

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS
ESCOLA POLITÉCNICA
CURSO DE CIÊNCIA DA COMPUTAÇÃO



**MINERAÇÃO DE DADOS APLICADA A PREVISÃO DE RESULTADOS DE
JOGOS DE FUTEBOL**

PEDRO HENRIQUE CANDIDO ROSA

GOIÂNIA
2022

PEDRO HENRIQUE CANDIDO ROSA

**MINERAÇÃO DE DADOS APLICADA A PREVISÃO DE RESULTADOS DE
JOGOS DE FUTEBOL**

Trabalho de Conclusão de Curso apresentado à Escola Politécnica da Pontifícia Universidade Católica de Goiás, como parte dos requisitos para a obtenção do título de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Sibelius Lellis Vieira

Banca examinadora: Prof. Ms. Joriver Rodrigues Canedo
Prof. Dr. José Elmo de Menezes

GOIÂNIA

2022

PEDRO HENRIQUE CANDIDO ROSA

MINERAÇÃO DE DADOS APLICADA A PREVISÃO DE RESULTADOS DE JOGOS
DE FUTEBOL

Trabalho de Conclusão de Curso aprovado em sua forma parcial pela Escola de Ciências Exatas e da Computação, da Pontifícia Universidade Católica de Goiás, para obtenção do título de Bacharel em Ciência da Computação, em ____/____/____.

Orientador: Prof. Dr. Sibelius Lellis Vieira

Examinador: Prof. Ms. Joriver Rodrigue Canedo

Examinador: Prof. Dr. José Elmo de Menezes

GOIÂNIA

2022

RESUMO

O futebol é um esporte de grande popularidade, e a busca por conhecimento a respeito dele se popularizou. A previsão é uma questão muito solicitada quando se fala em estatística a respeito do futebol, e ao mesmo tempo que solicitada, possui uma complexa dificuldade de realização. Com base nessa questão, este trabalho busca por meio de registros de jogos e da mineração de dados explorar análises e previsões a respeito da *premier league*, especificamente no âmbito que diz a respeito dos resultados que indiquem a entrada na zona da *champions league*, observando seu comportamento com diferentes perspectivas das ciências de dados e mineração de dados. Este trabalho apresenta experimentos com identificações de padrões visuais e experimentos que utilizam métodos de árvore de decisão e redes neurais, os quais apresentam bons resultados em termos de acurácia.

Palavras-chave: futebol, ciência de dados; mineração de dados, descoberta de conhecimento de base de dados, previsão.

ABSTRACT

Football is a very popular sport, and the search for knowledge about it has become popular. Forecasting is a very requested question when talking about football statistics, and at the same time it is requested, it has a complex realization difficulty. Based on this issue, this work seeks, through game records and data mining, to explore analyzes and forecasts regarding the premier league, specifically in the scope that concerns the results that indicate the entry into the champions league zone, observing its behavior with different perspectives from data science and data mining. This work presents experiments with identification of visual patterns and experiments that use decision tree methods and neural networks, which present good results in terms of accuracy.

Keywords: football, data science; data mining, database knowledge discovery, forecasting.

LISTA DE ABREVIATURAS

ARFF – *Attribute-Relation File Format*
CBF – Confederação Brasileira de Futebol
CSV – *Comma-Separated Values*
C-MEANS – *C-means clustering*
DNA – Ácido desoxirribonucleico
EY – Agência EY
FIFA – Federação internacional de Futebol
I.A – Inteligência artificial
INJ – Instituto Neymar Junior
KDD – *Knowledge Discovery in Databases*
KNN – *K-nearest neighbors*
K-MEANS – *K-means clustering*
MD – Mineração de dados
MDP – *Markov Decision Process*
MGF – Média de gols feitos
MGFC – Média de gols feitos em casa
MGFF – Média de gols feito fora de casa
MGS – Média de gols sofridos
MGSC – Média de gols sofridos em casa
MGSF – Média de gols sofridos fora de casa
MLP – MultilayerPerceptron
PIB – Produto Interno Bruto
RNA – Rede neural artificial
SOM – *Self Organizing Map*
SVM – *Support Vector Machine*
TBL – *team based learning*
UE – União Europeia
WEKA – *Waikato Environment for Knowledge Analysis*

LISTA DE ILUSTRAÇÕES

FIGURAS

Figura 1. Popularidade dos Sports	12
Figura 2. Receita econômica do futebol Europeu.....	14
Figura 3. Tipos de Machine Learning	17
Figura 4. KDD.....	18
Figura 5. Mineração de dados	18
Figura 6. Neurônio humano	20
Figura 7. Unidade de rede neural	20
Figura 8. Representação de uma rede neural	21
Figura 9. Representação de uma árvore de decisão.....	22
Figura 10. Interface do software Jupyter Notebook	23
Figura 11. Interface inicial do software Weka	24
Figura 12. Interface inicial do explorer do software Weka	24
Figura 13. Interface com a classificação do explorer do software Weka	25
Figura 14. Informações constante no site fbref.com	26
Figura 15. Modelo dos dados após extração do site fbref.com	26
Figura 16. Parte 1 – extração dos dados.....	27
Figura 17. Parte 2 – extração dos dados.....	28
Figura 18. Parte 2 – extração dos dados.....	28
Figura 19. Parte 3 – extração dos dados.....	29
Figura 20. Parte 4 – extração dos dados.....	29
Figura 21. Parte 5 – extração dos dados.....	29
Figura 22. Parte 6 – extração dos dados.....	30
Figura 23. Parte 7 – extração dos dados.....	30
Figura 14. Modelo dos dados após 1ª etapa do tratamento	31
Figura 15. Modelo dos dados após 2ª etapa do tratamento	32
Figura 26. Dados processados na ferramenta Weka	33
Figura 27. Pontuação e troca de treinador na temporada 2011/2012	34
Figura 28. Pontuação e troca de treinador na temporada 2012/2013	34
Figura 29. Pontuação e troca de treinador na temporada 2013/2014	35
Figura 30. Pontuação e troca de treinador na temporada 2014/2015	35
Figura 31. Pontuação e troca de treinador na temporada 2015/2016	36

Figura 32. Pontuação e troca de treinador na temporada 2016/2017	36
Figura 33. Pontuação e troca de treinador na temporada 2017/2018	37
Figura 34. Pontuação e troca de treinador na temporada 2018/2019	37
Figura 35. Pontuação e troca de treinador na temporada 2019/2020	38
Figura 36. Pontuação e troca de treinador na temporada 2020/2021	38
Figura 37. Média de pontuação por rodada nas temporadas 2011 a 2021	39
Figura 38. Média de gols feitos e sofridos na temporada 2011/2012	39
Figura 39. Média de gols feitos e sofridos na temporada 2012/2013	40
Figura 40. Média de gols feitos e sofridos na temporada 2013/2014	40
Figura 41. Média de gols feitos e sofridos na temporada 2014/2015	41
Figura 42. Média de gols feitos e sofridos na temporada 2015/2016	41
Figura 43. Média de gols feitos e sofridos na temporada 2016/2017	42
Figura 44. Média de gols feitos e sofridos na temporada 2017/2018	42
Figura 45. Média de gols feitos e sofridos na temporada 2018/2019	43
Figura 46. Média de gols feitos e sofridos na temporada 2019/2020	43
Figura 47. Média de gols feitos e sofridos na temporada 2020/2021	44

QUADROS

Quadro 1 – Tipos de técnicas relacionadas a análises para mineração de dados. ...	19
--	----

TABELAS

Tabela 1 – lista de atributos	32
Tabela 2 – relação de desempenho por algoritmo com 70 instancias	45
Tabela 3 – relação de desempenho por algoritmo com 200 instancias	45

SUMÁRIO

1. INTRODUÇÃO.....	10
1.1 Contextualização.....	10
1.2 Justificativa.....	10
1.3 Objetivo.....	10
1.4 Estrutura do trabalho.....	11
2. REFERENCIAL TEÓRICO	12
2.1 Popularidade	12
2.2 Impacto cultural.....	13
2.3 Impacto Social.....	13
2.4 Impacto Econômico.....	14
2.5 Estatística.....	15
2.6 Inteligência artificial.....	15
2.7 Mineração de dados.....	18
3. Materiais e Métodos.....	23
4. EXPERIMENTOS E RESULTADOS.....	26
4.1 Coleta de dados	26
4.2 Tratamento dos dados	31
4.3 Aplicação dos dados	33
4.4 Discussão e Resultados.....	33
5. CONCLUSÃO	46
REFERÊNCIAS	47

1. INTRODUÇÃO

1.1 Contextualização

O futebol é o esporte mais popular de todos os tempos, responsável por afetar diretamente pilares da sociedade como áreas sociais, culturais e econômicas (CAMPOS, 2013). Observando esse contexto em um desporto de tamanho representatividade, é necessário um grande desenvolvimento em vários aspectos profissionais de desempenho.

Conforme o crescimento do esporte, também é acrescida a tecnologia em auxílio de seu desenvolvimento, de forma a melhorar cada vez mais o esporte (ALMEIDA, 2020). Devido a estas questões, surgiram-se diversos campos de estudos desta combinação como as áreas que buscam analisar e prever os resultados e desempenho das equipes a fim de auxiliar nos mais diversos ambientes, como melhorar o desempenho de equipes dentro de campo, ou até mesmo utilizar em ambientes de apostas esportivas com a possibilidade de uma maior assertividade (COMO A TECNOLOGIA..., 2022).

Os estudos de técnicas de previsão vêm ganhando forças e causando impactos ao longo das últimas décadas, esse trabalho visa estudar um modelo para a previsão aplicada ao campeonato de futebol inglês conhecido como *Premier League*.

1.2 Justificativa

Observando o crescimento do mercado futebolístico e a profissionalização do esporte, é notado a necessidade de se analisar os dados e compreendê-los, pois é através da possibilidade de criação de previsões com maiores taxas de assertividades, possibilitando assim criar estratégias e fazer as melhores tomadas de decisões para conclusão do objetivo definido.

1.3 Objetivo

1.3.1 Objetivo geral

Aplicar técnicas de mineração de dados nos dados coletados dos últimos anos do campeonato inglês para prever os prováveis classificados para a zona da *Champions League*.

1.3.2 Objetivos específicos

Para se atingir o objetivo geral, propõem-se os seguintes objetivos específicos:

- Desenvolver um estudo através de uma pesquisa exploratória para os dados que impactam de forma direta no resultado das partidas;
- Explorar os conceitos de *ciência de dados* e técnicas de mineração de dados;

1.4 Estrutura do trabalho

Esse estudo apresenta-se estruturado em 5 (cinco) capítulos, sendo o capítulo 1 (um) referente a esta introdução. No capítulo 2, apresenta-se a base de estudo dos conceitos relacionados ao trabalho. No capítulo 3, são descritos os materiais e métodos utilizados para a implementação dos experimentos realizados. No capítulo 4, referencia-se os resultados obtidos. O Capítulo 5, apresenta a conclusão referente ao trabalho.

2. REFERENCIAL TEÓRICO

O futebol é um esporte coletivo de extensa lista de conteúdo relacionado a seu conceito, metodologia e propriedades que o constitui. Esta seção busca apresentar resumidamente seu conceito, valores culturais, sociais, econômicos e sua relação com análises estáticas interligadas a mineração de dados e inteligência artificial para análises e identificação de padrões e previsões.

2.1 Popularidade

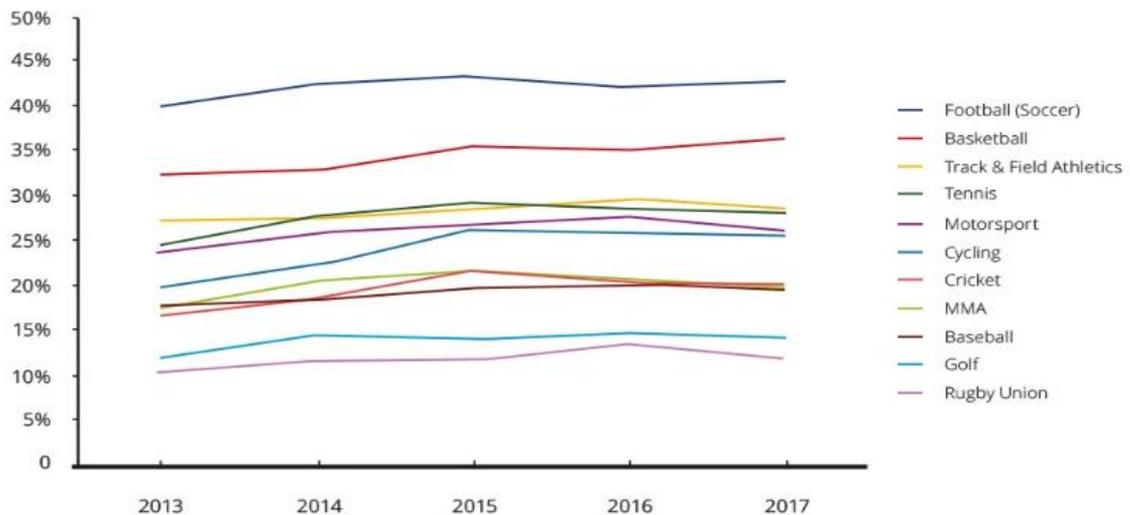
A popularidade do esporte não é fator único, pode elencar-se alguns fatores que contribuem para o sucesso do futebol como o surgimento de um esporte semelhante em diversas eras e locais diferente, ou a sua forma oficial tendo sido declarado na era dominante da coroa britânica o que facilitou sua exportação além das fronteiras, a sua simplicidade facilitou a sua compreensão e adesão pelo público, sua improvisação e presença de espírito dos torcedores aumentaram a sua atração e forma de ser visto, a atmosfera da rivalidade trazendo um tempero a mais ao jogo e a publicidade em torno dele criou-se uma círculo virtuoso do futebol (TEIXEIRA, 2018).

Um total de 3,572 bilhões de telespectadores, mais da metade da população global com quatro anos ou mais, sintonizou a maior competição do futebol mundial, de acordo com dados de audiência da cobertura oficial da Copa do Mundo da Federação internacional de Futebol (FIFA) 2018. (FIFA, 2018)

É um esporte com grande apelo global e interesse dos fãs. Na verdade, mais de 40% das pessoas com 16 anos ou mais nos principais centros populacionais do mundo se consideram interessadas ou muito interessadas em seguir o futebol, mais do que qualquer outro esporte. (NIELSIN COMPANY, 2018).

A empresa norte-americana *Nielsen Sports* fez uma pesquisa entre 2013 e 2017 para avaliar a relação de popularidade dos esportes ao longo dos anos entre os jovens. A pesquisa demonstrou uma tendencia superior relacionando o futebol, o qual se manteve no topo durante toda a pesquisa (Figura 1).

Figura 1. Popularidade dos Sports



Fonte: NIELSEN COMPANY, 2018

2.2 Impacto cultural

O futebol existe dentro de uma esfera sociocultural, e este regulamento está transformando o futebol como entidade cultural; um desporto que é reconhecido pela União Europeia (UE) como tendo um papel importante na formação da identidade, aproximando as pessoas e apoiando as comunidades locais em geral (CROLLEY; LEVERMORE; PEARSON, 2007).

Segundo a publicação de Sawyer (2018) para o jornal britânico *The Guardian*, “existem poucas entidades desportivas ou culturais capazes de atravessarem fronteiras internacionais, atravessando todas as culturas e impactando diretamente na forma como as sociedades se interpretam”.

2.3 Impacto Social

Existem inúmeros projetos sociais relacionado ao futebol, que buscam criar práticas saudáveis e aprendizado de habilidades importantes para a vida em sociedade e principalmente contribuir com a inclusão do indivíduo na sociedade. Esses projetos visão ajudar principalmente as famílias mais carentes. É notório observar que a contribuição de grande parte destes projetos é incentivada e iniciada por atletas profissionais que já atuaram ou ainda atua. Em grande maioria se sensibilizam e entendem a dificuldade de grande parte da população (GESTÃO ESPORTIVA, 2020).

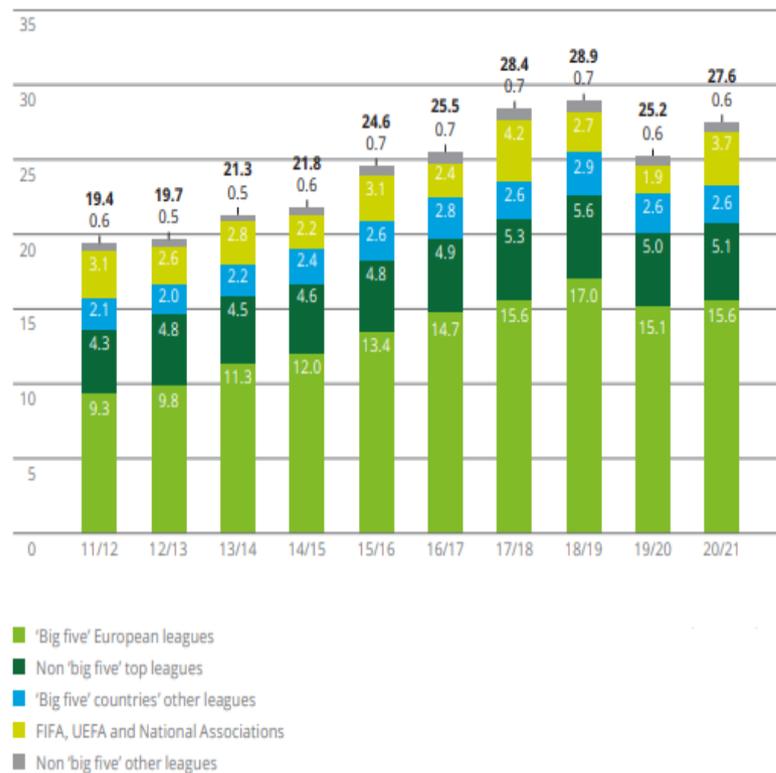
Esses projetos são responsáveis por afetar de maneira positiva milhares de vidas, gerando emprego e oportunidades, além de desenvolver habilidades. No Brasil o Instituto Neymar Jr (INJ) vem fazendo um grande trabalho com crianças e adolescentes em situação de vulnerabilidade e suas famílias, através de ações sociais, educacionais, culturais e esportiva, o qual atualmente não trabalha apenas com o futebol, mas que possui suas raízes de nascença conectada com o campo. Segundo o Instituto, “atualmente são atendidas milhares de crianças e jovens, em uma área de 8.400 m², com impacto em mais de 10.000 pessoas” (INSTITUTO NEYMAR..., [2022]?).

