



Análise de Apostas Esportivas no Futebol

A proposta consiste em uma análise de apostas esportivas no futebol, utilizando ciência de dados e modelos preditivos para identificar padrões que influenciam o sucesso das apostas. A partir de dados como valores investidos, odds, ganhos e resultados, buscamos classificar apostas em bem-sucedidas e malsucedidas, transformando dados brutos em insights estratégicos.

Sumário

1. Apresentação do Grupo	03
2. Premissas do Projeto	04
3. Definição do Produto Analítico	07
4. Manipulação e Visualização de Dados	13
5. Análise Exploratória de Dados com Python	18
6. Conclusões Obtidas na Análise Exploratória de Dados Python	20
7. Modelo de Aprendizado de Máquina Para Apostas no Futebol	21
8. Método de Avaliação do Modelo de Aprendizado de Máquina Para Apostas no Futebol	25
9. Conclusões Gerais	27
10. Acesso ao Projeto e todos os arquivos no GitHub	28

Apresentação do Grupo

O grupo é formado por quatro integrantes, cada um com responsabilidades específicas para realizar uma análise dos momentos mais favoráveis para apostas em futebol, baseada no desempenho histórico de diferentes jogos.

I

**Bruno Luigi dos Santos
Tobias**

Função: Revisor e Documentador



2

Guilherme Soares Frota

Função: Visualização de Dados

3

**Thábata Cristina Seixas
Balbino**

Função: Banco de Dados e Análise Estatística

4

**Rayan Christofer Gomes da
Silva**

Função: Banco de Dados e Análise Estatística

Premissas do Projeto



Definição de empresa e área de atuação

A empresa responsável pelos dados é a "Game Winner", uma plataforma de apostas esportivas que, embora hoje inativa, forneceu informações valiosas para análise de apostas em futebol. Esse mercado cresce com a popularidade do esporte e avanços tecnológicos, permitindo aos apostadores usar dados para estratégias mais informadas. O projeto explora esses dados para entender melhor os padrões e oportunidades no mercado de apostas.

Premissas do Projeto

Apresentação dos dados

O dataset "Analisando Apostas de Futebol" contém registros detalhados de apostas feitas em jogos de futebol pela empresa "Game Winner" durante 2022. No arquivo "futebol.csv", cada coluna fornece dados específicos:

- **Data:** Indica a data da partida.
- **Jogo:** Nome das equipes que se enfrentaram, facilitando a identificação do jogo.
- **Minutos:** Momento específico (em minutos) da partida em que a aposta foi realizada ou um evento ocorreu.
- **Investimento:** Valor apostado pelo usuário, em reais.
- **Odd:** Multiplicador para cálculo do ganho se a aposta for vencedora.
- **Ganho:** Retorno líquido ao apostador caso ganhe (valor do investimento descontado).
- **Resultado:** Status da aposta, sendo "Green" para vitória e "Red" para perda.

Data	Jogo	Minutos	Investimento	Odd	Ganho	Resultado
2022-07-01	Inglaterra x Israel	60 ao 70	10	1.57	5.700000000000001	Green
2022-07-01	Ayacucho x Cantolao	10 ao 20	10	1.5	5.0	Green
2022-07-01	Ayacucho x Cantolao	20 ao 30	10	1.5	5.0	Green
2022-07-01	Lara x UCV AC. 0.5 GOLS	Gols	10	1.61	6.100000000000001	Red
2022-07-01	Guairena x Asuncion	20 ao 30	10	1.57	5.700000000000001	Green
2022-07-01	Cordoba x Juniors	10 ao 20	10	1.57	5.700000000000001	Green
2022-07-01	Belgrano x Almagro	20 ao 30	10	1.61	6.100000000000001	Green
2022-07-01	Belgrano x Almagro	60 ao 70	5	1.53	2.6500000000000004	Green
2022-07-01	Mar x Colo Colo	60 ao 70	5	1.5	2.5	Green
2022-07-01	Cruzeiro x Vila	10 ao 20	5	1.5	2.5	Green
2022-07-01	Col x Sac	10 ao 20	5	1.57	2.8500000000000005	Green
2022-07-01	Brusque x Operario	60 ao 70	5	1.5	2.5	Green
2022-07-02	Huachipate x Unido	20 ao 30	10	1.61	6.100000000000001	Green
2022-07-04	Braga x Botafogo	10 ao 20	10	1.44	4.399999999999999	Green
2022-07-04	Colorado x Austin	00 ao 10	10	1.53	5.300000000000001	Green
2022-07-04	Dallas x Inter	00 ao 10	10	1.57	5.700000000000001	Green

A Figura 1 apresenta uma amostra do dataset "Futebol.csv", que ilustra a estrutura dos dados para ajudar a equipe e interessados a compreenderem as variáveis analisadas. Esse entendimento facilitará a identificação de padrões e insights para definir momentos estratégicos para apostas em jogos de futebol.

