

Cartola FC Data Analysis: Uma Ferramenta para Simulação, Análise e Visualização de Dados para o Fantasy Game Cartola-FC

Cartola FC Data Analysis: A simulation, analysis, and visualization tool based on Cartola FC Fantasy Game

Edson Mota Danilo Coimbra Maycon Peixoto
Universidade Federal da Bahia (UFBA) Universidade Federal da Bahia (UFBA) Universidade Federal da Bahia (UFBA)
Salvador, Brasil Salvador, Brasil Salvador, Brasil
edsonmottac@gmail.com coimbra.danilo@ufba.br maycon.leone@ufba.br

ABSTRACT

The search for computational models based on the prediction of scenarios has attracted the attention of many researchers over the years. Within this context, Fantasy Games work as social media environments that mix real information on a virtual world, offering a challenging scenario. This work aims to explore this scenario by introducing a simulation, analysis, and visualization tool, based on the Fantasy Game – Cartola FC. This tool automatically generates teams by using two distinct strategies based on statistical models for each tournament round, simulating an entire championship. From these theoretic models, we create a visualization tool to support user analysis. Lastly, we perform the planning of experiments to determine the influence of certain aspects of team's composition.

CCS CONCEPTS

• Information systems → Social networking sites;

KEYWORDS

Fantasy Games, Algorithms, Data Visualization, Simulation, Data Analysis

ACM Reference Format:

Edson Mota, Danilo Coimbra, and Maycon Peixoto. 2018. Cartola FC Data Analysis: Uma Ferramenta para Simulação, Análise e Visualização de Dados para o Fantasy Game Cartola-FC [1ex] *Cartola FC Data Analysis: A simulation, analysis, and visualization tool based on Cartola FC Fantasy Game*. In *SBSI'18: XIV Brazilian Symposium on Information Systems, June 4–8, 2018, Caxias do Sul, Brazil*. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3229345.3229366>

1 INTRODUÇÃO

Estabelecido como um dos esportes mais populares da atualidade [1], o futebol tem cumprido um importante papel no cenário Brasileiro. Seja por seu caráter inclusivo, promovendo a integração entre participantes em todas as faixas etárias e em diferentes níveis

sociais [1], seja pelo seu apelo comercial, visto a enorme quantia e importância financeira, ou ainda por seu apelo emocional, que torna cada um dos expectadores e torcedores um técnico em potencial.

Pensando nesta demanda, plataformas conhecidas como *Fantasy Football* (FF) têm obtido grande notoriedade nos últimos anos dentro de setor de entretenimento [5]. Além de possibilitar ao fã de esporte acompanhar de modo mais assíduo o futebol, o estilo FF (e os jogos de *Fantasy Game* de modo geral) tem uma característica social relevante. A socialização ocorre jogando com amigos, família, colegas de trabalhos, usuários aleatórios que extrapolam seus vínculos das comunidade reais para a criação de comunidades, ligas ou sociedades no mundo virtual FF [20], tornando FF aplicações com características de mídias sociais. Estas plataformas procuram direcionar o desejo destes expectadores no tocante a aplicação de suas estratégias em cenários similares a realidade dos campeonatos de futebol. Assim, expectores têm agora ao alcance de suas mãos a possibilidade de escalar equipes, implantar e testar estratégias com foco nos melhores resultados.

Um aspecto importante das plataformas de *Fantasy Football* consiste no fato de que sua representação fictícia sofre influência direta dos acontecimentos reais [5]. Deste modo, um jogador que se lesiona em uma determinada partida durante o campeonato real, poderá ou não estar disponível para uma próxima partida na plataforma *Fantasy Football*. Do mesmo modo, se um jogador tem desempenhos acima da média em uma determinada fase do campeonato real, este desempenho será reproduzido no cenário virtual, influenciando em aspectos como pontuação média ou preço do referido jogador. Esta relação de influência impele que os usuários das plataformas *Fantasy Football* tenham um razoável conhecimento acerca dos acontecimentos do campeonato. No Brasil, a mais famosa plataforma de *Fantasy Football* é conhecida como Cartola-FC¹ e foi criado pela Rede Globo de Televisão no ano de 2005. Atualmente, esta plataforma conta com aproximadamente 5.5 milhões de times [6], evidenciando a sua crescente presença no cenário nacional.

Diante de tal representatividade e concomitante aos avanços relacionados às tecnologias de informação, pesquisadores vêm vislumbrando possibilidades de aplicar modelos computacionais inteligentes na construção de estratégias mais assertivas para a análise de tendências que possam promover conhecimentos que contribuam na identificação das melhores estratégias e resultados [2, 5, 23]. Em geral, os trabalhos neste campo enfatizam o desenvolvimento de

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SBSI'18, June 4–8, 2018, Caxias do Sul, Brazil

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-6559-8/18/06.

<https://doi.org/10.1145/3229345.3229366>

¹Disponível em: <http://globoesporte.globo.com/cartola-fc/>. Acessado em 18/12/2017

algoritmos e estratégias de seleção baseadas em: i) análises estatísticas, voltadas à confecção de algoritmos determinísticos; ii) aplicação de técnicas baseadas em aprendizado de máquina [8], utilizando métodos estocásticos para identificar o melhor resultado diante de um conjunto de critérios de entrada; ou ainda iii) abordagens de visualização para transmitir informações de resultados em cada etapa do torneio [22], identificar padrões táticos e estratégicos de jogos [13] ou avaliar qual a melhor visualização para determinadas tarefas de usuários [3].

Este artigo tem como objetivo apresentar um modelo de análise e visualização de dados baseado em simulação. A partir de alguns *datasets* do Cartola FC, são empregados modelos estatísticos para implementação de algoritmos cujo foco é fornecer direcionamento sobre cenários mais rentáveis dentro do universo do jogo. Estes cenários são carregados a começar de uma semente ou conjunto de critérios iniciais e executados de forma autônoma permitindo analisar tendências e comparações com base em diferentes configurações.