2.4 Impacto Econômico

O futebol assim como os demais esportes possui participações na economia, através de diversos mecanismos como a venda de ingressos, camisas, direitos de transmissão e imagem, contratos, parcerias, geração de empregos entre outras maneiras de se gerar ou movimentar a economia. No Brasil o futebol é responsável por 0,72% do Produto Interno Bruto (PIB) nacional, representando um valor total de R\$ 52,9 bilhões, conforme relatório produzido pela Confederação Brasileira de Futebol (CBF) e Agência EY (EY) (CBF; EY, 2018).

A consultoria *Deloitte* fez o relatório que ilustra que a Europa mesmo diante de um cenário pandêmico as receitas do futebol voltaram a crescer e superaram o PIB de mais de 90 países atingindo a marca de €27,6 bilhões (R\$ 144,8 bilhões) na temporada de 2020/2021 (Figura 2).

Figura 2. Receita econômica do futebol Europeu



Fonte: DELLOITTE, 2022;

2.5 Estatística

A estatística pode ser definida como uma área da matemática voltada ao estudo de dados numéricos e casos, demonstrando uma previsibilidade da situação, buscando modelar e usar um determinado conjunto para alcançar um determinado objetivo, podendo ser utilizada em diversas áreas para diferentes aplicações. Para Wheelan (2017) diferente do cálculo, “Ela pode ser usada para explicar tudo, desde testes de DNA até a idiotice de jogar na loteria”.

Para Cordani (2003), a estatística possui uma importância para a tomada de decisão em vários campos do saber, principalmente nos dias de hoje, com o acúmulo de informações disponíveis.

A adaptabilidade e tomadas de decisões contribuem para o uso diversificado no dia a dia. “A estatística é uma ciência que usa a análise dos dados para testar as hipóteses estatísticas, verificar a força da evidência clínica e, assim, se existem associações entre grupos ou a veracidade de fenômenos de interesse.” (RODRIGUES; LIMA; BARBOSA, 2018).

2.6 Inteligência artificial

A Inteligência Artificial não possui uma data específica para seu nascimento, pois seu conceito e ideia foram popularizados bem antes do surgimento da tecnologia capaz de replicá-la. O desenvolvimento teve grande crescimento durante e a partir da segunda guerra mundial após diversos cientistas renomados começarem a se dedicar nesta área.

2.6.1 Inteligência artificial: história

Em 1943, Warren McCulloch e Walter Pitts apresentaram um artigo sobre as redes neurais e as estruturas de raciocínio artificiais que são capazes de imitar o sistema nervoso. Já em 1950, é fundado o primeiro laboratório de inteligência artificial por Allen Newell e Herbert Simon, e no mesmo ano Alan Turing apresenta um teste conhecido como jogo da imitação, o qual seria utilizado para avaliar se uma máquina é capaz de se passar por um humano em diálogo escrito e a partir deste teste começaram a surgir outros eventos impactantes para a pesquisa. Em 1956 John McCarthy juntamente com outros cientista computacionais da época na conferência de Dartmouth formalizou o termo inteligência artificial (I.A) além de avançar as pesquisas nesse campo (KLEINA, 2018).

2.6.2 Inteligência artificial: conceitos

Inteligência Artificial é um dos conceitos mais especulados quando se trata do assunto sobre inovações tecnológicas, pois seu conceito está relacionado a criação de entidades com capacidades, possuindo diversas abordagens para representar seu conceito (INTELIGÊNCIA ARTIFICIAL..., 2022).

A inteligência artificial pode ser observada e relacionada ao comportamento humano, como explica McCarthy (1956), “fazer a máquina comportar-se de tal forma que seja chamada inteligente caso fosse este o comportamento de um ser humano”. Para Russell (2013), “a I.A abrange uma enorme variedade de subcampos, do geral (aprendizagem e percepção) até tarefas específicas”.

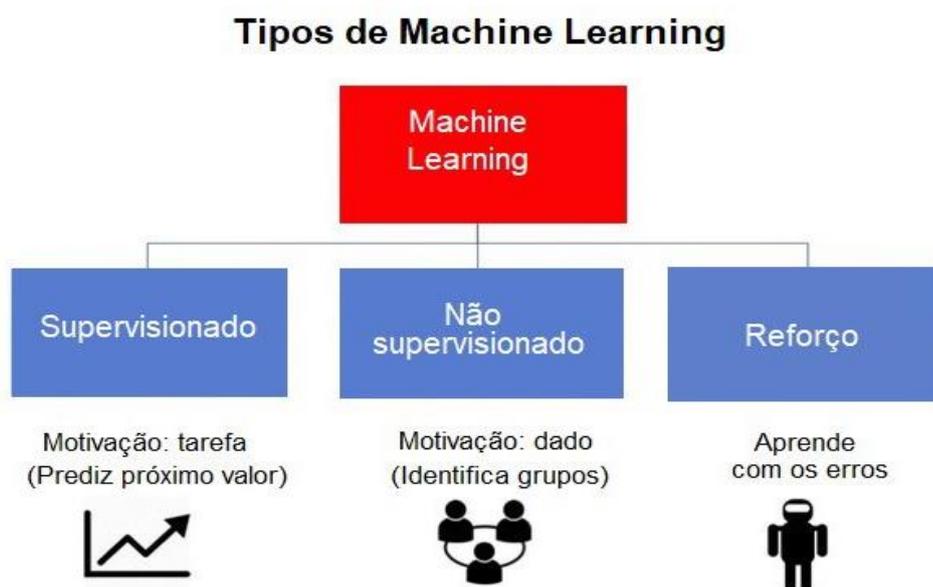
2.6.3 Inteligência artificial: aprendizagem de máquina

Existe uma ideia tradicional conhecida como “ninguém nasce sabendo”, e essa frase é uma representação de que para a execução de uma determinada atividade é necessário ensinar. No campo da inteligência artificial segue-se a mesma metodologia, sendo necessário treinar a máquina a um certo ponto em que ela possa se autodesenvolver. Russell e Norvig (2013) relatam que, “Um agente está aprendendo se melhorar o seu desempenho nas tarefas futuras de aprendizagem”.

Existem diversos tipos de aprendizagem como mostra na figura 3 que podem ser classificados em setores, exemplo, os aprendizados supervisionados que possuem conjunto de dados correspondente para classificar e analisar sua condição de previsibilidade utilizando algoritmos como árvores de decisão, Rede Neural Artificial (RNA), *Support Vector Machine* (SVM), *Team Based Learning* (TBL); ou os não supervisionados que não possuem esses dados rotulados e interpreta e organiza por conta própria utilizando de algoritmos como *C-means clustering* (C-means), *K-means clustering* (K-means), *K-nearest neighbors* (KNN), *Self Organizing Map* (SOM); ou por reforço o qual aprende com seus erros utiliza-se algoritmos como *Markov Decision Process* (MDP) e Q-learning (NEVES, 2020).

A aprendizagem pode assumir diferentes tipos, a depender de uma combinação de fatores como a natureza do agente, do componente a ser aperfeiçoado e da realidade disponível (RUSSEL; NORVIG, 2013).

Figura 3. Tipos de Machine Learning



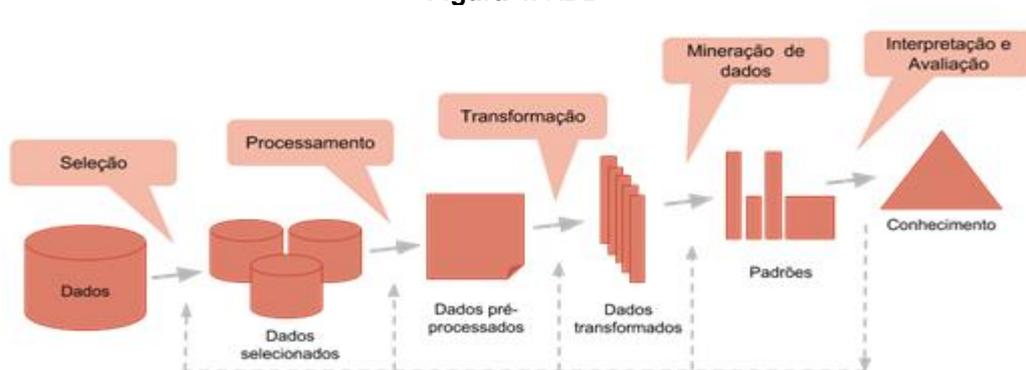
Fonte: NEVES, 2020;

2.7 Mineração de dados

A mineração de dados (MD) é um conjunto de métodos dentro de uma sequência de passos conhecido como *Knowledge Discovery in Databases* (KDD) (figura 4), que visa descobrir informações através da análise de um conjunto de dados, buscando encontrar anomalias, padrões ou analogias para prever resultados e tomadas de decisões (SILVA; PERES; BOSCARIOLI, 2016). Segundo Silva, Peres e Boscaroli (2016), “É a partir do estudo e da mineração de dados, a descoberta acontece, e então novo conhecimento é produzido, contribuindo para a melhoria de produtos, sistemas, processos, negócios etc”.

São utilizadas técnicas de estatísticas, inteligência artificial e ciência de dados para ajudar na descoberta (figura 5). A MD é utilizada em diversas tarefas como classificação, regressão, agrupamento e associação (SILVA; PERES; BOSCARIOLI, 2016).

Figura 4. KDD



Fonte: MOURA, [20219];

Figura 5. Mineração de dados



Fonte: DATA MINING, [2022?];

Essas ferramentas estão conceitualmente interligadas a diferentes tipos de análise, sendo elas a análise descritiva, responsável por mostrar como os dados estão no presente, descrevendo padrões, semelhanças e razões através dos dados, a preditiva a qual busca prever o futuro com base nos dados coletados no passado, indicando eventos ou estimando resultados futuros desconhecidos e a prescritiva indicando as melhores decisões que podem ser executadas filtrando os dados e observando as restrições, conforme o cenário apresentado. O quadro 1, demonstra a relação entre os tipos de análises e técnicas utilizadas em sua execução (MINERAÇÃO..., [2022?]).

Quadro 1 – Tipos de técnicas relacionadas a análises para mineração de dados.

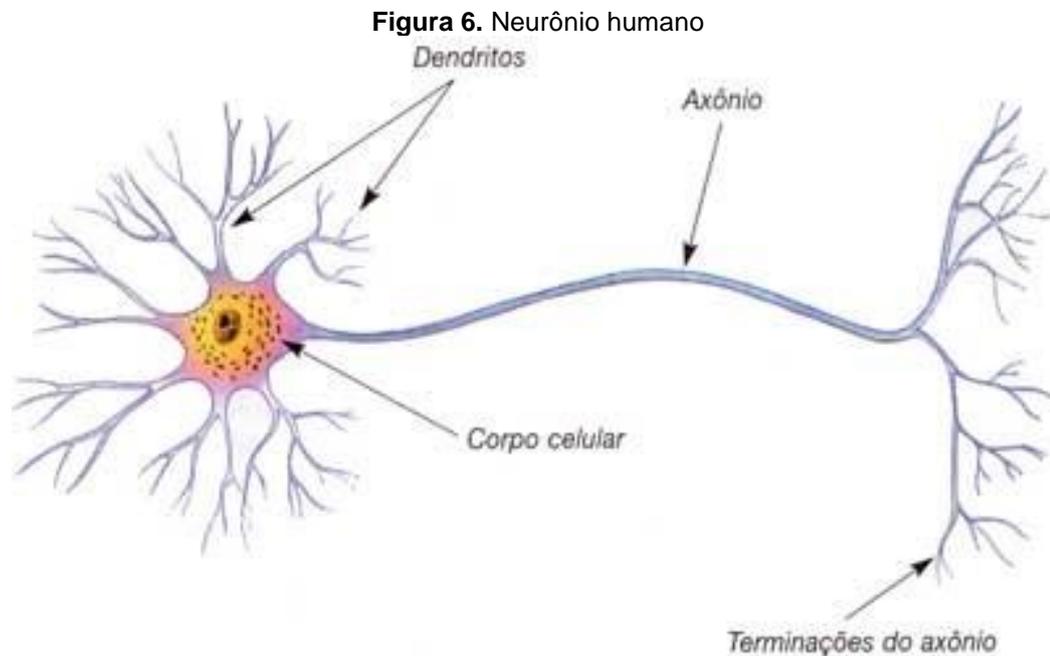
Modelagem	Técnica	Detalhes
Modelagem descritiva	Clustering	Grupos de afinidade
	Detecção de anomalias	Identifica valores discrepantes multidimensionais.
	Regras de associação	Detecta relações entre registros.
	Análise do componente principal	Detecta relações entre variáveis.
	Grupos de afinidade	Agrupar pessoas com interesses ou objetivos semelhantes
Modelagem preditiva	Regressão	Uma medida da força da relação entre uma variável dependente e uma série de variáveis independentes.
	Redes neurais	Programas de computadores que detectam padrões, fazem previsões e aprendem disso.
	Árvores de decisão	Diagramas na forma de árvores em que cada galho representa uma ocorrência provável.
	Máquinas de vetores de suporte	Modelos de aprendizagem supervisionada com seus algoritmos de aprendizagem associados.
Modelagem prescritiva	Análises preditivas e suas regras	Desenvolve regras do tipo se/então a partir de padrões e prevê resultados.
	Otimização de marketing	Simula, em tempo real, o mix de mídia mais vantajoso para alcançar o maior ROI possível.

Fonte: SAS, [2022?]

2.7.1 Técnica: Redes neurais

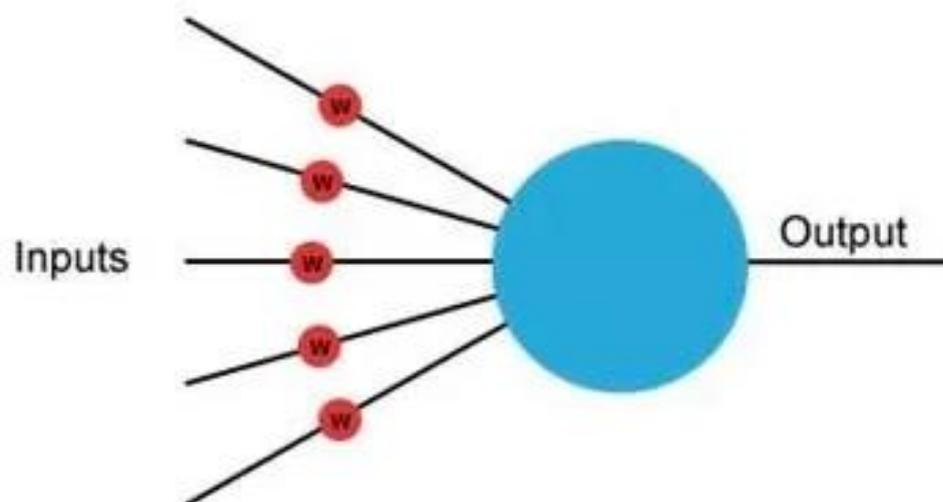
Redes neurais artificiais são modelos computacionais estudado dentro da IA que busca por meio de um modelo matemático simular a estrutura neural do cérebro

humano (figura 6 e 7), gerando um campo de neurônios interligados criando se um sistema capaz de aprender, aprimorar e tomar decisões inteligentes com ajuda humana reduzida (HAYKIN, 2001).



Fonte: GOMES, 2014;

Figura 7. Unidade de rede neural



Fonte: GOMES, 2014;

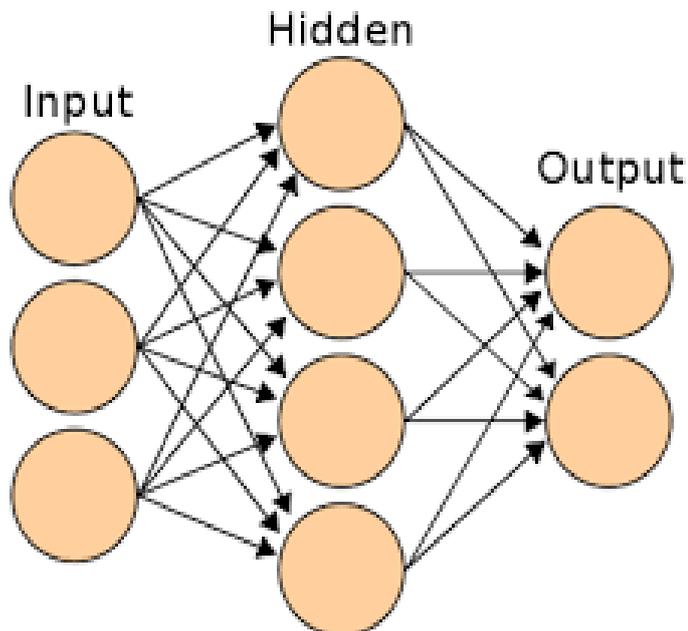
A RNA em sua composição por se tratar da tentativa de simular o conceito de processamento de informação como o cérebro, se difere do conceito convencional de

um computador digital, pois ele é altamente complexo, não-linear e paralelo (HAYKIN, 2001). Por se tratar de um sistema complexo de diversas definições, diversos especialistas da área como Braga, Carvalho e Ludemir (2000, p. 1) a define como:

Sistemas paralelos distribuídos composto por unidades de processamento simples que calculam determinadas funções matemáticas (...) disposta em uma ou mais camadas e interligadas por um grande número de conexões (...) associadas a pesos, os quais armazenam o conhecimento representado no modelo e servem para ponderar a entrada recebida por cada neurônio da rede.

As RNA's são métodos de camadas, as quais se dividem em camadas de entrada, oculta e saída, recebendo os dados na entrada, transforma os dados durante a camada oculta e gerando um dado útil para ser apresentado na saída como mostra a figura 8 (ALVES, 2020).

Figura 8. Representação de uma rede neural



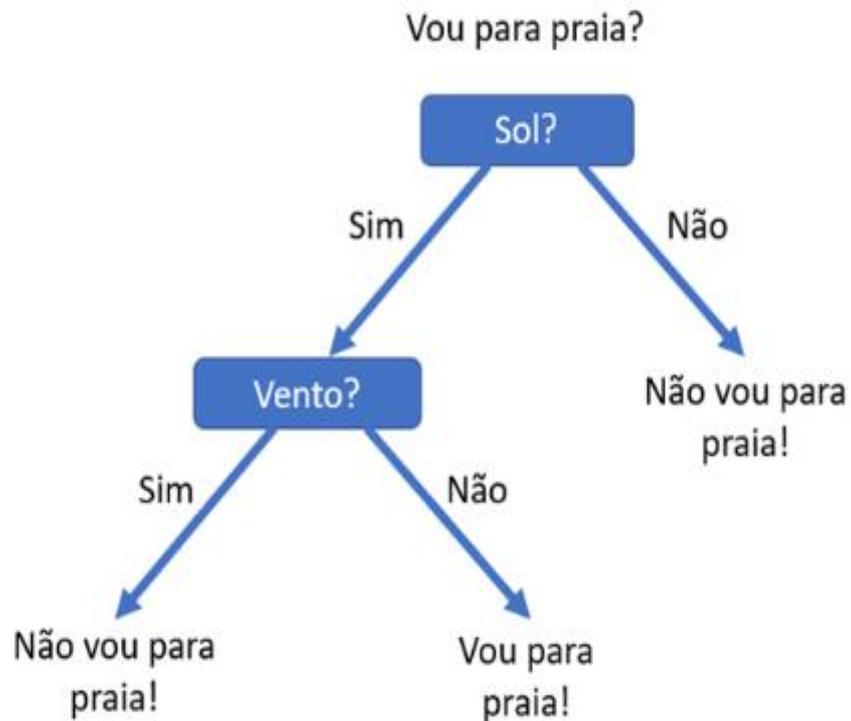
Fonte: GOMES, 2014

2.7.2 Técnica: árvore de decisão

Árvore de decisão é uma tabela de decisão sob a forma de uma árvore, e cada nó baseia sua tomada de decisão, buscando prever a melhor ação a se executar. São

algoritmos popularmente utilizados para tomadas de decisões simples, devido a suas características como facilidade de visualização visual e formação de seus nós que costumam seguir uma dinâmica básica de sim ou não, o que facilita sua tomada de decisão no fluxo como o exemplo da figura 9 (SACRAMENTO, 2021).

