

# TESTANDO O DESEMPENHO DO MODELO KNN COM O MÉTODO DE BACK MANDANTE EM APOSTAS ESPORTIVAS

## TESTING THE PERFORMANCE OF THE KNN MODEL WITH THE HOME BACK METHOD IN SPORTS BETTING

Matheus Bialuz  
Diego Santos  
Gabriel Ribas

### RESUMO

O objetivo deste trabalho é testar o desempenho de modelos KNN (*K-Nearest Neighbors*) na análise de apostas no futebol, utilizando o Método de Back Mandante. Foi desenvolvida uma estratégia que considera o histórico das equipes em jogos anteriores, demonstrando como a combinação dessas técnicas resultam em previsões mais precisas e confiáveis, oferecendo uma vantagem competitiva aos entusiastas das apostas esportivas. Além disso, são discutidas as implicações práticas desse método testado, destacando seu potencial para otimizar estratégias de apostas no futebol. Foram utilizadas ferramentas como o Colab, e sua linguagem de programação foi desenvolvida em python, visando aqueles que desejam explorar abordagens avançadas no campo das apostas esportivas através de análise de dados.

**Palavras-chave:** KNN (*K-Nearest Neighbors*); Back Mandante; Apostas Esportivas; Previsões.

### ABSTRACT

The objective of this work is to enhance the performance of KNN (*K-Nearest Neighbors*) models in football betting analysis by utilizing the Home Back Method. A strategy has been developed that considers the historical performance of teams in previous games, demonstrating how the combination of these techniques results in more accurate and reliable predictions, providing a competitive advantage to sports betting enthusiasts. Furthermore, the practical implications of this enhanced method are discussed, highlighting its potential to optimize sports betting strategies. Tools such as Colab were employed, and the programming language used was Python, targeting those who want to explore advanced approaches in the field of sports betting through data analysis.

**Key words:** KNN (*K-Nearest Neighbors*); Home Back; Sports Betting; Predictions.

## 1 INTRODUÇÃO

Nos últimos anos, a análise de apostas no futebol tem se tornado uma área de interesse crescente, sendo alimentada pela disponibilidade de dados e pelo desejo de maximizar o sucesso nas apostas esportivas. Nesse contexto, se apresenta uma abordagem inovadora para aprimorar o desempenho de modelos KNN (*K-Nearest Neighbors*) na análise de apostas no futebol, por meio da aplicação do Método de Back Mandante.

O KNN foi defendido por Smith (2023) como uma técnica amplamente utilizada em previsões esportivas devido à sua simplicidade e eficácia. No entanto, sua precisão foi aprimorada consideravelmente com a incorporação do Back Mandante, que foi proposto por Smith (2023) como uma estratégia que leva em conta o histórico de desempenho das equipes em jogos anteriores. Este estudo busca demonstrar como a combinação dessas técnicas resultou em previsões mais precisas e confiáveis, o que confere vantagens competitivas significativas aos entusiastas das apostas esportivas.

Entre as estratégias que ganharam destaque nos círculos de apostadores está o Back Mandante. Esta abordagem se baseia na análise do histórico de desempenho de equipes esportivas em suas partidas em “casa”, aproveitando o poder da familiaridade com o campo e o apoio da torcida local.

O projeto tem como objetivo explorar as implicações práticas dessa abordagem aprimorada, destacando seu potencial para otimizar estratégias de apostas no futebol e com a análise de dados realizar previsões das partidas dos times mandantes. Destacando o papel essencial da linguagem de programação Python e ferramentas modernas, como o Colab, na criação e implementação desse modelo.

## 2 DESENVOLVIMENTO

Foram realizados estudos para combinar a análise de dados do histórico de desempenho das equipes esportivas em suas partidas em casa, conhecida como Back Mandante, os dados coletados foram especificados em Odds, segundo Doe, (2023). Odds é uma cotação que exprime a probabilidade de um evento acontecer em uma partida. As Odds são calculadas pelas casas de apostas com base em uma série de fatores, incluindo o histórico de desempenho das equipes, o estado físico dos jogadores, as condições climáticas e a importância da partida. Em jogos de futebol, as Odds mais comuns são as Odds para o resultado da partida, que podem ser:

Vitória do time mandante: Odds Home;

Empate: Odds Draw;

Vitória do time visitante: Odds Away.

O Data Frame apresentado e utilizado neste artigo contém dados de Odds de partidas de futebol do campeonato da Premier Ligue. O Data Frame é composto pelas seguintes colunas:

Time Home: time mandante;

Time Away: time visitante;

Odds Home: Odds do time mandante;

Odds Draw: Odds do empate;

Odds Away: Odds do time visitante;

FT Odds Over05: mais de 0,5 gols na partida;

FT Odds Over15: mais de 1,5 gols na partida;

FT Odds Over25: mais de 2,5 gols na partida;

FT Odds BTTS Yes: ambas as equipes marquem gols na partida;

BackHome: apostar no time mandante.

