DOI: 10.53660/CONJ-1627-2E61

Utilizando Inteligência Artificial Explicável para formação de perfil de jogadores no Cartola FC

Using Explainable Artificial Intelligence to profile players in Cartola FC

Davi Lima da Cruz[1], José Vigno Moura Sousa[1], Dario Brito Calçada[1]*

RESUMO

O Cartola FC é um *fantasy game*, onde é possível realizar escalações em um time virtual usando como base os jogadores e técnicos de futebol da Série A do campeonato brasileiro. Existe uma dificuldade em obter bons resultados nas escalações pela volatilidade dos jogos e desempenho dos jogadores. O trabalho visa extrair e validar um modelo de perfil de jogador para a escalação no Cartola FC, utilizando técnica de Inteligência Artificial Explicável. Para criar o modelo foi usado a base de dados disponibilizada pela API da Globo, utilizando técnica de Redes de Regras de Associação Filtradas como principal meio de estruturar e entender os padrões da base de dados.

Palavras-chave: Cartola FC; Inteligência Artificial Explicável; Perfil de Jogadores; Redes de Regras de Associação Filtradas;

ABSTRACT

Cartola FC is a fantasy game in which it is possible to perform lineups in a virtual team using the soccer players and coaches of Serie A of the Brazilian championship as a base. There is a difficulty in getting good results in the lineups due to the volatility of the games and the performance of the players. The work aims to create and validate a player profile model for Cartola FC's lineup, using an Explainable Artificial Intelligence technique. The database provided by the Globo API was used to create the model, using the Filtered Association Rules Network technique as the main means of structuring and understanding the database standards.

Keywords: Cartola FC; Explainable Artificial Intelligence; Filtered Association Rules Networks; Players Profile

[1] Universidade Estadual do Piauí.
*E-mail: dariobcalcada@frn.uespi.br

INTRODUÇÃO

Cartola FC é um jogo do site da Globo criado em 2005. O jogo é do estilo *Football Manager and Fantasy*, nele é possível comandar o próprio clube, escalar jogadores e técnicos com base no que acontece na vida real dentro de campo. Por meio de pontuações que levam o nome de *scouts*, o usuário escolhe aqueles jogadores que apresentam bom rendimento e pontos na partida, que são transferidos para o Cartola (OUL, 2020). Ao início do jogo, são disponibilizadas a todos os usuários do jogo 100 cartoletas (moeda de valor da plataforma para a compra de jogadores). A abertura do mercado acontece dias antes de cada rodada do Brasileirão (Campeonato Brasileiro de Futebol) e o seu saldo disponível deverá ser utilizado para escalar indivíduos a fim de ter um time com 11 jogadores e 1 técnico, de acordo com o valor de mercado de cada um, não podendo ultrapassar o número de cartoletas disponível no seu saldo (OUL, 2020).

Assim, os competidores que participam desse jogo precisam colocar em prática seus conhecimentos estatísticos, acompanhar o desempenho de cada jogador no Campeonato Brasileiro do ano e estar constantemente informados sobre possíveis lesões sofridas pelos membros do time escalado. Tais fatores implicam na vitória ou derrota dos times, influenciando na compra de mais jogadores, análise do trabalho de técnicos, dentre outros fatores (BATISTA et al., 2019). Conseguir escalar um time de boa pontuação não consiste em uma tarefa fácil, pois os diversos fatores influenciam no bom rendimento, entendendo que tudo se baseia em análises da vida real.

Com o propósito de melhorar a descoberta de padrões relevantes, o uso de técnicas de Descoberta de Conhecimento em Base de Dados (KDD, do inglês *Knowledge-discovery in Databases*) é uma alternativa de grande potencial, com resultados que podem ser inovadores. Uma das técnicas utilizadas no KDD é a Mineração de Regras de Associação (ARM, do inglês *Association Rules Mining*), que visa a identificação de padrões em *datasets*. O processo de Mineração de Regras de Associação pode ser visto como um conjunto de ações genéricas que devem ser realizadas de acordo com os dados disponíveis. Um exemplo de uso da Mineração de Regras de Associação para a descoberta de conhecimento é a sua utilização na análise de dados referente a compras em um supermercado, que pode gerar uma regra como: {feijão, couve} ⇒ {linguiça}. Essa regra é utilizada para gerar a hipótese de que "clientes que compram feijão e couve tendem também a comprar linguiça". O exemplo ilustra uma das características mais atrativas das

Regras de Associação, na qual cada regra é expressa de uma forma muito fácil de ser compreendida quando formulada por *itemsets* de tamanho reduzido (SILVA et al., 2020).

A quantidade de Regras de Associação extraídas está diretamente ligada ao número de itens que formam a base de dados. Em um *dataset*, com uma quantidade elevada de elementos, o conjunto de Regras de Associação geradas torna-se cada vez maior, inviabilizando a análise de todas as regras. Por exemplo, em um conjunto de apenas 100 elementos gera-se 9.900 regras com *itemsets* unitários e 98.000.100 regras se os *itemsets* possuírem dois elementos, um crescimento exponencial. A utilização de redes para a análise das regras geradas é de grande auxílio no processo de identificação dos padrões no conjunto de dados. As Redes de Regras de Associação (ARNs, do inglês *Association Rules Networks*) tem como ideia central sintetizar, podar e integrar, no contexto dos objetivos específicos da pesquisa, as Regras de Associação descobertas pelo algoritmo de mineração (SILVA *et al*., 2020).