Figura 9. Representação de uma árvore de decisão



Fonte: COMO FUNCIONA O ALGORITMO..., [2022?];

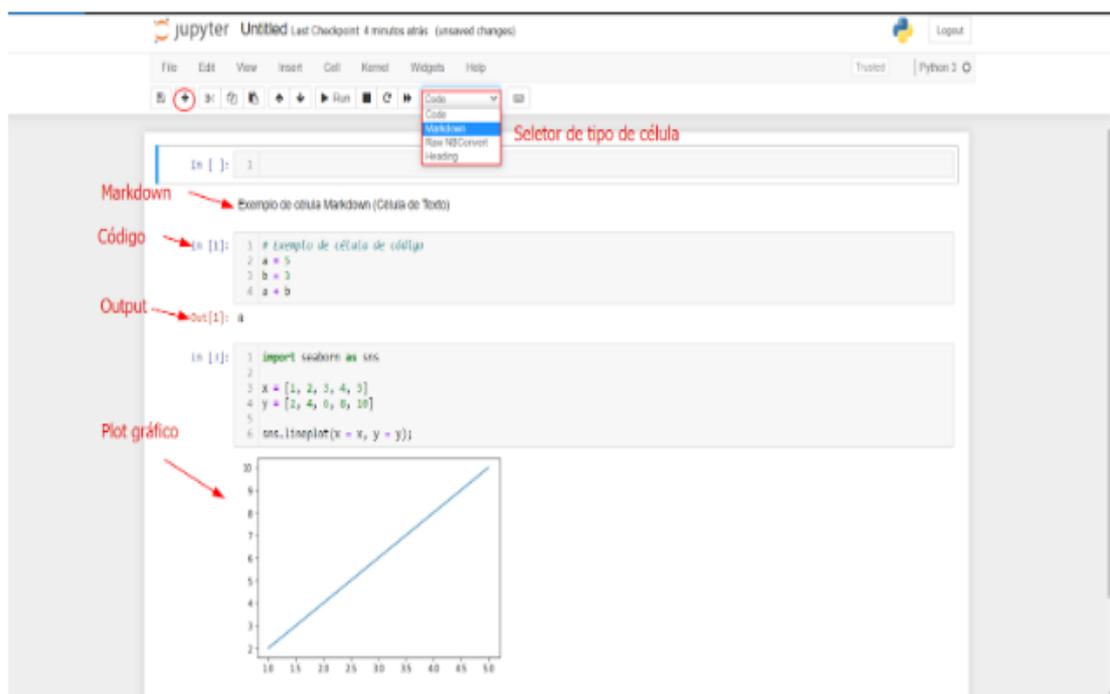
A árvore de decisão é utilizada comumente para classificação ou regressão, sendo um algoritmo de aprendizado supervisionado. Segundo Campos (2017), a popularidade das árvores de decisões é a fácil interpretação e explicabilidade comparada a outros algoritmos, o pouco esforço na preparação dos dados, a complexidade logarítmica e a capacidade de lidar com problemas com múltiplos rótulos.

3. Materiais e Métodos

Para a realização da pesquisa exploratória sobre as previsões das classificações para a *Champions League* através do campeonato inglês foram utilizadas ferramentas e códigos da linguagem Python para a coleta de dados, devido a sua capacidade de processamento e pacotes prontos que auxiliam no processamento e coleta.

Para os desenvolvimentos dos códigos em python, foi utilizado a plataforma Anaconda e seu produto conhecido como Notebook Jupyter (figura 10), um ambiente web que oferece uma abordagem de notebook permitindo criar blocos de textos mesclado ao desenvolvimento python com interpretador, permitindo assim observar as conclusões e soluções daquele devido arquivo desenvolvido como uma IDE com compilador.

Figura 10. Interface do software Jupyter Notebook



Fonte: MIRANDA, 2022;

Os códigos foram divididos em duas etapas, a primeira utilizado para a coleta é uma adaptação do código fonte do pacote ScraperFC do usuário do Github

oseymour, tais que as alterações correspondem a atualizações e implementações no código para a utilização nesta pesquisa. Na segunda etapa foi criado um código para manipular os dados coletados, selecionando, processando e o transformando para que seja utilizado conforme a requisição dos experimentos.

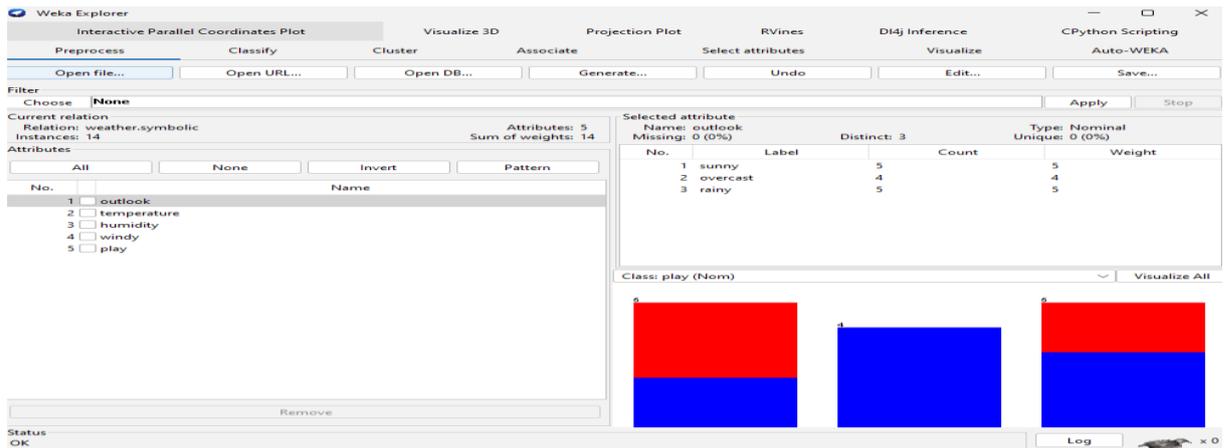
Desenvolvida na universidade de Waikato, a ferramenta *open source* utilizada é o *Waikato Environment for Knowledge Analysis* (Weka) (figura 11). Escrita na linguagem java, ela disponibiliza uma gama de algoritmos para DM.



Fonte: Autoria própria;

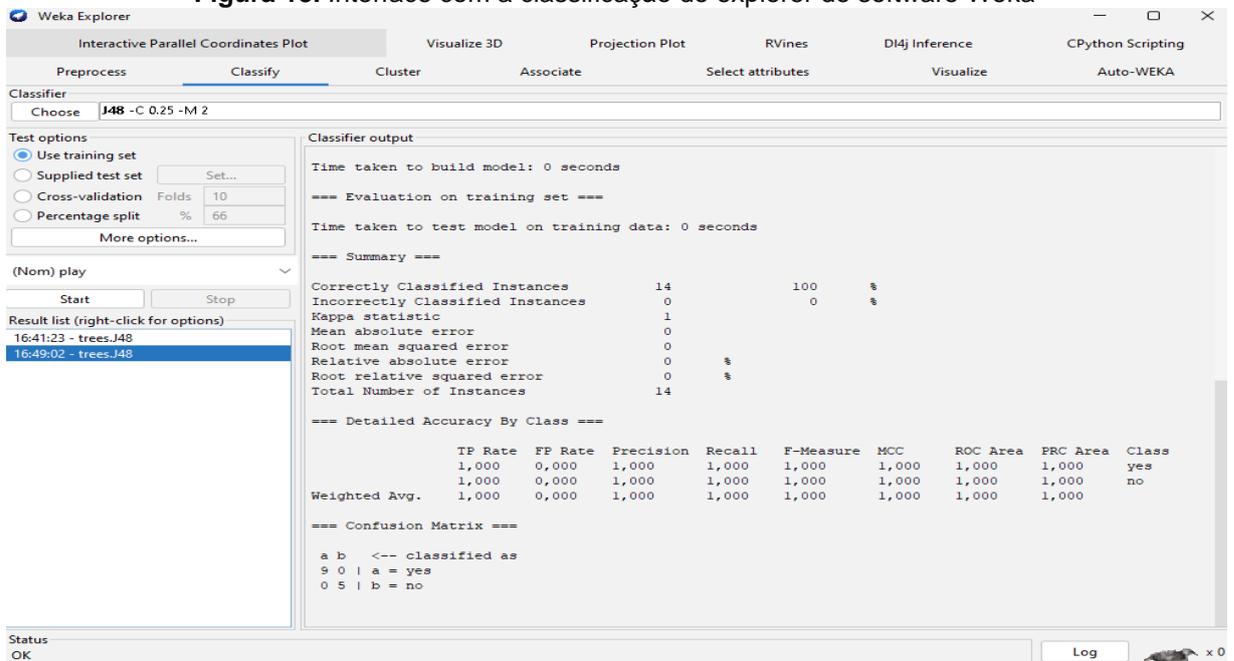
A interface inicial oferece 5 formas de aplicar a mineração de dados, cada uma com sua característica e maneira de usar. A opção utilizada para esse trabalho é a Explorer, interface que proporciona ao usuário um ambiente gráfico para a manipulação de dados e utilização de inúmeros algoritmos de forma interativa. Após instanciar os dados é possível verificar a quantidade de amostras e os atributos que serão analisados e classificados, possibilitando as escolhas do tipo de algoritmo e método de treinamento, gerando resultados dessas classificações como exemplos das figuras 12 e 13.

Figura 12. Interface inicial do explorer do software Weka



Fonte: Autoria própria;

Figura 13. Interface com a classificação do explorer do software Weka



Fonte: Autoria própria;

Os algoritmos utilizados para o experimento são baseados em árvore de decisão e RNA, popularmente utilizadas para prever resposta como sim ou não. Observando os algoritmos da árvore de decisão, utiliza-se o algoritmo J48, onde cada nó da árvore é definido após análise na base dados indicado, a qual, gera um caminho até a chegada do nó-folha e este é considerado a solução. Na RNA é utilizado o MultilayerPerceptron (MLP), o qual utiliza-se de camadas ocultas de neurônio interligadas com pesos, gerando resultados na camada de saída após seu treinamento nas camadas intermediárias.

4. EXPERIMENTOS E RESULTADOS

Este capítulo visa descrever as etapas executadas durante o experimento do trabalho proposto. O experimento inicia-se com coleta de dados de site de estatísticas e histórico de jogos de futebol, passa por uma etapa de tratamento e manipulação e depois é aplicado a mineração de dados para resolução do objetivo proposto. A proposta visa verificar a capacidade de um time entrar na zona de classificação para *champions league* observando os padrões dos primeiros 19 jogos.

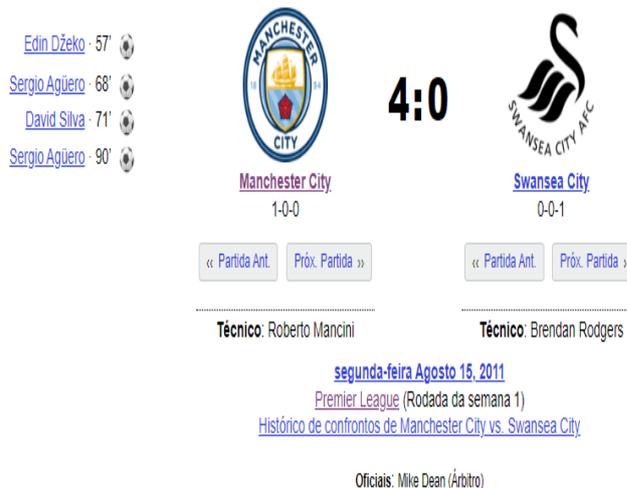
4.1 Coleta de dados

A coleta de dados é realizada através de um processo conhecido como *web scraping*, sendo extraídas 3800 partidas da *premier league* entre as temporadas de 2011-2012 a 2020-2021 para um arquivo *Comma-Separated Values (CSV)*, através do código adaptado supracitado no capítulo 3, o qual extraiu as informações das partidas do site, como mostra na figura 14 e 15.

Figura 14. Informações constante no site fbref.com

Manchester City vs. Swansea City Relatório da Partida – segunda-feira Agosto 15, 2011

[Premier League](#) (Rodada da semana 1)



Fonte: Autoria própria;

Figura 15. Modelo dos dados após extração do site fbref.com

Out[13]:

	Date	Matchweek	Home Team	Away Team	Home Goals	Away Goals	Home Ast	Away Ast	FBRef Home xG	FBRef Away xG	FBRef Home npxG	FBRef Away npxG	FBRef Home xA	FBRef Away xA	FBRef Home psxG	FBRef Away psxG	Away Formation	Home Formation
0	2011-09-18	5	Fulham	Manchester City	2	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	2012-01-15	21	Newcastle Utd	QPR	1	0	NaN	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	2012-01-21	22	Stoke City	West Brom	1	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	2011-10-22	9	Bolton	Sunderland	0	2	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	2011-08-27	3	Aston Villa	Wolves	0	0	0.0	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5	2011-12-20	17	Blackburn	Bolton	1	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	2011		Newcastle															

Fonte: Autoria própria;

Os dados extraídos na figura 15 derivam-se do código (figura 16 a 23) ScrapperFC modificado, utilizando pacotes do python como selenium, pandas e outros para a coleta esperada.

Figura 16. Parte 1 – extração dos dados

```
import datetime
from IPython.display import clear_output
import numpy as np
import pandas as pd
pd.set_option('display.max_columns', None)
from ScrapperFC.shared.functions import check_season
from selenium import webdriver
from selenium.webdriver.chrome.options import Options
from selenium.webdriver.common.by import By
from webdriver_manager.chrome import ChromeDriverManager
from urllib.request import urlopen
from urllib3.connection import HTTPConnection
import urllib3, socket
HTTPConnection.default_socket_options = (
    HTTPConnection.default_socket_options + [
        (socket.SOL_SOCKET, socket.SO_SNDBUF, 15000000), #1MB in byte
        (socket.SOL_SOCKET, socket.SO_RCVBUF, 15000000)
    ])

import requests, random
from requests.packages.urllib3.exceptions import InsecureRequestWarning
from bs4 import BeautifulSoup
from lxml import etree
from tqdm import tqdm
import time
```

Fonte: Autoria própria;

Figura 17. Parte 2 – extração dos dados

```
def scrape_matches(self, year, League, save=False):
    err, valid = check_season(year, League, 'FBRef')
    if not valid:
        print(err)
        return -1
    season = str(year-1)+'-'+str(year)
    failures = []
    datayears = [year]
    # initialize df
    cols = ['Date', 'Matchweek', 'Home Team', 'Away Team', 'Home Goals', 'Away Goals',
            'Home Ast', 'Away Ast', 'FBRef Home xG', 'FBRef Away xG', 'FBRef Home npxG',
            'FBRef Away npxG', 'FBRef Home xA', 'FBRef Away xA', 'FBRef Home psxG',
            'FBRef Away psxG']
    matches = pd.DataFrame(columns=cols)
```

Fonte: Autoria própria;

Figura 18. Parte 2 – extração dos dados

```
for years in datayears:
    season = str(years-1)+'-'+str(years)
    links = self.get_match_links(years, League)
    print("Scraping matches: %d." % years)
    for link in tqdm(links):
        flag = True
        interator=0
        while(flag):
            try:
                match = self.scrape_match(link, years, League)
                matches = matches.append(match, ignore_index=True)
                flag = False
            except Exception as e:
                failures.append([link, e])
                print('testando')
                for failure in failures:
                    print(failure, '\n')
                interator= interator+1
        if save:
            filename = '{}_{}_FBRef_matches.csv'.format(season, League.replace(' ', '_'))
            matches.to_csv(path_or_buf=filename, encoding='utf-16', index=False)
            filename2 = '{}_{}_FBRef_players.csv'.format(season, League.replace(' ', '_'))
            players.to_csv(path_or_buf=filename2, encoding='utf-16', index=False)
            print('Matches dataframe saved to ' + filename)
    time.sleep(20)
    return filename, filename2
```

Fonte: Autoria própria;

Figura 19. Parte 3 – extração dos dados

```

853     err, valid = check_season(year, League, 'FBRef')
854     if not valid:
855         print(err)
856         return -1
857     df = pd.read_html(Link)
858

```

Fonte: Autoria própria;

Figura 20. Parte 4 – extração dos dados

```

if League == 'EPL':
    if year >= 2008:
        spliton = '-Premier-League'
    else:
        spliton = '-Premiership'
elif League == 'La Liga':
    spliton = '-La-Liga'
elif League == 'Bundesliga':
    spliton = '-Bundesliga'
elif League == 'Serie A':
    spliton = '-Serie-A'
elif League == 'Ligue 1':
    if year >= 2003:
        spliton = '-Ligue-1'
    else:
        spliton = '-Division-1'

# Get date of the match
try:
    # Try this first. Assumes league name is one word
    date_elements = Link.split("/")[-1].split("-")[-4:-1]
    date = '-'.join(date_elements)
    date = datetime.datetime.strptime(date, '%B-%d-%Y').date()
except:
    # Assumes league name is two words
    date_elements = Link.split("/")[-1].split("-")[-5:-2]
    date = '-'.join(date_elements)
    date = datetime.datetime.strptime(date, '%B-%d-%Y').date()

```

Fonte: Autoria própria;

Figura 21. Parte 5 – extração dos dados

```

requests.packages.urllib3.disable_warnings(InsecureRequestWarning)
response = requests.get(Link,time.sleep(3), headers=headers, verify=False, timeout= 120)
soup = BeautifulSoup(response.content, "html.parser")
dom = etree.HTML(str(soup))
i=0
time.sleep(3)
while( ( response.status_code != 200 ) & ( i < 15 ) ):
    i=i+1
    response = requests.get(Link,time.sleep(3), headers=headers, verify=False, timeout= 120)
    soup = BeautifulSoup(response.content, "html.parser")
    dom = etree.HTML(str(soup))
    time.sleep(15)
i=0

matchweek = int(
    dom.xpath('//*[@id="content"]/div[2]/div[3]/div[2]/text())[0]\
    .split('Matchweek')[-1]\
    .replace(' ','')\
    .strip()
)