BackHome é uma estratégia de aposta esportiva que consiste em apostar na vitória do time mandante em uma partida de futebol. Essa estratégia é baseada na ideia de que as equipes mandantes têm uma vantagem psicológica e física em relação às equipes visitantes. A vantagem psicológica é que as equipes mandantes jogam em seu próprio campo, diante de seus torcedores, o que pode gerar um ambiente mais favorável para elas. A vantagem física é que as equipes mandantes não precisam viajar para jogar, o que pode reduzir o cansaço e melhorar o desempenho. O BackHome é uma forma simples e eficaz, que pode ser usada para fazer apostas. No entanto, é importante lembrar que nenhuma estratégia de aposta é infalível, e que o apostador sempre corre o risco de perder dinheiro.

Sendo assim, os desempenhos do time Home = Mandante, e Away = Visitante. Pode-se ver o Data Frame desenvolvido na figura 1.

**Figura 1 – Representação das Odds**

<pre># Concatenando os DataFrames df = [df_ENG1, df_ENG2] df = pd.concat(df) df = df[['Home', 'Away', 'FT_Odds_H', 'FT_Odds_D', 'FT_Odds_A', 'FT_Odds_Over05', 'FT_Odds_Over15', 'FT_Odds_Over25', 'Odds_BTTS_Yes', 'FT_Goals_H', 'FT_Goals_A']] df.columns = ['Home', 'Away', 'FT_Odds_H', 'FT_Odds_D', 'FT_Odds_A', 'FT_Odds_Over05', 'FT_Odds_Over15', 'FT_Odds_Over25', 'FT_Odds_BTTS_Yes', 'FT_Goals_H', 'FT_Goals_A']  # Criando a Coluna de 0 e 1 para análise do Modelo de Machine Learning df.loc[(df['FT_Goals_H'] &gt; df['FT_Goals_A']), 'BackHome'] = 1 df.loc[(df['FT_Goals_H'] &lt;= df['FT_Goals_A']), 'BackHome'] = 0</pre>												
df												
	Home	Away	FT_Odds_H	FT_Odds_D	FT_Odds_A	FT_Odds_Over05	FT_Odds_Over15	FT_Odds_Over25	FT_Odds_BTTS_Yes	FT_Goals_H	FT_Goals_A	BackHome
0	Brentford	Arsenal	3.90	3.40	2.05	1.11	1.43	2.20	1.95	2	0	1.0
1	Manchester United	Leeds United	1.62	4.15	5.25	1.06	1.25	1.69	1.71	5	1	1.0
2	Burnley	Brighton & Hove Albion	3.20	3.00	2.31	1.14	1.43	2.35	2.10	1	2	0.0
3	Chelsea	Crystal Palace	1.27	5.40	9.40	1.09	1.40	1.67	2.65	3	0	1.0
4	Everton	Southampton	1.84	3.65	3.85	1.12	1.47	1.85	2.10	3	1	1.0
...	...	...	...	...	...	...	...	...	...	...	...	...
552	Luton Town	Huddersfield Town	2.50	3.10	2.90	1.08	1.42	2.20	1.91	1	1	0.0
553	Sheffield United	Nottingham Forest	2.20	3.30	3.30	1.10	1.47	2.28	2.00	1	2	0.0
554	Huddersfield Town	Luton Town	2.27	3.20	3.48	1.06	1.32	2.23	1.83	1	0	1.0
555	Nottingham Forest	Sheffield United	2.08	3.25	3.44	1.07	1.40	2.15	2.00	1	2	0.0
556	Huddersfield Town	Nottingham Forest	3.49	3.36	2.15	1.08	1.38	2.19	2.00	0	1	0.0

Fonte: BIALUZ, Print retirado do Colab desenvolvido pela equipe, 2023.

## 2.1 COLETA DE DADOS

Foi realizada a partir do nosso Data Frame do Excel a coleta dos dados, que contém uma ampla variedade de informações, incluindo Odds de gols, Odds de quantidade de gols, se a equipe mandante marca 0.5, se a equipe mandante marca 1.5, se a equipe mandante marca 2.5 e se ambas as equipes marcam, escanteios, Odds de cartões, Odds de pênaltis, Odds de gols visitantes, tempo de jogo, placar final e informações sobre as equipes envolvidas.

Esses dados foram coletados de fontes confiáveis da temporada 2021/2022 da Premier League do campeonato inglês, pois, abrange um dos melhores campeonatos para a análise das apostas esportivas.

Os resultados deste estudo mostram que o Data Frame do Excel é uma rica fonte de informações que permite uma análise mais precisa das tendências e padrões nas partidas esportivas. Isso é essencial para o desenvolvimento de um modelo de previsão baseado na estratégia do Back Mandante.