O objetivo principal do presente trabalho foi validar o uso de técnica de extração de conhecimento em base de dados, identificando os padrões de bom desempenho no futebol brasileiro e criando uma modelagem do perfil de jogador a ser escalado em cada rodada (nas posições de Goleiro e Atacante), utilizando uma base de dados do Cartola FC dos anos 2014 a 2021 (com exceção ao ano de 2015). Foi utilizado o método de KDD com a técnica de Redes de Regras de Associação Filtradas (*Filtered*-ARNs, do inglês *Filtered-Association Rules Networks*) para identificar o conhecimento extraído da base de dados com o intuito de auxiliar na tomada de decisão na escalação dentro do jogo.

Este artigo está estruturado em 6 seções, incluindo esta introdução. Na Seção 2 são descritos os trabalhos relacionados que inspiraram a pesquisa. A fundamentação teórica é apresentada na Seção 3, abordando os principais conceitos utilizados. Na Seção 4 é descrito os materiais e métodos utilizados nos experimentos de estudo e extração de conhecimento pelo uso da técnica de inteligência artificial explicável (xAI, do inglês *Explainable Artificial Intelligence*). Os resultados são apresentados e avaliados na seção 5. E por fim, as conclusões são apresentadas na seção 7.

TRABALHOS RELACIONADOS

No trabalho de Ribeiro (2019) foram usados Redes Neurais Artificiais e Algoritmos Genéticos para obter o algoritmo que se comportasse de forma mais eficaz ao

escalar um time no Cartola FC, em duas rodadas do brasileirão, levando em consideração 3 cenários: com 70 cartoletas disponíveis, com 100 cartoletas disponíveis e com 150 cartoletas disponíveis. Além disso, foi feita a variação de 7 esquemas táticos em cada um desses cenários. As conclusões deste trabalho sugerem que os resultados que foram obtidos estão diretamente ligados ao limite de patrimônio disponível para escalação, pois tem influência direta na qualidade dos times na qualidade dos times escalados.

Em Viscondi *et al*. (2017) foram usados 3 algoritmos de *Machine Learning* com o intuito de prever os jogadores com maior pontuação na rodada do Cartola FC, sendo esses: *Extreme Gradient Boosting, Random Forest* e *Suport Vector Machine*. Além disso, foi utilizado uma técnica de aprendizagem não supervisionada K-*means*, tendo sido dividido os jogadores em grupos de características semelhantes para aprimorar a acurácia do modelo. Foram usadas 28 rodadas do campeonato brasileiro de futebol série A do ano de 2017 para treinar o modelo e 5 rodadas para os testes do modelo. Os resultados foram promissores, mas sem a descoberta do conhecimento que explicam as seleções realizadas.

Já em Sehnem *et al*. (2021), o objetivo do trabalho era analisar conjuntos de variáveis que poderiam ter mais influência sobre resultado de uma partida de futebol e utilizou de técnicas como cálculos de probabilidade e algoritmos de previsão, com a intenção de obter lucros em apostas. As técnicas utilizadas para o desenvolvimento foram Redes Bayesianas, com o algoritmo Naïve Bayes,e a de probabilidade, baseada no cálculo de Poisson. Os dados utilizados para treinamento foram do Campeonato Brasileiro, entre 2010 a 2017, sendo considerados os dados dos anos de 2018 e 2019 para testes. Os principais resultados atingidos foram de 53% de acerto do resultado de uma partida e as principais variáveis envolvidas foram força de ataque e força de defesa.

Após estudos dos trabalhos citados, percebe-se a ausência do uso de técnicas de xAI que podem promover a descoberta de conhecimento pertinente para aplicações relacionadas ao Cartola FC. Desta forma, a aplicação de uma técnica de xAI proporciona um estudo inédito nessa área. Para este trabalho, foi selecionada a técnica de Redes de Regras de Associação Filtradas que já são utilizadas várias aplicações, como na área da saúde por exemplo (CALÇADA *et al*., 2020, SILVA *et al*., 2020).

FUNDAMENTAÇÃO TEÓRICA

Neste tópico, serão apresentados os conceitos necessários para compreensão das técnicas de redes de regras de associação.

KDD

O KDD é um campo que se preocupa com o desenvolvimento de técnicas e métodos para trazer sentido aos dados. O problema básico abordado pelo KDD é o de transformar dados brutos e de baixo nível (que tipicamente extrapolam o tamanho para serem interpretados) em uma forma mais compreensível e de melhor interpretação. Para isso, o núcleo desse processo é a aplicação de técnicas de mineração de dados para a extração e descoberta de padrões (FAYYAD, 1996). O processo completo de KDD é dividido em 3 etapas: pré-processamento, mineração dos dados e pós-processamento.

O pré-processamento consiste na aquisição e manipulação dos dados que podem conter informações úteis para auxiliar na descoberta de conhecimento. Após a obtenção dos valores brutos, selecionam-se os atributos de interesse e é feito um tratamento dos dados. Nesta etapa podem ocorrer a remoção de valores duplicados, a conversão de dados simbólicos para numéricos, além de processos de normalização, redução de dimensões, identificação e tratamento de outliers (valores que não seguem o mesmo padrão de distribuição do conjunto) e dados faltantes. Essas etapas são necessárias para preparar o dataset para a mineração propriamente dita (SILVA et al., 2020).