```

Fonte: Autoria própria;

Figura 22. Parte 6 – extração dos dados

```

948 match = pd.Series()
949 match['Date'] = str(date)
950 match['Matchweek'] = matchweek
951 if year >= 2016:
952     match['Home Team'] = df[2].columns[0][0]
953     match['Away Team'] = df[2].columns[1][0]
954     match['Home Formation'] = df[0].columns[0].split("(")[-1].split(")")[0]
955     match['Away Formation'] = df[1].columns[0].split("(")[-1].split(")")[0]
956     if year >= 2018:
957         match['Home Goals'] = np.array(df[3][('Performance', 'Gls')][-1] + np.array(df[15][('Performance', 'OG')][-1]
958         match['Away Goals'] = np.array(df[10][('Performance', 'Gls')][-1] + np.array(df[8][('Performance', 'OG')][-1]
959         match['Home Ast'] = np.array(df[3][('Performance', 'Ast')][-1]
960         match['Away Ast'] = np.array(df[10][('Performance', 'Ast')][-1]
961         match['FBRef Home xG'] = np.array(df[3][('Expected', 'xG')][-1]
962         match['FBRef Away xG'] = np.array(df[10][('Expected', 'xG')][-1]
963         match['FBRef Home npxG'] = np.array(df[3][('Expected', 'npxG')][-1]
964         match['FBRef Away npxG'] = np.array(df[10][('Expected', 'npxG')][-1]
965         match['FBRef Home xA'] = np.array(df[3][('Expected', 'xA')][-1]
966         match['FBRef Away xA'] = np.array(df[10][('Expected', 'xA')][-1]
967         match['FBRef Home psxG'] = np.array(df[16][('Shot Stopping', 'PSxG')][-1]
968         match['FBRef Away psxG'] = np.array(df[9][('Shot Stopping', 'PSxG')][-1]
969     else:
970         match['Home Goals'] = np.array(df[3][('Performance', 'Gls')][-1] + np.array(df[5][('Performance', 'OG')][-1]
971         match['Away Goals'] = np.array(df[5][('Performance', 'Gls')][-1] + np.array(df[3][('Performance', 'OG')][-1]
972         match['Home Ast'] = np.array(df[3][('Performance', 'Ast')][-1]
973         match['Away Ast'] = np.array(df[5][('Performance', 'Ast')][-1]
974         match['FBRef Home xG'] = np.nan
975         match['FBRef Away xG'] = np.nan
976         match['FBRef Home npxG'] = np.nan
977         match['FBRef Away npxG'] = np.nan
978         match['FBRef Home xA'] = np.nan
979         match['FBRef Away xA'] = np.nan
980         match['FBRef Home psxG'] = np.nan
981         match['FBRef Away psxG'] = np.nan

```

Fonte: Autoria própria;

Figura 23. Parte 7 – extração dos dados

```

982     else:
983         match['Home Team'] = df[0].columns[0][0]
984         match['Away Team'] = df[0].columns[1][0]
985         match["Home Formation"] = np.nan
986         match["Away Formation"] = np.nan
987     try:
988         match['Home Goals'] = np.array(df[1][('Performance', 'Gls')])[-1] + np.array(df[3][('Performance', 'OG')])[-1]
989         match['Away Goals'] = np.array(df[2][('Performance', 'Gls')])[-1] + np.array(df[1][('Performance', 'OG')])[-1]
990         match['Home Ast'] = np.array(df[1][('Performance', 'Ast')])[-1]
991         match['Away Ast'] = np.array(df[3][('Performance', 'Ast')])[-1]
992     except:
993         match['Home Goals'] = np.array(df[1][('Performance', 'Gls')])[-1] + np.array(df[2][('Performance', 'OG')])[-1]
994         match['Away Goals'] = np.array(df[2][('Performance', 'Gls')])[-1] + np.array(df[1][('Performance', 'OG')])[-1]
995         match['Home Ast'] = np.array(df[1][('Performance', 'Ast')])[-1]
996         match['Away Ast'] = np.array(df[2][('Performance', 'Ast')])[-1]
997
998     match['FBRef Home xG'] = np.nan
999     match['FBRef Away xG'] = np.nan
1000     match['FBRef Home npxG'] = np.nan
1001     match['FBRef Away npxG'] = np.nan
1002     match['FBRef Home xA'] = np.nan
1003     match['FBRef Away xA'] = np.nan
1004     match['FBRef Home psxG'] = np.nan
1005     match['FBRef Away psxG'] = np.nan
1006     return match
    
```

Fonte: Autoria própria;

4.2 Tratamento dos dados

Observando a fração dos dados (figura 15) e como os experimentos trabalharam, é necessário tratar e modelar esses dados para que se enquadre no modelo adequado de aplicabilidade. Para tal alinhamento é necessário tratar os dados em brancos, transformar as informações obtidas a respeito das partidas, como transformar os gols das partidas em estatísticas de vitória, empate ou derrota, definir as rodadas para a respectiva linha correspondente e verificar se o time entrou na zona de classificação da *champions league* no respectivo ano, entre outras alterações, como mostra na figura 24 e 25, onde os dados da figura 15, foram transformados para o objetivo do experimento, delimitando 20 atributos especificados na tabela 1.

Figura 24. Modelo dos dados após 1ª etapa do tratamento

	Team	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	Ano	SN
0	Arsenal	E	D	D	V	D	V	D	V	V	V	V	V	E	V	V	D	V	E	V	D	D	D	E	V	V	V	V	V	V	D	V	V	D	E	E	V	2011	S		
1	Arsenal	E	E	V	V	E	D	V	D	V	D	E	V	E	E	D	V	V	V	V	V	E	D	D	E	V	V	V	D	E	V	V	V	V	E	V	V	V	2012	S	
2	Arsenal	D	V	V	V	V	E	V	V	V	D	V	V	V	E	D	E	V	V	V	V	V	E	V	D	E	V	D	E	V	D	E	D	V	V	V	V	V	2013	S	
3	Arsenal	V	E	E	E	V	E	D	E	V	V	D	D	V	V	D	V	E	V	V	D	V	V	V	D	V	V	V	V	V	V	V	E	E	V	D	E	V	2014	S	
4	Arsenal	D	V	E	V	V	D	V	V	V	V	E	D	E	V	V	V	D	V	V	E	E	D	E	V	V	D	D	E	V	V	V	E	E	E	V	E	V	2015	S	
5	Arsenal	D	E	V	V	V	V	V	E	V	E	E	V	V	V	D	D	V	V	E	V	V	D	D	V	V	D	V	D	E	V	D	V	V	D	V	V	2016	N		
6	Arsenal	V	D	D	V	E	V	V	D	V	V	D	V	V	V	D	E	E	V	E	V	E	E	D	V	D	V	D	D	D	V	D	V	V	D	V	D	V	2017	N	
7	Arsenal	D	D	V	V	V	V	V	V	E	E	E	V	V	E	V	D	V	E	D	V	D	V	V	D	V	V	V	E	V	D	V	D	V	D	D	E	V	2018	N	
8	Arsenal	V	V	D	E	E	V	E	V	D	E	E	D	E	E	D	V	D	E	E	D	V	E	E	E	E	V	V	D	V	D	V	V	V	E	D	V	D	V	2019	N
9	Arsenal	V	V	D	V	D	D	V	D	E	D	D	D	E	D	V	V	V	E	V	V	E	D	D	V	D	V	E	V	E	D	V	E	D	V	V	V	V	2020	N	
10	Chelsea	E	V	V	V	D	V	V	V	D	D	V	D	V	V	V	E	E	E	D	V	V	E	E	E	D	V	D	V	D	E	V	V	E	D	E	V	D	V	2011	N

Fonte: Autoria própria;

Figura 25. Modelo dos dados após 2ª etapa do tratamento

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	SN
0	E	D	D	V	D	V	D	V	V	V	V	V	E	V	V	D	V	E	V	S
1	E	E	V	V	E	D	V	D	V	D	E	V	E	E	D	V	V	V	V	S
2	D	V	V	V	V	E	V	V	V	D	V	V	V	E	D	E	V	V	S	
3	V	E	E	E	V	E	D	E	V	V	D	D	V	V	D	V	E	V	S	
4	D	V	E	V	V	D	V	V	V	V	E	D	E	V	V	V	D	V	S	
5	D	E	V	V	V	V	V	E	V	E	E	V	V	V	D	D	V	V	N	
6	V	D	D	V	E	V	V	D	V	V	D	V	V	V	D	E	E	V	N	
7	D	D	V	V	V	V	V	V	E	E	E	V	V	E	V	D	V	E	N	
8	V	V	D	E	E	V	E	V	D	E	E	D	E	E	D	V	D	E	N	
9	V	V	D	V	D	D	V	D	E	D	D	D	E	D	V	V	V	E	N	
10	E	V	V	V	D	V	V	V	D	D	V	D	V	V	V	E	E	E	N	

Fonte: Autoria própria;

Tabela 1 – lista de atributos

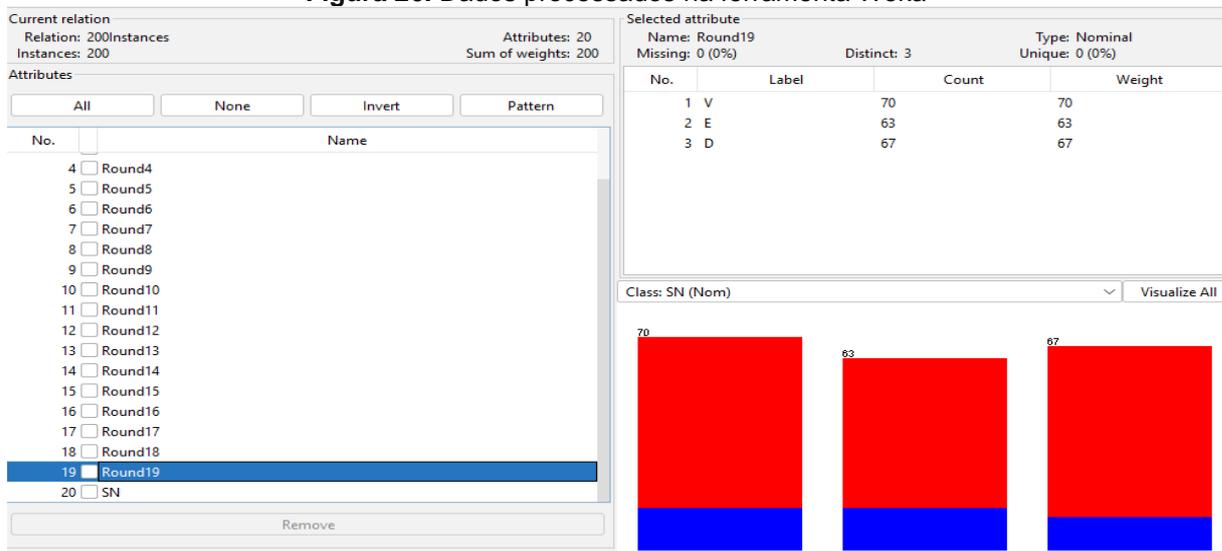
ATRIBUTO	TIPO DE DADO
PARTIDA 1	V/E/D
PARTIDA 2	V/E/D
PARTIDA 3	V/E/D
PARTIDA 4	V/E/D
PARTIDA 5	V/E/D
PARTIDA 6	V/E/D
PARTIDA 7	V/E/D
PARTIDA 8	V/E/D
PARTIDA 9	V/E/D
PARTIDA 10	V/E/D
PARTIDA 11	V/E/D
PARTIDA 12	V/E/D
PARTIDA 13	V/E/D
PARTIDA 14	V/E/D
PARTIDA 15	V/E/D
PARTIDA 16	V/E/D
PARTIDA 17	V/E/D
PARTIDA 18	V/E/D
PARTIDA 19	V/E/D
CLASSIFICADO	S/N

Fonte: Autoria própria;

4.3 Aplicação dos dados

Nesta etapa ocorre a fase conhecida como mineração de dados, a qual será guiada utilizando o algoritmo J48 E MLP através do software Weka. Durante o processo foram utilizados os dados manipulados citados no capítulo 4.2, contidos em um arquivo CSV, o qual o software reconhece, compreende e separa os dados e atributos (figura 26) e instancia um arquivo *Attribute-Relation File Format* (ARFF) padrão do Weka.

Figura 26. Dados processados na ferramenta Weka



Fonte: Autoria própria;

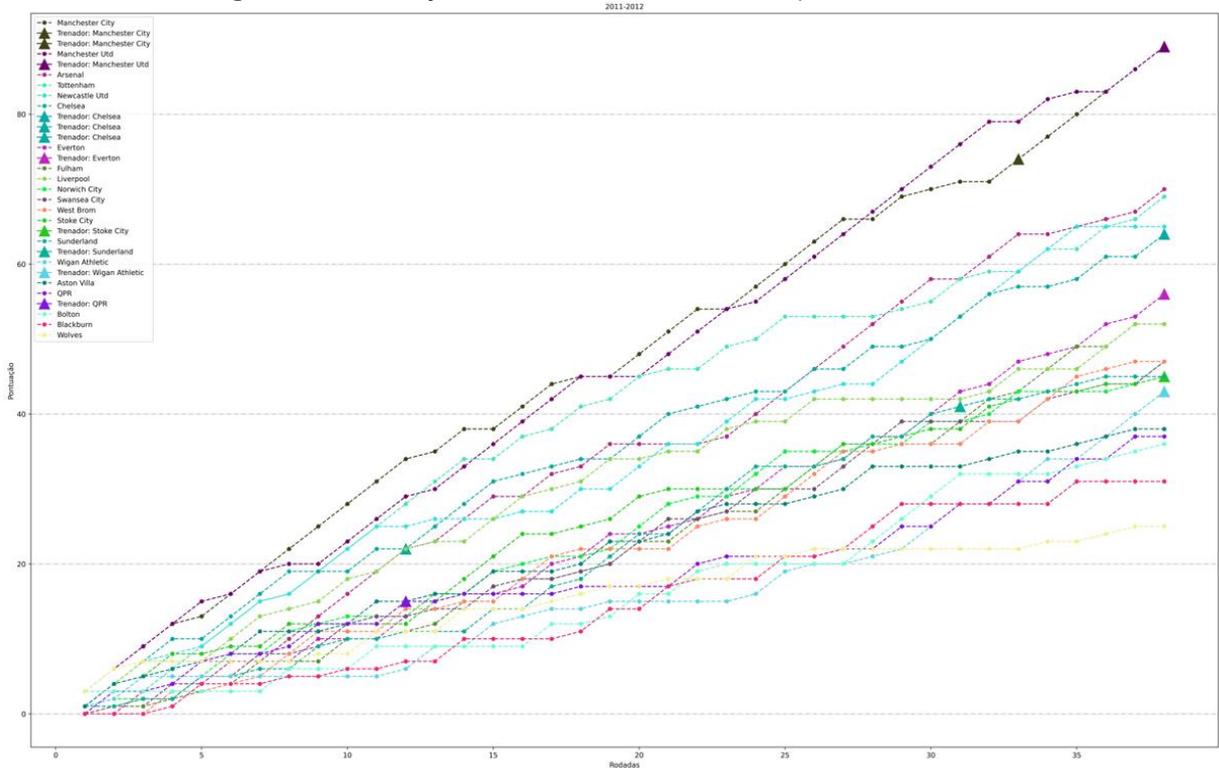
Os algoritmos J48 e MLP utilizados na MD, além de demonstrar a capacidade de aplicabilidade do experimento, fornecem informações utilizadas para comparação e estudo das classificações de sim e não, indicando suas vantagens e desvantagens.

4.4 Discussão e Resultados

Durante o processo de análise e processamento dos dados coletados, foram realizados experimentos secundários com o objetivo de analisar fatores que poderiam estar correlacionados a classificação dos times, observando comportamentos como média de pontuação, Média de gols feitos (MGF), Média de gols sofridos (MGS), Média de gols feitos em casa (MGFC), Média de gols feito fora de casa (MGFF), Média de gols sofridos em casa (MGSC), Média de gols sofridos fora de casa (MGSF), intervalos de vitórias, saltos de desempenhos e trocas de

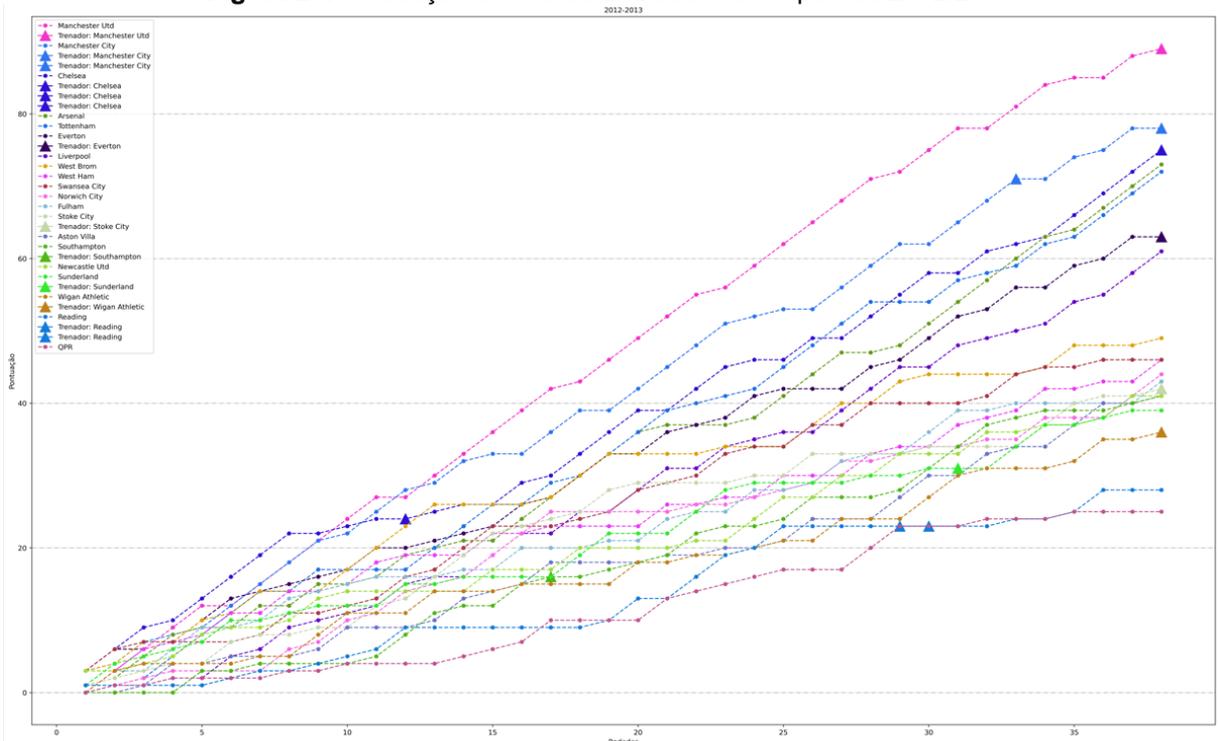
treinadores, gerando imagens a partir desses dados (figuras 17 a 37), buscando identificar visualmente possíveis padrões.