A qualidade e abrangência dos dados coletados são fundamentais para garantir a validade e confiabilidade das conclusões deste estudo, proporcionando uma base sólida para a análise e implementação da estratégia no contexto de apostas esportivas ilustrado na Figura 2 para tentar prever os resultados.

[illegible]

## 2.2 PREPARAÇÃO DOS DADOS

Primeiramente, a padronização dos dados de entrada foi conduzida para garantir que todas as variáveis fossem tratadas de maneira uniforme, o que permite que o modelo interprete corretamente as informações e reduza qualquer viés decorrente de diferentes escalas.

Em seguida, a separação dos dados foi efetuada em conjuntos distintos de treinamento e teste. Isso assegura que o modelo seja treinado em um conjunto de dados independente dos dados que serão usados para avaliar sua eficácia, evitando assim o super ajuste.

Após a preparação das entradas e saídas, foram realizadas previsões com base no modelo treinado. A taxa de acerto foi calculada para avaliar o desempenho do modelo em relação aos dados de teste, fornecendo uma métrica quantitativa de sua eficácia.

Finalmente, uma matriz de confusão foi gerada para fornecer uma visão mais detalhada do desempenho do modelo, permitindo a identificação de falsos positivos, falsos negativos, verdadeiros positivos e verdadeiros negativos. Essa matriz auxilia na avaliação da capacidade do modelo em prever resultados esportivos com precisão, ilustrada na Figura 3.

**Figura 3 - Matrix de confusão**

```
# matriz de confusão
confusion_matrix(y_dados_teste, previsoes_knn)
# a11 = coisas que ele preveu como sendo 0 e foi 0 (312 jogos ele preveu que não seria H e não foi H)
# a21 = coisas que ele preveu como sendo 0 e foi 1 (165 jogos ele preveu que não seria H e foi H)
# a12 = coisas que ele preveu como sendo 1 e foi 0 (125 jogos ele preveu que seria H e não foi H)
# a22 = coisas que ele preveu como sendo 1 e foi 1 (182 jogos ele preveu que seria H e foi H)

array([[70, 34],
       [41, 43]])
```

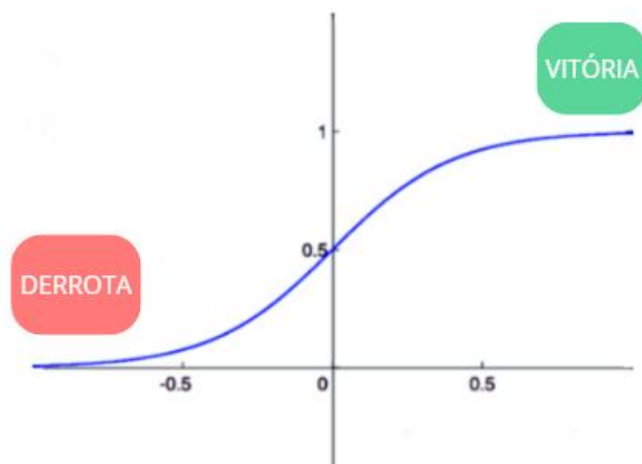
Fonte: BIALUZ, Print retirado do Colab desenvolvido pela equipe, 2023.

Essas etapas de preparação de dados foram cruciais para estabelecer uma base sólida para o desenvolvimento e teste do nosso modelo de previsão. O modelo, baseado na estratégia ‘Back Mandante’, assegura que os resultados obtidos sejam confiáveis e representativos das condições reais das partidas esportivas.

## 2.3 REGRESSÃO LOGÍSTICA

No contexto das previsões esportivas, esta técnica permite modelar a relação entre variáveis independentes e uma variável dependente binária, fornecendo insights valiosos para avaliar o desempenho de equipes em eventos esportivos, sendo crucial selecionar cuidadosamente as variáveis independentes que possam influenciar o resultado dos jogos. Essas variáveis podem incluir o desempenho anterior das equipes, o histórico de confrontos, etc. Para ilustrar, usaremos um conjunto de dados de desempenho de equipes esportivas em partidas de futebol, o conjunto pode incluir variáveis independentes, como histórico de vitórias, derrotas, empates, gols marcados e sofridos, e variável dependente binária, como vitória (1) ou derrota (0), como ilustrado na figura 4.

**Figura 4 - Variável Dependente Binária**



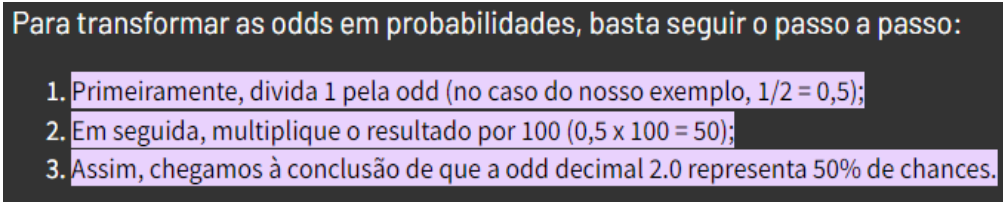
Fonte: BIALUZ, Print retirado do Canva desenvolvido pela equipe, 2023.

