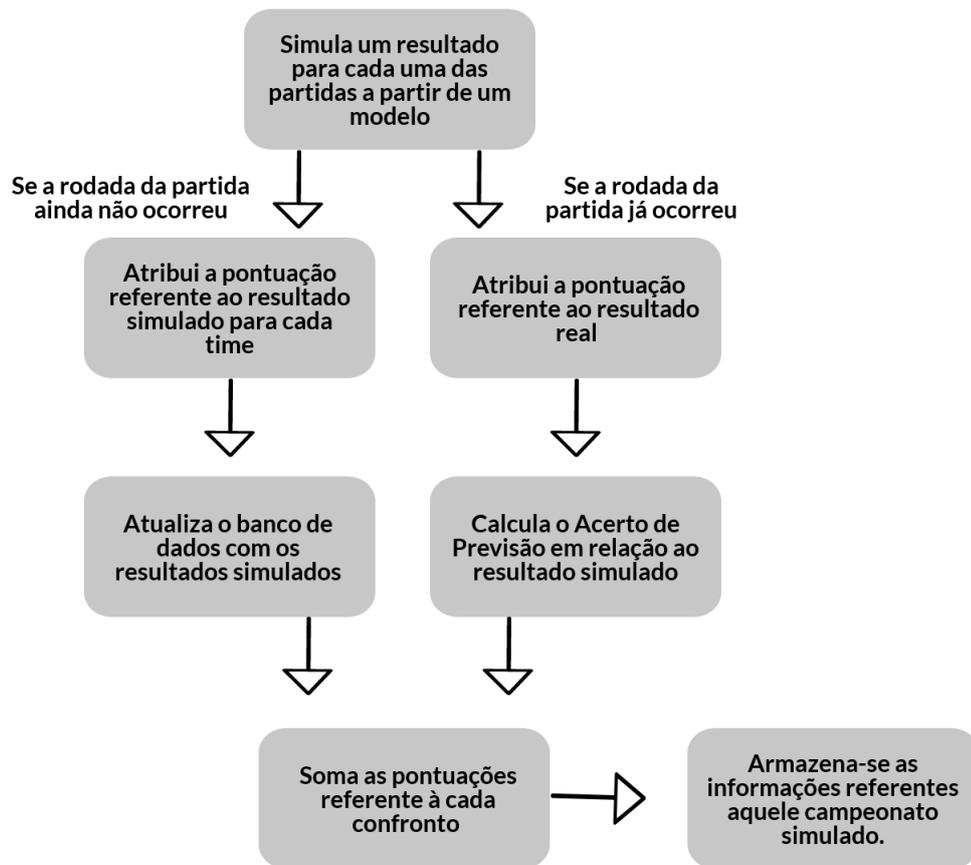


Figura 2 – Diagrama da Estrutura de Simulação do Campeonato.

2.5.3 Sobre os dados

Todos os dados utilizados para a aplicação desses modelos são reais e foram coletados do site OGol (acessado em 14/11/2017) via *web scraping* realizado no software R. Os dados foram estruturados via expressão regular para possibilitar a execução dos algoritmos.

3 Resultados

Após a implementação dos modelos no software estatístico R, é possível obter resultados que indicam as previsões de resultado para cada um dos jogos. Dessa forma, se baseando nessas previsões, pode-se construir a simulação do campeonato em si e por meio dessa estrutura, obter previsões sobre o campeonato. Para que isso funcione, basta ter em princípio o funcionamento do campeonato, no caso do Campeonato Brasileiro de Futebol tem-se o sistema de pontos corridos como citado na metodologia. Após a simulação de cada jogo, a pontuação resultante será atribuída ao clube dado o resultado da simulação, por exemplo, se o time A e o time B estão na disputa e a simulação resultar em vitória do time A, atribui-se a ele 3 pontos no campeonato e nenhum ao time B, caso o resultado seja um empate atribuiria 1 ponto para ambos os times e caso fosse derrota do time A, o time B seria bonificado com os 3 pontos da vitória.

Na simulação de todo o campeonato ainda tem-se uma situação distinta. Com o campeonato em andamento, pode-se obter os resultados relacionados à competição (Chances de título, rebaixamento e etc) considerando os resultados que já ocorreram. Nesse ponto, se por exemplo o campeonato já ocorreu até a décima quinta rodada, os resultados simulados somente serão considerados após essa rodada. Esse sistema ajuda a obter estatísticas cada vez mais precisas com o decorrer do campeonato e claro, ao fim, obter o verdadeiro valor, dado que não serão mais consideradas resultados de simulações e sim, os verdadeiros resultados.

Entretanto, é importante ainda continuar o processo de simulação das rodadas que já ocorreram, isso possibilita obter uma medida de como as previsões estão em relação aos resultados reais. Essa medida ajuda a comparar modelos e avaliar a qualidade das previsões que dado seu valor pode-se tomar que a previsão nem ajude a obter informações sobre o jogo. Essa medida, também poderá determinar o quanto informações acrescentadas por meio da priori do modelo, em futuros trabalho, estão ajudando na previsão dos resultados e se está causando alguma diferença no modelo.

3.1 Modelo 1 - Confrontos Diretos

Com base nos resultados obtidos pelo Modelo de Confrontos Diretos serão analisados a seguir as chances de cada um dos times participantes do campeonato, conquistar o título, conquistar vaga para a Copa Libertadores ou ser rebaixado para a Série B. Os resultados apresentados em tabelas serão dados por 4 momentos, o primeiro logo antes do início do campeonato, o segundo após a décima rodada, o terceiro quando se termina o primeiro turno e por fim ao fim da rodada 26 (rodada mais recente na obtenção dos dados).

Assim, analisando a chance de cada time participante ser campeão, a tabela abaixo apresenta essas probabilidades em forma de percentual:

Tabela 2 – Chances de conquistar o título ao longo do Campeonato Brasileiro de Futebol 2017 - Modelo 1.

Time	Inicial	Rodada 10	Rodada 19	Rodada 26
Atlético-GO	3.24	0.06	0.00	0.00
Avaí	0.70	0.01	0.00	0.00
Bahia	3.43	0.26	0.00	0.00
Botafogo	3.44	2.33	0.03	0.22
Atlético-MG	12.16	1.99	0.06	0.00
Atlético-PR	4.19	1.89	0.07	0.00
Coritiba	1.48	0.55	0.01	0.00
Chapecoense	1.70	0.55	0.00	0.00
Corinthians	10.21	47.44	80.15	81.20
Cruzeiro	2.95	1.32	0.03	0.33
Flamengo	4.83	2.71	0.47	0.26
Fluminense	1.47	0.79	0.01	0.00
Grêmio	5.74	18.85	10.21	8.64
Palmeiras	12.96	10.47	1.68	2.18
Ponte Preta	3.40	1.57	0.02	0.00
Santos	17.77	7.25	7.11	7.17
São Paulo	4.24	0.59	0.00	0.00
Sport-PE	3.01	0.57	0.15	0.00
Vasco	1.81	0.79	0.00	0.00
Vitória	1.27	0.01	0.00	0.00

Por meio da Tabela 2, nota-se que em relação ao título, inicialmente os times com maior probabilidade de conquistar o título eram o Santos (17,7%), o Palmeiras (12,96%) e o Atlético-MG (12,16%). Outro time a se destacar é o Corinthians com (10,21%).

Já levando em consideração a ocorrência de 10 rodadas, tem-se que as probabilidades já sofrem mudança, o Corinthians que antes tinha somente 10% passa a ter 47,44% de chances de ser campeão; o Grêmio que inicialmente tinha somente 5,7% passa a ter 18,8% e se torna o segundo com maior probabilidade de ganhar o título.

O Santos que antes possuía 17,7% de probabilidade passa a ter 7,25% enquanto o Palmeiras sofre uma redução de 12% para 10,4%. Desse forma, pode-se notar também que a maior parte dos demais times passam a ter probabilidades inferiores à 2% sendo acrescidos as exceções somente o Flamengo e o Botafogo.

Já ao fim do primeiro turno do campeonato, ou seja, na décima nona rodada, onde todos já se enfrentaram uma vez e reinicia-se a sequência de confrontos com mandos de campos invertidos, tem-se que o impacto da boa campanha do Corinthians o faz obter chances ainda maiores de ser campeão, passando de 47% para 80,15%. Em sequência vem o Grêmio que reduz suas chances de 18% para 10,2% seguido pelo Santos que permanece na casa dos 7% de chance de ser campeão. O Palmeiras se torna o outro time a ter probabilidade superior a 1%, com apenas 1,68%.

Nessa rodada, já se obtêm times com chances nulas de conquista do título sendo eles o Atlético-GO, Avaí, Bahia, Chapecoense, São Paulo, Vasco e Vitória.

Após 7 rodadas, na rodada 26, o movimento em relação ao fim do primeiro turno é menor. Nesse caso o Corinthians teve suas chances ampliadas para 81,20%, o Grêmio teve chances reduzidas para 8,64% enquanto o Santos permaneceu na por volta dos 7%. O Palmeiras aumenta suas chances para 2,18%. Nesse caso permanecem somente esses times e o Botafogo, Cruzeiro e Flamengo com chances de título, todos os demais passam a ter probabilidade nula ou inferior à 0,00%.

Outra meta de interesse dos clubes é a conquista da vaga para a Taça Libertadores da América, um torneio de grande valor internacional. Como definido por regulamento, o Campeonato Brasileiro de Futebol de 2017 tem inicialmente 6 vagas à disposição e elas são dadas aos 6 primeiros colocados no resultado final. Assim as probabilidades dos times obterem vaga foram obtidas pela frequência relativa de simulações na qual estiveram em alguma das 6 primeiras posições.

Abaixo segue uma tabela indicando essas probabilidades no decorrer do campeonato. Nesse caso, assim como na probabilidades de títulos, foram considerados o momento inicial ao campeonato, após a décima rodada, ao fim do primeiro turno e na rodada mais recente na coleta dos dados, a rodada 26

Tabela 3 – Chances de conquistar vaga à Taça Libertadores ao longo do Campeonato Brasileiro de Futebol 2017 - Modelo 1.

Time	Inicial	Rodada 10	Rodada 19	Rodada 26
Atlético-GO	26.71	4.12	0.05	0.07
Avaí	10.08	0.50	0.06	0.06
Bahia	26.15	10.57	8.69	2.19
Botafogo	25.67	30.55	14.87	50.77
Atlético-MG	53.04	30.28	20.81	9.45
Atlético-PR	29.85	28.27	22.75	11.34
Coritiba	16.14	15.17	10.31	0.17
Chapecoense	19.60	16.54	4.74	3.66
Corinthians	48.78	93.62	99.98	99.99
Cruzeiro	26.29	26.25	24.54	65.29
Flamengo	33.31	38.99	57.99	61.27
Fluminense	15.47	18.69	11.42	1.51
Grêmio	37.44	80.04	96.32	96.74
Palmeiras	56.89	68.08	78.03	87.31
Ponte Preta	26.21	28.08	8.90	1.76
Santos	63.11	59.07	93.05	97.00
São Paulo	29.98	15.17	1.09	0.85
Sport-PE	24.71	14.69	35.29	3.41
Vasco	17.35	19.39	10.38	5.79
Vitória	13.22	1.93	0.73	1.37

¹Note que os percentuais somam 600%, visto que os times disputam 6 vagas para a Libertadores.