Figura 27. Pontuação e troca de treinador na temporada 2011/2012



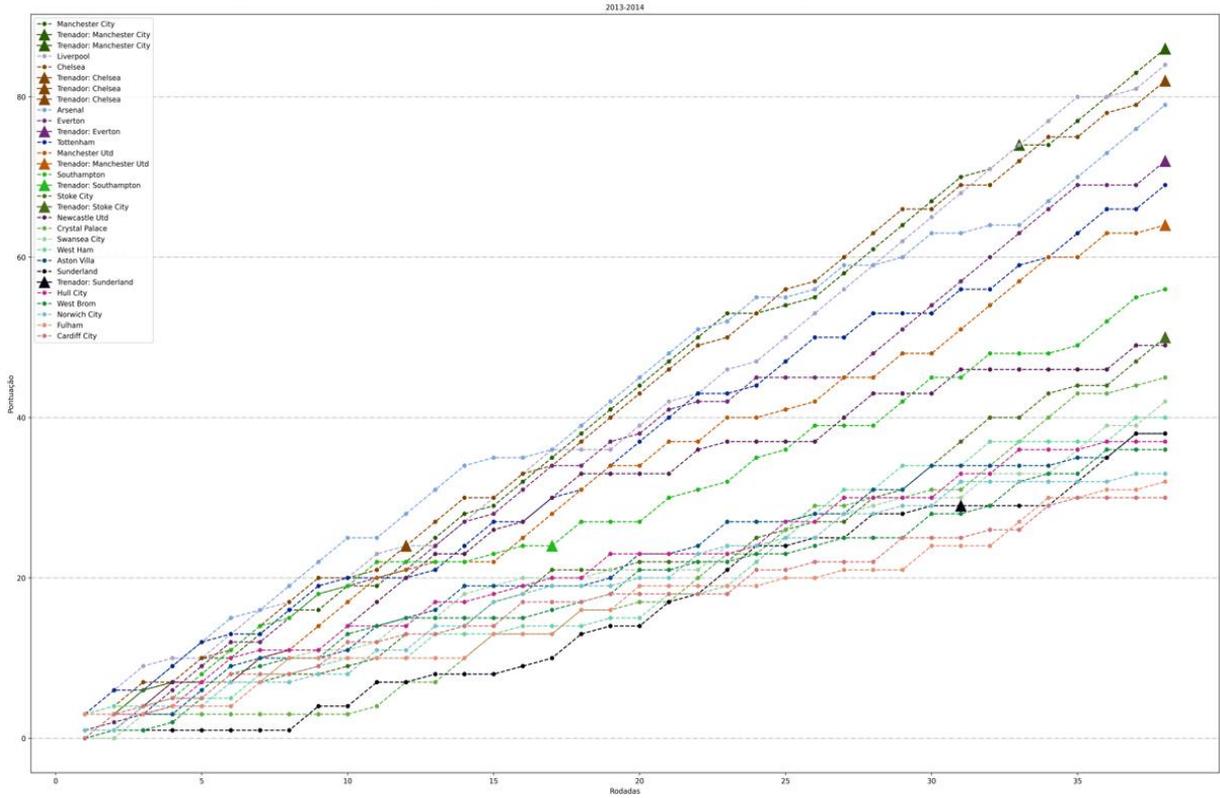
Fonte: Autoria própria;

Figura 28. Pontuação e troca de treinador na temporada 2012/2013



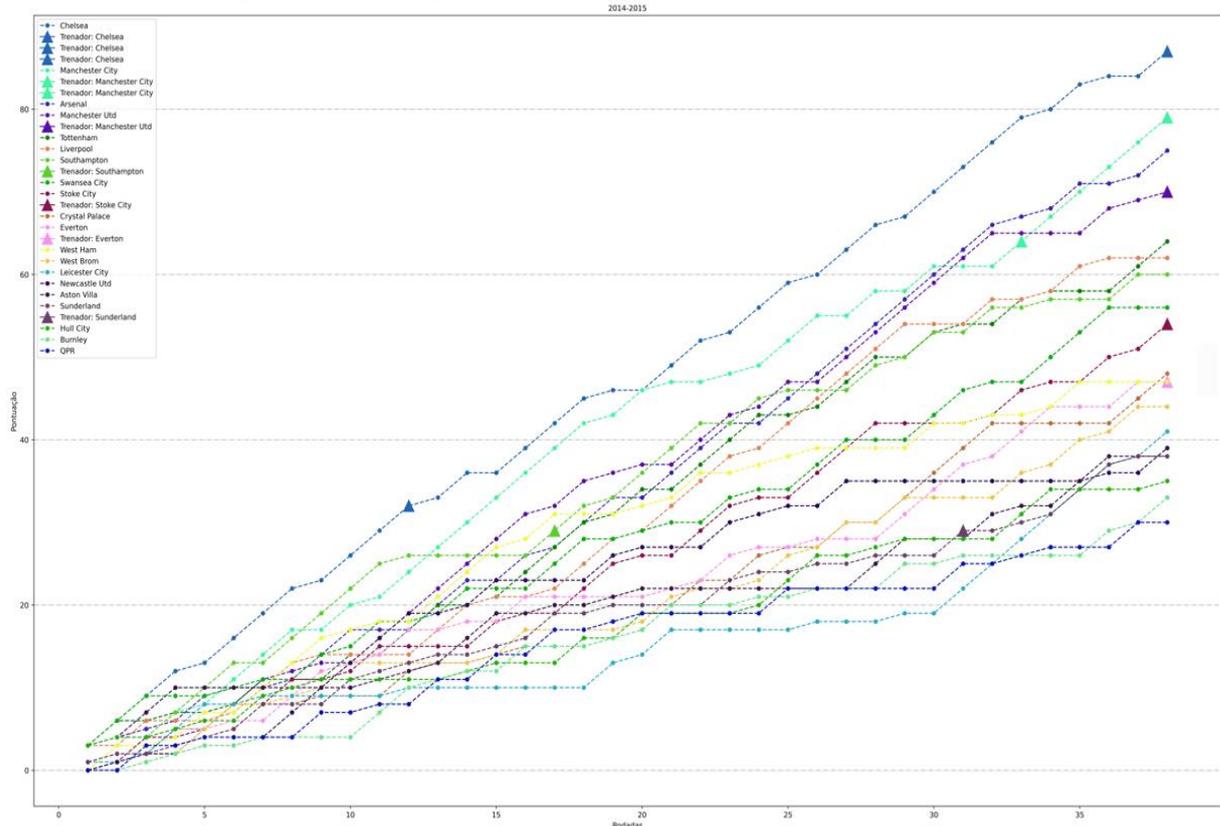
Fonte: Autoria própria;

Figura 29. Pontuação e troca de treinador na temporada 2013/2014



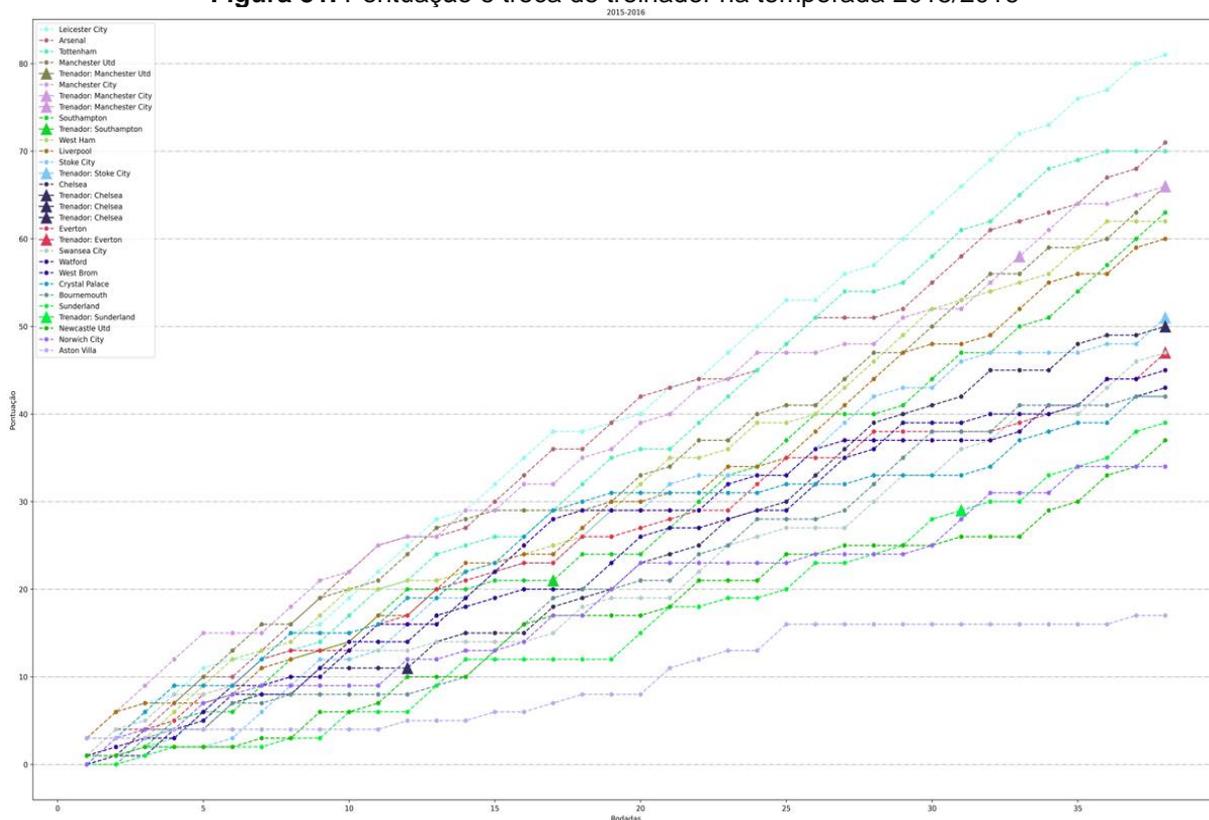
Fonte: Autoria própria;

Figura 30. Pontuação e troca de treinador na temporada 2014/2015



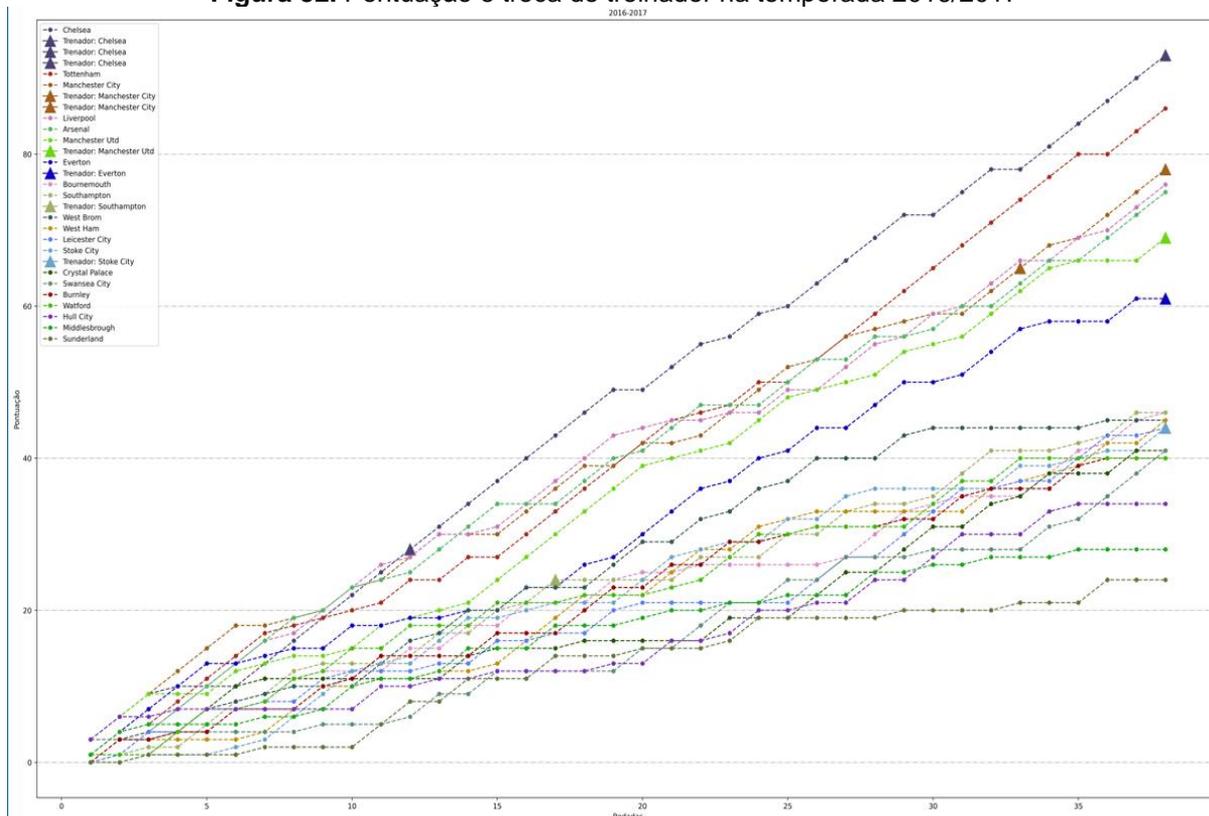
Fonte: Autoria própria;

Figura 31. Pontuação e troca de treinador na temporada 2015/2016



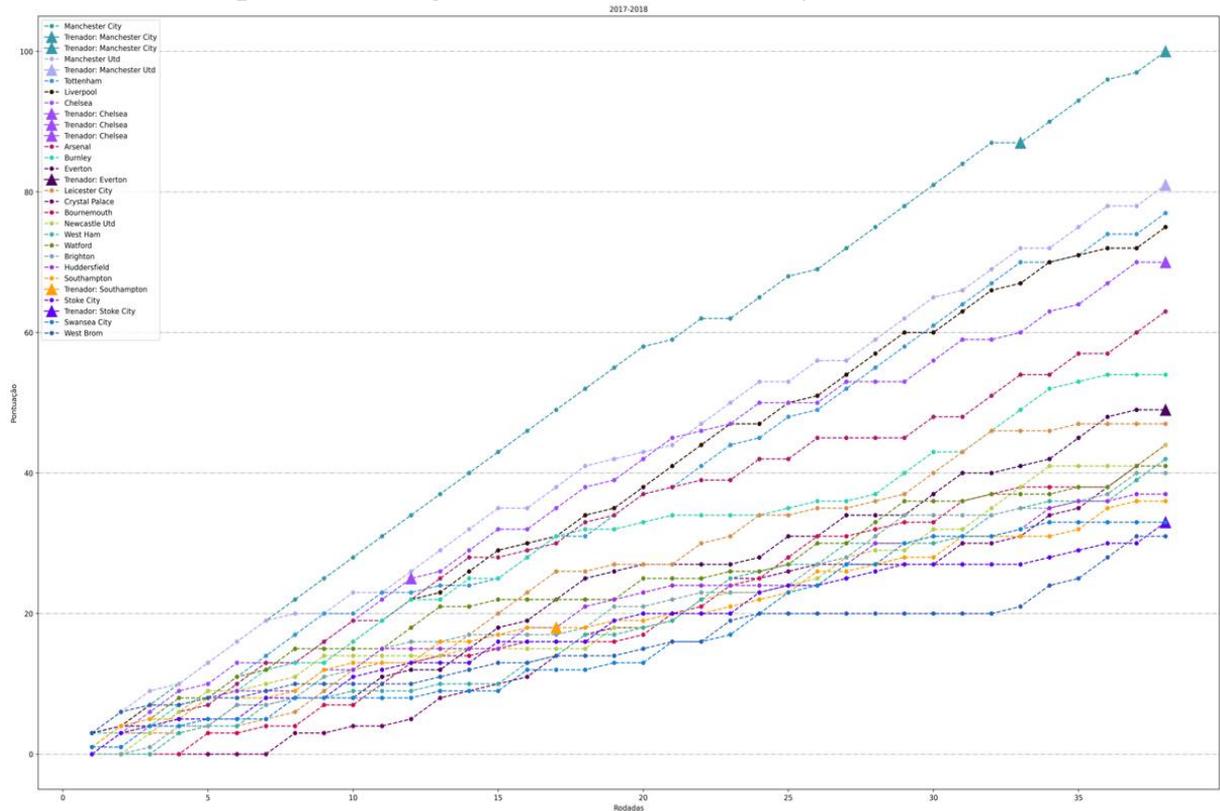
Fonte: Autoria própria;