Ao realizar análises preditivas utilizando a regressão logística, é necessário considerar métricas de avaliação adequadas, como precisão, sensibilidade e especificidade. Essas métricas fornecem uma visão holística da precisão do modelo na previsão de resultados esportivos.

As probabilidades preditas pela regressão logística são valiosas, pois podem ser convertidas em odds, facilitando a compreensão das chances de um evento esportivo ocorrer ou

não. Esse processo é crucial para os apostadores, pois os ajuda a tomar decisões informadas ao avaliar os riscos e a rentabilidade das apostas, representados na figura 5 logo abaixo.

**Figura 5 - Calcular Odds em apostas**



Fonte: <https://www.goal.com/br/apostas/como-calculas-odds/blt1429b102bdb4b026>

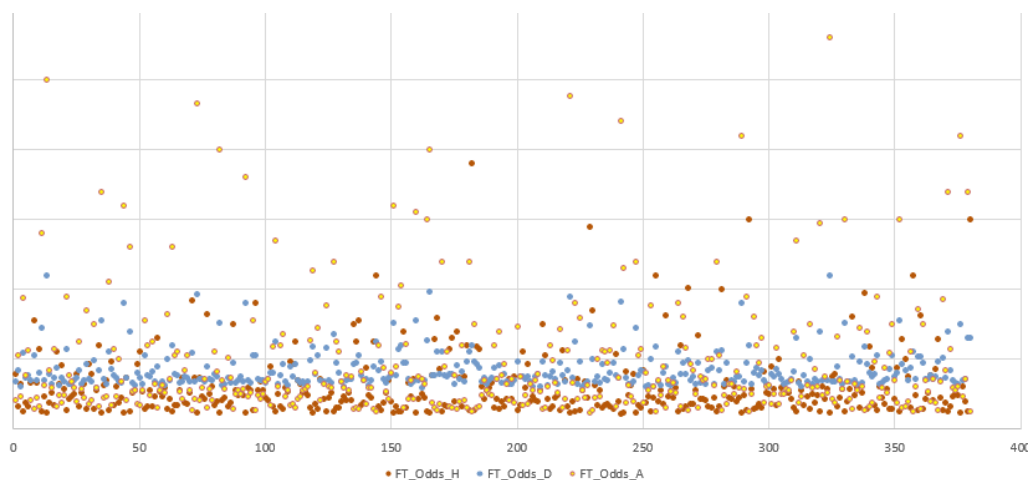
A aplicação da regressão logística em previsões esportivas visa não apenas prever resultados, mas também melhorar a compreensão dos fatores que influenciam o desempenho das equipes. Isso permite aos analistas e apostadores ajustar estratégias com base em insights dados pelo modelo, buscando maior assertividade nas previsões.

É importante ressaltar que a regressão logística não é uma garantia absoluta de precisão nas previsões, uma vez que os resultados esportivos são influenciados por variáveis imprevisíveis. Contudo, seu uso sistemático e análise contínua dos dados podem fornecer vantagens significativas na tomada de decisões no contexto das apostas esportivas, dessa forma, a aplicação da regressão logística nas previsões esportivas se revela como uma ferramenta valiosa, fornecendo insights aprofundados e embasamento analítico para os envolvidos nesse campo, promovendo uma abordagem mais informada e estratégica na compreensão e antecipação de resultados em eventos esportivos.

### 3 RESULTADOS

No nosso gráfico de dispersão, realizamos a seguinte análise: as odds do mandante, as odds do visitante e as odds do empate são apresentadas como variáveis que geram vários pontos no gráfico. Observamos que quanto mais próximo do eixo X, maior é a tendência de vitória, enquanto que quanto mais distante, maior é a tendência de derrota. Os dados foram coletados de 188 jogos do campeonato da Premier League, conforme apresentado na Figura 4.

**Figura 6 - Gráfico de Dispersão**



Fonte: SANTOS, Diego. Print retirado do Excel desenvolvido pela equipe, 2023.



Como foi observado a análise da Figura 5, os times da casa em sua maioria ganham, destacando a tendência de que as equipes que jogam em seus próprios estádios geralmente têm uma vantagem competitiva. Também foi reconhecido que há uma porcentagem de empates. Isso significa que, apesar da tendência de vitória para os times da casa e de derrota para os times visitantes, uma parcela dos jogos termina em empate.