Premissas do Projeto

Objetivos e Metas

O **objetivo** deste projeto é analisar o dataset "Analisando Apostas de Futebol" para identificar padrões e momentos ideais para apostas em partidas de futebol. Os principais objetivos incluem:

- Análise dos dados históricos para entender tendências de sucesso nas apostas,
- Identificação de momentos estratégicos no jogo para apostas mais bem-sucedidas,
- Avaliação da performance das apostas, relacionando odds, investimentos e resultados.

As **metas** específicas são: realizar uma análise exploratória dos dados, desenvolver um modelo preditivo, criar visualizações informativas, documentar o processo e preparar uma apresentação final. Esses objetivos e metas orientam o projeto para uma análise aprofundada e conclusiva sobre apostas em futebol.

Definição do produto analítico

A Etapa 2 do projeto tem como objetivo definir as diretrizes para a análise de dados, incluindo os métodos e ferramentas a serem utilizados. A análise do database "futebol.csv" será realizada em Python, utilizando suas bibliotecas, métodos estatísticos e modelos de aprendizado de máquina. Os tópicos dessa seção são apresentados a seguir.

Bibliotecas e Repositório GitHub

Para a execução colaborativa do trabalho, foram escolhidas as seguintes bibliotecas Python:

Pandas: Manipulação e análise de dados em DataFrames.

NumPy: Operações numéricas e manipulação de arrays para cálculos estatísticos.

Matplotlib e Seaborn: Visualização de dados por meio de gráficos informativos.

Scikit-learn: Aplicação de métodos de aprendizado de máquina e divisão de dados para modelos preditivos.

Statsmodels: Análise estatística profunda e construção de modelos de regressão.

Um repositório no GitHub, chamado

"**PROJ_APL_2-AN-LISE-DE-APOSTAS-ESPORTIVAS-NO-FUTEBOL**", será criado para versionamento do código e colaboração entre a equipe, organizando scripts, notebooks Jupyter e documentos de suporte para facilitar o acesso e edição.

Definição do produto analítico

DEFINIÇÃO DA BASE DE DADOS E ANÁLISE EXPLORATÓRIA

A base de dados utilizada será o arquivo "futebol.csv", que contém informações sobre apostas em partidas de futebol. A análise será realizada em Python, empregando suas bibliotecas e métodos estatísticos para extrair insights. A análise exploratória inicial incluirá:

- **Estatísticas descritivas:** Cálculo de métricas como média, mediana, desvio padrão e contagens de valores únicos.
- **Distribuição dos dados:** Visualização dos investimentos, odds e ganhos com histogramas e boxplots.
- **Correlação entre variáveis:** Avaliação das relações entre variáveis, como investimento e ganho, usando gráficos de dispersão e matrizes de correlação.



Definição do produto analítico

Tratamento da base de dados (Preparação e Treinamento)

O tratamento da base de dados incluirá as seguintes etapas:

- ❑ **Remoção de entradas ausentes:** Excluir entradas com valores faltantes para garantir a integridade dos dados.
- ❑ **Conversão de tipos de dados:** Ajustar as colunas para o formato correto, como converter a coluna 'Data' para o tipo datetime.
- ❑ **Normalização:** Normalizar os dados, se necessário, para facilitar a comparação entre variáveis.
- ❑ **Divisão dos dados:** Separar a base de dados em conjuntos de treino e teste para validação do modelo.



Definição do produto analítico

Bases Teóricas dos Métodos Analíticos

A análise estatística preditiva será fundamentada nas seguintes abordagens teóricas:

- **Modelos de Regressão:** A regressão linear será utilizada para entender a relação entre as variáveis de interesse (como investimento e ganho) e prever resultados com base em novas entradas.
- **Análise de Classificação:** Métodos como regressão logística poderão ser aplicados para classificar se uma aposta resultará em sucesso ou falha com base nas características do jogo.
- **Avaliação de Desempenho:** A validação cruzada será utilizada para avaliar a eficácia dos modelos, garantindo que eles se generalizem bem para dados não vistos.

