**Resumo** | **Predicting the potential ability of football players in the Football Manager game**

O jogo Football Manager é um produto da empresa Sports Interective e é lançado anualmente, de maneira geral, no final do ano. Tal mídia é altamente reconhecida nos países europeus, sendo vendido quase 3 milhões de unidades na versão de 2020. Tal sucesso não é à toa, já que Football Manager é um jogo de altíssima qualidade apresentando diversas Inteligências Artificiais. Dentre elas pode-se ressaltar a área de táticas, relacionamentos, scouting [8] e desenvolvimento da base.

Este projeto focou em entender mais sobre a inteligência artificial de jogadores da base e seus potenciais. Dessa forma, buscou-se entender mais em como o jogo calcula potencial de jogadores de menores de 21 anos e quais eram os modelos e variáveis usadas para isso. Para entender mais isso buscou-se dois artigos que explicam como é feita a previsão [1][2].

A metodologia utilizada no projeto foi uma tradicional de projetos de Machine Learning, na qual há análise exploratória, pré-processamento, treinamento de modelos, avaliação e melhorias, porém se utilizou a técnica de sobreamostragem, por conta da baixa porcentagem de jogadores com habilidade para utilizar no treinamento.

Neste problema, é necessária a classificação binária. Esta tarefa é feita usando modelos de Árvores de Decisão (DT), Regressão Logística (LR) e Máquina de Vetores de Suporte (SVM).

Uma DT é uma representação gráfica, que é um fluxograma de nós baseado em hierarquia [3], de todas as soluções possíveis para uma decisão baseada em certas condições, podendo ser *entropy* ou *gini* [4]. As árvores de decisão são uma forma relativamente simples e transparente de classificação, embora o cálculo às vezes possa ser excessivamente complexo e demorado, sendo modelos de aprendizado supervisionado.

LR é diferente da regressão linear no sentido de que o valor previsto deve ser categórico, no entanto também utiliza matemática para achar relação entre os recursos [5]. Um gráfico da classificação LR com dados de amostra rotulados. O algoritmo LR fornece pontuações de probabilidade para características diferentes, mas tem um desempenho ruim com muitos recursos irrelevantes e correlacionados.

SVM é um modelo linear que cria um limite de decisão para separar os dados em classes. É semelhante ao LR, mas o SVM tenta encontrar a melhor margem que separa as classes e isso reduz o risco de erro nos dados. Isto ele faz através dos vetores de suporte são simplesmente as coordenadas da observação individual. SVM é uma fronteira que melhor segrega as duas classes [6].

As redes neurais podem ser modelos alternativos. Ao lado das redes neurais, o XGboost foi considerado para esta tarefa. O Xgboost usa o algoritmo de árvore de decisão de aumento de gradiente e geralmente obtém bons resultados com ele, devido ao reforço do gradiente [7]

Após o balanceamento das classes, os modelos DT, LR e SVM superaram consideravelmente a linha de base. Com uma linha de base de aproximadamente 0,5, o modelo LR e SVM atingiu uma precisão em torno de 0,8. O modelo DT foi o melhor modelo de pontuação, com uma precisão de 0,91. O modelo DT balanceado foi utilizado para extrair os atributos mais importantes.

Utilizando esses resultados, ficou evidente que as principais características que determinam um alto potencial em um jogador é o equilíbrio, técnica e bravura, principalmente. Esses resultados podem ser utilizados por clubes para contratar jogadores com essas características ou investirem em treinamentos específicos na divisão de base visando lucrar com seus jovens.

**Referências**

[1] V. W., WILLIAM. ***Predicting the potential ability of football players in the Football Manager game***. Disponível em <<http://williamvanwijk.nl/wp-content/uploads/2022/07/Thesis_FM_Predict_potential_ability.pdf>>. Acesso em 25 out. 2022.

[2] M. R. L., ERIC. ***O uso do jogo eletrônico Football Manager e do portal Transfermarkt para mensurar o potencial de futebolistas promissores***. Disponível em <<https://repositorio.unesp.br/bitstream/handle/11449/235740/rocha-lima_em_dr_rcla.pdf?sequence=5>>. Acesso em 28 nov. 2022.

[3] SACRAMENTO, G. ***Árvore de Decisão: entenda esse algoritmo de Machine Learning***. Disponível em <<https://blog.somostera.com/data-science/arvores-de-decisao>>. Acesso em 29 nov. 2022.

[4] FREITAS, T. ***Entendendo as árvores de decisão em Machine Learning***. Disponível em <<https://sigmoidal.ai/entendendo-as-arvores-de-decisao-em-machine-learning/>>. Acesso em 29 nov. 2022.

[5] TIBC. ***O que é regressão logística?***.Disponível em <<https://www.tibco.com/pt-br/reference-center/what-is-logistic-regression>>. Acesso em 29 nov. 2022.

[6] ADDAN, D. ***Support Vector Machine***. Disponível em <<https://www.inf.ufpr.br/dagoncalves/IA07.pdf>>. Aceso em 29 nov. 2022.

[7] Data Science Team. ***XGBoost – O que é isso?***. Disponível em <<https://datascience.eu/pt/programacao/xgboost-2/>>. Acesso em 29 nov. 2022.

[8] AJAY. ***How Do Football Scouts Work?***. Disponível em <<https://sportsfinding.com/how-do-football-scouts-work/104257/>>. Acesso em 29 nov. 2022.