Além da introdução, o restante deste artigo está dividido do seguinte modo: Na Seção 2 são apresentados alguns conceitos fundamentais presentes neste estudo, abordando, com base na literatura, temas como: a implementação e utilização dos ambientes de simulação baseados em dados históricos, além de uma análise acerca da visualização de dados, tratando mais especificamente da representação de séries temporais. Na Seção 3 é apresentada a ferramenta Cartola FC Data Analysis, um ambiente para simulação e visualização baseado no *Fantasy Football* Cartola FC. Subsequentemente, na Seção 4 é realizado um planejamento de experimento através da realização de um planejamento fatorial 2^k e na Seção 5 são discutidos e analisados os resultados. Por fim, na Seção 6 são apresentadas as conclusões.

2 FANTASY GAME

Esta seção discute alguns temas deste estudo por meio da identificação e análise de trabalhos correlatos. Subdividido em duas subseções, procura-se aqui analisar os aspectos pertinentes às iniciativas nas plataformas de *Fantasy Football*, passando pelo ambiente de simulação e algoritmos e por fim, as análises baseadas em visualização de dados.

2.1 Aplicação de modelos computacionais na análise e predição em aplicações de Fantasy Football

A análise de dados tem cumprido um papel relevante no tocante à descoberta e entendimento de diferentes domínios e contextos informacionais. Ao analisar grandes conjuntos de dados, pesquisadores têm a possibilidade, não só de traçar um panorama dos acontecimentos passados, mas também, de realizar estimativas, observando tendências e comportamentos que forneçam *insights* sobre a possibilidade de ocorrências futuras. Neste contexto, as plataformas *Fantasy Football* oferecem um cenário desafiador. A natureza imprevisível dos acontecimentos dos jogos e os consequentes resultados têm motivado pesquisas na direção de especular sobre a possibilidade de mitigar tais incertezas a partir de modelos computacionais [2]. Abordagens voltadas para a estatística, matemática e computação

são apenas alguns campos nos quais é possível identificar trabalhos relacionados a predições em plataformas de *Fantasy Football* disponíveis ao redor do globo [2, 5, 7, 18].

Corroborando com esta argumentação, o trabalho de [2] fornece uma análise das projeções semanais de jogadores da plataforma *Fantasy Football* da NFL [14], avaliando sua acurácia e atuando na implementação de modelos baseados em regressão linear como fim de influenciar positivamente na qualidade destas projeções. Embora o método não tenha demonstrado estatisticamente significativos aumentos na acurácia, para o autor, o aumento alcançado já representa ganhos de produtividade que justificam o seu uso na seleção de jogadores e sua evolução em trabalhos futuros. Ainda com foco na NFL, no trabalho de [18], os autores utilizaram redes neurais como forma de analisar e prever os resultados da NFL. Redes também foram utilizadas em [15] na criação de um ranking matematicamente justificado para as ligas do *American College Football*. Outra abordagem foi proposta em [5], no qual os autores propõem um método de predição estatística baseado em técnicas de mineração de dados com foco na plataforma *Fantasy Premier League*. A análise dos autores incide sobre um expressivo conjunto de dados contendo estatísticas de futebol que remontam o início da Liga Premier em 1999 [17]. Com base nos resultados das análises, os autores propõem um algoritmo de mineração utilizando um sistema de pontuação própria que fornece aos usuários da plataforma uma indicação acerca dos jogadores com maior probabilidade de sucesso em um dado contexto. A análise da literatura demonstrou ainda que muitos esforços têm sido envidados à construção de modelos que se estendem para além do mundo virtual. A aplicação de métodos e modelos computacionais com foco na predição dos resultados de partidas tem também forte representatividade no mundo real, como pode ser visto em [8, 11, 15, 19, 23], apenas para citar alguns exemplos.

No escopo dos trabalhos analisados durante esta pesquisa, foi possível identificar que, em certa medida, há um consenso sobre as possibilidades relacionadas à aplicação de modelos computacionais na predição de resultados em jogos virtuais baseados em *Fantasy Football*. Embora alguns projetos tenham pouca representatividade estatística em seus resultados, um leve aumento de qualidade na seleção de jogadores durante o torneio pode representar um aumento mais significativo no resultado das partidas [2]. Esta característica torna as plataformas de *Fantasy Game* um ambiente bastante apropriado no tocante à aplicação de modelos, métodos e técnicas que elevem as probabilidades de sucesso dos jogadores.

2.2 Visualização de dados para análise e identificação de tendências

Os trabalhos que empregam a visualização como mecanismo de análise de dados de jogos virtuais divergem significativamente quanto às abordagens. [13] desenvolveram um conjunto de visualizações para identificar os padrões de ocorrência de estratégias em jogos de batalhas em arena de vários jogadores online (do inglês, *Multiplayer Online Battle Arena* – MOBA), incluindo trajetórias dos personagens e séries temporais com eventos táticos dos times. Dentro do contexto dos jogos de *Fantasy Game*, dois trabalhos destacam-se: No primeiro, foi desenvolvida uma visualização hierárquica que adapta seu formato, dinamicamente, de acordo com a semântica

da informação proveniente de competições de times baseadas em chaves [22]. O segundo estudo realizou avaliações com usuários para determinar a facilidade de uso e satisfação de três abordagens de visualização: uma representação tabular, a mesma representação anterior acrescida de gráficos de barra e uma terceira composta por uma Treemap. A última obteve resultados melhores em termos de satisfação, mesmo sendo mais complexa [3]. Em suma, os trabalhos supramencionados evidenciam que o desenvolvimento das visualizações depende de fatores como os tipos de dados disponibilizados, qual metáfora visual e interações são mais adequadas às tarefas, e qual informação deseja-se comunicar ao usuário. Os trabalhos de [21] e [12] apresentam classificações de visualizações detalhando tais fatores.

Para entender o comportamento visual da ferramenta proposta, faz-se necessária uma breve descrição dos dados provenientes do Cartola FC, conforme mencionado pelos autores em trabalhos analisados nesta pesquisa.

