



**INSTITUTO FEDERAL DE EDUCAÇÃO CIÊNCIA E TECNOLOGIA DO
TRIÂNGULO MINEIRO - Campus Ituiutaba
CURSO SUPERIOR DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

GABRIEL DE ARAUJO

**CONSTRUÇÃO DE UM SOFTWARE PARA ANÁLISE DE DADOS DE
PARTIDAS DE FUTEBOL COLETADOS ATRAVÉS DO PROCESSO
DE SCOUT**

ITUIUTABA, MG

2019

GABRIEL DE ARAUJO

**Construção de um software para análise de dados de partidas de futebol coletados
através de processo de scout**

Projeto de pesquisa apresentado ao Instituto Federal de Educação, Ciência e Tecnologia do Triângulo Mineiro, Campus Ituiutaba, como requisito parcial para conclusão do Curso de Bacharelado em Ciência da Computação.

Orientador: Prof. Dr. Rodrigo Grassi Martins

ITUIUTABA, MG

2019

Lista de Figuras

1 Fluxograma do sistema proposto 4

Lista de Tabelas

1 Cronograma de atividades do projeto. 6

Sumário

1	INTRODUÇÃO	1
1.1	OBJETIVO GERAL	2
1.2	OBJETIVOS ESPECÍFICOS	2
2	REFERENCIAL TEÓRICO	3
3	METODOLOGIA	4
4	RECURSOS	5
5	CRONOGRAMA	6
	REFERÊNCIAS	7

1 INTRODUÇÃO

O futebol é uma das modalidades esportivas mais famosas e disputadas do mundo. Trata-se de um esporte cujo objetivo transpor uma bola entre as balizas, que são as extremidades do campo, utilizando basicamente toques com os pés. Vence a partida a equipe que atingir o objetivo – que são chamados de gols - mais vezes na partida (SFEIR, 2011). Uma das principais razões pelo futebol se tornar uma febre mundial é o fácil entendimento das regras, o baixo custo e o fato de ser uma das modalidades mais empolgantes no meio esportivo. A última copa do mundo que foi jogada na Rússia no ano de 2018 atingiu uma audiência televisiva de mais de 3,5 bilhões de pessoas, batendo recorde de audiência, e somente a final entre França e Croácia atraiu 1,1 bilhões de telespectadores (CHADE, 2018).

O alto nível de interesse das pessoas por esse esporte gera não apenas telespectadores, mas também muita movimentação financeira em torno dessa modalidade. A copa do mundo de 2018 teve lucro para a FIFA de 5,35 bilhões (FIFA, 2018). A movimentação financeira envolve patrocínios a clubes e seleções nacionais, venda de ingressos, produtos licenciados e transmissões por veículos de comunicação. Além é claro de transações de transferências envolvendo jogadores. Esse tipo de movimentação financeira atrai vários investidores que visão lucrar com o esporte. Sistemas computacionais que trabalham com a previsão de resultados e que auxiliam a minimizar os riscos e maximizar os lucros tornam-se então uma importante ferramenta de trabalho para o dia a dia do futebol (PERIN; VUILLEMOT; FEKETE, 2013).

O processo de scout é amplamente utilizado esportes, principalmente no futebol, para o registro, observação e análise do desempenho técnico e tático de equipes em partidas. Em esportes, o scout pode ser definido como uma técnica que consiste em analisar a partida, os momentos, os lances de um jogo para verificar o rendimentos das equipes. O scout é o objeto de estudo em várias modalidades esportivas: basquete, vôlei, handebol, futebol americano, beisebol e futebol (MARTINS et al., 2017). A coleta de dados é utilizada não apenas para conhecimento do rendimento da própria equipe, mas também para o estudo de táticas e técnicas de equipes adversárias. Esses dados possibilitam mensurar quais são as principais características de uma equipe, identificando quais são suas principais jogadas, seus principais jogadores, sua organização tática e técnica, porém não existe uma padronização de quais dados devem ser coletados, o que deixa o processo muito abrangente (DUARTE; SOARES; TEIXEIRA, 2015). O objetivo proposto é criar um software que seja capaz de acessar base de dados e compilar e projetar esses dados de maneira que auxilie o usuário a tomar decisões.

1.1 OBJETIVO GERAL

O objetivo deste trabalho é construir um software que consiga acessar diferentes bases de dados e compilar e projetar esses dados de maneira que auxilie o usuário no processo de tomada de decisão.