Com base na Tabela 3, tem-se que inicialmente os times com maiores probabilidades de conquistar a seis vagas são: o Santos com 63%, Palmeiras (56%), Atlético-MG (53%), Corinthians (48,7%), Grêmio (37%) e Flamengo com 33%. Logo atrás, numa situação de disputa por vaga vem o São Paulo (30%) e o Atlético-PR (29,85%).

Também há os times com chances menores de concorrer à uma dessas vagas sendo eles o Avaí com apenas 10%, Vitória (13%) e o Fluminense (15,4%). É importante destacar que como o Atlético-GO e o Bahia não disputaram o Campeonato Brasileiro Série A após 2015, os dados desses times foram dados pela priori que nesse caso, por ser não informativa, tornou a chance de resultado desses times aleatória. Com o decorrer do campeonato, pode-se notar que esses times começam a ter o comportamento esperado dada a situação de que são recém promovidos para a primeira divisão.

Dando prosseguimento, considerando após os resultados da rodada 10, tem-se que o Corinthians toma a ponta com 93,6% de chance de conseguir uma vaga, seguido pelo Grêmio com 80%, Palmeiras (68%), Santos (59%), Flamengo (39%), Atlético-MG e Botafogo por volta dos 30%. Nessa rodada, alguns clubes já começam a ter chances mínimas de obter uma vaga, como o Avaí com 0,5%.

No fim do primeiro turno, o Corinthians começa a ter confirmação de sua vaga com

99,98% de chances, o Grêmio vem logo em seguida com 96,32% de chance, o Santos com 93% sendo seguidos pelo Palmeiras (78%), Flamengo (58%), Sport (35%). Na disputa por vaga ainda tem o Cruzeiro eo Atlético-PR com 24% e 22%, respectivamente. Ao fim do turno, mais dois times passam a ter chances inferior à 1% sendo eles o Atlético-GO e o Vitória.

Já na rodada 26, o Corinthians praticamente confirma sua vaga com 99,99%, o Santos e Grêmio também se aproximam da confirmação com 97% e 96,7%, respectivamente. O Palmeiras vem logo a seguir com 87,3%, acompanhado por Cruzeiro que aumenta suas chances para 65%, Flamengo (61,27%) e Botafogo com 50%.

O Atlético-PR com 11% é o último time com mais de 10% de chance de conquista uma vaga, já o Sport acaba tendo suas chances reduzidas para 3%. O Coritiba e o São Paulo tornam-se times com chance de vaga inferior à 1%.

A última meta a ser analisada é o rebaixamento para a Série B (ou segunda divisão). Nessa edição do campeonato, os quatro últimos colocados são rebaixados para que deem suas vagas no campeonato para o quatro primeiros colocados na Série B. Dessa forma, a chance é obtida de forma semelhante à da Taça Libertadores, sendo a chance de estar em alguma das quatro últimas colocações. Assim como nos casos anteriores, serão calculadas para os mesmos quatro momentos específicos do campeonato.

Tabela 4 – Chances de ser rebaixado ao longo do Campeonato Brasileiro de Futebol 2017 - Modelo 1.

Time	Inicial	Rodada 10	Rodada 19	Rodada 26
Atlético-GO	26.15	60.17	89.47	79.29
Avaí	50.12	85.10	87.34	80.49
Bahia	26.46	39.01	24.64	31.48
Botafogo	25.95	14.77	15.72	0.57
Atlético-MG	8.59	13.32	10.89	10.07
Atlético-PR	21.85	15.97	9.98	9.11
Coritiba	36.75	26.04	20.00	69.43
Chapecoense	30.91	25.34	34.85	21.72
Corinthians	9.69	0.12	0.00	0.00
Cruzeiro	24.93	16.78	8.09	0.08
Flamengo	18.26	8.24	1.07	0.10
Fluminense	39.25	23.35	19.31	37.64
Grêmio	16.01	0.71	0.02	0.00
Palmeiras	6.75	1.97	0.18	0.00
Ponte Preta	25.05	15.87	24.53	36.03
Santos	5.96	3.70	0.02	0.00
São Paulo	22.00	29.11	63.23	47.06
Sport-PE	25.93	28.51	3.90	23.11
Vasco	36.36	21.60	20.43	14.77
Vitória	43.03	70.32	66.33	39.05

¹Note que os percentuais somam 400%, visto que 4 times serão rebaixados.

Inicialmente, os times com maiores chances de serem rebaixados são o Avaí (50%), Vitória (43%), Fluminense (39,2%) e com chances em torno de 36% estão o Coritiba e o Vasco. Vale ressaltar como anteriormente a situação de não informação do Atlético-GO e do Bahia, que inicialmente é aleatória por esse modelo. Nesse caso, o Santos e Palmeiras iniciam o campeonato com chance em torno de 6%.

Após a décima rodada, o Avaí aumenta suas chances para 80% enquanto o Vitória tem 70%, o Atlético-GO já passa de 26 para 60% de chances e o Bahia de 26 para 39%. Com chances a se destacar estão o São Paulo e o Sport em torno 29% de serem rebaixados. Nessa rodada o Corinthians e o Grêmio passam a ter chance inferior a 1%.

No início do segundo turno, o Atlético-GO passa a ser o time com maior chance de ser rebaixado com 89,47% seguido pelo Avaí com 87,3%, logo após vem o Vitória com 66% e o São Paulo que aumenta de aproximadamente 30% na décima rodada para 63,2% na décima nona. Outros destaques são a Chapecoense com 34,85%, Bahia e Ponte Preta (24%).

Na rodada 26, o Atlético-GO reduz suas chances para 79,3% enquanto o Avaí vai para 80,5%, o Coritiba passa a ter 69,43% de chance sendo acompanhado pelo São Paulo com 47%. Nessa rodada, vale ressaltar ainda que o Vitória, Fluminense, Ponte Preta e Bahia passam a ter chances superiores à 30%.

3.2 Modelo 2 - Últimas Partidas

Nessa seção serão analisados os resultados obtidos por meio do segundo modelo sugerido, o modelo de Últimas Partidas. Da mesma forma como foram apresentados os resultados do modelo 1, serão apresentados os resultados deste. Inicialmente será analisada as probabilidades de meta para os times que disputam o campeonato em 4 momentos: antes do início do campeonato, ao fim da rodada 10, ao fim do primeiro turno e na rodada 26.

Para iniciar a análise, segue a tabela com os resultados referentes as chances de título dos clubes no decorrer do campeonato. Na tabela a seguir são apresentados esses resultados.

Tabela 5 – Chances de conquistar o título ao longo do Campeonato Brasileiro de Futebol 2017 - Modelo 2.

Time	Inicial	Rodada_10	Rodada_19	Rodada_26
Atlético-GO	0.34	0.00	0.00	0.00
Avaí	0.52	0.00	0.00	0.00
Bahia	2.92	0.16	0.00	0.00
Botafogo	2.07	0.72	0.00	0.42
Atlético-MG	8.60	1.11	0.00	0.01
Atlético-PR	0.64	0.20	0.03	0.01
Coritiba	7.18	1.86	0.04	0.00
Chapecoense	1.94	0.39	0.00	0.00
Corinthians	16.35	61.87	83.20	83.10
Cruzeiro	9.87	1.75	0.14	0.51
Flamengo	13.26	4.48	0.77	0.52
Fluminense	1.73	0.64	0.01	0.00
Grêmio	10.20	19.37	12.58	6.73
Palmeiras	4.57	2.23	0.56	1.33
Ponte Preta	3.11	0.75	0.00	0.00
Santos	6.70	3.36	2.50	7.36
São Paulo	2.29	0.16	0.00	0.00
Sport-PE	2.13	0.42	0.15	0.01
Vasco	1.31	0.47	0.02	0.00
Vitória	4.27	0.06	0.00	0.00

Analisando as chances de título logo antes do início do campeonato, pode-se notar que nesse modelo o Corinthians já inicia o campeonato como favorito com 16,35% de chance sendo seguido pelo Flamengo com 13,26% de chance. Em decorrência seguem o Grêmio com 10,2%, Cruzeiro (9,87%), Atlético-MG (8,6%) e Coritiba com 7,18%. O fato do Coritiba estar presente nessa lista é da natureza dos dados (Nesse modelo, antes do início do Brasileiro, são considerado somente resultados de torneios realizados no ano de 2017, antes do início do Brasileiro). Já com as menores chances estão o Atlético-GO, o Avaí e Atlético-PR, todos com chance inferior à 1%.

Ao fim da décima rodada, onde foi obtida novamente as chances, tem-se que o Corinthians aumenta suas chances para 61,87%. Já o Grêmio aumenta as suas chances para 19,4% sendo nesse momento, os únicos com chances acima de 10%. Em sequência vem o Flamengo que cai para 4,5%, Santos com 3,36% e o Palmeiras com 2,23%.

Ao fim do primeiro turno, as chances pouco se alteram em relação a ordem dos times. O Corinthians continua liderando, nesse caso, com 83% de chances enquanto o Grêmio tem suas chances reduzidas para 12,5%, o Santos com 2,5% de chances completa o grupo de times com mais de 1% de chances de título. Nessa ponto do campeonato diversos times já possuem chances praticamente nulas.

Ao fim da rodada 26, o Corinthians continua com chances em torno de 83% enquanto o Santos, com 7,36%, ultrapassa o Grêmio que tem chances reduzidas para 6,73%. O Palmeiras ainda surge como o outro time com chance acima de 1% de ser campeão. Nesse caso, Flamengo, Cruzeiro e Botafogo surgem com chances remotas, mesmo inferiores à 1%.

Abaixo segue a análise das chances obtidas pelo Modelo 2 para disputa de uma das vagas da Taça Libertadores.

Tabela 6 – Chances de conquistar vaga à Taça Libertadores ao longo do Campeonato Brasileiro de Futebol 2017 - Modelo 2.

Time	Inicial	Rodada_10	Rodada_19	Rodada_26
Atlético-GO	6.09	0.49	0.02	0.09
Avaí	7.89	0.99	0.35	0.53
Bahia	25.53	10.56	10.06	3.06
Botafogo	20.97	23.98	12.88	55.17
Atlético-MG	46.02	28.11	12.55	10.57
Atlético-PR	9.76	13.34	14.86	10.85
Coritiba	43.91	38.94	19.86	0.64
Chapecoense	18.62	15.41	4.69	5.27
Corinthians	61.61	97.69	99.98	100.00
Cruzeiro	50.23	39.50	34.43	61.53
Flamengo	56.74	58.76	68.63	61.21
Fluminense	17.48	22.56	16.05	2.25
Grêmio	48.86	85.51	97.65	94.37
Palmeiras	32.96	39.13	62.92	75.10
Ponte Preta	24.16	24.03	9.09	1.68
Santos	40.21	49.22	84.73	95.52
São Paulo	20.66	8.63	0.69	0.83
Sport-PE	19.09	17.62	36.12	6.68
Vasco	14.75	18.91	12.04	9.71
Vitória	34.46	6.62	2.40	4.94

¹Note que os percentuais somam 600%, visto que os times disputam 6 vagas para a Libertadores.

Por meio da Tabela 6, antes do início do campeonato brasileiro, as chances de

conquistar uma das 6 vagas para a Taça Libertadores é dada seguinte forma. O Corinthians inicia o campeonato com 61,6% de chances, seguido pelo Flamengo (56,7%), Cruzeiro (50,2%), Grêmio (48,8%) e Atlético-MG com 46%. Logo a seguir vem o bloco de times que ficam na disputa de vagas composto por Coritiba (43,9%), Santos (40,2%) e Vitória com 34,4%.