Na etapa de mineração, são aplicadas técnicas para extração de padrões para aplicação estudada. Métodos de inteligência artificial, estatística ou pesquisa operacional podem ser utilizados a fim de que o conhecimento seja identificado de forma otimizada. Neste trabalho foi utilizado o processo de Mineração de Regras de Associação. Por fim, na etapa de pós-processamento, os resultados obtidos na parte de mineração são analisados. Várias técnicas podem ser usadas a fim de que o conhecimento seja identificado de modo mais eficiente, como o uso de Redes. Neste trabalho foram construídas as Redes de Regras de Associação Filtradas (*Filtered*-ARNs) para otimização do processo de identificação automática de conhecimento de modo que os padrões gerados tenham maior probabilidade de serem relevantes ao domínio estudado (CALÇADA *et al.*, 2018).

REGRAS DE ASSOCIAÇÃO

Uma regra de associação caracteriza o quanto a presença de um conjunto de elementos em uma base de dados tem como consequência a presença de algum outro conjunto distinto de elementos nos mesmos registros. Desse modo, o objetivo das regras

de associação é encontrar tendências que possam ser usadas para entender e explorar padrões de comportamento dos dados (CALÇADA, 2019).

Definição 1: seja $I = \{i_1,i_2,...,i_n\}$ um conjunto de objetos denominados itens que podem assumir valores binários 0 ou 1 (falso ou verdadeiro), que representam a presença ou não de um objeto em particular. Seja T um conjunto de transações, em que cada transação D corresponde a um conjunto a um conjunto de itens tal que $D \subseteq I$. Considera-se ainda que um conjunto de itens A está contido numa transação D, se todos os itens do conjunto tiverem valor "verdadeiro" na transação, ou seja, fizeram parte dessa mesma transação. Uma Regra de Associação R pode ser representada por uma expressão no formato: $A \Rightarrow B$, com $A \subseteq I$, $B \subseteq I$ e $A \cap B = \emptyset$. E ainda é possível tratar as variáveis quantitativas ou qualitativas, criando intervalos de valores e utilizando-as, posteriormente, como variáveis binárias. A é denominado de antecedente (LHS – *Left Hand Side*) da regra e B o consequente (RHS – *Right Hand Side*) (CALÇADA; REZENDE, 2019).

Definição 2: : para cada regra ($LHS \Rightarrow RHS$), extraída de um conjunto de transações T, e calculado um valor de suporte (sup), apresentado na **Eq. (1)**, que verifica a força de associação entre LHS e RHS (probabilidade da ocorrência de LHS RHS); e um valor de confiança (conf), apresentado na **Eq. (2)**, que mede a força da implicação lógica da regra (probabilidade condicional de RHS dado LHS) (CALÇADA; REZENDE, 2019).

$$sup(LHS \Rightarrow RHS) = P(LHS \cup RHS) \quad (1)$$

$$conf(LHS \Rightarrow RHS) = P(LHS|RHS) \quad (2)$$

REDES DE REGRAS DE ASSOCIAÇÃO

Como os algoritmos de regras de associação são capazes de extrair todas as regras de associação de acordo com um suporte mínimo e valor mínimo de confiança, o número de regras extraídas geralmente supera a capacidade de interpretação do usuário. O conhecimento gerado muitas vezes é inconclusivo ou não consegue ser aplicado. Sendo assim, a busca de métodos que produzam um resultado de fácil interpretação e de extrema importância em aplicações (CALÇADA, 2019).

Redes de Regras de Associação (ARNs, do inglês *Association Rules Networks*) possuem uma estrutura para sintetizar, podar, e analisar um conjunto de Regras de Associação para a construção de hipóteses candidatas. De uma perspectiva de descoberta

de conhecimento, a ARN possibilita uma análise orientada para o contexto, o que facilita o entendimento dos relacionamentos e a utilidade das informações dos dados. Do ponto de vista matemático, ARNs são instâncias de Hipergrafos Direcionados (PANDEY *et al.*, 2009).