Figura 32. Pontuação e troca de treinador na temporada 2016/2017



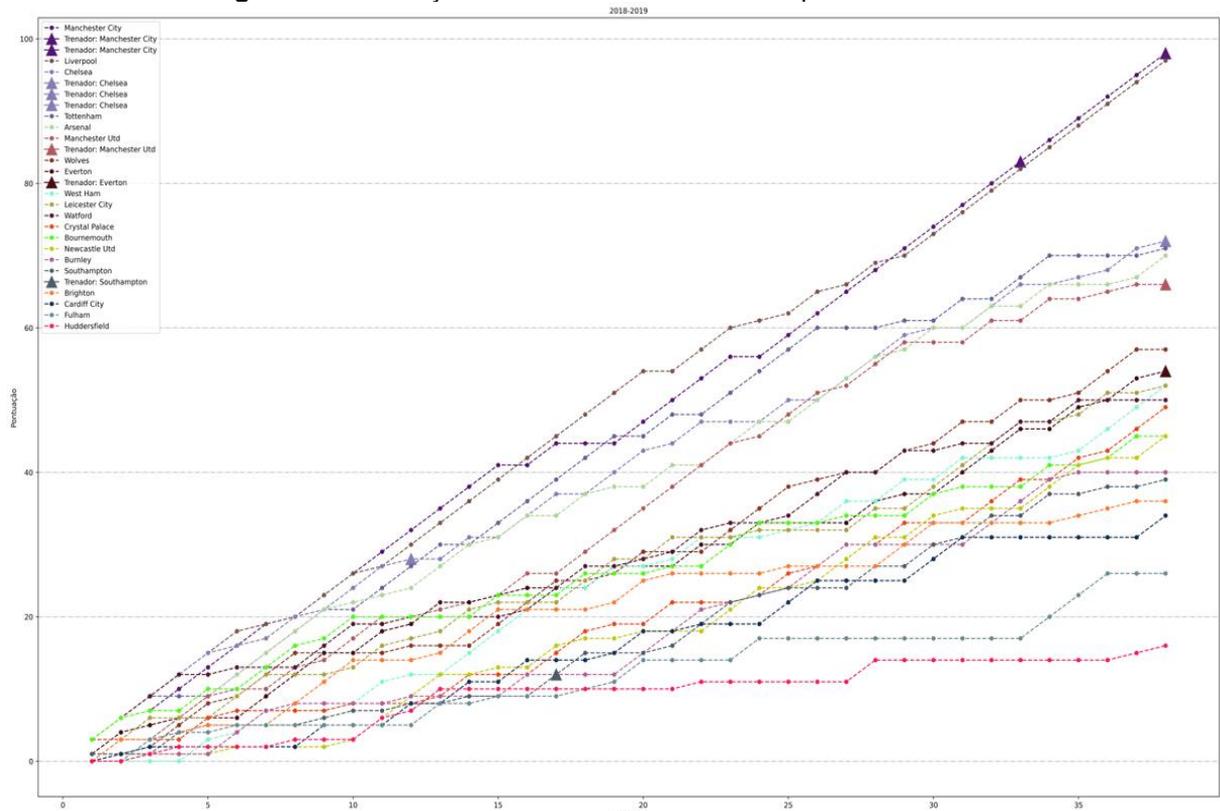
Fonte: Autoria própria;

Figura 33. Pontuação e troca de treinador na temporada 2017/2018



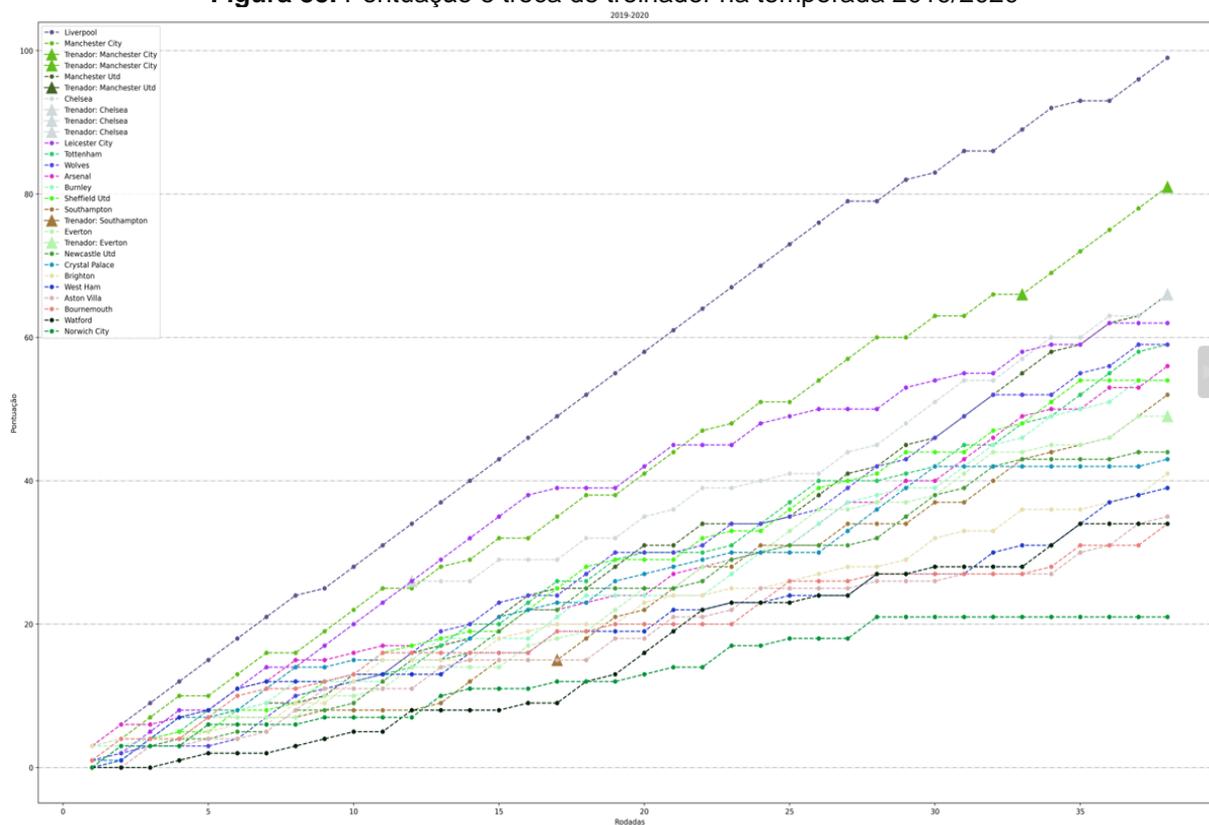
Fonte: Autoria própria;

Figura 34. Pontuação e troca de treinador na temporada 2018/2019



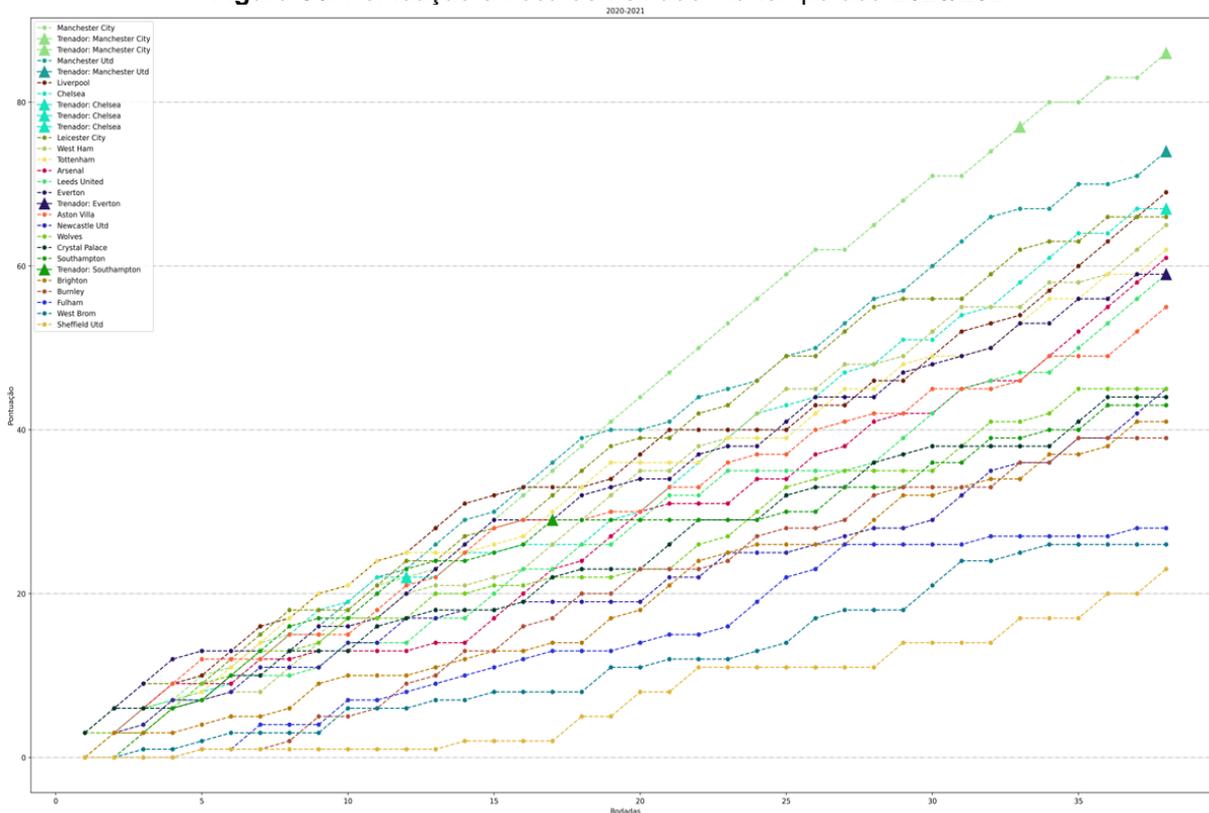
Fonte: Autoria própria;

Figura 35. Pontuação e troca de treinador na temporada 2019/2020



Fonte: Autoria própria;

Figura 36. Pontuação e troca de treinador na temporada 2020/2021



Fonte: Autoria própria;

Figura 37. Média de pontuação por rodada nas temporadas 2011 a 2021
2011 - 2021

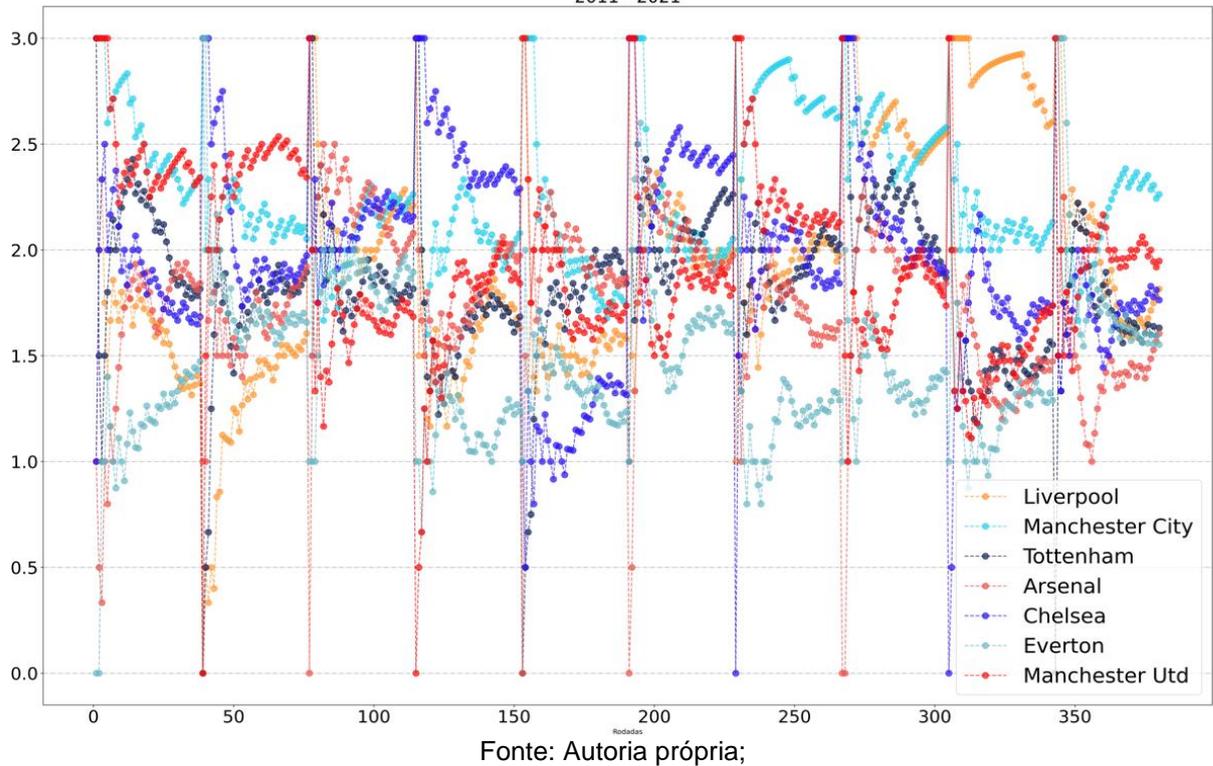


Figura 38. Média de gols feitos e sofridos na temporada 2011/2012

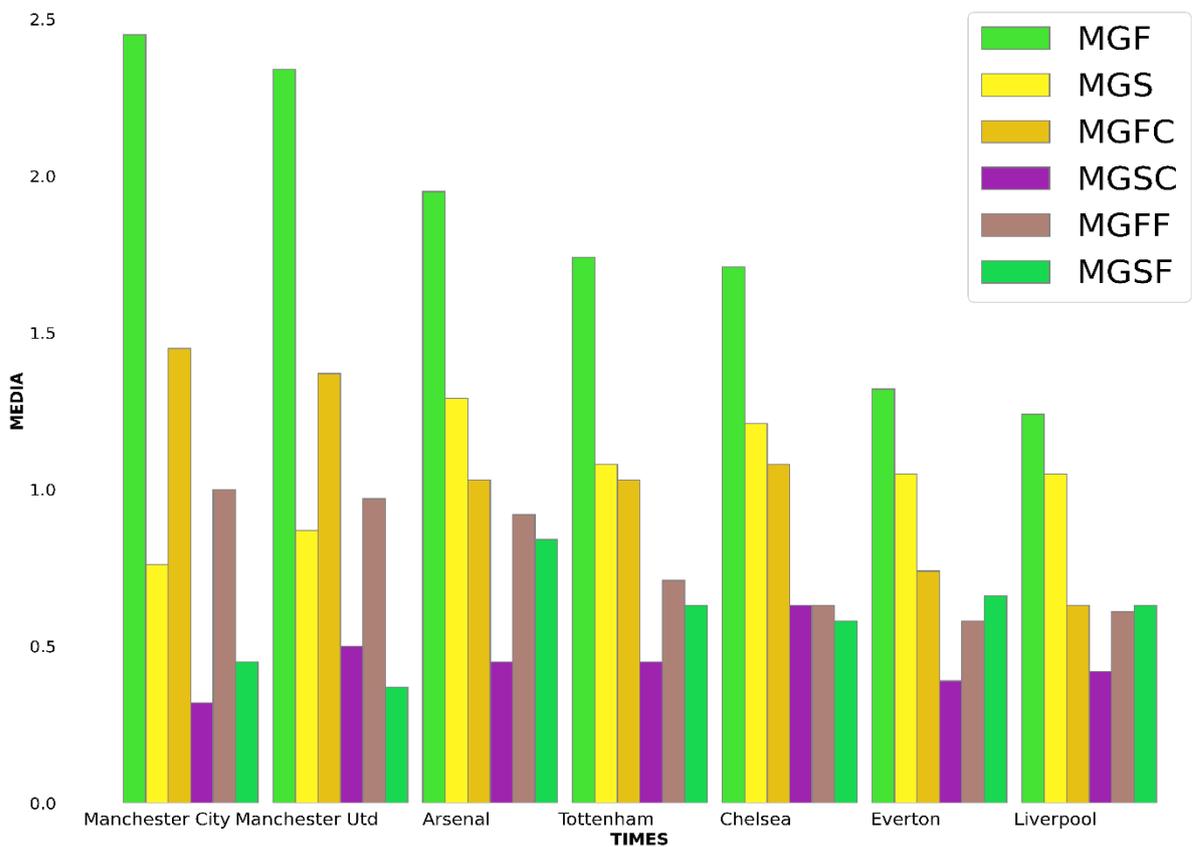
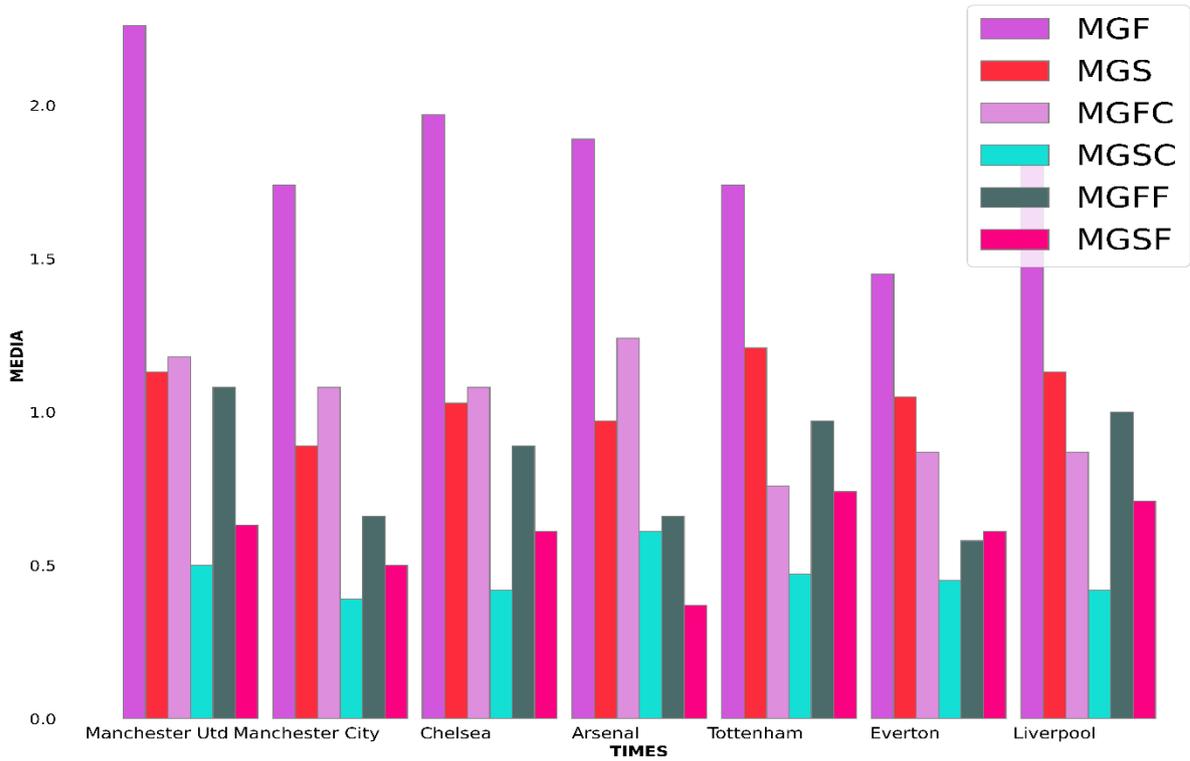
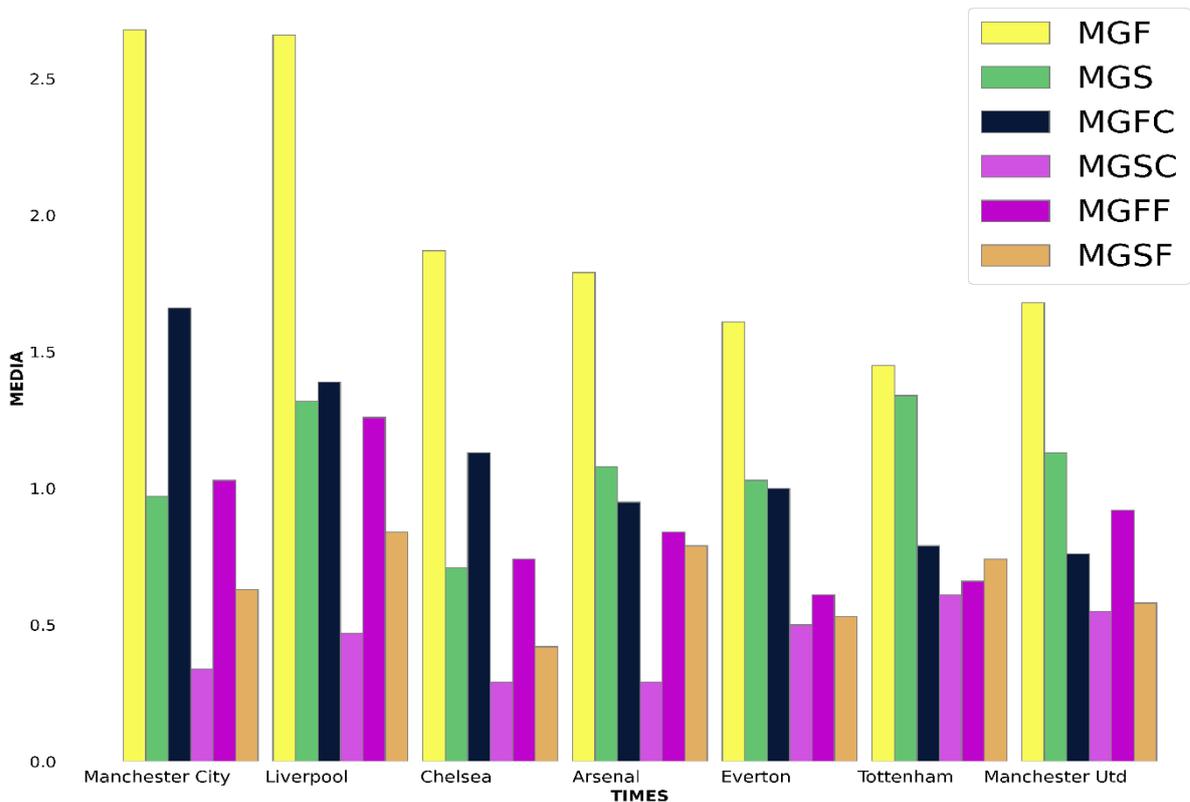