Ao fim da décima rodada, o Corinthians aumenta suas chances para 97,7% estando bem próximo de conseguir uma vaga, enquanto o Grêmio aumenta suas chances para 85,5%. O Flamengo sai de 56% para 58,7% enquanto o Santos aumenta suas chances para 49,2%. Logo após esse grupo, vem o Cruzeiro com 39,5%, Palmeiras (39,1%) e o Coritiba com 38,9%.

No início do retorno (segunda metade do campeonato), o Corinthians aumenta suas chances para 99,9% enquanto o Grêmio vai a 97,6%. O Santos aumenta suas chances para 84,7%, enquanto o Flamengo e o Palmeiras possuem chances entre 60% e 70%. Com chances menores, em torno de 35% estão o Sport e o Cruzeiro na disputa de vaga. Todos os demais times tem chances inferiores à 20%.

Já na última rodada analisada, na rodada 26, o Corinthians em todas as simulações fica entre os 6 primeiros, tendo sua chance de vaga para a taça Libertadores de 100%. As chances do Grêmio caem de 97 para 94% enquanto as do Santos sobem de 84 para 95%. O Palmeiras aumenta suas chances para 75% enquanto o Flamengo e Cruzeiro têm chances em torno de 61%. O Botafogo dispara de 12% para 55% de chances de conseguir uma vaga. Fora esses times, somente o Atlético-MG e o Atlético-PR tem chances superiores à 10% de terminar o campeonato entre os 6 primeiros colocados.

Por fim, tem-se a análise das chances de ser rebaixado para a segunda divisão (estar entre os 4 últimos colocados). Nesse caso a tabela abaixo informa as chances no decorrer do campeonato.

Tabela 7 – Chances de ser rebaixado ao longo do Campeonato Brasileiro de Futebol 2017 - Modelo 2.

Time	Inicial	Rodada_10	Rodada_19	Rodada_26
Atlético-GO	59.04	85.68	95.78	82.19
Avaí	52.75	81.96	80.34	65.00
Bahia	24.37	37.79	24.22	35.08
Botafogo	27.90	16.40	18.66	0.65
Atlético-MG	10.53	14.90	19.51	13.69
Atlético-PR	48.03	29.50	16.80	14.09
Coritiba	12.01	9.35	12.44	61.53
Chapecoense	32.53	26.80	37.64	25.25
Corinthians	4.82	0.01	0.00	0.00
Cruzeiro	8.45	9.00	5.31	0.30
Flamengo	6.50	3.38	0.81	0.17
Fluminense	33.66	18.78	14.31	38.60
Grêmio	8.98	0.28	0.01	0.00
Palmeiras	17.21	8.20	0.87	0.09
Ponte Preta	24.49	17.74	24.91	44.80
Santos	12.71	5.48	0.06	0.00
São Paulo	29.88	40.67	72.45	55.49
Sport-PE	30.99	24.37	4.46	20.98
Vasco	37.66	22.44	19.63	15.16
Vitória	17.49	47.27	51.79	26.93

¹Note que os percentuais somam 400%, visto que 4 times serão rebaixados.

Baseando-se na Tabela 7, será analisado a seguir as chances de rebaixamento de cada time no decorrer do campeonato. Dessa forma, as chances iniciais de rebaixamentos (logo antes do início do campeonato), tem com maiores chances o Atlético-GO e o Avaí com 59 e 52,7%, respectivamente. Logo a seguir vem os Atlético-PR (48%), o Vasco (37,6%) e a Fluminense (33,6%). Outros times com chances próximas de 30% são a Chapecoense, o Sport e o São Paulo.

Ao fim da décima rodada, o Atlético-GO e o Avaí aumentam ainda mais suas chances para 85% e 82%, respectivamente. Nesse ponto o Vitória salta de 17% no início do campeonato para 47% de chances seguido pelo São Paulo que eleva suas chances de rebaixamento para 40%. O Bahia vem próximo com 37%. Nesse momento do campeonato o Atlético-PR, a Chapecoense, o Sport e o Vasco ainda tinham chances superiores à 20%.

Na metade do campeonato, as chances já se consolidam mais. O Atlético-GO passa para 95% de chance enquanto o Avaí recua para 80%. O São Paulo aumenta suas chances para 72,4% enquanto o Vitória passa a marca dos 50% de chance. Nesse momento, com chances que se destacam também vem a Chapecoense (37%), e a Ponte Preta e o Bahia com (24%). Nesse ponto, as chances de rebaixamento do Corinthians atigem 0,00%.

Ao fim da rodada 26, o Atlético-GO recua suas chances para 82% assim como o

Avaí que recua para 65%. O Coritiba aumenta suas chances de 12 para 61% enquanto o São Paulo reduz para 55%. Logo a seguir, vem a Ponte Preta com 44,8%, o Fluminense com 38,6% e o Bahia com 35%.

Nessa rodada, mais dois times se unem ao Corinthians com chances aproximadamente nulas de rebaixamento, sendo eles o Grêmio e o Santos. Porém times como o Palmeiras, Flamengo, Cruzeiro e Botafogo tem chances inferiores à 1%.

3.3 Comparação dos Modelos

Outra forma de análise que pode ser obtida das previsões é a chance de meta rodada a rodada e dessa forma ter uma ideia mais contínua de como se comportou determinada chance ao longo do campeonato. Nesse caso, dado os dois modelos, ainda será realizada a comparação para determinado time e meta de como os modelos se comportaram ao longo do campeonato. Para melhor dinâmica na análise, são analisados somente os times que estavam na disputa direta da meta (Título, vaga para a Libertadores ou rebaixamento).

Abaixo segue a análise dos times com maior chance de serem campeões. Nesse caso a linha contínua representa o modelo de Últimas Partidas enquanto a linha tracejada o de Confrontos Diretos.

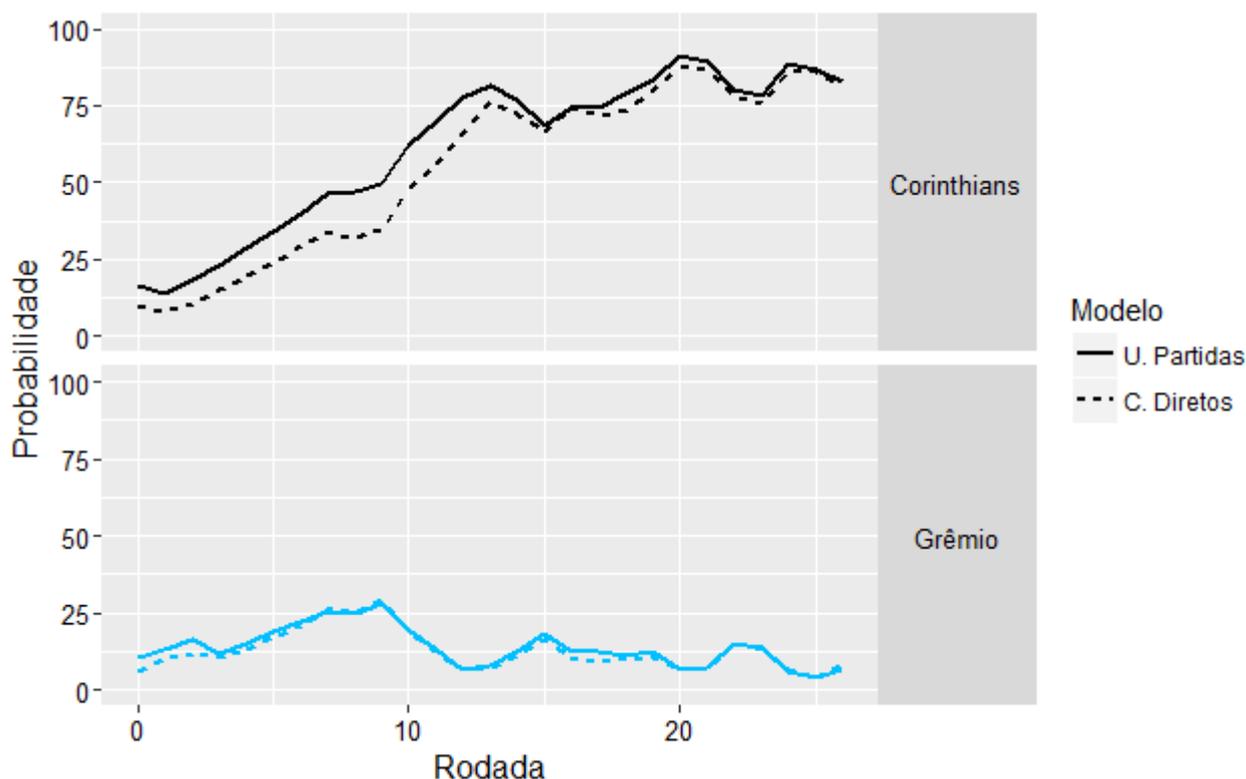


Figura 3 – Comparação dos Modelos de Previsão de Título para Corinthians e Grêmio. Campeonato Brasileiro de Futebol 2017.

Nota-se que o início do campeonato para o Corinthians o ajudou muito para aumentar suas chances de título, em seu caso também é possível notar que a estimativas de chances pelo modelo de confrontos diretos é inferior à curva de últimas partidas. Isso se dá pelo fato de que o modelo de últimas partidas considera sua boa campanha em campeonato estadual esse ano.

Outro time que foi considerado favorito de chances para o título, o Grêmio, tam-

bém tem suas chances apresentadas. No caso do Grêmio, pode notar que suas chances, independente do modelo estavam sempre próximas e da mesma forma que o Corinthians, o modelo de últimas partidas favoreceu seus resultados. Nota-se também que a partir da nona rodada suas chances caem e após isso, o time aparenta manter-se fora da disputa do título.

A seguir são analisados as chances de times em relação à vaga para Libertadores da América (estar entre os seis primeiros colocados). Nesse caso, iniciaremos por Palmeiras e Santos.