As bases de dados utilizadas pela aplicação Cartola FC são multidimensionais, constituídas por 21 dimensões (atributos), dentre as principais: média de pontos, pontos na rodada x , preço na rodada x , variação de preço, cartões (amarelo e vermelho), finalizações (no gol, na trave), defesas (difíceis e pênalti), faltas, roubadas de bola, assistência, entre outras. Possui também 1026 instâncias, cada uma representando cada participante de cada time em cada rodada do campeonato. Importante ressaltar que “participante” refere-se a um jogador ou técnico e que algumas dimensões estão relacionadas a participantes específicos, como por exemplo, o atributo *defesa de pênalti* que está relacionado exclusivamente a jogadores da posição goleiro, assim como o atributo *finalizações* está mais relacionado aos atacantes. Com exceção das dimensões de *preço* e *pontos*, as demais são dados reais, provenientes dos confrontos dos times, jogos e jogadores em cada rodada do campeonato. O jogo possui ainda uma moeda virtual que possibilita a montagem de um time próprio por meio da compra/venda de qualquer jogador de qualquer time.

Com o intuito de fornecer uma descrição visual dos dados obtidos do Cartola FC, utilizou-se uma versão customizada da ferramenta de visualização PEx-Sensors[16] por conter duas visualizações multidimensionais tradicionais, as quais auxiliam no processo de análise, identificando possíveis comportamentos como padrões ou tendências. A Figura 1 ilustra, do lado esquerdo, a visualização da base de dados de 2015 [4] utilizando a técnica de projeção multidimensional Lamp [10]. As instâncias são representadas pelos participantes, mapeados como pontos no exemplo da figura. A distância entre esses pontos reflete relações de similaridade, quanto menor a distância entre eles, mais próximos e, portanto, mais similares. Se muito distantes, eles possuem pouca ou nenhuma relação. Uma projeção multidimensional visa reduzir a dimensionalidade do espaço original para o espaço visual. A Lamp foi escolhida por ser uma técnica robusta, preservando tanto aspectos globais quanto locais da estrutura original dos dados. À direita da Figura 1 é projetada a mesma base de dados com a técnica de coordenadas paralelas (CP). Essa técnica representa cada dimensão como um eixo vertical e as instâncias como uma polilinha que cruza cada eixo em uma posição proporcional ao valor daquela instância naquela dimensão.

Para exemplificar uma análise com ambas visualizações, selecionou-se na projeção multidimensional a região com maior concentração

de pontos (retângulo vermelho) com o intuito de descobrir o motivo para tal concentração (Figura 1). Tal ação seleciona também automaticamente as mesmas instâncias (polilinhas azuis) nas coordenadas paralelas, ao lado direito, denominada visualização co-ordenada. Dessa maneira, é possível perceber que os participantes selecionados têm valores abaixo da média para todas as dimensões na CP, o que nos leva a concluir que tratam-se de participantes com poucos jogos do campeonato, nos quais não tiveram muita ação/participação, característica também presente nos jogadores reservas que entram nos últimos minutos dos jogos.

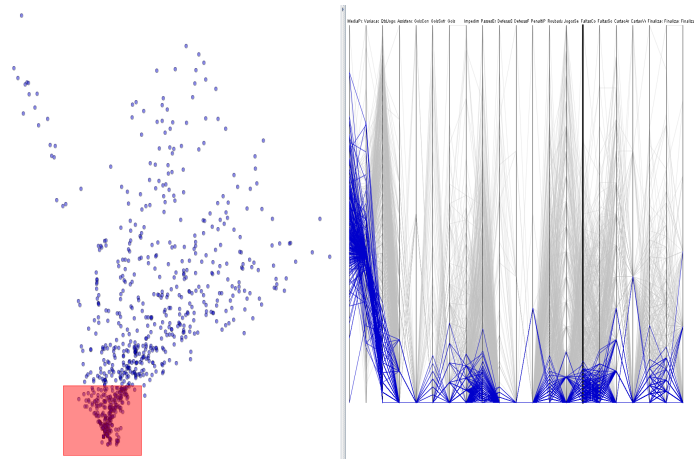


Figura 1: Visualizações multidimensionais analisando o ano de 2015 do Cartola FC.

A Figura 2 ilustra uma análise da projeção multidimensional que utiliza a cor como atributo visual para mapear os valores das dimensões. Inicialmente, na região marcada pela letra (A) observa-se que a cor mapeia a posição tática dos jogadores, conforme a legenda de cor categórica localizada no canto inferior esquerdo. Por conseguinte, foi identificado e selecionado (retângulo laranja) um grupo de pontos cor-de-rosa, todos goleiros. Após, na região marcada pela letra (B), tem-se a mesma projeção, mas a escala de cor foi alterada para representar valores contínuos/sequenciais: azul (mínimo) para amarelo (máximo), nas dimensões mais relacionadas aos goleiros, com o objetivo de entender o motivo da disposição espacial dos pontos. Selecionando *defesas difíceis*, nota-se que o maior valor (amarelo mais saturado) está no ponto mais à esquerda. Ao passar o ponteiro do mouse sobre esse mesmo ponto identifica-se que corresponde ao goleiro Marcelo Lomba, jogador da Ponte Preta. Ao ser selecionado o atributo de *defesas de pênalti*, nota-se que o ponto com maior valor está no topo, no caso o goleiro Weverton do Atlético-PR.

No contexto deste trabalho, objetiva-se utilizar metáforas visuais dinâmicas para melhor auxiliar o usuário em sua análise durante o processo de simulação. Assim, conforme o processo de simulação é executado, animações de séries temporais são apresentadas para acompanhar os resultados dos algoritmos, ilustrando tendências que, em última análise, esboçam o comportamento dos algoritmos variando no tempo.