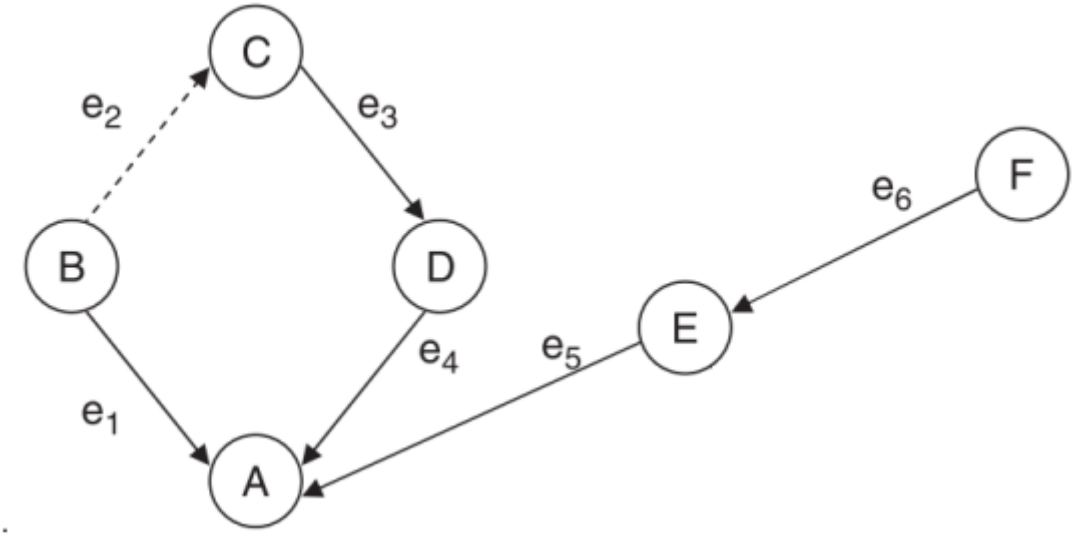
A ideia central da ARN é que as Regras de Associação descobertas pelo algoritmo de mineração podem ser sintetizadas, podadas, e integradas no contexto de objetivos específicos da pesquisa. Em particular, se houver uma variável de interesse ("alvo" ou "objetivo"), pode-se formar uma Rede com as variáveis mais relevantes e relacionadas ao objetivo, e, em seguida, elaborar uma estrutura que pode ser testada usando métodos estatísticos ou seja, acoplar uma tarefa de mineração de dados com análise estatística. Na prática, a Mineração de Regras de Associação envolve centenas de variáveis, portanto, a estratégia de poda que faz parte das ARNs pode ser usada para remover inconsistências locais entre as variáveis, como ciclos. Resumindo, ARNs oferecem as seguintes características (CALÇADA, 2019):

- Poda no contexto: usa-se ARN para podar regras no contexto de um objetivo específico. Alterando o objetivo resultará na poda de regras diferentes.
- Estrutura de rede: ARNs fornecem um mecanismo para determinar a relação entre as variáveis relevantes e o objetivo utilizando a construção de uma Rede. Isso pode
- ajudar na análise dos efeitos de mudanças ocorridas de modo direto e indireto na mineração das Regras de Associação.
- Geração e avaliação de hipóteses: ARN pode servir como uma ponte entre as saídas geradas pela Mineração de Regras de Associação e sua avaliação.

Na Figura 1, é apresentada um exemplo de ARN, na qual foi selecionado o item "A" como objetivo. Seleciona-se então todas as regras que possuem "A" como consequente, neste caso apenas as regras (B⇒A), (D⇒A) e (E⇒A). Assim, os itens "B", "D" e "E" são modelados no nível 1 da ARN e passam a ser os objetivos, passando a executar o algoritmo de forma recursiva até que não restem mais itens como objetivos nos níveis mais altos da abordagem. Neste caso, foram modeladas as regras que possuem "B" como objetivo, depois as regras que possuem "D" e então "E", "C" e "F", respectivamente. Nesse exemplo, não existem regras que possuem "F" como subsequente. A hiper-aresta "e_2" em destaque será umas das eliminadas no processo de

poda, pois mesmo possuindo o item “C” como consequente, o item “B” já estava inserindo na ARN em um nível abaixo, inviabilizando esta regra.

Figura 1 – Exemplo de ARN com hiper-aresta reversa



Fonte: Veras, V. N. R., Calçada, J. C. M., Bezerra, S. M. G. and Calçada, D. B. (2022)

REDES DE REGRAS DE ASSOCIAÇÃO FILTRADAS

Para que as regras geradas tenham uma maior probabilidade de representar um conhecimento verdadeiro, elaborou a construção das Redes de Regras de Associação Filtradas, que consiste é um grafo direcionado construído para modelar todas as regras a fim de descrever um item selecionado. O resultado é um gráfico que explica o item e permite que o usuário construa hipóteses com base nesse item selecionado. Este item selecionado é chamado de “item objetivo”, pois se torna o alvo da exploração do *dataset*. A *Filtered*-ARN oferece as seguintes características: (1) filtragem das regras, (2) poda no contexto, (3) estrutura de rede e (4) geração de hipóteses para avaliação. A filtragem é realizada com o uso de medidas objetivas assimétricas a fim de excluir as regras em que não existe influência entre o antecedente e o consequente da regra. A poda é feita de acordo com o item objetivo, no qual a rede é modelada considerando apenas as regras que estão correlacionadas direta ou indiretamente a esse item (CALÇADA; REZENDE, 2019). Nesse trabalho, para a filtragem das regras de associação, foram utilizadas as medidas *Added Value* e *Gain*.

- ***Added Value* [-1..0..1]**: a medida *Added Value* (AV) indica o quanto a frequência do consequente aumenta na presença do antecedente, ou seja, mede o ganho de RHS na presença de LHS. Se AV for positivo, então a frequência de RHS aumenta na presença de LHS. Sendo AV negativo, a frequência de RHS diminui na

presença de LHS. Se AV for nulo, tem se uma coincidência aleatória, ou seja, a frequência de LHS não altera a frequência de RHS.

$$AV = conf(LHS \Rightarrow RHS) - sup(RHS) \quad (3)$$

- ***Gain* [0..1]**: É uma medida que dá um trade-off entre suporte e confiança, auxiliando na seleção das regras de acordo com a frequências da mesma em relação à confiança mínima.

$$Gain = (sup(LHS \cap RHS) - minconf).sup(LHS) \quad (4)$$

A utilização de medidas assimétricas como *Added Value* e *Gain* na filtragem das regras é umas das diferenças principais entre a *Filtered*-ARN e a ARN sendo que, na *Filtered*-ARN, o usuário pode visualizar um conjunto de itens que provem uma influência estatística em vez de elementos que apenas se relacionam com o item objetivo. Assim, a *Filtered*-ARN apresenta regras que indicam hipóteses com comprovação de dependência entre antecedente e consequente (CALÇADA; REZENDE, 2019).