Foi desenvolvida simulações utilizando o Colab para auxiliar na análise dos resultados em um modelo de previsão binária, entra-se com os valores das Odds do time mandante, inseridas através de 7 inputs, determinando assim 2 tipos de resultados diferentes, o primeiro resultado possível utilizado pelo algoritmo foi “Fazer Back Mandante” pois o algoritmo através dos resultados anteriores e concatenações determinou essa melhor escolha visto na Figura 6, ou “Ficar Fora desse Jogo” tendo como segunda resposta determinada pelos dados treinados, visto na Figura 7.

**Figura 7 - Fazer Back Mandante**

```
# Verificando um Jogo Especifico
FT_Odds_H = input("Digite a Odd do Mandante: ")
FT_Odds_A = input("Digite a Odd do Visitante: ")
FT_Odds_D = input("Digite a Odd do Empate: ")
FT_Odds_Over05 = input("Digite a Odd do Over05: ")
FT_Odds_Over15 = input("Digite a Odd do Over15: ")
FT_Odds_Over25 = input("Digite a Odd do Over25: ")
FT_Odds_BTTS_Yes = input("Digite a Odd do BTTS: ")

Odds = [FT_Odds_H,FT_Odds_A,FT_Odds_D,FT_Odds_Over05,FT_Odds_Over15,FT_Odds_Over25,FT_Odds_BTTS_Yes]

# Transformando a Lista em uma Matriz
odds_jogo = np.array(Odds)

previsao_jogo_knn = odds_jogo.reshape(1, -1)
previsao_jogo_knn = knn_dados.predict(previsao_jogo_knn.astype(float))

if previsao_jogo_knn == 1:
    print('');print("Fazer Back Mandante")
else:
    print('');print("Ficar de Fora desse jogo!")

Digite a Odd do Mandante: 1.65
Digite a Odd do Visitante: 4.5
Digite a Odd do Empate: 4.25
Digite a Odd do Over05: 1.06
Digite a Odd do Over15: 1.5
Digite a Odd do Over25: 2.7
Digite a Odd do BTTS: 1.5

Fazer Back Mandante
```

*Fonte: BIALUZ, Print retirado do Colab desenvolvido pela equipe, 2023.*

**Figura 8 - Ficar Fora desse Jogo**

```
# Verificando um Jogo Especifico
FT_Odds_H = input("Digite a Odd do Mandante: ")
FT_Odds_A = input("Digite a Odd do Visitante: ")
FT_Odds_D = input("Digite a Odd do Empate: ")
FT_Odds_Over05 = input("Digite a Odd do Over05: ")
FT_Odds_Over15 = input("Digite a Odd do Over15: ")
FT_Odds_Over25 = input("Digite a Odd do Over25: ")
FT_Odds_BTTS_Yes = input("Digite a Odd do BTTS: ")

Odds = [FT_Odds_H,FT_Odds_A,FT_Odds_D,FT_Odds_Over05,FT_Odds_Over15,FT_Odds_Over25,FT_Odds_BTTS_Yes]

# Transformando a Lista em uma Matriz
odds_jogo = np.array(Odds)

previsao_jogo_knn = odds_jogo.reshape(1, -1)
previsao_jogo_knn = knn_dados.predict(previsao_jogo_knn.astype(float))

if previsao_jogo_knn == 1:
    print('');print("Fazer Back Mandante")
else:
    print('');print("Ficar de Fora desse jogo!")

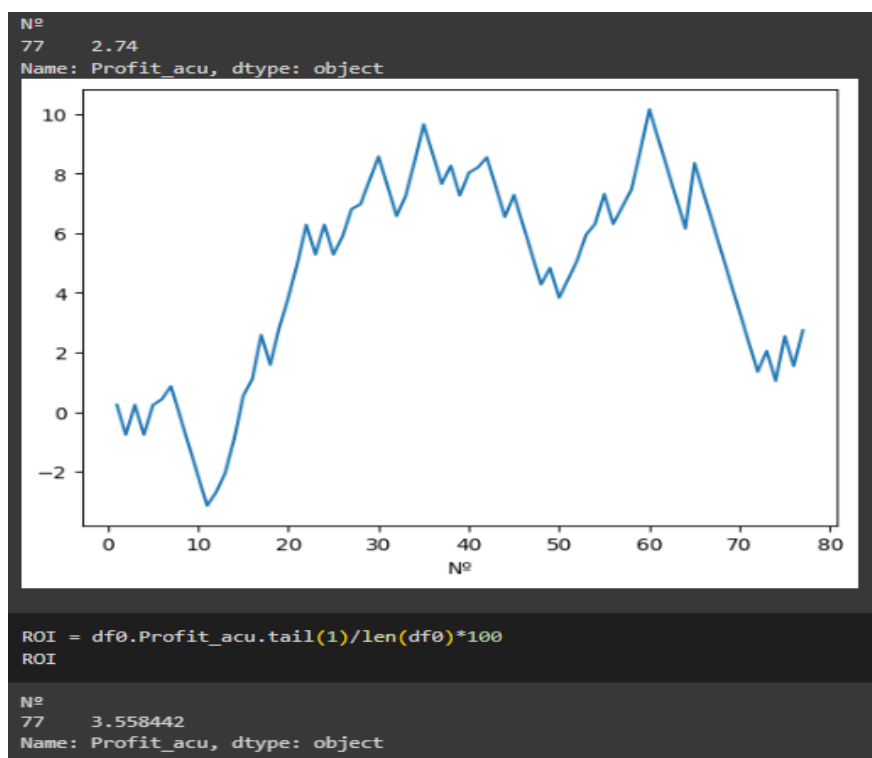
Digite a Odd do Mandante: 6.0
Digite a Odd do Visitante: 1.48
Digite a Odd do Empate: 4.5
Digite a Odd do Over05: 1.55
Digite a Odd do Over15: 4
Digite a Odd do Over25: 9.5
Digite a Odd do BTTS: 1.78

Ficar de Fora desse jogo!
```