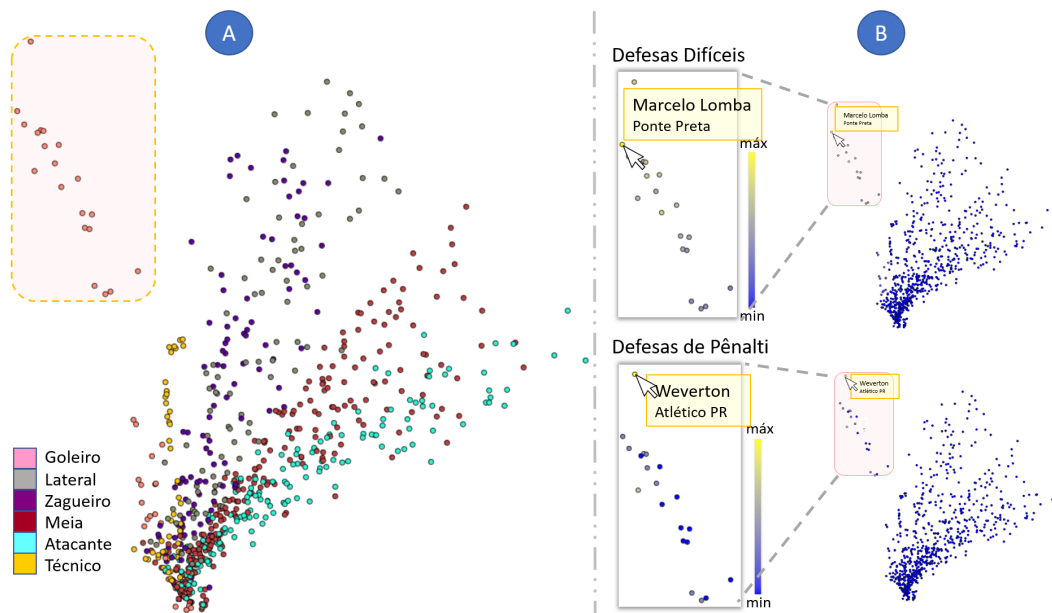


Figura 2: Projeções multidimensionais explicitando goleiros da base de 2015.

3 DESCRIÇÃO DA FERRAMENTA CARTOLA FC DATA ANALYSIS

Nesta seção será apresentada a Cartola FC Data Analysis, uma ferramenta cujo objetivo é permitir a realização de simulações a partir de uma análise histórica dos dados obtidos da plataforma do Cartola FC.

A motivação para a confecção desta proposta centra-se em dois aspectos fundamentais, o primeiro refere-se à possibilidade de analisar a influência do uso de modelos computacionais baseados em métodos estatísticos no tratamento destes conjuntos de dados, avaliando a possibilidade de inferir comportamentos e tendências que auxiliem nas escolhas e decisões dos usuários. O segundo aspecto consiste em analisar os resultados obtidos a começar dos processos de simulação, identificando fatores e níveis, avaliando, a partir de um planejamento de experimento, como suas interações influenciam no resultado final do processo. Neste contexto, o projeto foi desenvolvido de modo que estes fatores e seus respectivos níveis fossem incluídos na ferramenta, permitindo a troca da semente em função do experimento desejado. O planejamento e análise do experimento são descritos detalhadamente nas próximas seções.

A Figura 3 apresenta a visão geral do sistema. Inicialmente, os dados são extraídos do Cartola FC por meio de sua API, em formato JSON. Na etapa seguinte, esses dados são pré-processados e transformados em estruturas compreensíveis, as quais viabilizam a criação de estruturas visuais específicas para tais estruturas. Por fim, essas estruturas visuais são integradas à mecanismos de interação, inclusão dos parâmetros e *design* da interface, compondo a visualização final do sistema. O usuário atua como peça essencial do sistema, utilizando a interação para compreender e analisar o comportamento da técnica de simulação.



Figura 3: Visão geral do sistema.

O processo de simulação consiste, fundamentalmente, em coletar os dados, agora estruturados, e sobre estes, reproduzir processamentos sucessivos. A cada rodada, novos jogadores são escalados com base em critérios e métricas implementadas nos algoritmos de seleção, compondo a equipe que irá atuar na rodada subsequente. Durante o processamento, a visualização demonstra, por meio de um gráfico de linha, o comportamento das equipes selecionadas a cada rodada do torneio, indicando o sucesso ou fracasso de determinada estratégia. A Figura 4 ilustra a tela inicial da ferramenta.

Analisando a Figura 4, pode-se observar o ambiente de simulação em operação. Na parte superior direita da imagem, marcado pela letra (A), são identificadas as rodadas do campeonato. A cada rodada, uma equipe é escalada e uma pontuação é obtida. Ao expandir uma determinada rodada é possível identificar os jogadores e suas respectivas pontuações.

Na área marcada pela letra (B), é possível selecionar o ano sob o qual se deseja executar a simulação. Ao selecionar um determinado ano, o software deverá restringir os processos de simulação aos dados do campeonato realizado no referido ano. Ainda na área demarcada pela letra (B), têm-se a opção de escolher a visualização a partir de gráficos sobrepostos. Este recurso viabiliza a realização de comparações entre diferentes cenários de simulação. O processo é realizado mantendo o gráfico anterior com mais transparência, enquanto um novo gráfico, com cores mais saturadas é projetado representando uma nova simulação.

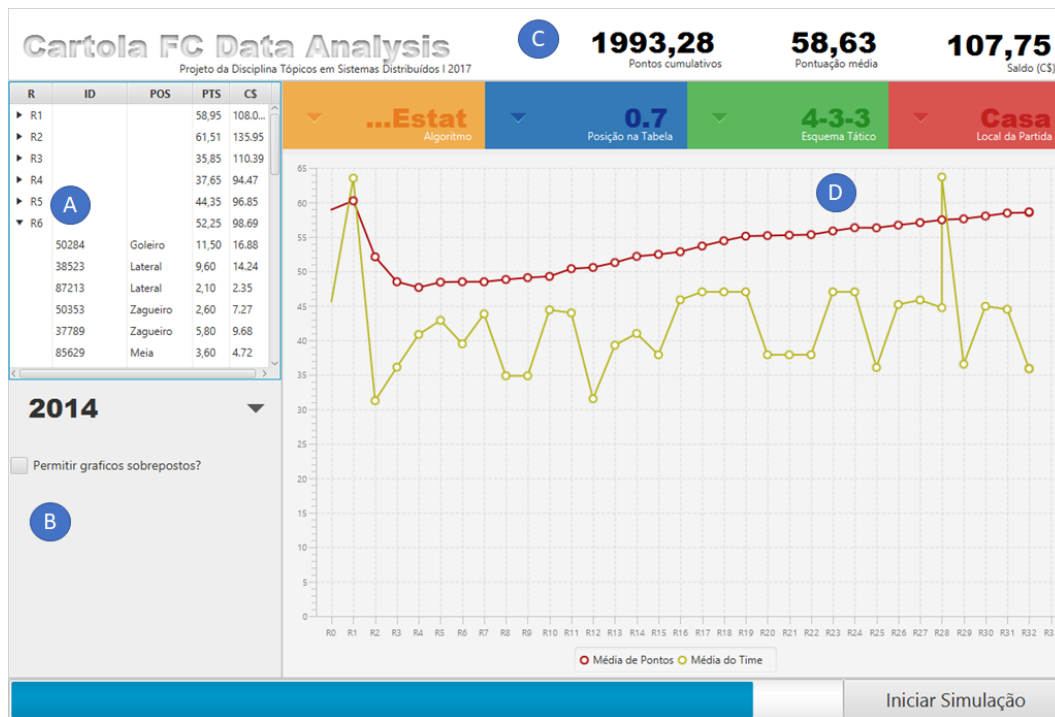


Figura 4: Tela inicial da ferramenta de simulação Cartola FC Data Analysis.