METODOLOGIA

Na etapa inicial do processo de mineração de dados, foram coletados os dados de um *dataset* disponível on-line[2], contendo registros do cartola FC dos anos de 2014 a 2021. Em análise inicial dos dados, foi verificado que a padronização dos dados divergia a cada ano, sendo, portanto, necessário uma seleção dos anos a serem analisados. Desta forma, foi desenvolvido um script para padronização do *dataset*.

Após a execução da padronização, foi constatado que o ano de 2015 possuía atributos que se incrementavam a cada rodada, comportamento esse totalmente diferente dos outros anos e que eram esperados para a utilização das técnicas. Sendo assim, foi decidido pela retirada dos dados do ano de 2015 do estudo.

Com o conjunto de dados já estabelecido, passou-se para a etapa de pré-processamento. Na etapa de pré-processamento foi feita classificação de cada dos atributos do conjunto de dados, levando em consideração a amplitude de valor de cada

[2] https://github.com/henriquepgomide/caRtola/tree/master/data

um. Todos os atributos foram submetidos a classificação por quartis (A, B, C, D), começando de menor valor à um valor máximo registrado em cada atributo, como descrito nas Tabelas 1 e 2.

Tabela 1 - Categorização dos atributos dos atacantes.

	A	B	C	D
FS	$0 \leq X > 3$	$3 \leq X > 6$	$6 \leq X > 9$	$9 \leq X \geq 12$
PE	$0 \leq X > 3$	$3 \leq X > 6$	$6 \leq X > 9$	$9 \leq X \geq 12$
A	$0 \leq X > 1$	$1 \leq X > 2$	$2 \leq X > 3$	$3 \leq X \geq 4$
FT	$0 \leq X > 0,5$	$0,5 \leq X > 1$	$1 \leq X > 1,5$	$1,5 \leq X \geq 2$
FF	$0 \leq X > 1,5$	$1,5 \leq X > 3$	$3 \leq X > 4,5$	$4,5 \leq X \geq 6$
G	$0 \leq X > 1$	$1 \leq X > 2$	$2 \leq X > 3$	$3 \leq X \geq 4$
I	$0 \leq X > 1,75$	$1,75 \leq X > 3,5$	$3,5 \leq X > 5,25$	$5,25 \leq X \geq 7$
PP	$0 \leq X > 0,25$	$0,25 \leq X > 0,5$	$0,5 \leq X > 0,75$	$0,75 \leq X \geq 1$
RB	$0 \leq X > 2$	$2 \leq X > 4$	$4 \leq X > 6$	$6 \leq X \geq 8$
FC	$0 \leq X > 2,5$	$2,5 \leq X > 5$	$5 \leq X > 7,5$	$7,5 \leq X \geq 10$
GC	$0 \leq X > 0,25$	$0,25 \leq X > 0,5$	$0,5 \leq X > 0,75$	$0,75 \leq X \geq 1$
CA	$0 \leq X > 0,5$	$0,5 \leq X > 1$	$1 \leq X > 1,5$	$1,5 \leq X \geq 2$
CV	$0 \leq X > 0,25$	$0,25 \leq X > 0,5$	$0,5 \leq X > 0,75$	$0,75 \leq X \geq 1$

Fonte: próprios autores (2022)

Tabela 2 - Categorização dos atributos dos goleiros.

	A	B	C	D
PE	$0 \leq X > 1,75$	$1,75 \leq X > 3,5$	$3,5 \leq X > 5,25$	$5,25 \leq X \geq 7$
FD	$0 \leq X > 0,25$	$0,25 \leq X > 0,5$	$0,5 \leq X > 0,75$	$0,75 \leq X \geq 1$
FC	$0 \leq X > 0,25$	$0,25 \leq X > 0,5$	$0,5 \leq X > 0,75$	$0,75 \leq X \geq 1$
GC	$0 \leq X > 0,25$	$0,25 \leq X > 0,5$	$0,5 \leq X > 0,75$	$0,75 \leq X \geq 1$
CA	$0 \leq X > 0,25$	$0,25 \leq X > 0,5$	$0,5 \leq X > 0,75$	$0,75 \leq X \geq 1$
CV	$0 \leq X > 0,25$	$0,25 \leq X > 0,5$	$0,5 \leq X > 0,75$	$0,75 \leq X \geq 1$
SG	$0 \leq X > 0,5$	$0,5 \leq X > 1$	$1 \leq X > 1,5$	$1,5 \leq X \geq 2$

DD	$0 \leq X > 2$	$2 \leq X > 4$	$4 \leq X > 6$	$6 \leq X \geq 8$
DP	$0 \leq X > 0,5$	$0,5 \leq X > 1$	$1 \leq X > 1,5$	$1,5 \leq X \geq 2$
GS	$0 \leq X > 1,5$	$1,5 \leq X > 3$	$3 \leq X > 4,5$	$4,5 \leq X \geq 6$

Fonte: próprios autores (2022)

Para o atributo “Pontos”, ao invés de obter uma classificação em relação a todos os dados do atributo disponíveis no *dataset*, o mesmo foi classificado em quartis em relação a cada rodada para melhor análise do desempenho em cada momento do Brasileirão. O atributo “Pontos” recebe uma nova nomenclatura e passa a ser chamado de “CLASSE”, sendo classificado de 1 a 4, sendo 1 o grupo de atletas com pior desempenho da rodada e 4 o grupo com maior desempenho daquela rodada. A classificação dos itens em quartis foi feita porque os atributos têm valor numérico, sendo assim, classificá-los faz com que grupos de valores possam representar um mesmo elemento, e que as regras de associação representem o comportamento de influência entre as variáveis de modo mais eficiente (CALÇADA *et al*., 2020).