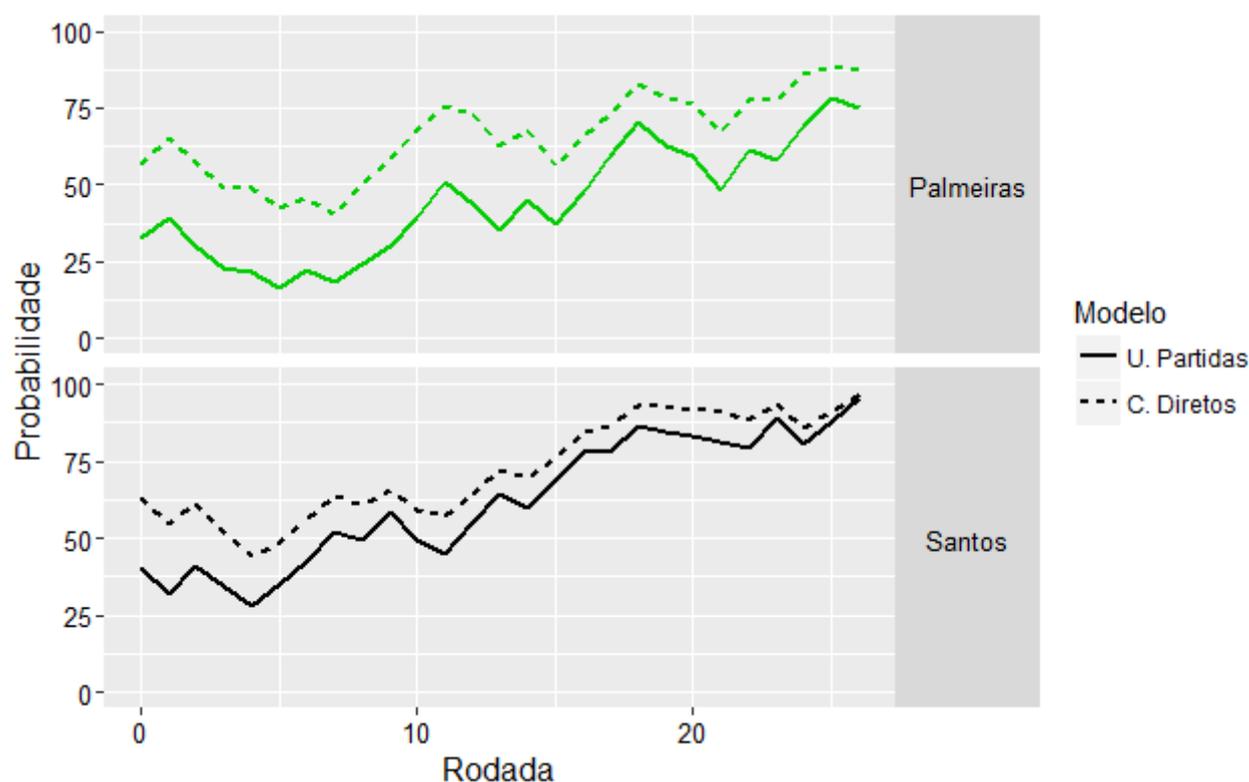


Figura 4 – Comparação dos Modelos de Previsão de Libertadores para Palmeiras e Santos. Campeonato Brasileiro de Futebol 2017.

Pode-se notar que tanto o Palmeiras como o Santos são times em que o modelo de confrontos diretos os atribui maior chance. Isso se dá pela boa campanha dos times nos dois últimos campeonatos brasileiros, sendo na edição de 2016, o Palmeiras terminou como campeão e Santos finalizou como vice.

Em relação as chances de conquistar vaga na Libertadores, ambos os times começam o campeonato com chances razoáveis, mas a medida que o resultado decorre, pode-se notar que as chances dos times aumentam gradualmente. Sendo a chance dos dois times alcançando valores acima de 75% em ambos os modelos.

Já analisando o Cruzeiro e Flamengo, pode-se notar que nesse caso, ambos são

times em que o modelo de última partidas atribui melhores chances de obter vaga para Libertadores. Nesse caso, o Cruzeiro sofre no decorrer da rodadas um queda gradual em suas chances que só voltam a realmente melhorarem a partir da rodada 20. Por outro lado, o Flamengo aumenta suas chances no início do campeonato e começa a variar em torno de 75% de chances de obter uma vaga.

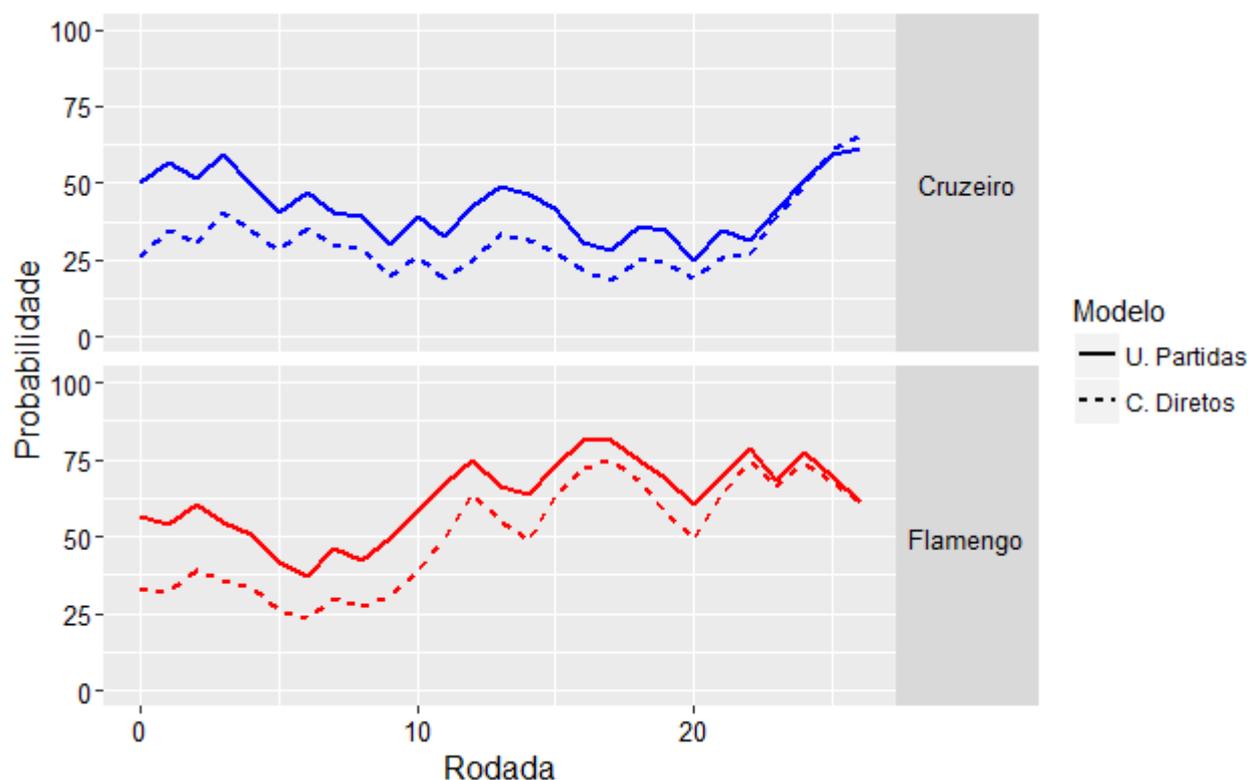


Figura 5 – Comparação dos Modelos de Previsão de Libertadores para Cruzeiro e Flamengo. Campeonato Brasileiro de Futebol 2017.

O rebaixamento, sendo também um ponto em que os times se preocupam, foi avaliado para os times que tiveram maiores chances de estar entre os quatro últimos. Nesse caso, são avaliados o Atlético-GO, o Avaí, o São Paulo e o Vitória.

Pode-se notar um caso interessante, o Atlético-GO por não ter disputado o Campeonato Brasileiro Série A durante 2015 e 2016, não teve nenhuma ocorrência de resultados pelo modelo de Confrontos Diretos. Assim, no seu caso, esse modelo iniciou-se em 26%, mas que dada a sequência de resultados ocorridos, as chances de rebaixamento aumentam gradualmente. Já o modelo de últimas partidas, inicia-se com uma chance de rebaixamento maior e dado isso, tem um aumento inferior no decorrer do campeonato.

Já o Avaí, inicia o campeonato com as duas curvas bem próximas e com chances de rebaixamento até alta. Isso se dá pelo má desempenho do time nos últimos anos no Campeonato Brasileiro e pela má campanha em campeonatos regionais. No decorrer do

campeonato as curvas se afastam, sendo que o modelo de confrontos diretos ainda mantém as chances de rebaixamento maiores.

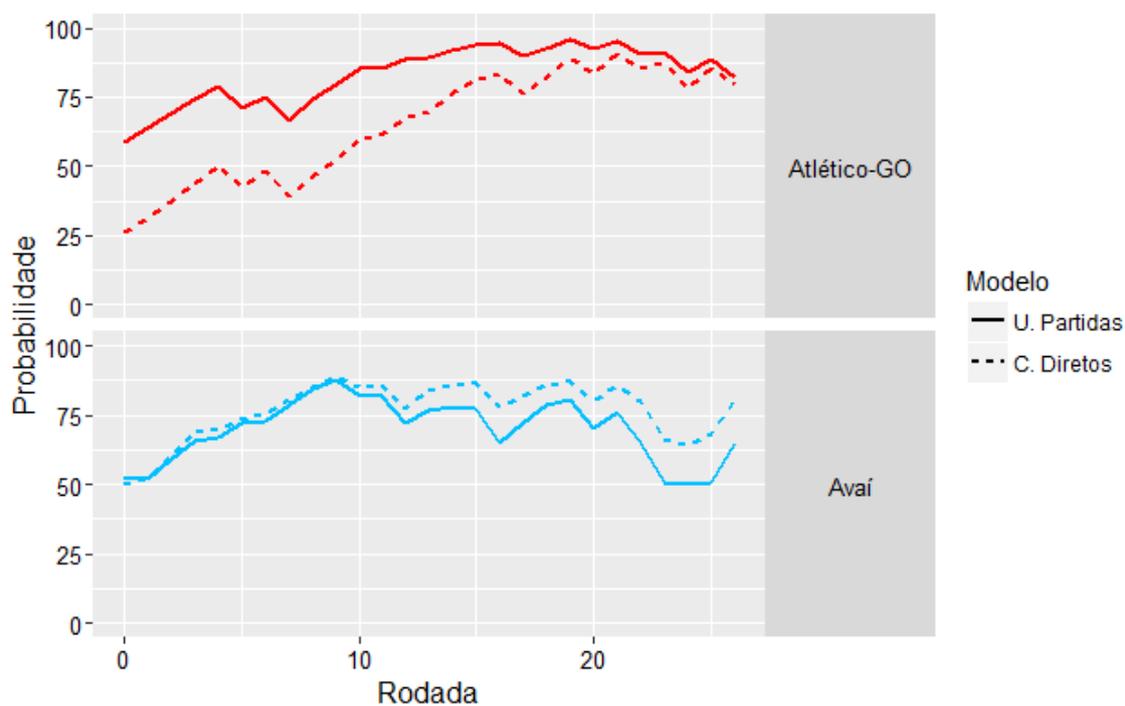


Figura 6 – Comparação dos Modelos de Previsão de Rebaixamento para Atlético-GO e Avaí. Campeonato Brasileiro de Futebol 2017.