No topo da janela, na área marcada pela letra (C), pode-se verificar a presença de três indicadores numéricos, a saber: (i) Pontos cumulativos, que calculam a pontuação do time acumulando os valores a cada rodada; (ii) Média da pontuação, que tem como função calcular a pontuação média obtida por rodada; (iii) Saldo, que apresenta a variação do preço durante o processo de simulação.

Por fim, na parte central, delimitada pela letra (D), observa-se a visualização principal da ferramenta. Conforme a simulação é executada, os indicadores no topo da janela são ajustados e estas variações são representadas por linhas distintas de modo animado na região central. Este modelo possibilita a visualização em tempo de execução da trajetória e o desempenho das equipes selecionadas pelo algoritmo. Este recurso pode contribuir para um maior entendimento acerca do comportamento dos dados durante as rodadas frente aos critérios e restrições fornecidos como entrada para o processo de simulação.

A configuração dos critérios de entrada para a execução da simulação pode ser observada na parte superior do gráfico. Estes critérios são de fato os fatores e níveis que serão analisados durante o planejamento do experimento.

Utilizando as caixas de seleção, é possível indicar (i) O tipo de algoritmo que se deseja executar, variando entre aleatório e estatístico; (ii) Os pesos a serem dados aos participantes (jogadores ou técnico) em função da atuação dos seus respectivos clubes no campeonato, neste contexto, os valores dos pesos podem variar entre 0,3 e 0,7; (iii) O esquema tático adotado pelo algoritmo, variando entre “4-4-2” e “4-3-3”, e por fim, (iv) A preferência pela escalação de participantes que irão jogar “em casa” ou “fora de casa”, referindo-se ao local onde a partida irá ocorrer. Com base nestes parâmetros, os algoritmos criam seus times a cada rodada

do campeonato, utilizando para isso os atributos relevantes a todos os participantes: pontos, preço e posição.

3.1 Algoritmo aleatório e estatístico

A fim de avaliar a eficácia da utilização de modelos estatísticos na escalação das equipes, foram implementados dois algoritmos distintos. Um algoritmo denominado aqui por aleatório, cuja função é selecionar jogadores aleatoriamente ou utilizando o mínimo de critérios necessários para a escalação das equipes segundo as regras do Fantasy Game Cartola FC, e, um segundo algoritmo aqui descrito como estatístico. Neste, foram aplicados cálculos estatísticos sobre os conjuntos de dados, a fim de selecionar as equipes que seriam supostamente superiores as equipes geradas pelo algoritmo aleatório.

Assim, o algoritmo de seleção aleatória foi desenvolvido enfatizando o atendimento de um único critério que seria o valor disponível para aquisição de jogadores. Na plataforma Cartola FC, é fornecido a cada usuário iniciante um montante de 100 *cartoletas* (nome dado aos valores monetários na plataforma), que deverá ser distribuído na escalação da equipe. Após a inicialização do algoritmo, uma instância do conjunto de dados filtrados por um determinado ano é inicializada. Subsequentemente, o valor disponível é distribuído igualmente entre os onze jogadores mais um técnico. O processo então localiza jogadores dentro do orçamento previsto e compõe a equipe que irá atuar na rodada seguinte.

Já o algoritmo de seleção estatística foi desenvolvido mantendo o enfoque na análise dos dados. Conforme os processos de simulação são executados, o algoritmo realiza cálculos como média dos jogadores, posições mais valiosas, entre outros aspectos encontrados nos conjuntos de dados, a fim de tentar maximizar as chances de sucesso na escalação das equipes durante as rodadas. Como ocorre

no algoritmo aleatório, a cada final de rodada, a equipe é “vendida” e uma nova escalação é realizada com base nos novos valores disponíveis. Os aspectos mais relevantes do algoritmo de seleção estatística são descritos brevemente a seguir.

Input: Lista de partidas restritas ao ano selecionado
Output: Lista de jogadores selecionados para a próxima rodada
 $NúmeroDeRodadas(R) \leftarrow Quantidade;$
 $List:Players \leftarrow Partidas(ANO = ?);$
 $List:Position \leftarrow PosiesDeJogo;$
while $RodadaAtual \leq NúmeroDeRodadas$ **do**
 $Saldo \leftarrow AtualizaSaldo();$
 $Players \leftarrow CalculaMediaJogadoresPorRodada();$
 $Players \leftarrow AplicaFatorBaseadoNaPosicaoDoClube(0.3 \text{ ou } 0.7);$
 $Position \leftarrow CalculaPosicoesMaisRentaveis();$
 Itera sobre as posições para iniciar a seleção dos jogadores
 foreach $Position P_i \in P$ **do**
 if $ObtemMelhorMedia() \text{ E } Saldo \leq ValorDisponivel$ **then**
 $Result \leftarrow SelecionaJogador();$
 end
 end
end

Algoritmo 1: Algoritmo de seleção estatístico.

Analisando o pseudocódigo, pode-se observar que a partir de uma lista de entrada contendo as partidas realizadas em um determinado ano, são extraídos os jogadores e então, o processo de análise é inicializado.

O processo começa com a análise do saldo atual que consiste no saldo restante das operações de compra e venda das equipes, somando-se a isso, o resultado das partidas jogadas. Subsequentemente, são realizados os cálculos com base nas rodadas anteriores. O cálculo da média dos jogadores consiste em calcular, com base na sua pontuação, o desempenho médio dos jogador durante o torneio, assumindo a premissa de que $(R \leq Rodada\ Atual)$. Sobre esta média, incide ainda o fator aqui descrito por *Fator Tabela*, cujos níveis podem variar entre 0.3 e 0.7. Este fator pode reduzir ou aumentar a média do jogador com base na posição do seu clube no campeonato. Assim, o desempenho do clube passa a ter influência direta nas médias dos jogadores. Tende-se a acreditar que ao escolher jogadores cujos clubes estão em melhor situação na tabela do campeonato, as chances de pontuação deste jogador serão maiores.