*Fonte: BIALUZ, Print retirado do Colab desenvolvido pela equipe, 2023.*

Uma análise do ROI das apostas realizadas com base nas previsões do modelo será conduzida. Segundo o Doe (2023). O ROI (*Return on Investment*), quando aplicado a apostas, é uma métrica que avalia o desempenho financeiro de um apostador ao longo do tempo. Em termos simples, o ROI das apostas é uma medida que indica o retorno financeiro que um apostador obteve com base em seus investimentos em apostas esportivas. Em: Revista de Análise de Apostas Esportivas, 12(4), 112-125. Esta métrica crítica medirá a eficácia global das estratégias de apostas em termos de retorno financeiro mostrado na figura 8.

**Figura 9 - ROI**



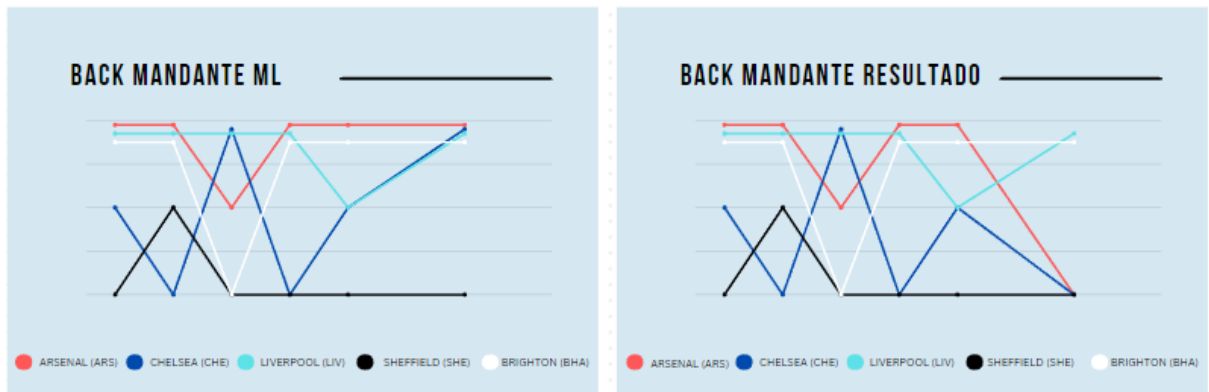
Fonte: BIALUZ, Print retirado do Colab desenvolvido pela equipe, 2023.

Uma Taxa de Retorno sobre o Investimento (ROI) de 2.74 significa que, em média, para cada unidade de moeda (ou unidade de aposta) investida nas suas apostas esportivas, você está obtendo um retorno de 2.74 unidades de moeda. Em termos percentuais, isso equivale a um ROI de 274% se as apostas se concretizarem. Refletindo um desempenho relevante nas apostas, onde nossos ganhos superaram de forma sutil os investimentos.

Apostas simuladas foram criadas no site Bet365 colocando assim em pratica o nível de conhecimento e treinamento do algoritmo, das 5 apostas criadas, foi obtido sucesso em 3 delas, onde nos jogos Liverpool x West Ham e Brighton x Bournemouth o algoritmo orientou a fazer Back Mandante e realizado assim o sucesso com o resultado, na partida entre Sheffield x Newcastle o algoritmo orientou a ficar de fora desse jogo e continuou com 100% de aproveitamento pois com o “Não fazer back mandante” então acertou o resultado, porém 2 das 5 partidas o algoritmo falhou, reduzindo assim a porcentagem de acerto, onde de 100% caiu para 60%, constatando a acurácia do algoritmo, que ficou em 59,77%. No entanto foi considerado a análise das apostas que foi representado no gráfico Back Mandante disponibilizado na Figura 9.