Figura 39. Média de gols feitos e sofridos na temporada 2012/2013



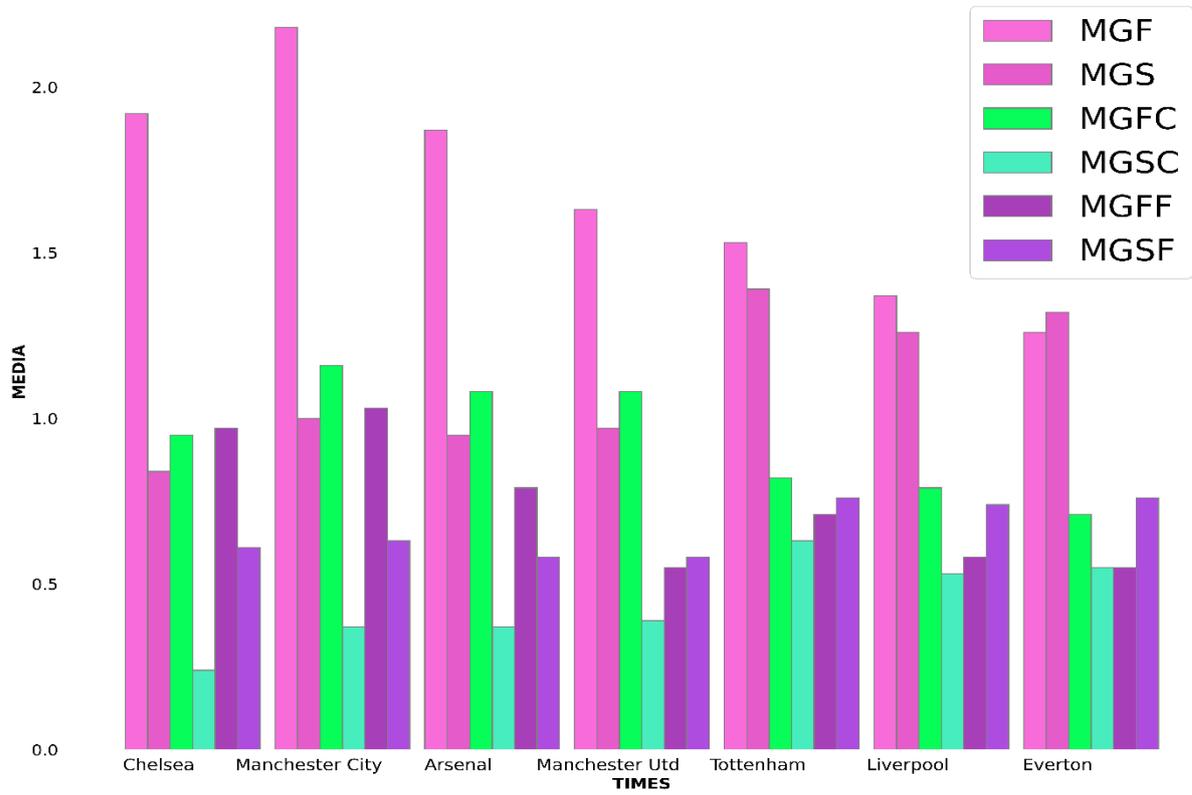
Fonte: Autoria própria;

Figura 40. Média de gols feitos e sofridos na temporada 2013/2014



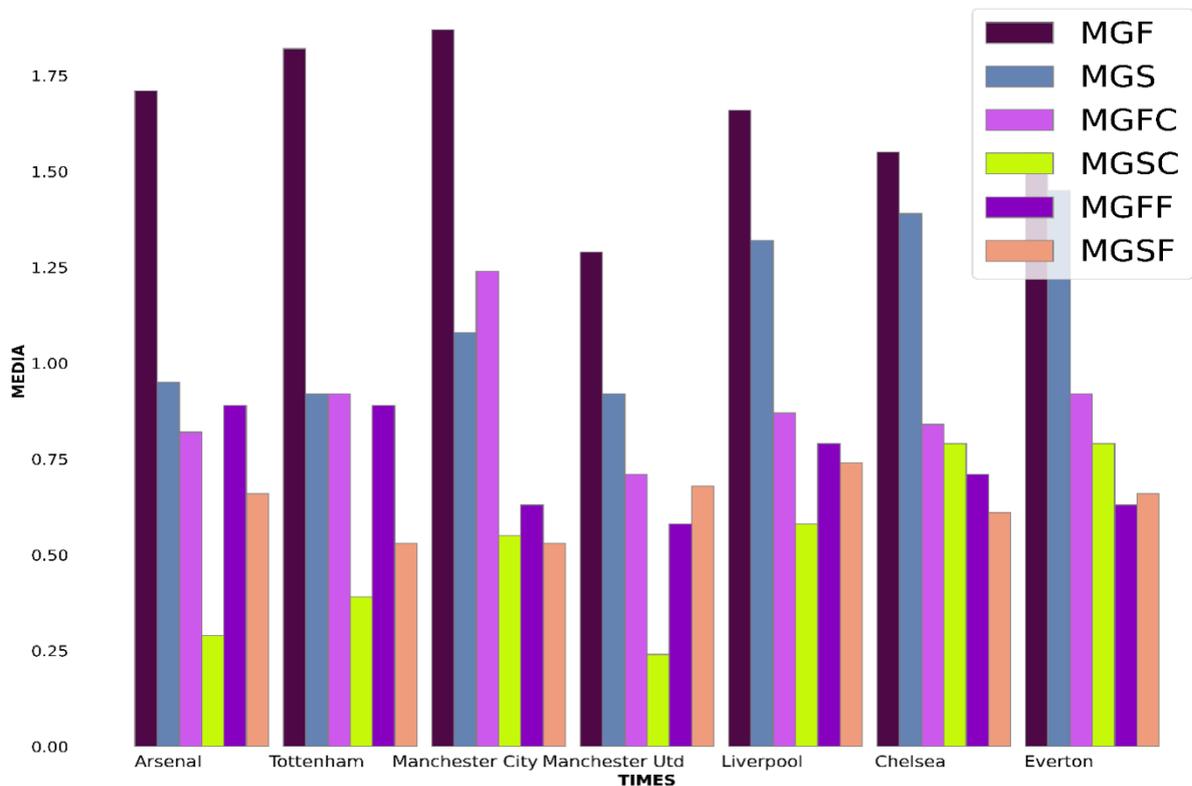
Fonte: Autoria própria;

Figura 41. Média de gols feitos e sofridos na temporada 2014/2015



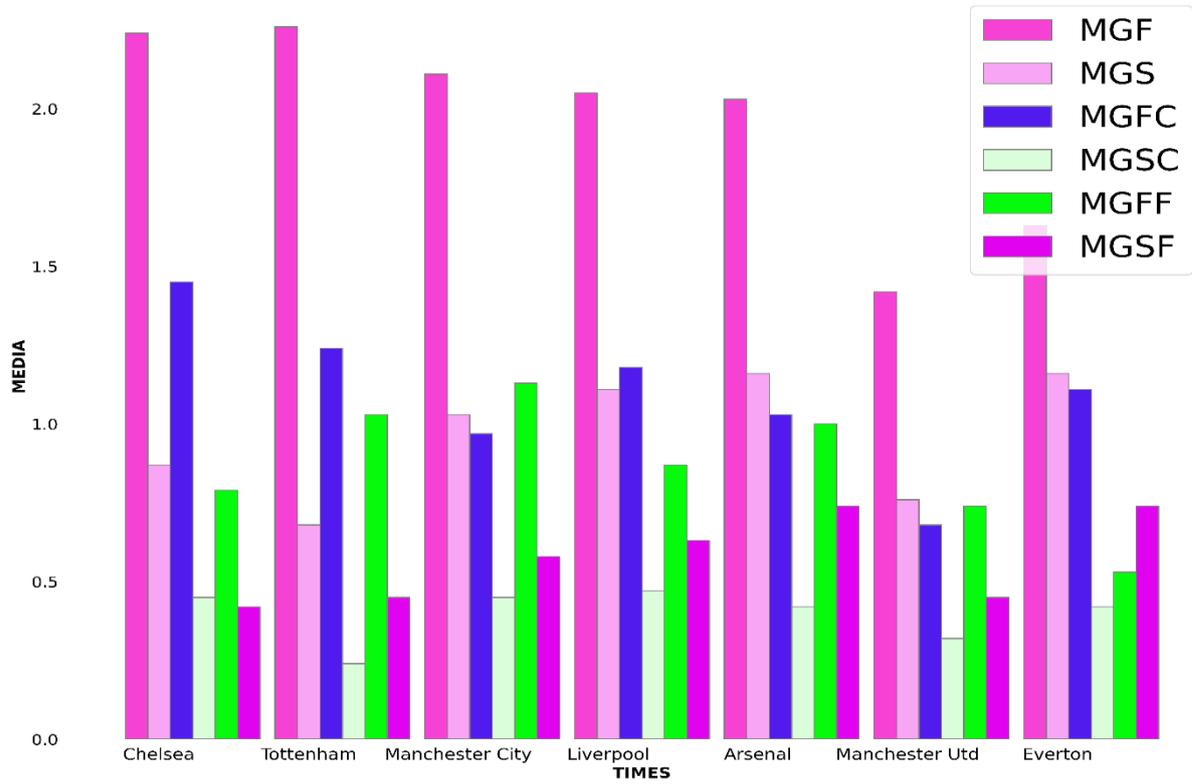
Fonte: Autoria própria;

Figura 42. Média de gols feitos e sofridos na temporada 2015/2016



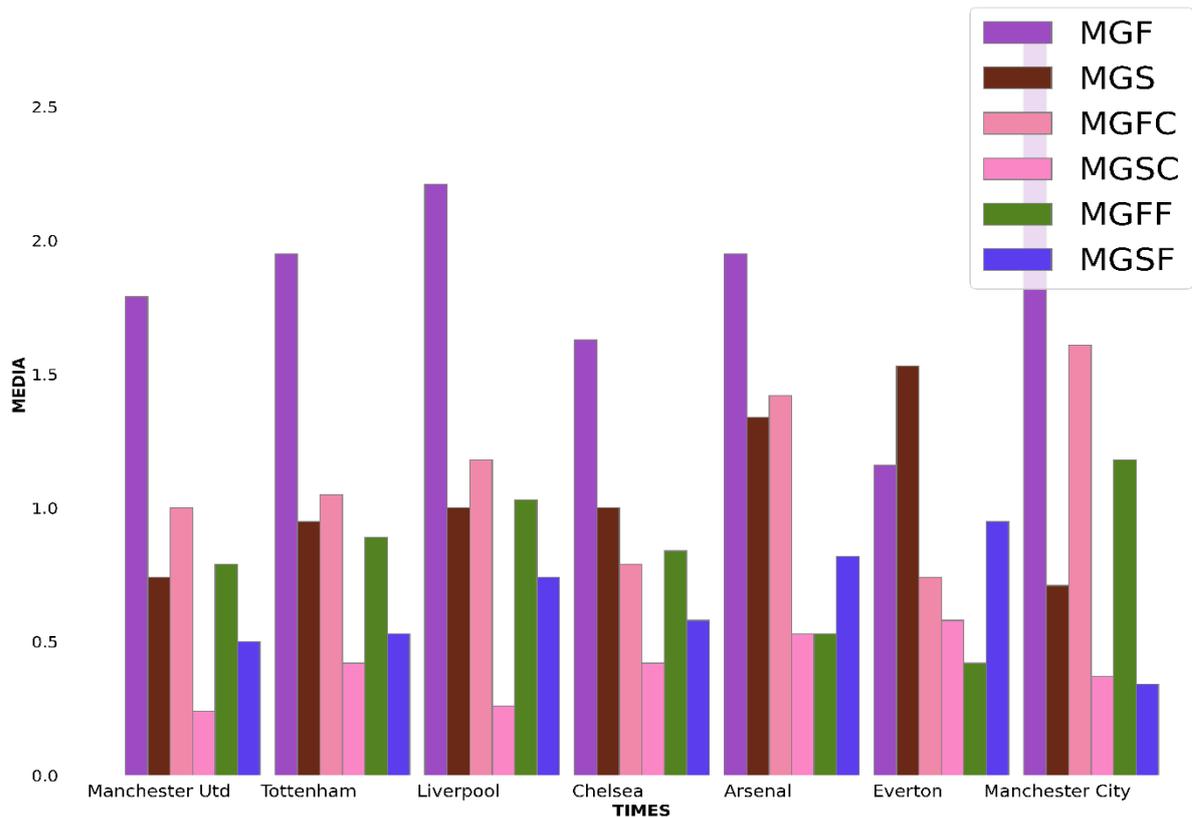
Fonte: Autoria própria;

Figura 43. Média de gols feitos e sofridos na temporada 2016/2017



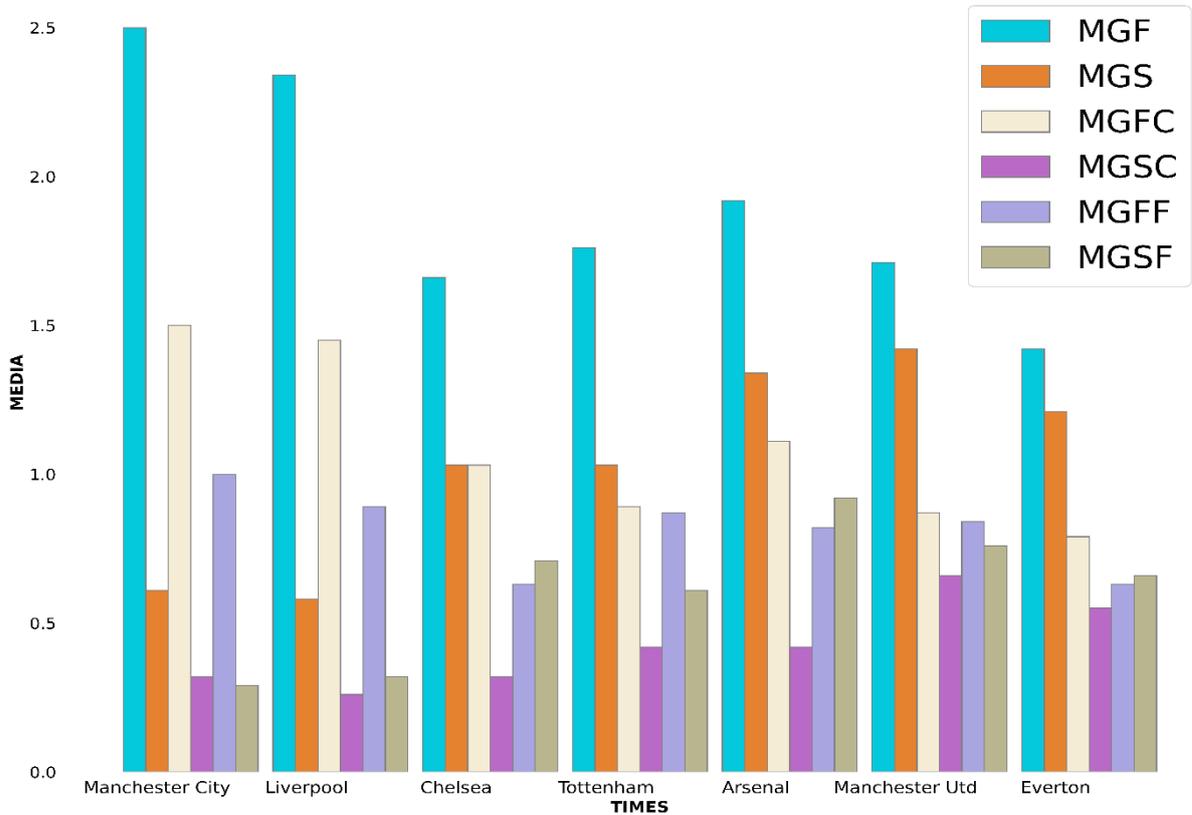
Fonte: Autoria própria;