1.2 OBJETIVOS ESPECÍFICOS

Os objetivos específicos deste trabalho são:

- comparar as bases de dados existentes e compilar essas informações de maneira a projetar uma base de dados única;
- fornecer acesso ao usuários de informações referentes aos quatro elementos do processo de scout: espaço, tempo, jogador e fundamento;
- permitir ao usuário projetar e visualizar os relatórios com a compilação das informações proveniente da base de dados.

2 REFERENCIAL TEÓRICO

Em uma partida de futebol o número de anotações que se pode fazer para que seja possível sua correta descrição é bastante elevado (CONSTANTINO; FENTON; NEIL, 2013). Se for levado em consideração outros fatores extracampo temos novamente um alto número de componentes envolvidos. Dessa maneira a definição de quantas e quais características serão utilizadas para se realizar a predição dos resultados de partidas se torna bastante complexa (TAX; JOUSTRA, 2015). Na literatura é possível encontrar trabalhos que levam em consideração uma série de fatores tais como dados de fundamentos obtidos por scout como em Sfeir (2011), outros como em Owramipur e Mozneb (2013) são utilizados fatores de logísticas tais como distância percorrida entre duas partidas subsequentes (TAX; JOUSTRA, 2015) ou fatores psicológicos como em Duarte, Soares e Teixeira (2015).

O processo scout apesar de bastante difundido e utilizado não é padronizado no que diz respeito ao número de fundamentos que serão coletados (PENDHARKAR; KHOSROWPOUR; RODGER, 2000). Em Brooks (BROOKS; KERR; GUTTAG, 2016) é utilizado um conjunto de 7 fundamentos obtidos por scout e um conjunto de 6 características de logísticas para se obter uma correta predição. Já em Ulmer e Fernandez (2013) são utilizados 9 fundamentos obtidos por scout. Em Igiri (2015) são utilizadas como características 23 fundamentos coletados por scout. Em Hucaljuk e Rakipović (2011) são utilizados 10 fundamentos obtidos por scout. Tax e Joustra (2015) utiliza 19 características obtidas por scout, 12 de logística e 12 baseadas em sites de apostas. Enquanto que Duarte, Soares e Teixeira (2015) utiliza um conjunto de 12 características obtidas por scout e 12 características envolvendo aspectos psicológicos das equipes.

É importante ressaltar mais uma vez que não existe uma padronização para as características envolvidas com os trabalhos existentes na literatura. Pode-se afirmar, entretanto que as características estão ligadas ao tipo de informação que se necessita coletar sobre um ou mais aspectos que envolvem uma partida de futebol (TAX; JOUSTRA, 2015). Não é possível ainda afirmar que uma outra metodologia é superior a outra, mas que são complementares. Para o caso desse software o mesmo deverá ser capaz de: comparar as bases de dados existentes e compilar essas informações de maneira a projetar uma base de dados única.

3 METODOLOGIA

Toda a execução do projeto será realizada por microcomputador presente no parque computacional da instituição. Serão utilizadas as linguagens HTML, PHP, Java e JavaScript e o software Weka, que já se mostraram ser uma boa escolha em desenvolvimento de softwares esportivos scout.

A construção do software será modularizada, conforme podemos observar na figura 1. A montagem da base de dados será baseada nas bases de dados existentes na literatura e deverá ser capaz de armazenar diferentes modelagens em uma única base. O primeiro módulo, selecionar, é o responsável por selecionar: a base de dados e partida desejada. O próximo módulo, visualizar, é o responsável por apresentar os dados do jogo através dos seus quatro elementos do processo de scout: espaço (imagens de campo), tempo (escolha do tempo da partida), jogador(escalação) e fundamento (tabelas de dados). Já o terceiro e o quarto módulo, otimizar e relatório permite ao usuário projetar e visualizar os relatórios com a compilação das informações proveniente da base de dados.