Por fim, é realizado o cálculo das posições. Analisa-se as posições avaliando-se o desempenho individual e então particiona-se o valor disponível proporcionalmente com base nesta aferição de desempenho para cada posição. Têm-se então *goleiro* = $x\%$; *atacante* = $y\%$; e assim sucessivamente. Conforme o processo é executado, os resultados são visualizados nos indicadores e rastreados através do gráfico.

4 PLANEJAMENTO DO EXPERIMENTO

O planejamento do experimento consiste em definir um conjunto de configurações que possa ser analisado com o fim de obter a maior precisão estatística possível em um dado cenário [9]. Para realizar o planejamento do experimento deve ser definida (i) a variável de resposta que baseia-se no elemento que se deseja aferir; (ii) os fatores ou conjunto de variáveis que afetam a referida variável de

resposta e, subsequentemente, (iii) os níveis, que consistem nos valores que um determinado fator pode assumir.

No contexto deste trabalho, foi definida como variável de resposta a pontuação média do jogador ao longo de todo o torneio. Analogamente, foram definidos os fatores que influenciam no desempenho do sistema: [Fator A - Algoritmo] = {Aleatório, Estatístico}; [Fator B - Fator Tabela] = {0,3, 0,7}, que consiste no fator aplicado em função da posição do jogador na tabela do campeonato; [Fator C - Esquema Tático] = {4-4-2, 4-3-3}, referindo-se a formação dos jogadores em campo; [Fator D - Local] = {Casa, Fora}, referindo-se a seleção de jogadores de equipes locais ou visitantes. Os fatores e níveis do sistema foram definidos através de valores mais prováveis e utilizados na literatura, permitindo avaliar a completude do efeito que essas variáveis provocam ao ambiente. Todos os experimentos foram analisados a partir do intervalo de confiança², da média e do desvio padrão aferidos. Esses parâmetros são utilizados como base para o cálculo da soma dos quadrados, resultando na influência de cada fator nas variáveis de resposta.

4.1 Resultado das replicações

Com base no planejamento do experimento foram realizadas as replicações. Este processo tomou como premissa os conjuntos de dados disponíveis, assim, para cada torneio em que a simulação foi executada, foram aferidos os resultados. Os conjuntos de dados utilizados possuem representação dos anos de 2014 a 2017, perfazendo, portanto, um total de quatro replicações para cada um dos 16 experimentos projetados. Para cada um destes experimentos, foram utilizados intervalos de confiança de 95% segundo a tabela T-Student. Esta medida permite analisar, a partir dos resultados, a dispersão dos dados segundo a probabilidade adotada. Com base nos valores obtidos durante a execução dos experimentos, foram calculadas as variações totais, bem como a influência dos fatores na variável de resposta conforme é descrito em [9].

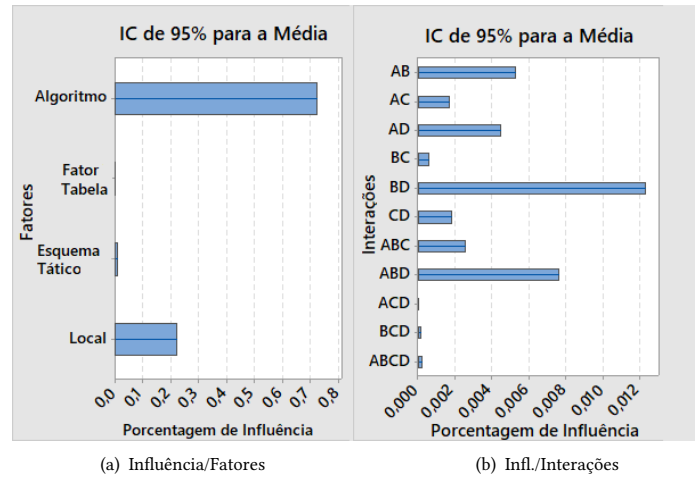


Figura 5: Análise dos resultados.

Analisando a Figura 5(a), é possível observar que o fator algoritmo, cujos níveis podem variar entre aleatório e estatístico, possui um percentual de influência de 72%. Observado-se este resultado

²Utilização da distribuição *t-student* com replicações de 4 bases de dados e $\alpha = 0, 05\%$.

frente aos demais fatores, percebe-se que a variação do algoritmo possui níveis de influência preponderantes no tocante aos resultados obtidos durante os processos de simulação. Com menor intensidade, mas com significativa representatividade, o fator Local obteve níveis de influência de 22%. Este resultado demonstra sua relevância, na medida em que, a variação deste fator exerce quase metade da influência em relação ao algoritmo. Com valores menos expressivos em relação aos demais fatores estão o esquema tático com 0,011% e o Fator Tabela, alcançando um percentual de influência 0,003%. Ao examinar os efeitos destes fatores sobre a pontuação média obtida durante o processo de simulação, é possível identificar além da influência, a tendência em relação a cada um dos níveis. A Figura 6 ilustra esta análise.

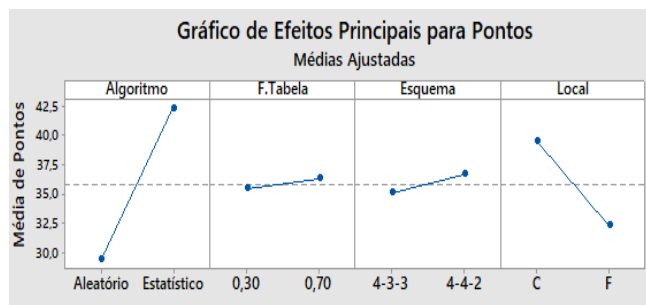


Figura 6: Análise dos efeitos na pontuação média.

Observa-se no gráfico que a pontuação média utilizando o algoritmo aleatório é inferior, evoluindo progressivamente em direção ao algoritmo estatístico. Do mesmo modo, pode-se observar que o fator Local, mantém um efeito descendente, representando maior pontuação em situações nas quais a simulação utiliza jogadores locais (casa) em detrimento aos visitantes (fora). Os demais resultados não demonstraram efeito representativo.