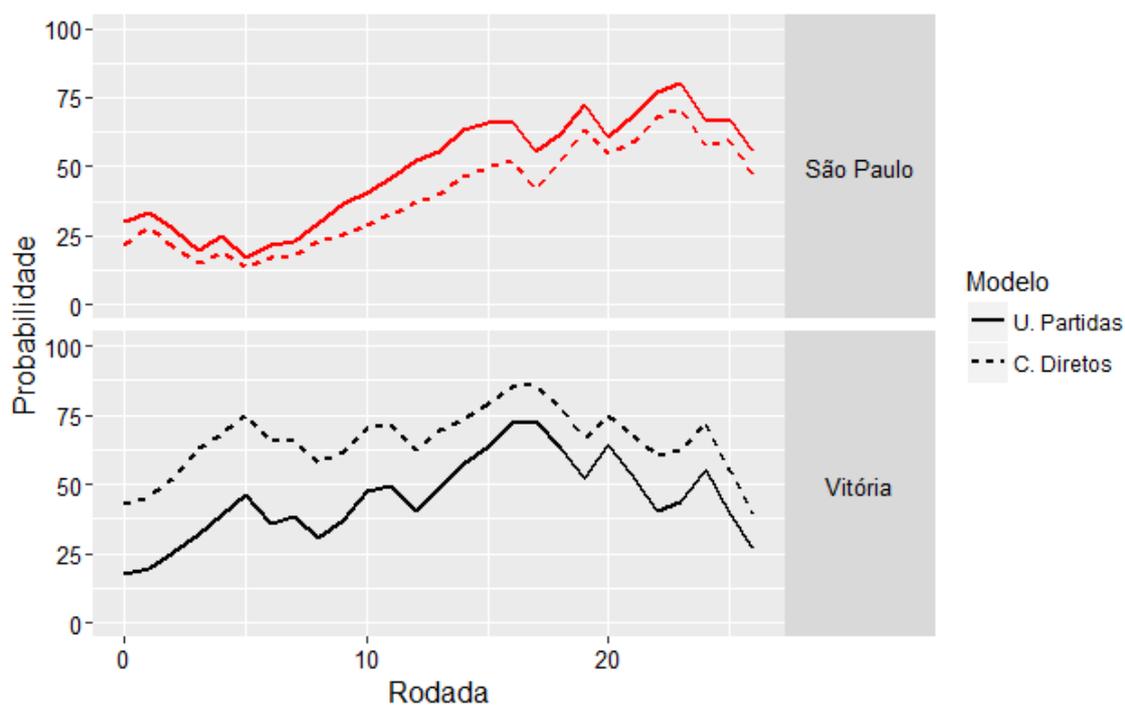


Figura 7 – Comparação dos Modelos de Previsão de Rebaixamento para São Paulo e Vitória. Campeonato Brasileiro de Futebol 2017.

O São Paulo e o Vitória, foram outros dois times que tiveram risco de rebaixamento alto no decorrer do campeonato. Dado isso, pode-se observar o comportamento de suas chances de rebaixamento ao longo do campeonato. Nesse caso, o São Paulo iniciou o campeonato com chances pequenas, mas nota-se que a partir da quinta rodada essas chances começam a crescer até atingir a marca de 75% pelo modelo de últimas partidas. Nas últimas rodadas também é possível ver essa chance se reduzindo.

Já no caso do Vitória, pode-se notar que as chances começam baixas e vão aumentando a medida que o campeonato ocorre de forma gradual e tendo suas variações. Também é possível ver que as chances reduzem nas últimas rodadas. Nesse caso o modelo de Confrontos Diretos atribui maior chance de rebaixamento.

Após a exposição dos resultados, é importante ter uma noção do diagnóstico da previsão desses modelos. Nesse ponto, foi calculada para cada confronto do campeonato a proporção de simulações em que o resultado simulado foi igual ao resultado oficial da partida, o que decorreu no campeonato.

3.4 Diagnóstico dos Modelos

Como definido na metodologia, para avaliar a qualidade dos modelos, foi usada a medida Percentual de Acertos de Previsão. Dessa forma após cada confronto simulado até a rodada 26, a última rodada considerada ocorrida para esse relatório, foi obtido a proporção de resultados que foram previstos de forma correta em relação ao resultado que ocorreu na realidade.

Dessa forma, cada partida possui um percentual correspondente sendo possível assim encontrar as partidas em que os modelos mais acertaram e as partidas em que mais erraram. Essa medida também é usada para comparar a qualidade dos dois modelos e avaliar qual teve resultados mais próximos da realidade.

Iniciando a análise, segue abaixo um histograma com a distribuição do percentual de acertos de previsão para cada um dos modelos:

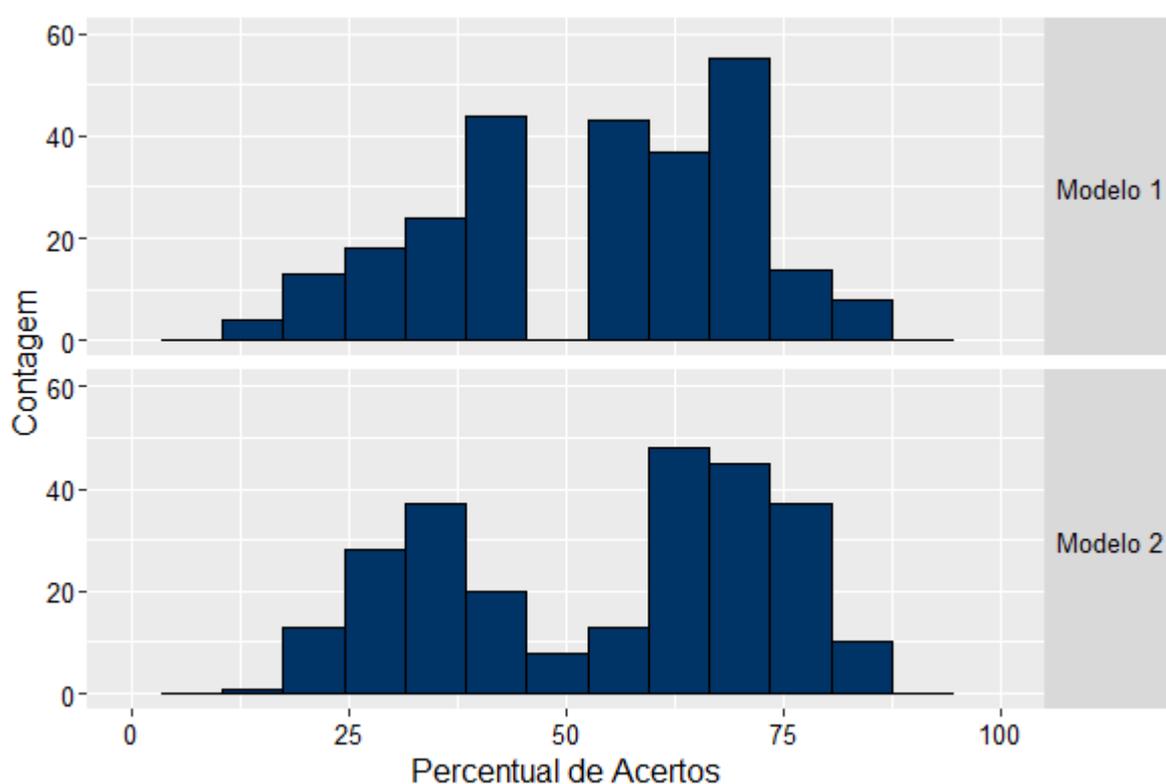


Figura 8 – Comparativo das Distribuições da Proporção de Acertos de Previsão. Jogos do campeonato brasileiro de futebol de 2017. Até a rodada 26.

Com base na Figura 6, é notável que os modelos possuem um comportamento semelhante, no caso, o modelo 1 possui em sua distribuição a ausência de observações em torno de 50%. No modelo 2 isso se dá de forma semelhante, porém há poucas observações, tendo uma distribuição bimodal. Em ambos os modelos existe uma predominância de

ocorrência entre valores entre 35 e 40% e em percentual de acertos entre 60 e 70%.

Para melhor avaliar os modelos, foram calculadas algumas medidas descritivas para cada um dos modelos e essas são apresentadas na tabela abaixo:

Tabela 8 – Medidas descritivas do Percentual de Acerto dos Modelos. Jogos do campeonato brasileiro de futebol de 2017. Até a rodada 26.

Medida	Modelo 1	Modelo 2
Média	53.23	54.55
Mínimo	14.40	13.19
Primeiro Quartil	39.57	35.34
Mediana	57.48	61.13
Terceiro Quartil	66.88	71.15
Máximo	86.18	84.97
Desvio Padrão	17.98	19.12

Inicialmente é válido ressaltar que os modelos possuem médias próximas, nesse caso o modelo 1 tem média de acerto de 53,23% enquanto o modelo 2 possui média de 54,55%. Em relação a mediana, os valores do percentual de acerto aumentam, sendo que metade dos confrontos tem valor superior à 57,48% no modelo 1 enquanto no modelo 2 esse valor é de 61,13%.

Em relação ao máximo, pode-se notar que nessa medida, o modelo de confrontos diretos tem um valor maior que o modelo de últimas partidas. O valor mínimo nesse caso é menor para o modelo 2, indicando uma maior amplitude que o modelo 1. Analisando o desvio padrão, pode-se notar que ele é superior para o modelo 2 sendo de 19,12%, enquanto no modelo 1 é de 18%.

Como o objetivo desses modelos é a previsão de um resultado entre três possíveis, tem-se que 1/3 pode ser tomado como o valor de chute ao acaso, ou seja, a probabilidade de acerto sem conhecimento algum. Assim outra forma de avaliar esses modelos é com o percentual de confrontos que tiveram percentual de acertos superior à 1/3. Essa informação indicaria o percentual de confrontos em que tem-se um ganho de informação. Abaixo seguem os percentuais obtidos.

Tabela 9 – Percentuais de Confrontos que tiveram Acerto de Previsão superior à 1/3. Jogos do campeonato brasileiro de futebol de 2017. Até a rodada 26.

Modelo	Percentual
Confrontos Diretos	81.15
Últimas Partidas	78.08

Nesse caso, o modelo 1 (Confrontos Diretos) teve um percentual superior à 80% sendo ligeiramente maior que o percentual do modelo 2 (Últimas Partidas) que teve um percentual ganho de informação em 78% dos confrontos. Assim dado que os modelos estão

sempre próximos também é válido estudar a relação entre eles, visando avaliar se existe algum comportamento simultâneo entre os dois modelos.

Dessa forma, é apresentado abaixo um gráfico de dispersão dos confrontos, nesse caso, cada ponto é um confronto ocorrido desde início do campeonato até a rodada 26 e em cada eixo seu respectivo percentual de acerto pelos modelos 1 e 2.

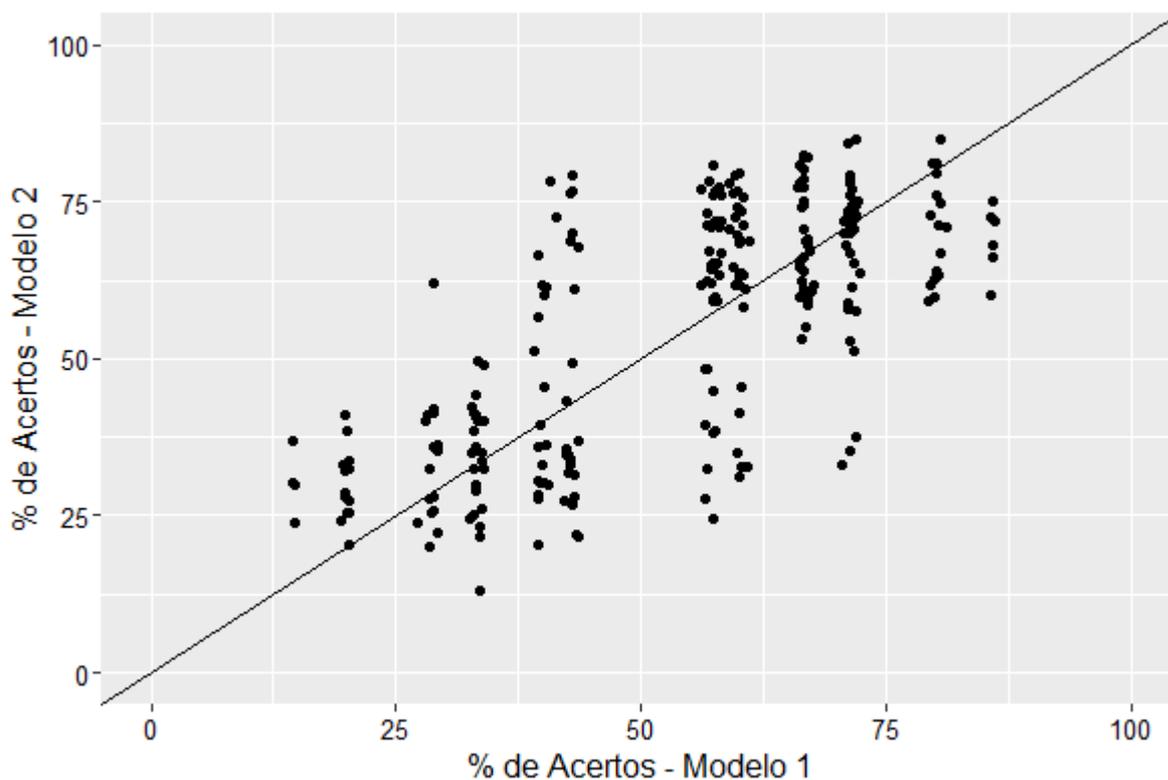


Figura 9 – Comparativo das Distribuições da Proporção de Acertos de Previsão. Jogos do campeonato brasileiro de futebol de 2017. Até a rodada 26.