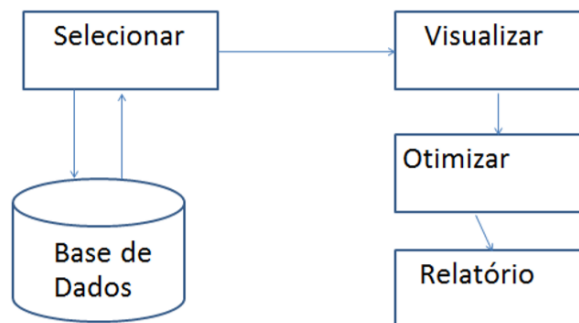


Figura 1: Fluxograma do sistema proposto

4 RECURSOS

Para a realização do projeto, foram utilizados os seguintes recursos:

- Um notebook;
- Eclipse IDE;
- A ferramenta WEKA.

5 CRONOGRAMA

Abaixo encontra-se o cronograma de atividades que serão desenvolvidas neste projeto até a defesa do TCC.

Atividades	Fev/2019	Mar/2019	Abr/2019	Mai/2019	Jun/2019	Jul/2019	Ago/2019	Set/2019	Out/2019	Nov/2019	Dez/2019
Levantamento Bibliográfico	X	X									
Montagem da base de dados			X	X							
Estudo da ferramenta Weka				X	X						
Preparação dos Dados						X	X				
Execução dos testes							X	X	X		
Escrita de artigo e relatório									X	X	
Banca											X

Tabela 1: Cronograma de atividades do projeto.

REFERÊNCIAS

- BROOKS, J.; KERR, M.; GUTTAG, J. Using machine learning to draw inferences from pass location data in soccer. **Statistical Analysis and Data Mining: The ASA Data Science Journal**, Wiley, v. 9, n. 5, p. 338–349, jun 2016.
- CHADE, J. **China faz Copa de 2018 ter audiência recorde de 3,5 bilhões de pessoas**. 2018. Disponível em: <https://esportes.estadao.com.br/noticias/futebol, china-faz-copa-de-2018-ter-audiencia-recorde-de-3-5-bilhoes-de-pessoas,70002654539>. Acesso em: 17 jun 2019.
- CONSTANTINOU, A. C.; FENTON, N. E.; NEIL, M. Profiting from an inefficient association football gambling market: Prediction, risk and uncertainty using bayesian networks. **Knowledge-Based Systems**, Elsevier BV, v. 50, p. 60–86, sep 2013.
- DUARTE, L. M. da S.; SOARES, C.; TEIXEIRA, J. **Previsão de resultados de jogos de futebol**. Dissertação (Mestrado) — Faculdade da Engenharia da Universidade do Porto, 2015.
- FIFA. **FIFA Financial Report 2018**. 2018. Disponível em: <https://resources.fifa.com/image/upload/xzshsoe2ayttyquuxhq0.pdf>. Acesso em: 17 jun 2019.
- HUCALJUK, J.; RAKIPOVIĆ, A. Predicting football scores using machine learning techniques. **2011 Proceedings of the 34th International Convention MIPRO**, may 2011.
- IGIRI, C. P. Support vector machine–based prediction system for a football match result. **IOSR Journal of Computer Engineering**, v. 17, n. 3, p. 21–26, jun 2015.
- MARTINS, R. G. et al. Exploring polynomial classifier to predict match results in football championships. **Expert Systems with Applications**, Elsevier BV, v. 83, p. 79–93, oct 2017.
- OWRAMIPUR, P. E. F.; MOZNEB, F. S. Football result prediction with bayesian network in spanish league-barcelona team. **International Journal of Computer Theory and Engineering**, October 2013.
- PENDHARKAR, P. C.; KHOSROWPOUR, M.; RODGER, J. A. Application of bayesian network classifiers and data envelopment analysis for mining breast cancer patterns. **Journal of Computer Information Systems**, v. 40, n. 4, p. 127–132, 2000.
- PERIN, C.; VUILLEMOT, R.; FEKETE, J. D. Soccerstories: A kick-off for visual soccer analysis. **IEEE Transactions on Visualization and Computer Graphics**, v. 19, n. 12, p. 2506–2515, Dec 2013. ISSN 1077-2626.
- SFEIR, M. N. Laws of the game (adapted from fifa 2010-11). **World Literature Today**, v. 85, n. 3, p. 38–39, may 2011.
- TAX, N.; JOUSTRA, Y. Predicting the dutch football competition using public data: A machine learning approach. Unpublished, 2015.
- ULMER, B.; FERNANDEZ, M. **Predicting Soccer Match Results in the English Premier League**. Tese (Doutorado) — Stanford University, 2013.