Conclusão

Projeto 3 – Ciência dos dados

O dataset do Fifa 18 contém uma série de informações para todos os jogadores disponíveis dentro do jogo. Entre essas informações, podemos citar dados referentes a habilidade do jogador nas diferentes áreas do jogo, como ataque, defesa, drible, poder, entre outros dados de habilidade. Porém, dentre os dados apresentados, três acabam chamando mais atenção do que os demais, Overall, que seria basicamente uma nota dada ao jogador com base em todos os dados de habilidade, Potencial, que se trata de uma “previsão” que o jogo cria considerando a capacidade do jogador de desenvolver seus atributos e alcançar um valor de Overall maior do que o inicial, e o Valor de mercado, que se trata, basicamente, do valor de venda do jogador no jogo.

Dentre todos os dados fornecidos, o de Valor de Mercado é o mais “intrigante”, visto que o jogo não explicita a maneira com que este valor é definido, por isso, decidimos tratar como pergunta como o jogo define tal valor.

Consideramos de início testar se as variáveis Overall, Potencial e Value (Valor de mercado) apresentavam uma correlação, para tentar encontrar a forma que o Value era definido. Visto que a relação entre os valores se mostrava aproximadamente exponencial, utilizamos o logaritmo dos valores para obter a relação linear. Como resultado da regressão, obtivemos um valor de “R-squared” ou R², que se trata basicamente de quão bem o modelo se encaixa nos dados analisados, onde temos um valor alto de determinação do modelo se encaixar nos dados analisados, de 70,1%. Além disso, obtivemos valores baixos de p-value, ou seja, é baixa a probabilidade de aceitarmos a hipótese nula, o que significa que nosso modelo aparenta estar de acordo com os dados analisados.

Para confirmar o que obtivemos no último passo, analisamos os resíduos da nossa regressão, que trata da diferença entre os valores reais dados pelo dataset e os valores estimados pelo plano. O resultado se aproximou de uma normal, mostrando novamente que nosso modelo se encaixa no que foi previsto. Nesta análise dos resíduos, é possível observar também, que, pela existência de uma frequência grande de resíduos pequenos, jogadores “supervalorizados” apresentariam uma diferença grande entre o valor obtido pelo modelo e o valor real.

Após estas análises, testamos com um jogador, Gabriel Jesus, e o valor obtido pelo modelo se aproximou muito do valor real, demonstrando que nosso modelo é válido, visto que ele é um jogador considerado uma promessa dentro do jogo com base em seu Potencial, de forma que a maior parte dos jogadores nesta mesma categoria apresentam grande proximidade entre o valor calculado e o valor real. Além disso, adicionamos uma opção para facilitar ao usuário a utilização do programa, onde existe um campo para inserção do nome de um jogador do jogo Fifa 18, que retorna o valor obtido pelo modelo e o real, permitindo a comparação.

No mais, o objetivo de se obter um modelo que demonstrasse o cálculo por trás do Value ou Valor de Mercado dentro do jogo foi concluído. Foi feito o uso do logaritmo dos dados para aperfeiçoar o resultado, com base na distribuição da relação linear. Para melhorar o modelo, teríamos que encontrar outros dados presentes no dataset que também apresentassem uma boa correlação com os já utilizados, para possivelmente obter melhores resultados no que se dizem os jogadores “supervalorizados”, que possuem um cálculo diferente para o valor de mercado.