Por meio dessa figura é possível notar uma leve tendência linear entre os dois modelos, porém é notável uma grande dispersão entre os modelos, tendo em vista que há confrontos, onde o acerto em um modelo é alto enquanto no outro é baixo.

Por esse gráfico se nota também que o modelo 1 possui um comportamento entre numa margem de valores de forma mais discreta, isso se dá pelo fato de que pela formulação do modelo, são poucos os confrontos diretos entre os clubes desde 2015 no Campeonato Brasileiro e isso restringe as probabilidades retornadas pela simulação da distribuição Dirichlet. Enquanto o modelo 2, acumula mais informações dado que conta todos os jogos de cada time, na temporada 2017, para a simulação, via Dirichlet, da probabilidade de cada um dos confrontos.

Buscando avaliar se algum dos modelos possui a capacidade de melhorar a previsão com o decorrer das rodadas é apresentado um gráfico com o percentual de acerto das

previsões para cada um das rodadas em cada um dos modelos.

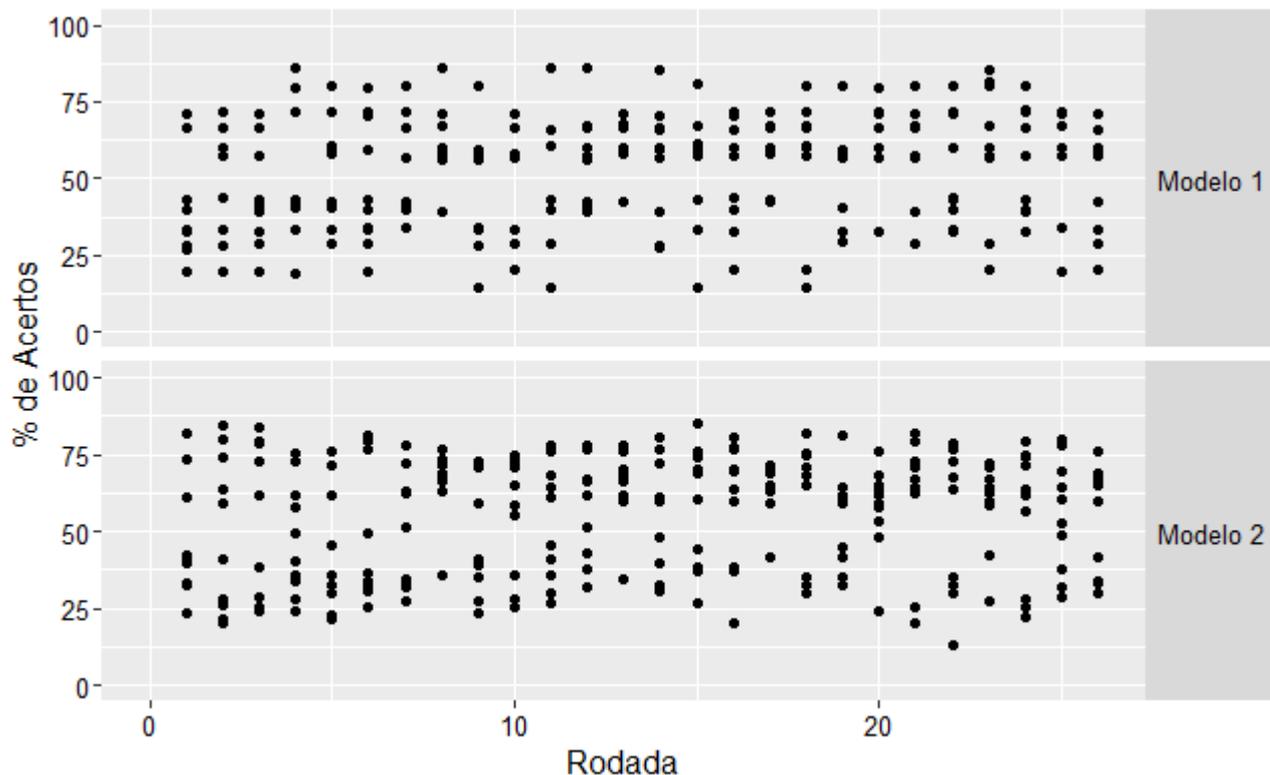


Figura 10 – Comparativo das Distribuições da Proporção de Acertos de Previsão em relação à rodada. Jogos do campeonato brasileiro de futebol de 2017. Até a rodada 26.

Assim por meio do gráfico acima, pode-se notar que em ambos os casos o acerto de previsão ocorre de forma aleatória, ou seja, não há um ganho notável em previsão dada a rodada do confronto que ela está prevendo.

Abaixo segue, como exemplo, os confrontos em que os modelos tiveram o melhor desempenho e também os confrontos em que os modelos tiveram o pior desempenho, pelo Percentual de Acerto de Previsão (PAP).

Tabela 10 – Confrontos com Maiores e Menores Percentuais de Acerto de Previsão para cada Modelo. Jogos do campeonato brasileiro de futebol de 2017. Até a rodada 26.

Rodada	Confrontos	PAP - Modelo 1	PAP - Modelo 2	Resultado Real
15	Flamengo x Palmeiras	0.14	0.37	2 x 2
22	Atlético-GO x Corinthians	0.34	0.13	1 x 0
8	Coritiba x Corinthians	0.86	0.72	0 x 0
15	Avaí x Corinthians	0.81	0.85	0 x 1

Dentre os confrontos, percebe-se que o confronto de pior desempenho no modelo

1 foi o Flamengo e Palmeiras, no qual se decorreu um empate sendo esse tido como o resultado menos esperado pelo modelo 1, no entanto pelo modelo 2 esse confronto teve resultado acima de 0,33. O mesmo decorreu para o confronto Atlético-GO e Corinthians onde o modelo 2 teve seu pior desempenho enquanto o modelo 1 se mantinha na marca de 33%.

Já em relação aos confrontos com boa performance dos modelos, em ambos os confrontos os dois modelos conseguem agregar informação, sendo o Avaí e Corinthians que decorreu em vitória do segundo, nesse ambos os modelos tem percentual de acerto de previsão acima de 80%, e o Coritiba e Corinthians que teve como desfecho o empate.

4 Conclusão

Finalmente, pode-se ter uma noção de como funcionou os modelos em relação aos resultados reais dos jogos, assim com base nas previsões obtidas e no diagnóstico de cada modelo, pode-se perceber que ambos os modelos agregam uma boa previsão. Também é notável ao longo da análise que cada um dos modelos valoriza um ponto em específico e ambos são modelos em que a estrutura permite entrar dados com diversos filtros.

Nesse trabalho, como para o modelo 1 foram considerados somente os jogos de 2015 e 2016 realizados no Campeonato Brasileiro Série A, a tendência é que os times em que tiveram boas campanhas nesses anos sejam favorecidos pelo modelo 1, isso pôde ser visto pelos probabilidades do Palmeiras e Santos, times que possuem bons retrospectos no campeonato brasileiro dos últimos anos.

Já o modelo 2, possui a tendência de beneficiar os times que possuíram boas pré-temporadas (disputas em campeonatos antes do campeonato Brasileiro), assim inicialmente, os times que possuem o ano de 2017 com melhores resultados se favorece nas previsões.

Com base nos resultados obtidos, pode-se ter uma noção do comportamento dos modelos em relação aos resultados reais dos jogos. Assim com base nas previsões obtidas e no diagnóstico de cada modelo, pode-se perceber que ambos os modelos agregam uma boa previsão. Também é notável ao longo da análise que cada um dos modelos valoriza um ponto em específico e ambos são modelos em que a estrutura permite entrar dados com diversos filtros.

Nesse trabalho, como para o modelo 1 foram considerados somente os jogos de 2015 e 2016 realizados no Campeonato Brasileiro Série A, a tendência é que os times em que tiveram boas campanhas nesses anos sejam favorecidos por esse modelo, isso pôde ser visto, por exemplo, pelas probabilidades do Palmeiras e Santos, times que possuem bons retrospectos no campeonato brasileiro dos últimos anos.

Já o modelo 2, possui a tendência de beneficiar os times que possuíram boas pré-temporadas (disputas em campeonatos antes do campeonato Brasileiro). Assim inicialmente, os times que possuem o ano de 2017 com melhores resultados se favorece nas previsões.

Com base nos resultados, pode-se perceber que ambos os modelos possuem uma eficiência e um ganho na previsão de resultados, mesmo cada um considerando pesos diferentes a um certo retrospecto. Dessa forma a escolha de um modelo a se usar pode ser direcionada por outro ponto. Em critério de eficiência, o modelo de confrontos diretos se atualiza somente na metade do campeonato (após o fechamento do primeiro turno) e possui um algoritmo mais eficiente. Já o modelo de últimas partidas se atualiza rodada à

rodada e com isso se torna computacionalmente mais oneroso.

Outra perspectiva possível é com base no volume de informações utilizadas. Como o modelo 2 consegue agregar informações de partidas independente do clube oponente, a sua predição pode se basear em um volume maior de dados. Em contrapartida, o modelo 1 baseia-se somente nos confrontos diretos, restringindo o volume de informações que são consideradas para previsão.

Por fim, para esse trabalho, foram feitas as restrições indicadas para cada modelo (o modelo de confrontos diretos considera resultados do Campeonato Brasileiro de 2015 e 2016 enquanto o modelo de últimas partidas considera os resultados que ocorreram em 2017 até o confronto), porém pode-se alterar essas restrições facilmente. Por se tratarem de modelos bayesianos, ambos podem obter previsões, sem considerar nenhuma informação de retrospecto, fazendo a previsão dos jogos apenas com a informação à priori. Além disso outras informações do jogo, como por exemplo, o mando de campo, desfalque de jogadores e qualidade do elenco podem ser facilmente agregadas ao modelo por meio da distribuição à priori dos parâmetros.

Referências

- ARRUDA, M. L. Poisson, bayes, futebol e defineti. 2000. Citado na página 16.
- BRUNELLO, G. H. V.; NAKANO, E. Y. Bayesian inference on proportional elections. 2015. Citado na página 15.
- EHLERS, R. S. *Introdução à Inferência Bayesiana*. [S.l.: s.n.], 2013. Citado na página 17.
- MACEDO, P. A. P. Predição de resultados no campeonato brasileiro 2012 série a. 2014. Citado na página 16.
- OGOL. acessado em 14/11/2017. <<http://www.ogol.com.br/>>. Citado na página 30.
- PAULINO, C. D. *Estatística Bayesiana*. [S.l.: s.n.], 2003. Citado na página 17.
- SUZUKI, A. K. Modelagem estatística para previsão esportiva: Uma aplicação no futebol. 2015. Citado na página 16.
- TEAM, R. C. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. [S.l.: s.n.], 2017. Citado na página 16.