**Figura 10 – Resultados das partidas**



*Fonte: BIALUZ, Print retirado do Canva desenvolvido pela equipe, 2023.*

Foi obtido a seguinte análise dos times: Liverpool, Sheffield e Brighton estavam em uma sequência de acertos, enquanto as previsões do nosso algoritmo continuaram. No entanto, Arsenal e Chelsea não continuaram com seu trajeto, o que resultou na queda da porcentagem de acerto e gerou uma variação no gráfico.

#### **4 ALTERANDO OS VALORES DE VIZINHANÇA NO KNN**

Nos testes realizados, ao considerar o aumento do valor de K no algoritmo KNN, são explorados os benefícios e malefícios dessa alteração. A ampliação de K para valores maiores, como 10 ou 15, é submetida a uma análise minuciosa.

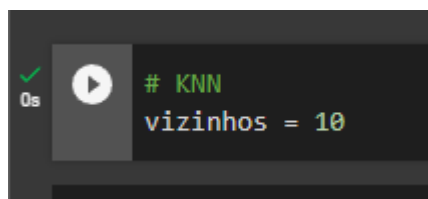
Os benefícios potenciais incluem a suavização das fronteiras de decisão, permitindo uma visão mais abrangente do conjunto de dados e uma redução do impacto de ruídos ou outliers. Entretanto, os malefícios associados a esse aumento de K residem na perda de detalhes relevantes nos dados, resultando em uma simplificação excessiva das fronteiras de decisão.

Portanto, o incremento de K é crucial ponderando cuidadosamente entre os benefícios de suavização e estabilidade versus os malefícios da perda de detalhes importantes, buscando um equilíbrio que melhor se adeque ao contexto e à natureza dos dados em análise.

##### **4.1 KNN = 10**

Com treinamentos e testes desempenhados temos o  $K = 10$ , o algoritmo considera um número maior de vizinhos para tomar decisões, ocorrendo assim, fronteiras de decisões mais suaves, pois o modelo considerará mais pontos classificando novos dados, ilustrado na figura 11.

**Figura 11 - K 10**



*Fonte: BIALUZ, Print retirado do código no Colab desenvolvido pela equipe, 2023.*

Tendo em vista que aumentar o K se reduz a influência de pontos individuais e também possíveis ruídos no conjunto de dados, considerando mais vizinhos, o impacto de pontos atípicos ou mal rotulados é minimizado, exibido na figura 12.

**Figura 12 - Treinando KNN = 10**

```
# Definições de Parâmetros e dados de entrada e resultado para treinamento
knn_dados = KNeighborsClassifier(n_neighbors=vizinhos, metric="minkowski", p=2)
knn_dados.fit(x_dados_treinamento, y_dados_treinamento)
```

▼ KNeighborsClassifier

KNeighborsClassifier(n\_neighbors=10)

Fonte: BIALUZ, Print retirado do código no Colab desenvolvido pela equipe, 2023.

Por outro lado, ocorre um aumento excessivo de K levando a uma perda de detalhes. Valores maiores de K podem não capturar padrões mais específicos nos dados, resultando assim, em um modelo de média-precisão. Aumentando-se o K implica-se também em um aumento computacional, pois o algoritmo precisa calcular as distâncias para um número maior de vizinhos.

#### 4.2 KNN = 15

Seguindo a mesma etapa de treinamento do KNN = 15 obtém-se resultados parecidos, mas na perspectiva das apostas esportivas, onde se acredita que um valor menor de K (como 3) seja ideal, aumentar K para 15 resulta em uma suavização extrema das decisões de classificação disponibilizados na figura 13.

**Figura 13 - K 15**

✓ 0s # KNN  
vizinhos = 15

Fonte: BIALUZ, Print retirado do código no Colab desenvolvido pela equipe, 2023.

Levando assim a um modelo menos preciso, diminuindo detalhes importantes nos quais é crucial na previsão precisa dos resultados esportivos.

Obtendo também os parâmetros e dados de entrada e resultado nos treinamentos do KNN = 15, ilustrados na figura 14.

**Figura 14 - Treinando KNN = 15**

```
# Definições de Parâmetros e dados de entrada e resultado para treinamento
knn_dados = KNeighborsClassifier(n_neighbors=vizinhos, metric="minkowski", p=2)
knn_dados.fit(x_dados_treinamento, y_dados_treinamento)
```

▼ KNeighborsClassifier

KNeighborsClassifier(n\_neighbors=15)

Fonte: BIALUZ, Print retirado do código no Colab desenvolvido pela equipe, 2023.