Finalizando a etapa de pré-processamento, foram separados os dados por posição, onde foram coletados para o estudo apenas os registros de goleiros e atacantes, bem como os seus atributos particulares demonstrados por meio das Tabelas 3 e 4, respectivamente. Os dados foram manipulados no formato CSV, onde depois foram transformados em arquivo “.arff”.

Tabela 3 - Dados coletados de atacantes.

ITEM	DESCRIÇÃO
FS	faltas sofridas
PE	passes errados
A	assistências
FT	finalizações na trave
FF	finalizações para fora
G	gols
I	impedimentos
PP	pênaltis perdidos
RB	roubadas de bola

FC	faltas cometidas
GC	gols contra
CA	cartões amarelo
CV	cartões vermelhos
Pontos	pontuação do jogador

Fonte: próprios autores (2022)

Tabela 4 - Dados coletados de goleiros.

ITEM	DESCRIÇÃO
PE	passes errados
FD	finalizações defendidas
FC	faltas cometidas
GC	gols contra
CA	cartões amarelo
CV	cartões vermelhos
SG	jogos sem sofrer gols
DD	defesas difíceis
DP	defesas de pênalti
GS	gols sofridos
Pontos	pontuação do jogador

Fonte: próprios autores (2022)

A partir dos *datasets* de goleiro e atacante foram extraídas as Regras de Associação com o uso do Algoritmo *Apriori*-TID implementado em Java. Para a obtenção de todas as regras possíveis, os valores mínimos de confiança e suporte foram zerados. Foram geradas duas *Filtered*-ARNs, uma para atacantes e outra para goleiros, ambas com atributo alvo "[CLASSE]=4", a fim de que sejam descobertas as principais características dos jogadores de maior pontuação em cada uma dessas posições.

Com os dados resultantes do pré-processamento, com a finalidade de comparação de resultados gráficos, foram geradas árvores de decisão para o atributo de "Classe" dos conjuntos de dados de atacantes e goleiros, utilizando o WEKA, um *software open*

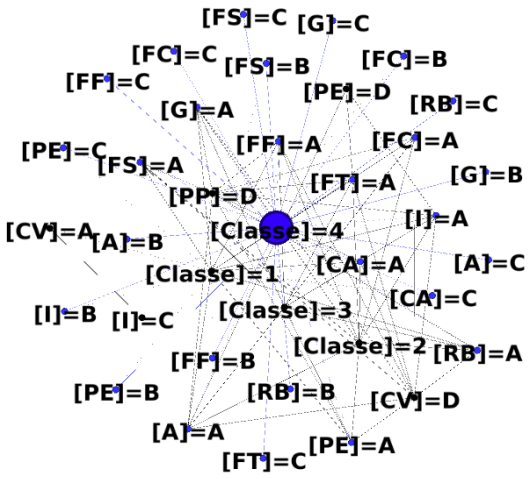
source. WEKA é uma coleção de algoritmos de aprendizado de máquina para tarefas de mineração de dados que contém ferramentas para pré-processamento de dados, classificação, regressão, *clustering*, regras de associação e visualização. Este estudo empírico, no entanto, trata apenas de um subconjunto de algoritmos classificadores.

RESULTADO E DISCUSSÕES

Com os *datasets* prontos, foi executado o algoritmo *Apriori*-TID que teve como resultado a extração de 1.245 regras para atacante e 554 regras para goleiro. Logo após a extração, foram aplicados os filtros de *Gain* e *Added Value* para excluir regras que não tinham uma influência estatística comprovada sobre os elementos e formando as *Filtered*-ARNs de cada posição. Após o filtro, foram obtidas, para posição de atacante e goleiro, a quantidade de 641 e 262 regras, respectivamente.

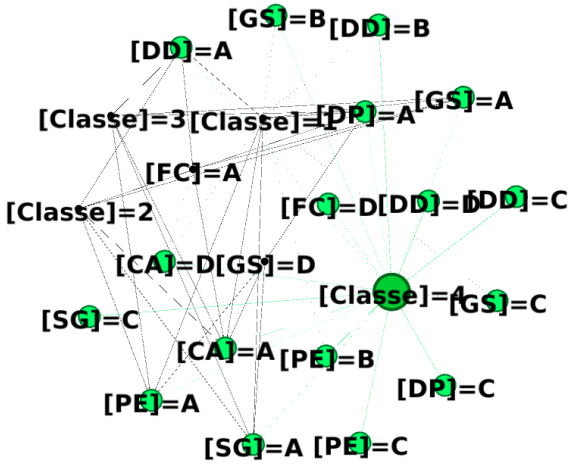
Ambas as redes filtradas são mostradas nas Figuras 2 e 3. Com esse intuito, observa-se os nós diretamente ligados ao nó alvo, denominados de nós de nível 1, pois os mesmos influenciam diretamente no resultado que se deseja analisar.