5 Apêndice

Seguem as programações em R utilizadas para executar os modelos propostos neste trabalho.

5.1 Programação - Extração dos Dados

```

### Modelo 1
timeA=c(3129,2229,2230,2615,2231,2233,3195,2234,2235,2236,2240,2241,
2243,2248,2251,2254,2256,2257,2258,2259)
timeB=c(3129,2229,2230,2615,2231,2233,3195,2234,2235,2236,2240,2241,
2243,2248,2251,2254,2256,2257,2258,2259)

### CRIACAO DA LISTA DOS CONFRONTOS
k=0
lista=matrix(NA, nrow = 380,ncol=2)
for(i in 1:20){
for(j in 1:20){
if(i!=j){
k=k+1
lista[k,]=c(timeA[i],timeB[j])
}}}

### ALGORITMO DE COLETA DOS DADOS (WEBSCRAPING)
k=0
dados=NULL

### ALGORITMO ADAPTADO COM O TEMPO DE COLETAS
a=Sys.time()
b=Sys.time()
while(length(lista[,1])>length(dados)){
a=Sys.time()
while(a>b+1800){
i=0
while(i<17){
i=i+1

```

```

k=k+1
dados[[k]]=readLines(paste0("http://www.ogol.com.br/
xray.php?grp=1&id_comp=0&ond=&epoca_ini=144&epoca_fim=0&equipa_id=",
lista[k,1],"&equipa_vs_equipa_id=",
lista[k,2],"&player_detail="))
print(paste0("Conclui-se:",round(k/380*100), "%"))
print(unlist(lapply(dados, length)))
b=Sys.time()
}
}
}

##### _____ MONTANDO BANCO DE DADOS
### PROGRAMACAO DO TCC
### DESENVOLVIMENTO DO BANCO DE DADOS
library(readxl)

### ABRINDO O BANCO DE DADOS
load("C:/Users/Bruno Matos/Documents/UnB - Estatística/
TCC/6. Dados e Algoritmo/Arquivos Aleatorios/Dados_Brutos_TCC.RData")

### AJUSTANDO AS VARIÁVEIS DE APOIO (SIGLA DOS TIMES E CODIGO DOS MESMOS)
siglas<-read_excel("C:\\Users\\Bruno Matos\\Documents\\UnB -Estatística\\
TCC\\6. Dados e Algoritmo\\Dados.xlsx",sheet = "ID")
timeA=siglas$cod
timeB=siglas$cod
siglaA=siglas$Sigla
siglaB=siglas$Sigla

k=0
lista=data.frame(TA=rep(NA, times=380),
TB=rep(NA, times=380),
SA=rep(NA, times=380),
SB=rep(NA, times=380))
for(i in 1:20){
for(j in 1:20){
if(i!=j){

```

```
k=k+1
lista[k,]=c(timeA[i],siglaA[i],timeB[j],siglaB[j])
}
}
}

### CRIANDO A BASE DE DADOS FINALIZADA DE TODOS OS CONFRONTOS
result=data.frame(timeA=lista[,2],
VA=rep(NA, times=380),
Emp=rep(NA, times=380),
VB=rep(NA, times=380),
timeB=lista[,4],
codA=lista[,1],
codB=lista[,3])

### DESENVOLVENDO A BASE EM MEIO DE CONDICIONAIS POR
### CAUSA DOS JOGOS SEM OCORRENCIA NO PERIODO
for(i in 1:length(dados)){
if(length(dados[[i]])==93){
result$VA[i]=0
result$Emp[i]=0
result$VB[i]=0
}
if(length(dados[[i]])==139){
confronto=dados[[i]][102]
conf=strsplit(confronto,",")
conf2<-conf[[1]][c(4,6,8)]
conf3<-as.numeric(gsub(pattern = ",",replacement = "",x = conf2))
result$VA[i]=conf3[1]
result$Emp[i]=conf3[2]
result$VB[i]=conf3[3]
}
}

### CONTABILIZANDO SOMENTE OS JOGOS DO BRASILEIRAO SERIE A OU B
result2=data.frame(timeA=lista[,2],
VA=rep(NA, times=380),
```

```

Emp=rep(NA, times=380),
VB=rep(NA, times=380),
timeB=lista[,4],
codA=lista[,1],
codB=lista[,3])

for(i in 1:length(dados)){
  if(length(dados[[i]])==93){
    result2$VA[i]=0
    result2$Emp[i]=0
    result2$VB[i]=0
  }
  if(length(dados[[i]])==139){
    xx=dados[[i]][124]
    regex<-"Brasileirão"
    gg<-gregexpr(regex,xx)
    ss<-substring(xx,gg[[1]][1],gg[[1]][1]+200)
    a<-strsplit(ss,split = "%")
    regex2<-"tbl_stats"
    gg2 <- gregexpr(regex2,ss)
    result2$VA[i]<-substring(ss,gg2[[1]][2]+11,gg2[[1]][2]+11)
    result2$Emp[i]<-substring(ss,gg2[[1]][3]+11,gg2[[1]][3]+11)
    result2$VB[i]<-substring(ss,gg2[[1]][4]+11,gg2[[1]][4]+11)
  }
}

for(i in 1:380){
  if(is.na(result2$VB[i])){
    result2$VB[i]=0
  }
  if(is.na(result2$Emp[i])){
    result2$Emp[i]=0
  }
  if(is.na(result2$VA[i])){
    result2$VA[i]=0
  }
}

result2$VA=as.numeric(result2$VA)

```

```
result2$VB=as.numeric(result2$VB)
result2$Emp=as.numeric(result2$Emp)

### RETIRAR RESULTADOS ENTRE O INICIO DO CAMPEONATO E 30/05

result2$atual=rep(NA, times=380)
for(i in 1:380){
  if(sum(result2$VA[i],result2$Emp[i],result2$VB[i])%2!=0){
    result2$atual[i]=1
  }
  if(sum(result2$VA[i],result2$Emp[i],result2$VB[i])%2==0){
    result2$atual[i]=0
  }
}

### Modelo 2
library(readxl)
library(data.table)

#_____WEBSCRAPING PARA OBTENCAO DOS DADOS BRUTOS
### DEFINICAO DOS OBJETOS
time=c(3129,2229,2230,2615,2231,2233,3195,2234,2235,2236,2240,2241
,2243,2248,2251,2254,2256,2257,2258,2259)

### CRIACAO DA LISTA DOS CONFRONTOS

### ALGORITMO DE COLETA DOS DADOS (WEBSCRAPING)
k=0
page1=NULL
page2=NULL

### ALGORITMO ADAPTADO COM O TEMPO DE COLETAS
a=Sys.time()
b=Sys.time()
while(length(time)>k){
a=Sys.time()
```

```

while(a>b+1800){
  i=0
  while(i<16){
    i=i+2
    k=k+1
    if(file.exists(paste0("Time_",k,"_p1.rds"))){print("O arquivo está baixado")}
    else{
      page1<-readLines(paste0("http://www.ogol.com.br/team_results.php?grp=0&
ond=&compet_id_jogos=0&epoca_id=146&ano=2017&ano_fim=2017&id=",
time[k],"&menu=results&type=year&epoca_id_fim=146&comfim=0&page=1"))
      saveRDS(page1,file=paste0("Time",k,"p1.rds"))
      page2<-readLines(paste0("http://www.ogol.com.br/team_results.php?grp=0&
ond=&compet_id_jogos=0&epoca_id=146&ano=2017&ano_fim=2017&id=",
time[k],"&menu=results&type=year&epoca_id_fim=146&comfim=0&page=2"))
      saveRDS(page2,file=paste0("Time",k,"p2.rds"))}
      print(paste0("Conclui-se:",round(k/20*100), "%"))
      print(unlist(lapply(page1, length)))
      b=Sys.time()
    }
  }
}

#save(dados, file="Coleta dos Resultados.RData")
save(dados, file="Coleta dos Resultados2.RData")

#-----MANIPULACAO DA BASE DE DADOS
dados2=rep(NA,times=40)
j=1
for(k in 1:20){
  for(i in 1:2){
    eval(parse(text=paste0("time_",k,"_pag_",i,"<-readRDS(file='Time",k,"p",i,".rds')")))
    eval(parse(text=paste0("prov<-time_",k,"_pag_",i,"[97]")))
    dados2[j]<-prov
    j=j+1
  }
}

todos<-NULL

```

```
banco=NULL
geral=NULL
lista=NULL
for(i in 1:length(dados2)){
  x<-strsplit(dados2[i],"Resultados")
  x2=strsplit(x[[1]][2],"sign")
  for(j in 2:length(x2[[1]])){
    lista[j-1]<-regmatches(x2[[1]][j],gregexpr(">[^<]+<",x2[[1]][j]))
    banco[[j-1]]<-unlist(lista[j-1])
    todos[[j-1]]<-banco[[j-1]]
  }
  geral[[i]]<-todos
}

save(geral, file="Lista com os Resultados.RData")

## Dados de Confrontos anteriores
## Dados para armazenamento dos resultados
infos<-read_excel("C:\\Users\\Bruno Matos\\Documents\\
UnB - Estatística\\TCC\\6. Dados e Algoritmo\\M2 - Padrao Coleta.xlsx",
sheet = "Sheet2")

names(geral)<-infos$time

base<-NULL
for(i in 1:length(geral)){
  for(j in 1:length(geral[[i]])){
    if(length(geral[[i]][[j]])==10){
      A1<-as.vector(geral[[i]][[j]][2])
      A2<-as.vector(names(geral)[i])
      A3<-as.vector(geral[[i]][[j]][1])
      A4<-as.vector(geral[[i]][[j]][5])
      A5<-as.vector(geral[[i]][[j]][7])
      data_temp <- data.table(
        data = gsub(">(.*<","\\1",A1),
        time = gsub(">(.*<","\\1",A2),
        resultado= gsub(">(.*<","\\1",A3),
        oponente = gsub(">(.*<","\\1",A4),
        Campeonato =gsub(">(.*<","\\1",A5)
```

```

)

}
if(length(geral[[i]][[j]])==11){
A1<-as.vector(geral[[i]][[j]][2])
A2<-as.vector(names(geral)[i])
A3<-as.vector(geral[[i]][[j]][1])
A4<-as.vector(geral[[i]][[j]][5])
A5<-as.vector(geral[[i]][[j]][8])
data_temp <- data.table(
data = gsub(">(.*<", "\\1",A1),
time = gsub(">(.*<", "\\1",A2),
resultado= gsub(">(.*<", "\\1",A3),
oponente = gsub(">(.*<", "\\1",A4),
Campeonato =gsub(">(.*<", "\\1",A5)
)
}
base<-rbindlist(list(base,data_temp),fill=T)
}
}