### 4.3 TAXA DE ACERTO

Ao aumentar o valor de K no algoritmo KNN, observa-se um notável aumento na taxa de acerto, essa mudança pode ser atribuída a uma consideração mais ampla dos vizinhos ao tomar decisões de classificação. Com um K maior, o modelo considera uma quantidade mais significativa de vizinhos para prever novos dados, permitindo uma interpretação mais abrangente dos padrões subjacentes nos dados de treinamento. Essa ampliação no número de vizinhos considerados parece permitir ao modelo capturar informações mais representativas e relevantes para a classificação precisa de novos exemplos, resultando em um aumento direto na acurácia das previsões. Essa mudança indica a importância do ajuste do parâmetro K na otimização do desempenho do modelo KNN para esse conjunto de dados específico, evidenciados e representados nas imagens a seguir, Figuras 15, 16, 17 e 18.

**Figura 15 - Acurácia de acerto K = 10**

```
# taxa de acerto
accuracy_score(y_dados_teste, previsoes_knn)

0.6063829787234043
```

Fonte: BIALUZ, Print retirado do código no Colab desenvolvido pela equipe, 2023.

**Figura 16 - Taxa de acerto K = 10**

```
# calculando a taxa de acerto
taxa_de_acerto = (confusion_matrix(y_dados_teste, previsoes_knn)[1][1]/jogos_entrada)*100
taxa_de_acerto

58.06451612903226
```

Fonte: BIALUZ, Print retirado do código no Colab desenvolvido pela equipe, 2023.

**Figura 17 - Acurácia de acerto K = 15**

```
# taxa de acerto
accuracy_score(y_dados_teste, previsoes_knn)

0.6223404255319149
```

Fonte: BIALUZ, Print retirado do código no Colab desenvolvido pela equipe, 2023.

**Figura 18 - Taxa de acerto K = 15**

```
# calculando a taxa de acerto
taxa_de_acerto = (confusion_matrix(y_dados_teste, previsoes_knn)[1][1]/jogos_entrada)*100
taxa_de_acerto

58.9041095890411
```

Fonte: BIALUZ, Print retirado do código no Colab desenvolvido pela equipe, 2023.

## 5 CONSIDERAÇÕES FINAIS

No início, o projeto pareceu complexo em relação à coleta de dados, que requer atenção especial. Ao entrar em contato com os dados, a melhor abordagem é extrair uma análise mais aprofundada do Data frame, com o objetivo de aumentar a probabilidade de acerto do algoritmo. Em termos de viabilidade do projeto, o trabalho é satisfatório, uma vez que o algoritmo está sendo executado e treinado corretamente, apresentado resultados com uma porcentagem de acerto de 59,77%. Testes reais foram realizados e alterações no próprio código fonte foram realizados, visando assim, a curiosidade pelo modelo KNN e suas variações na taxa de acerto e nas acurácias com impressionantes 62%, 2.23% a mais do que o modelo testado e apresentado nesse artigo.

Para maximizar o potencial do projeto, é interessante realizar atualizações constantes dos dados, uma vez que um jogador pode ou não continuar sua carreira na temporada atual, além disso, considerar a inclusão de dados secundários para contribuir em uma precisão ainda maior nas previsões.

## REFERÊNCIAS

Andreas C. Müller, S. G., 2016. *Introduction to Machine Learning with Python*. Gravenstein Highway North, Sebastopol, CA: O'Reilly Media.

Field, A., 2009. *Descobrendo a Estatística usando o SPSS*. 2ª Edição ed. Porto alegre: ARTMED.

Géron, A., 2019. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow, Concepts, Tools, and Techniques to Build Intelligent Systems*. Gravenstein Highway North, Sebastopol, CA: O'Reilly Media.

Grus, J., 2016. *Data Science do Zero*. Rio de Janeiro: Alta Books.

Joseph F Hair Jr, W. C. B. B. J. B. R. E. A. R. L. T., 2009. *Analise Multivariada de Dados*. Porto Alegre: Bookman.

Caroline Damaceno, G. C. B. M. J. R. P. P. V. S. d. C. R. S. D., 2023. *app.fiepr.org.br*. [Online] Available at:

<http://app.fiepr.org.br/revistacientifica/index.php/conhecimentointerativo/article/view/801/698>

[Acesso em 23 Setembro 2023].

BET365. Esportes. Disponível em:. Acesso em:22/09/2023.

GRAND VIEW RESEARCH. Sports Betting Market Size, Share & Trends Analysis Report By Platform (Online, Offline), By Type (Fixed Odds Wagering, eSports Betting), By Sports Type (Football, Basketball), By Region, And Segment Forecasts, 2023.

futpythontrader. (20 de Setembro de 2023). *futpythontrader*. Fonte: GitHub: <https://github.com/futpythontrader/YouTube>

Bataglin, B. (13 de Novembro de 2023). *goal.com*. Fonte: GOAL +18 | Conteúdo Comercial | : <https://www.goal.com/br/apostas/como-calculador-odds/blt1429b102bdb4b026>