Figura 2 – *Filtered*-ARNs de Atacantes com "[CLASSE]=4".



Fonte: próprios autores (2022)

Figura 3 – *Filtered*-ARNs de Goleiros com "[CLASSE]=4".

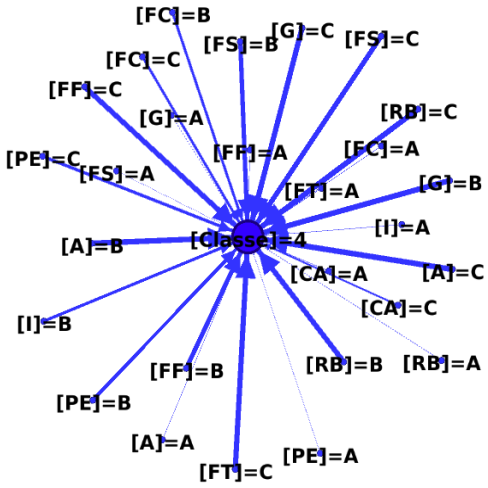


Fonte: próprios autores (2022)

A *Filtered*-ARN com "[CLASSE]= 4"como nó alvo para atacantes, apresenta 27 itens de nível 1 da rede (Figura 4). Ao analisar os atacantes com desempenho elevado, pode-se observar que fatores que mais influenciam são uma mediana de gols ([G]= B ou C) e de assistências ([A]= B ou C), mas também destaca um atacante com bons números de roubadas de bola ([RB]= B ou C). A característica de roubada de bola é um item diferencial que pode ser elencado no momento de seleção de jogador para montagem do time do Cartola FC. Com este conhecimento pode-se obter melhores resultados de pontuação.

Foi constatado também na rede que os atributos finalizações na trave, finalizações para fora e faltas sofridas tem pouca influenciam no “[CLASSE] = 4”, uma vez que quase todas as categorias ocorrem no Nivel 1, com ou sem antecedentes. Esse resultado vem ao analisar que os jogadores “[CLASSE] = 4” tem podem ter qualquer classificação desses itens.

Figura 4 – Nível 1 da Filtered-ARN com nó alvo “[CLASSE]=4” para atacantes.



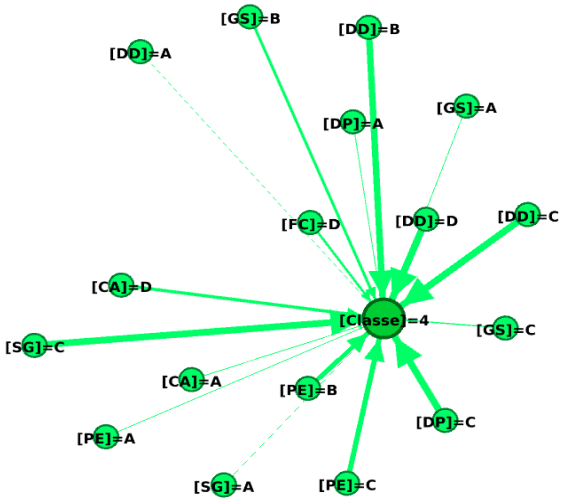
Fonte: próprios autores (2022)

Também é interessante destacar que os gols de um atacante quando são classificados como “[G] = A” possuem antecedentes que devem influenciar para que o jogador tenha o desempenho “[CLASSE]=4”. Já quando o atacante tem seus gols classificados como “[G] = B ou C”, não possui antecedentes, o que indica que a partir dessa classificação o jogador já estará dentro do grupo de bons desempenhos da rodada.

Já na *Filtered*-ARN com “[CLASSE]=4” como nó alvo para goleiros (Figura 5), pode-se verificar que o atributo “[DD]”, relacionado a defesas difíceis, quando possui valor não nulo tem sempre influência no desempenho do jogador desta posição. Em categorias superiores como “[DD] = C ou D”, os nós não possuem antecedentes, mostrando que esse atributo consegue gerar influência sem depender de outros.

Existem ainda atributos que não tinham influência no resultado do desempenho da pontuação do jogador mesmo estando em níveis mais elevados. É possível observar nas *Filtered*-ARNs de atacantes e goleiros, alguns atributos que geralmente deveriam classificar os jogadores na classe 4 como é o caso de “[RB]‘= D”, “[A] = D” e “[G] = D” para os atacantes e “[DP] = D” e “[SG] = D” para goleiros.

Figura 5 – Nível 1 da *Filtered*-ARN com nó alvo "[CLASSE]=4" para goleiros.



Fonte: próprios autores (2022)

Analisando os resultados do algoritmo da árvore de decisão para atacantes e goleiros, ao qual foi selecionado o atributo Classe, disponíveis on-line[3,4], percebeu-se uma pobreza no conhecimento gerado e a dificuldade na sua análise, em decorrência do tamanho das árvores. A árvore para os dados de atacantes possui 82 folhas com uma altura 109, e para o *dataset* de goleiro com 49 folhas com altura 65. O resultado gráfico das árvores não possibilita a identificação de fatores importantes diretamente conectados a jogadores de maior pontuação no Cartola FC.