Figura 44. Média de gols feitos e sofridos na temporada 2017/2018



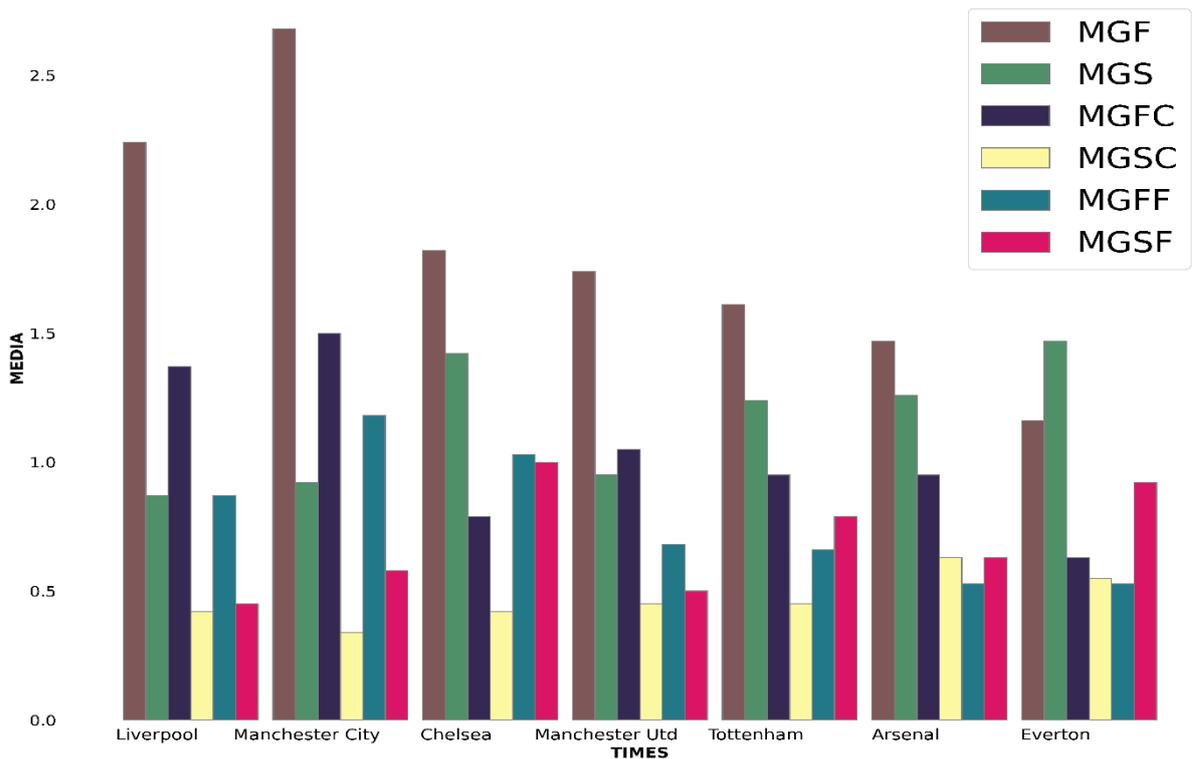
Fonte: Autoria própria;

Figura 45. Média de gols feitos e sofridos na temporada 2018/2019



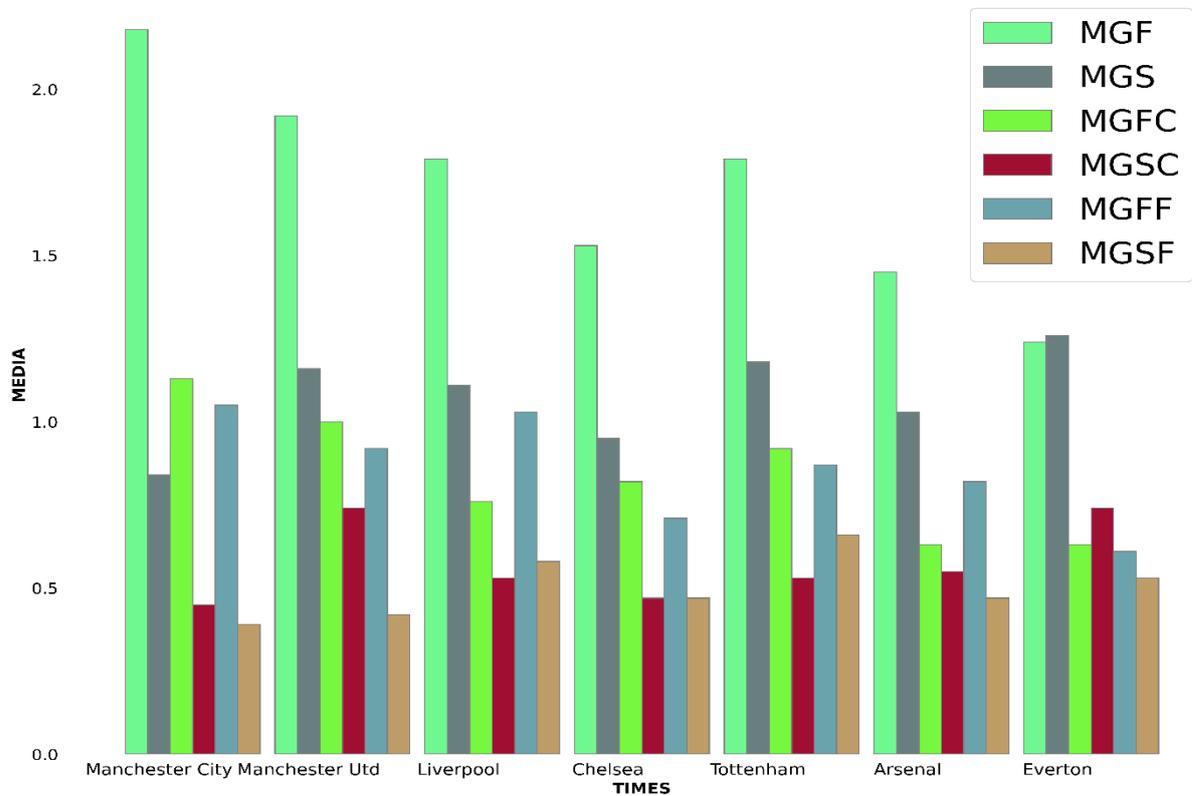
Fonte: Autoria própria;

Figura 46. Média de gols feitos e sofridos na temporada 2019/2020



Fonte: Autoria própria;

Figura 47. Média de gols feitos e sofridos na temporada 2020/2021



Fonte: Autoria própria;

Os dados e padrões desse experimento secundário observados nas figuras 17 a 37, possibilitam a identificação visual de padrões a serem observados e caminhos a serem seguidos e evitados, demonstrando o comportamento e rendimento dos times dentro de campo, como defensivamente ou no ataque através de dados como MGF e MGS, ou se a troca de treinador foi eficaz para o aumento do rendimento dos clubes, ou qual sequência de vitórias desempenham um papel importante na classificação de um time. Diante do exposto acima, são apenas análises visuais interpretativas de comportamentos e padrões projetados ao longo da pesquisa.

Esta pesquisa também realiza experimentos utilizando a modelagem de dados supracitados no capítulo 4.2 e 4.3 com dois conjuntos de dados, o qual um dos conjuntos utiliza os 7 times que se mantiveram durante os 10 anos coletados do campeonato, o qual será identificado como grupo A e o outro com todos os times que participaram do campeonato no respectivo ano, que correspondera como grupo B.

O experimento que visa analisar e prever o comportamento de conjuntos que levaram o time a zona da *champions league*, utilizou-se de algoritmos de MD

supracitados no capítulo 4.3. Para a análise dos conjuntos, foram utilizados apenas os dados do primeiro turno do campeonato, que corresponde a 19 rodadas e esquematizados (figura 26) da seguinte maneira:

- Resultado por rodada (Vitória, Empate ou Derrota);
- Classificação (Sim ou Não).

Nas tabelas 2 e 3 foram explicitas as relações de desempenho de cada grupo com base nos algoritmos e conjuntos de atributos descrito na tabela 1:

Tabela 2 – relação de desempenho por algoritmo com 70 instancias

Algoritmo	Capacidade de classificação	acurácia	tempo
J48	70%	0,458	0s
MultilayerPerceptron	70%	0,390	0.67s

Fonte: Autoria própria;

Tabela 3 – relação de desempenho por algoritmo com 200 instancias

Algoritmo	Capacidade de classificação	acurácia	tempo
J48	94%	0,963	0.1s
MultilayerPerceptron	99.5%	0,984	3s

Fonte: Autoria própria;

Na tabela 2 foram classificados os dados do grupo A, o que gerou uma menor taxa de classificação em relação aos dados do grupo B, em decorrência da quantidade de instancias, enquanto na tabela 2 possuía 70 instancias, a tabela 3 totalizou 200 linhas para teste, com taxas de previsibilidade acima de 90%, demonstrando a viabilidade da aplicação do experimento realizado nesta pesquisa exploratória.

Analisando a estas perspectivas o aumento das instancias e dados obtidos influenciam de maneira integral a capacidade de classificação dessas instancias, permitindo a piora ou melhora na capacidade de previsão.

5. CONCLUSÃO

A proposta apresentada por essa pesquisa exploratória relata resultados interessantes para o tema em questão, utilizando-se de dados e ferramentas de acesso na web, para a previsão dos classificados para a *champions league* do campeonato inglês.

Esta pesquisa mostrar a relação entre a necessidade de dados e a classificação deles para obtenção de resultados satisfatórios, demonstrando a necessidade e relação da coleta e tratamentos desses dados para uma mineração de dados qualitativa.

Apesar dos resultados obtidos, estudos envolvendo maiores conjuntos de amostras ou diferentes métodos de análise, como levar em consideração, o mando de campo, ou o time adversário para determinar o tamanho da vitória, podem gerar resultados interessante com relação ao desempenho da previsibilidade. Métodos e informações que podem ser exploradas em trabalhos futuros.

REFERÊNCIAS

ALMEIDA, C. M. Tecnologia no esporte mais popular do mundo: o Futebol. In: CIMM. [S. I.], 2020. Disponível em: <https://www.cimm.com.br/portal/artigos/20043-tecnologia-no-esporte-mais-popular-do-mundo-o-futebol>. Acesso em: 10 novembro 2022.

ALVES, P. M. Inteligência Artificial e Redes Neurais. In: IPEA. [S. I.], 2020. Disponível em: <https://www.ipea.gov.br/cts/pt/central-de-conteudo/artigos/artigos/106-inteligencia-artificial-e-redes-neurais>. Acesso em: 25 outubro 2022.

ÁRVORE DE DECISÃO. In: WIKIPÉDIA, a enciclopédia livre. [S. I.], 2021. Disponível em: https://pt.wikipedia.org/w/index.php?title=%C3%81rvore_de_decis%C3%A3o&oldid=61769702. Acesso em: 10 outubro 2022.

BRAGA, A. P.; CARVALHO A. C.; LUDEMIR, T. B. Redes Neurais Artificiais: Teoria e aplicações. Rio de Janeiro: LTC, 2000.

CAMPOS, C. C. EDITORIAL. In: BIBLIOTECA DIGITAL FGV [S. I.], 2013. Disponível em: https://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/11316/fgvprojetos_caderno_futebol.pdf?sequence=5&isAllowed=y. Acesso em: 10 novembro 2022.

COMO A TECNOLOGIA ajuda o futebol atual. In: LUDOPÉDIO. [S. I.], 2022. Disponível em: <https://ludopedio.org.br/arquibancada/como-a-tecnologia-ajuda-o-futebol-atual/>. Acesso em: 10 novembro 2022.

COMO FUNCIONA O ALGORITMO Árvore de Decisão. In: DIDÁTICA TECH. [S. I., 2022?]. Disponível em: <https://didatica.tech/como-funciona-o-algoritmo-arvore-de-decisao/>. Acesso em: 20 outubro 2022.

CONFEDERAÇÃO BRASILEIRA DE FUTEBOL (CBF); AGÊNCIA EY (EY). IMPACTO DO FUTEBOL BRASILEIRO. [S. I.], 2018. Disponível em:

https://conteudo.cbf.com.br/cdn/201912/20191213172843_346.pdf. Acesso em: 15 novembro 2022.

CORDANI, L. K. Algumas considerações sobre a inferência estatística. <https://nilsonjosemachado.net/lca20.pdf> Acesso em: 15 novembro 2022.

CROLLEY, L.; LEVERMORE, R.; PEARSON, G. For business or pleasure? A discussion of the impact of European Union law on the economic and socio-cultural aspects of football. *European Sport Management Quarterly*. v. 2, p. 276-295, 2002.

DATA MINING: entenda tudo sobre mineração de dados. In: LABONE. [S. l., 2022?]. Disponível em: <https://www.laboneconsultoria.com.br/o-que-e-data-mining/>. Acesso em: 10 novembro 2022.

DELOITTE. Annual Review of Football Finance 2022. [S. l.], 2022. Disponível em: <https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/sports-business-group/deloitte-uk-annual-review-of-football-finance-2022.pdf>. Acesso em: 15 novembro 2022.

FEDERAÇÃO INTERNACIONAL DE FUTEBOL (FIFA). more-than-half-the-world-watched-record-breaking-2018-world-cup. [S. l.], 2018. Disponível em: <https://www.fifa.com/tournaments/mens/worldcup/2018russia/media-releases/more-than-half-the-world-watched-record-breaking-2018-world-cup>. Acesso em: 10 novembro 2022.

GESTÃO ESPORTIVA - Futebol e Responsabilidade Social: Uma Parceria Importante. In: FUTEBOL INTERATIVO. [S. l.], 2020. Disponível em: <https://ludopedio.org.br/arquibancada/como-a-tecnologia-ajuda-o-futebol-atual/>. Acesso em: 10 novembro 2022.

GOMES, I. Visão geral de Redes Neurais. In: MEDIUM. [S. l.], 2014. Disponível em: <https://medium.com/turing-talks/aprendizado-por-reforço-1-introdução-7382ebb641ab>. Acesso em: 25 outubro 2022.

HAYKIN, S. Redes neurais: princípios e prática. 2.ed. Porto Alegre: Bookman, 2001.

INSTITUTO NEYMAR JR (INJ). QUEM SOMOS. [S. I., 2022?]. Disponível em: <https://www.institutoneymarjr.org.br/instituto/>. Acesso em: 15 novembro 2022.

INTELIGÊNCIA ARTIFICIAL. In: WIKIPÉDIA, a enciclopédia livre. Flórida: Wikimedia Foundation, 2022. Disponível em: <https://pt.wikipedia.org/w/index.php?title=Intelig%C3%Aancia_artificial&oldid=64379374>. Acesso em: 24 outubro 2022.

KLEINA, N. A história da inteligência artificial. In: TECMUNDO. [S. I.], 2018. Disponível em: <https://www.tecmundo.com.br/mercado/135413-historia-inteligencia-artificial-video.htm>. Acesso em: 25 outubro 2022.

MINERAÇÃO de Dados O que é e qual sua importância?. In: SAS. [S. I.], 2020. Disponível em: https://www.sas.com/pt_br/insights/analytics/mineracao-de-dados.html. Acesso em: 20 outubro 2022.

MIRANDA, J. V. Jupyter Notebook: Exemplos de Códigos e Como Usar. In: ALURA. [S. I.], 2022. Disponível em: <https://www.alura.com.br/artigos/conhecendo-o-jupyter-notebook>. Acesso em: 10 novembro 2022.

MOURA, K. Ciclo de vida dos dados #1. In: MEDIUM. [S. I.], 2022. Disponível em: <https://medium.com/@kvmoura/kdd-process-9b8e3062142>. Acesso em: 10 novembro 2022.

NEVES, E. C. Aprendizado por Reforço #1— Introdução. Turing Talks. In: MEDIUM. [S. I.], 2020. Disponível em: <https://medium.com/turing-talks/aprendizado-por-reforço-1-introdução-7382ebb641ab>. Acesso em: 25 outubro 2022.

NIelsen COMPANY. Favorito dos fãs: A Popularidade Global do Futebol está em ascensão. [S. I.], 2018. Disponível em: <https://www.nielsen.com/pt/insights/2018/fan-favorite-the-global-popularity-of-football-is-rising/>. Acesso em: 10 novembro 2022.

NIELSEN COMPANY. Popularidade dos Sports. [S. l.], 2018. Disponível em: <https://www.nielsen.com/wp-content/uploads/sites/2/2019/04/world-football-report-2018.pdf/>. Acesso em: 10 novembro 2022.

RODRIGUES, C. F. S.; LIMA, F. J. C.; BARBOSA, F. T. Importância do uso adequado da estatística básica nas pesquisas clínicas. In: SCIELO. [S. l.], 2017. Disponível em: <https://www.scielo.br/j/rba/a/N5PgBCrzhDkfRbX8QXsctHx/?lang=pt#>. Acesso em: 20 outubro 2022.

RUSSELL, S.; NORVIG, P. Inteligência artificial. 3. ed. Rio de Janeiro: Elsevier, 2013.

SAWYER, M. The Art of Football review – the cultural impact of our national sport. In: The Guardian. 2018. Disponível em: <https://www.theguardian.com/artanddesign/2018/jun/16/the-art-of-football-liverpool-celebrates-game-cultural-impact>. Acesso em: 11 novembro 2022.

SILVA, L. A; PERES, S. M; BOSCARIOLI, C. Introdução à mineração de dados: com aplicações em R. 1. ed. Rio de Janeiro: Elsevier, 2016.

TEIXEIRA, R. Por que o futebol é o esporte mais popular do mundo?. [S. l.], 2018. Disponível em: https://www.correiobraziliense.com.br/app/noticia/revista/2018/06/20/interna_revista_correio,689793/por-que-o-futebol-e-o-esporte-mais-popular-do-mundo.shtml. Acesso em: 10 novembro 2022.



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS
GABINETE DO REITOR

Av. Universitária, 1059 • Setor Universitário
Caixa Postal 96 • CEP 74605-010
Goiânia • Goiás • Brasil
Fone: (62) 3946.1030
www.pucgoias.edu.br • reitoria@pucgoias.edu.br

RESOLUÇÃO nº 038/2020 – CEPE

ANEXO I

APÊNDICE ao TCC

Termo de autorização de publicação de produção acadêmica

O(A) estudante Pedro Henrique Cândido Rosa do Curso de Ciência da Computação, matrícula 2017.2.0028.0052-0, telefone: 62994919958 e-mail pedro.h.candido.rosa@gmail.com, na qualidade de titular dos direitos autorais, em consonância com a Lei nº 9.610/98 (Lei dos Direitos do Autor), autoriza a Pontifícia Universidade Católica de Goiás (PUC Goiás) a disponibilizar o Trabalho de Conclusão de Curso intitulado Mineração de dados aplicada a resultados de jogos de futebol, gratuitamente, sem ressarcimento dos direitos autorais, por 5 (cinco) anos, conforme permissões do documento, em meio eletrônico, na rede mundial de computadores, no formato especificado (Texto(PDF); Imagem (GIF ou JPEG); Som (WAVE, MPEG, AIFF, SND); Vídeo (MPEG, MWV, AVI, QT); outros, específicos da área; para fins de leitura e/ou impressão pela internet, a título de divulgação da produção científica gerada nos cursos de graduação da PUC Goiás.

Goiânia, 14 de setembro de 2022.

Assinatura do autor:

Pedro Henrique Cândido Rosa

Nome completo do autor: Pedro Henrique Cândido Rosa

Assinatura do professor-orientador:

Sibélius Lellis Vieira

Nome completo do professor-orientador: Sibélius Lellis Vieira