No tocante às interações, que consistem na análise dos resultados obtidos a partir da interação de dois ou mais fatores, observa-se alguns aspectos importantes. Considerando o gráfico da Figura 5(b), é possível observar que as interações entre os fatores “Fator Tabela” e “Local” têm maior representatividade em relação às demais interações. Ainda que o algoritmo tenha significativa influência quando analisado individualmente, a interação entre os fatores que o envolvem demonstrou menos representatividade. Todavia, ao analisar os efeitos destes fatores em relação a pontuação média, observa-se que o algoritmo mantém-se como o elemento principal. A Figura 7 ilustra os efeitos destas interações.

Pode-se observar, portanto, que em todos os casos as interações com o algoritmo demonstraram um efeito ascendente em relação ao algoritmo estatístico. As demais interações não demonstraram um efeito representativo. Os resultados serão analisados na próxima seção.

5 DISCUSSÃO

Com base no planejamento do experimento delineado na seção anterior, pode-se observar a relevância do fator algoritmo representado nos resultados obtidos durante os processos de simulação. Ao

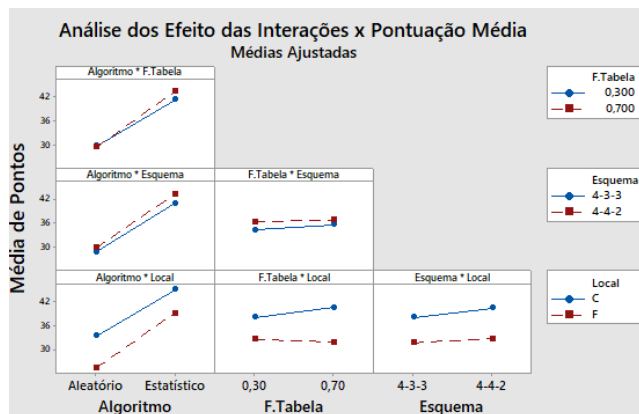
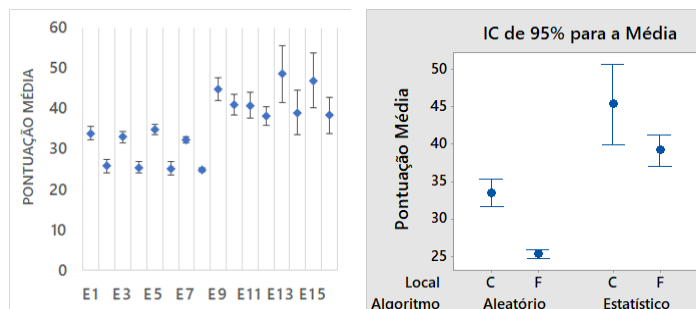


Figura 7: Análise dos Efeitos das Interações x Pontuação Média.

variado de aleatório para estatístico, houve um aumento da pontuação média na ordem de 29,39 para 42,23. Analogamente, ao analisar o desempenho do fator Local, um aumento significativo também é observado quando a variação ocorre entre os níveis “casa” com 32,24 e “fora”, com 39,38 respectivamente. Os demais fatores não obtiveram significativa representatividade. Ao analisar a dispersão dos valores da amostra em torno da média, pode-se perceber mais claramente as variações evidenciadas durante o experimento, observando especialmente, o momento em que se dá a substituição do algoritmo de aleatório para estatístico. A Figura 8(c) ilustra esta análise.



(c) Dispersão do erro padrão em torno da média

Figura 8: Análise dos resultados.

Embora as interações tenham em geral baixa representatividade no contexto dos experimentos analisados, como pode ser visto na Figura 5(b), foi possível observar a ocorrência de variações na pontuação final a partir do cruzamento entre certos fatores. Ao analisar o fator “algoritmo” frente ao fator “Local”, pode-se constatar que utilizando o algoritmo estatístico com a configuração do nível do fator Local em “C (casa)”, tem-se um ganho de desempenho de cerca de +- 35,35% em função do algoritmo aleatório. Ao substituir o nível do fator local por “F (fora)”, tem-se um ganho de desempenho na ordem de +- 67,96%. No entanto, neste segundo caso, embora o

ganho percentual seja maior, a pontuação média obtida é inferior a primeira situação. A Figura 8(d) ilustra melhor este cenário.

Por meio deste experimento, foi possível observar o comportamento dos fatores, demonstrando suas correlações e como tais variações afetam o resultado final. O planejamento de experimentos cumpre um papel fundamental neste sentido, na medida em que retrata diferentes cenários permitindo que, por exemplo, fatores com baixa representatividade durante os experimentos possam ser reanalisados, a fim de promover melhorias no tocante à sua contribuição, aprimorando o desempenho no resultado final.

6 CONCLUSÃO

A aplicação de modelos computacionais com fins na predição de cenários variados é tema recorrente na literatura. Autores de diferentes áreas têm realizado esforços na confecção de métodos e técnicas que contribuam para maximizar o grau de exatidão das aplicações em diferentes contextos. Analogamente, o campo da visualização de informações fornece um apoio importante neste sentido, contribuindo para o entendimento dos conjuntos de dados analisados, ilustrando tendências que podem auxiliar a construção e ou melhoria destes mesmos modelos. Neste trabalho foi desenvolvida uma ferramenta com foco na simulação, análise e visualização de dados baseados na rede social *Fantasy Game* Cartola FC. A partir do conjunto de dados de entrada, foram aplicados modelos estatísticos no desenvolvimento de algoritmos cujo foco é atuar na mineração dos dados, fornecendo cenários mais rentáveis dentro do universo do jogo.

A visualização computacional implementada consiste de séries temporais dinâmicas, utilizando animação para atualizar a cada rodada o comportamento do algoritmo. Por fim, foi realizado o planejamento e a execução de um experimento empregando o planejamento fatorial 2^k por meio do modelo de regressão linear. Com foco na variável de resposta, neste caso a pontuação média obtida por rodada, foram definidos os fatores e níveis a serem analisados. A análise permitiu indicar os níveis de influência destes fatores sobre a referida variável de resposta. Os experimentos demonstraram que com o algoritmo proposto neste artigo obtêm-se ganhos significativos na pontuação do *fantasy game* em comparação ao modelo de seleção aleatória.