RESULTADO E DISCUSSÕES

A elaboração desse trabalho foi de grande valia para o estudo das técnicas de Inteligência Artificial Explicável, auxiliando principalmente na descoberta de conhecimento sobre o desempenho de jogos digitais. Tais processos podem colaborar na automatização, tomada de decisões e reconhecimento de padrões e tendências por meio das Redes de Regras de Associação Filtradas.

[3] https://drive.google.com/file/d/1X6DM2HeVy6MQ2fFMLSGMZdvV_RRCk8Jb/view?usp=sharing

[4] https://drive.google.com/file/d/12HntdAuvNAxj9aB0IHAy2uvf_oH8BLhD/view?usp=sharing

Sobre o rendimento com os dados do Cartola FC, percebeu-se pela análise dos resultados que alguns atributos considerados de extrema relevância para classificação dos jogadores não obrigatoriamente precisam estar nos níveis mais elevados (categoria D), uma vez que jogadores de "[CLASSE]" = 4 sofrem influência principalmente de atributos intermediários (categoria B e C). A técnica de *Filtered*-ARNs conseguiu produzir conhecimento a respeito do perfil dos atributos dos jogadores com maior pontuação. Esse conhecimento pode ser utilizado para a elaboração de um time com maior possibilidade de ficar em primeiro lugar na classificação geral do jogo, uma vez que será formado por jogadores com mais possibilidades de terem pontuação mais elevada.

Para trabalhos futuros pode-se utilizar a técnica de *Filtered*-ARN em outros jogos digitais, bem como incluir atributos relacionados a parte financeira das cartoletas, com o objetivo de geração automatizada de times com a distribuição do dinheiro digital disponível. Outras Redes também podem ser analisadas para que mais conhecimento possa ser gerado a respeito do Cartola FC e de outros jogos similares.

AGRADECIMENTOS

Os autores agradecem a todos que disponibilizaram os dados utilizados nessa pesquisa e ao Grupo de Estudo e Desenvolvimento de Aplicações Inteligentes (GEDAI) da Universidade Estadual do Piauí, que proporcionou toda a estrutura necessária para que os experimentos fossem realizados e a pesquisa concluída.

REFERÊNCIAS

BATISTA, A.; DEMARCHI, A.; ROCHA, L. V. . Gamification as a strategy content marketing at fantasy game cartola fc, **Revista Temática,** 2019. https://periodicos.ufpb.br/ojs/index.php/tematica/article/download/48905/28508/.

CALÇADA, D. B.; CALÇADA, J. C. M.; SILVA, J. D.; REZENDE, S. O. (2020). COVID-19 diagnostic parameters: knowledge discovery using Filtered-Association Rules Networks, **Journal of Global Innovation 2**, 2020.

CALÇADA, D. B.; DE PÁDUA, R.; REZENDE, S. O. Asymmetric Objective Measures applied to Filter Association Rules Networks, **XLIV Latin American Computer Conference (CLEI) Asymmetric**, São Paulo, pp. 258–267, 2018.

CALÇADA, D. B.; REZENDE, S. O. Filtered-ARN: Asymmetric objective measures applied to filter Association Rules Networks, **CLEI Electronic Journal 22**(3): 1–16.

CALÇADA, D. B., 2019. Redes de regras de associação filtradas e multialvo, Master's thesis.

FAYYAD, U.; PIATETSKY-SHAPIRO, G. S. P.. From data mining to knowledge discovery in databases, 1996. https://ojs.aaai.org//index.php/aimagazine/article/view/1230.

PANDEY, G., CHAWLA, S., POON, S., ARUNASALAM, B., DAVIS, J. G. Association rules network: Definition and applications. Statistical Analysis and Data Mining: The ASA Data Science Journal, v. 1, n. 4, p. 260-279, 2009.

RIBEIRO, L. E. d. S. Predição de escalações para o jogo cartolafc utilizando aprendizado de máquina e otimização, 2019. https://repositorio.ufu.br/handle/123456789/26681.

SEHNEM, R.; FROZZA, R.; BAGATINI, D.; PERANCONI, D. Análise de variáveis em partidas de futebol: Previsão de resultados com naïve bayes e poisson, 2021, **Anais do XVIII Encontro Nacional de Inteligência Artificial e Computacional**, SBC, Porto Alegre, RS, Brasil, pp. 13–24. URL: https://sol.sbc.org.br/index.php/eniac/article/view/18237

SILVA, J. D.; CALÇADA, J. C. M.; REZENDE, S. O.; CALÇADA, D. B. Automatic identification of knowledge related to dengue cases in the state of piauí in public databases using filtered-association rules networks, **Revista de Informatica Teorica e Aplicada** 27(3): 40–49, 2020.

UOL. Cartola FC: como funciona, o que é, como é a pontuação e outras dúvidas, 2020. Available at https://www.uol.com.br/esporte/faq/cartola-fc-como-funciona-o-que-e-como-e-a-pontuacao-e-outrashtm.

VERAS, V. N. R.; CALÇADA, J. C. M.; BEZERRA, S. M. G.; CALÇADA, D. B. Descoberta de padrões no tratamento da hanseníase no estado do Piauí com uso de técnica de Inteligência Artificial Explicável, **Revista Concilium** 22(4): 527–541, 2022.

VISCONDI, G.; JUSTO, D.; GARCÍA, N. Aplicação de aprendizado de máquina para otimização da escalação de time no jogo cartola FC, 2017.

Recebido em: 10/08/2022

Aprovado em: 12/09/2022

Publicado em: 21/09/2022