```

5.2 Programação - Modelo Confrontos Diretos

```
### PACOTES UTILIZADOS
```

```
library(readxl)
```

```
library(gtools)
```

```
## Dados de Confrontos anteriores
```

```
dados<-read.csv2("C:\\Users\\Bruno Matos\\Documents\\UnB -
  Estatistica\\TCC\\6. Dados e Algoritmo\\M1 - Dados.csv")
```

```
## Dados para armazenamento dos resultados
```

```
dados2<-read_excel("C:\\Users\\Bruno Matos\\Documents\\UnB - Estatistica\\TCC\\
  6. Dados e Algoritmo\\M1 - Padrao Coleta.xlsx",sheet = "Sheet2")
```

```
## Dados para a atualizacao no decorrer do campeonato(resultados que ocorreram)
```

```
dados3<-read_excel("C:\\Users\\Bruno Matos\\Documents\\UnB -
  Estatistica\\TCC\\6. Dados e Algoritmo\\M1 - Resultados Atualizados.xlsx")
```

```
### CRIANDO COLETORES DE INFORMACOES
modelo1<-function(rod,repet){
  ## Banco de dados para hospedar posicao
  pos=rbind.data.frame(rep(0,20))
  names(pos)<-as.character(levels(dados$timeA))

  ## Banco de dados para hospedar pontos conquistados
  pts=rbind.data.frame(rep(0,20))
  names(pts)<-as.character(levels(dados$timeA))

  ## Banco de dados para hospedar numero de vitorias
  nvit=rbind.data.frame(rep(0,20))
  names(nvit)<-as.character(levels(dados$timeA))

  ## Banco de dados para hospedar acerto da previsao
  Prev=rbind.data.frame(rep(NA,380))
  names(Prev)<-as.character(dados$conf)

  ## Transformando os resultados em uma lista semelhante a do algoritmo
  resul.teste=data.frame(VA=rep(NA,380),Emp=rep(NA,380),
  VB=rep(NA,380),timeA=rep(NA,380),
  Emp=rep("TIE",380),timeB=rep(NA,380),conf=rep(NA,380),rodada=dados3$rodada)
  prob<-data.frame(p.VA=rep(NA,380),p.Emp=rep(NA,380),p.VB=rep(NA,380))

  ## Criando Variavel para hospedar o Acerto Medio de Previsao
  dados$AMP<-rep(0,380)

  ## SIMULANDO OS RESULTADOS DOS JOGOS DADO A PRIORI
  for(k in 1:repet){
    set.seed(k)
    for(i in 1:length(dados$codB)){
      prob[i,]<-rdirichlet(1,alpha=c(dados$VA[i]+1,dados$Emp[i]+1,dados$VB[i]+1))
      multi<-rmultinom(1,1,prob=c(prob[i,1],prob[i,2],prob[i,3]))
      resul.teste[i,1]<-multi[1]
      resul.teste[i,2]<-multi[2]
      resul.teste[i,3]<-multi[3]
      resul.teste[i,4]<-as.character(dados$timeA[i])
      resul.teste[i,6]<-as.character(dados$timeB[i])
    }
  }
}
```

```

resul.teste[i,7]<-paste0(dados$timeA[i],":",dados$timeB[i])

if(i<190){
dados$VA[i+190]=dados$VA[i]+resul.teste[i,1]
dados$Emp[i+190]=dados$Emp[i]+resul.teste[i,2]
dados$VB[i+190]=dados$VB[i]+resul.teste[i,3]
}
}

## ADICIONANDO OS RESULTADOS SIMULADOS AOS DADOS (PARA O RETORNO)
dados2<-data.frame(time=as.character(names(table(dados2$time))),
pts=rep(0,times=20),nvit=rep(0,times=20))

resul2<-resul.teste[,c(4,1:3,6:8)]
dados3a<-dados3[,c(1:5,9,10)]
rodada=rod
if(rodada>0){
resultados<-rbind.data.frame(resul2[resul2$rodada>rodada,],
dados3a[dados3a$rodada<=rodada,])

## COMPARANDO O PREVISTO COM O OCORRIDO
for(l in 1:(rodada*10)){
dados$AMP[l]<-dados3a[l,2:4]==resul.teste[l,1:3]
}
}
else{resultados<-resul2}

## MONTANDO A CONFIGURACAO DE CLASSIFICACAO
for(i in 1:length(resultados$VA)){

# SE TIME A VENCER
if(resultados$VA[i]==1){
for(j in 1:length(dados2$time)){
if(resultados$timeA[i]==dados2$time[j]){
dados2$pts[j]<-dados2$pts[j]+3
dados2$nvit[j]<-dados2$nvit[j]+1
}
else{dados2$pts[j]<-dados2$pts[j]
dados2$nvit[j]<-dados2$nvit[j]}
}
}
}
}

```

```
# SE TIME B VENCER
if(resultados$VB[i]==1){
for(j in 1:length(dados2$time)){
if(resultados$timeB[i]==dados2$time[j]){
dados2$pts[j]<-dados2$pts[j]+3
dados2$nvit[j]<-dados2$nvit[j]+1
}
else{dados2$pts[j]<-dados2$pts[j]
dados2$nvit[j]<-dados2$nvit[j]}
}
}
# SE OCORRER EMPATE
if(resultados$Emp[i]==1){
for(j in 1:length(dados2$time)){
if(resultados$timeA[i]==dados2$time[j]){
dados2$pts[j]<-dados2$pts[j]+1
dados2$nvit[j]<-dados2$nvit[j]
}
if(resultados$timeB[i]==dados2$time[j]){
dados2$pts[j]<-dados2$pts[j]+1
dados2$nvit[j]<-dados2$nvit[j]
}
else{dados2$pts[j]<-dados2$pts[j]
dados2$nvit[j]<-dados2$nvit[j]}
}
}
}
pts[k,]=dados2$pts
pos[k,]=rank(order(order(dados2$pts,dados2$nvit, decreasing = T)),
  ties.method = "random")
nvit[k,]=dados2$nvit
names(pos)<-as.character(dados2$time)
names(pts)<-as.character(dados2$time)
names(nvit)<-as.character(dados2$time)
Prev[k,]<-dados$AMP
}
return(list(pts=pts,pos=pos,nvit=nvit,APrev=Prev))
}
```

5.3 Programação - Modelo Últimas Partidas

```
### PACOTES UTILIZADOS
library(readxl)
library(gtools)

## Dados dos resultados das ultimas partidas
base<-read.csv2("C:\\Users\\UnB
\\Desktop\\Bruno\\M2 - Base de restrospecto_Final.csv")

## Dados de para hospedar as probabilidades
dados<-read.csv2("C:\\Users\\UnB\\Desktop\\Bruno\\M2 - Dados.csv")

## Dados para armazenamento dos resultados
dados2<-read_excel("C:\\Users\\UnB\\Desktop
\\Bruno\\M1 - Padrao Coleta.xlsx",sheet = "Sheet2")

## Dados para a atualizacao no decorrer do campeonato(resultados que ocorreram)
dados3<-read_excel("C:\\Users\\UnB\\Desktop\\
Bruno\\M1 - Resultados Atualizados.xlsx")

### TRATANDO A QUESTAO DAS DATAS DAS BASES
base$data<-as.Date(as.character(base$data), format = "%d/%m/%Y")
dados$data<-as.Date(as.character(dados$data), format = "%d/%m/%Y")

### CRIANDO COLETORES DE INFORMACOES
## Banco de dados para hospedar posicao
pos=rbind.data.frame(rep(0,20))
names(pos)<-as.character(names(table(dados3$timeA)))

## Banco de dados para hospedar pontos conquistados
pts=rbind.data.frame(rep(0,20))
names(pts)<-as.character(names(table(dados3$timeA)))

## Banco de dados para hospedar numero de vitorias
nvit=rbind.data.frame(rep(0,20))
names(nvit)<-as.character(names(table(dados3$timeA)))

## Banco de dados para hospedar acerto da previsao
```

```
Prev=rbind.data.frame(rep(NA,380))
names(Prev)<-as.character(dados3$conf)

## Transformando os resultados em uma lista semelhante a do algoritmo
resul.teste=data.frame(VA=rep(NA,380),Emp=rep(NA,380),
VB=rep(NA,380),timeA=rep(NA,380),
Emp=rep("TIE",380),timeB=rep(NA,380),conf=rep(NA,380),rodada=dados3$rodada)
prob<-data.frame(p.VA=rep(NA,380),p.Emp=rep(NA,380),p.VB=rep(NA,380))

## Criando Variavel para hospedar o Acerto Medio de Previsao
dados$AMP<-rep(0,380)

modelo2<-function(rod,repet){
## SIMULANDO OS RESULTADOS DOS JOGOS DADO A PRIORI
for(k in 1:repet){
set.seed(k)
simul<-data.frame(time=dados2$time, matrix(0,nrow = 20,ncol = 3))
for(i in 1:length(dados$codB)){
if(i<=rod2*10){
provA<-base[base$time==dados$timeA[i] & base$data<dados$data[i],]
provB<-base[base$time==dados$timeB[i] & base$data<dados$data[i],]
TA<-table(provA$resultado)
TB<-table(provB$resultado)

### Obtendo os thetas
PA<-rdirichlet(1,alpha=TA+1)
PB<-rdirichlet(1,alpha=TB+1)
}

if(i>rod2*10){
TA<-as.numeric(table(provA$resultado)+
rev(simul[simul$time==provA$time,2:4])[1,])
TB<-as.numeric(table(provB$resultado)+
rev(simul[simul$time==provB$time,2:4])[1,])

### Obtendo os thetas
PA<-rdirichlet(1,alpha=TA+1)
PB<-rdirichlet(1,alpha=TB+1)
}
}
```



```
## MONTANDO A CONFIGURACAO DE CLASSIFICACAO
for(i in 1:length(resultados$VA)){

# SE TIME A VENCER
if(resultados$VA[i]==1){
for(j in 1:length(dados2$time)){
if(resultados$timeA[i]==dados2$time[j]){
dados2$pts[j]<-dados2$pts[j]+3
dados2$nvit[j]<-dados2$nvit[j]+1
}
else{dados2$pts[j]<-dados2$pts[j]
dados2$nvit[j]<-dados2$nvit[j]}
}
}
# SE TIME B VENCER
if(resultados$VB[i]==1){
for(j in 1:length(dados2$time)){
if(resultados$timeB[i]==dados2$time[j]){
dados2$pts[j]<-dados2$pts[j]+3
dados2$nvit[j]<-dados2$nvit[j]+1
}
else{dados2$pts[j]<-dados2$pts[j]
dados2$nvit[j]<-dados2$nvit[j]}
}
}
# SE OCORRER EMPATE
if(resultados$Emp[i]==1){
for(j in 1:length(dados2$time)){
if(resultados$timeA[i]==dados2$time[j]){
dados2$pts[j]<-dados2$pts[j]+1
dados2$nvit[j]<-dados2$nvit[j]
}
if(resultados$timeB[i]==dados2$time[j]){
dados2$pts[j]<-dados2$pts[j]+1
dados2$nvit[j]<-dados2$nvit[j]
}
else{dados2$pts[j]<-dados2$pts[j]
dados2$nvit[j]<-dados2$nvit[j]}
}
```

```
}  
}  
}  
pts[k,]=dados2$pts  
pos[k,]=rank(order(order(dados2$pts,dados2$nvit, decreasing = T)),  
ties.method = "random")  
nvit[k,]=dados2$nvit  
names(pos)<-as.character(dados2$time)  
names(pts)<-as.character(dados2$time)  
names(nvit)<-as.character(dados2$time)  
Prev[k,]<-dados$AMP  
}  
return(list(pts=pts,pos=pos,nvit=nvit,APrev=Prev))  
}
```