Espera-se que este artigo contribua para a utilização de mecanismos de simulação e visualização em diferentes contextos, para diferentes redes sociais deste tema de esportes, com foco não só na predição de cenários, mas também na análise comportamental dos conjuntos de dados, um aspecto bastante evidente durante a construção deste trabalho.

7 AGRADECIMENTOS

Os autores agradecem a FAPEMIG, FAPESB, CAPES, CNPq. Em especial, ao MCTI-UFBA pelo apoio financeiro por meio do Edital PROP/CI/PROP/ – PROPESQ/UFBA 004/2016.

REFERÊNCIAS

- [1] Siobhan C. BuddJean and Christophe Egea. 2017. The Evolution of Sport in Society. *Sport and Oral Health* (2017). https://doi.org/10.1007/978-3-319-53423-7_1
- [2] Nathan Dunnington. 2015. *Fantasy Football Projection Analysis*. Ph.D. Dissertation. University of Oregon. http://economics.uoregon.edu/wp-content/uploads/sites/4/2015/03/Dunnington_Thesis_2015.pdf
- [3] Rob Euman and José Abdelnour-Nocera. 2013. *Data Visualisation, User Experience and Context: A Case Study from Fantasy Sport*. Springer Berlin, 146–155. https://doi.org/10.1007/978-3-642-39265-8_16
- [4] Cartola FC. 2017. Disponível em <https://github.com/henriquepgomide/cartola/tree/master/data/2015>. Acessado em 19/12/2017.
- [5] Anurag Gangal Vesit, Mumbai Abhishek, Talnikar Vesit, Mumbai Aneesh, Dalvi Vesit, Mumbai Vidya, Zope Vesit, Mumbai Aadesh, and Kulkarni Vesit. 2015. Analysis and Prediction of Football Statistics using Data Mining Techniques. *International Journal of Computer Applications* 132, 5 (2015), 975–9887.
- [6] Globo.com. 2017. virou-rotina-cartola-bate-novo-recorde-de-times-na-rodada-10-55-milhoes @ globoesporte.globo.com. <http://globoesporte.globo.com/cartola-fc/noticia/2017/06/virou-rotina-cartola-bate-novo-recorde-de-times-na-rodada-10-55-milhoes.html>
- [7] Kou Yuan Huang and Wen Lung Chang. 2010. A neural network method for prediction of 2006 World Cup Football Game. *Proceedings of the International Joint Conference on Neural Networks* (2010). <https://doi.org/10.1109/IJCNN.2010.5596458>
- [8] Josip Hucaljuk and Alen Rakipovic. 2011. Predicting football scores using machine learning techniques. *2011 Proceedings of the 34th International Convention MIPRO* 48 (2011), 1623–1627.
- [9] Raj Jain. 1991. *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling*. John Wiley & Sons Chichester.
- [10] P. Joia, D. Coimbra, J. A. Cuminato, F. V. Paulovich, and L. G. Nonato. 2011. Local Affine Multidimensional Projection. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (Dec 2011), 2563–2571. <https://doi.org/10.1109/TVCG.2011.220>
- [11] a. Joseph, N. E. Fenton, and M. Neil. 2006. Predicting football results using Bayesian nets and other machine learning techniques. *Knowledge-Based Systems* 19, 7 (2006), 544–553. <https://doi.org/10.1016/j.knsys.2006.04.011>
- [12] D. A. Keim. 2002. Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics* 8, 1 (Jan 2002), 1–8. <https://doi.org/10.1109/2945.981847>
- [13] Quan Li, Peng Xu, Yeuk Yin Chan, Yun Wang, Zhipeng Wang, Huamin Qu, and Xiaojuan Ma. 2017. A Visual Analytics Approach for Understanding Reasons behind Snowballing and Comeback in MOBA Games. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 211–220. <https://doi.org/10.1109/TVCG.2016.2598415>
- [14] National Football League. 2017. NFLFantasyFootball. <http://www.nfl.com/fantasyfootball>
- [15] Juyong Park and M. E.J. Newman. 2005. A network-based ranking system for US college football. *Journal of Statistical Mechanics: Theory and Experiment* 10 (2005), 239–252. <https://doi.org/10.1088/1742-5468/2005/10/P10014> arXiv:physics/0505169
- [16] Fernando V Paulovich, Marli L Moraes, Rafael Mitsuo Maki, Marystela Ferreira, Osvaldo N Oliveira Jr, and Maria Cristina F de Oliveira. 2011. Information visualization techniques for sensing and biosensing. *Analyst* 136, 7 (2011), 1344–1350.
- [17] Premier League. 1992. <https://www.premierleague.com/>. <https://www.premierleague.com/>
- [18] Michael C. Purucker. 1996. Neural network quarterbacking. *IEEE Potentials* 15, 3 (1996), 9–15. <https://doi.org/10.1109/45.535226>
- [19] a. P. Rotshtein, M. Posner, and a. B. Rakityanskaya. 2005. Football predictions based on a fuzzy model with genetic and neural tuning. *Cybernetics and Systems Analysis* 41, 4 (2005), 619–630. <https://doi.org/10.1007/s10559-005-0098-4>
- [20] Donald P. Roy and Benjamin D. Goss. 2007. A conceptual framework of influences on fantasy sports consumption. *Marketing Management Journal* 17, 2 (2007), 96.
- [21] B. Shneiderman. 1996. The eyes have it: a task by data type taxonomy for information visualizations. *Proceedings 1996 IEEE Symposium on Visual Languages* (1996), 336–343. <https://doi.org/10.1109/VL.1996.545307> arXiv:arXiv:1011.1669v3
- [22] D. Tan, G. Smith, B. Lee, and G. Robertson. 2007. AdaptiveTree: Adaptive Tree Visualization for Tournament-Style Brackets. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (Nov 2007), 1113–1120. <https://doi.org/10.1109/TVCG.2007.70537>
- [23] A Tsakonas, G Dounias, S Shtovba, and V Vidyuk. 2002. Soft computing-based result prediction of football games. In *The First International Conference on Inductive Modelling (ICIM'2002)*. Lviv, Ukraine.