Definição do produto analítico

Cálculo da Acurácia – Parte I

A acurácia dos modelos será avaliada em percentual pela fórmula:

$$Acuracia (\%) = \left(\frac{\text{Numero de previsoes corretas}}{\text{total de resultados reais}} \right) \times 10$$

Onde:

➤ **Número de Previsões Corretas:** Soma dos verdadeiros positivos e verdadeiros negativos.

Definição do produto analítico

Cálculo da Acurácia – Parte 2

O desvio padrão será calculado para complementar a análise das taxas de acertos e erros, indicando a dispersão das previsões em relação à média.

$$\text{Desvio Padro} = \sqrt{\frac{\sum (xi - \underline{x})^2}{N}}$$

Onde:

Manipulação e Visualização de Dados

As bibliotecas Pandas, Seaborn e Matplotlib foram importadas para manipulação e visualização dos dados. O dataset foi carregado em um DataFrame e padronizado, incluindo a remoção de espaços extras nos nomes das colunas para garantir uniformidade.

Figura 1: dataset “Futebol.csv” carregado em um dataframe.

	Data	Jogo	Minutos	Investimento	Odd	Ganho	Resultado
0	2022-07-01	Inglaterra x Israel	60 ao 70	10	1.57	5.7	Green
1	2022-07-01	Ayacucho x Cantolao	10 ao 20	10	1.50	5.0	Green
2	2022-07-01	Ayacucho x Cantolao	20 ao 30	10	1.50	5.0	Green
3	2022-07-01	Lara x UCV AC. 0,5 GOLS	Gols	10	1.61	6.1	Red
4	2022-07-01	Guairena x Asuncion	20 ao 30	10	1.57	5.7	Green
...
1559	2022-10-31	Verona x Roma	Gols	10	4.33	33.3	Green
1560	2022-10-31	Elfsborg x Helsingborg	Gols	20	1.50	10.0	Red
1561	2022-10-31	Verona x Roma	Gols	10	1.50	5.0	Red
1562	2022-10-31	Midtjylland x Odense	Gols	10	1.50	5.0	Red
1563	2022-10-31	Ceará x Fluminense	Gols	20	1.50	10.0	Red

1564 rows x 7 columns

Figura 2: Informações do dataframe.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1564 entries, 0 to 1563
Data columns (total 7 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Data            1564 non-null  object
1   Jogo            1564 non-null  object
2   Minutos        1473 non-null  object
3   Investimento    1564 non-null  int64
4   Odd             1564 non-null  float64
5   Ganho          1564 non-null  float64
6   Resultado       1564 non-null  object
dtypes: float64(2), int64(1), object(4)
memory usage: 85.7+ KB
```


Manipulação e Visualização de Dados

A coluna "Data" foi convertida para o formato datetime para facilitar análises temporais, e "Jogo" foi padronizada para minúsculas e sem espaços extras para simplificar a identificação de duplicatas. Valores nulos na coluna "Minutos" foram preenchidos como "desconhecido", e termos inconsistentes foram ajustados para uniformidade. Registros duplicados foram consolidados somando os valores de "Investimento" e "Ganho" e eliminando redundâncias, preservando a integridade dos dados financeiros.

Figura 3: Processo de padronização e limpeza dos dados.

	Data	Jogo	Minutos	Investimento	Odd	Ganho	Resultado	
412	2022-08-05	guarani x gremio	desconhecido	10	1.50	5.0	Green	
431	2022-08-06	botafoogo x ceará	desconhecido	10	1.50	5.0	Green	
458	2022-08-07	danubio x maldonado	desconhecido	10	1.50	5.0	Green	
670	2022-08-22	arg jrs x patronato	50 a 60	10	1.53	5.3	Green	
752	2022-08-26	sagan tusu x fukuoka	20 a 30	10	1.66	6.6	Green	
1214	2022-10-05	flamengo x inter	20 a 30	10	1.50	5.0	Green	
1303	2022-10-11	copenhagen x city	20 a 30	10	1.57	5.7	Green	
1306	2022-10-11	milan x chelsea	00 a 10	10	1.66	6.6	Green	
1307	2022-10-11	psg x benfica	00 a 10	10	1.61	6.1	Green	
1308	2022-10-11	dortmund x sevilla	20 a 30	10	1.61	6.1	Green	
1309	2022-10-11	wigan x blackburn	60 a 70	40	1.50	20.0	Red	
1310	2022-10-11	patriotas x dep. pasto	00 a 10	10	1.66	6.6	Green	
1311	2022-10-11	patriotas x dep. pasto	20 a 30	10	1.57	5.7	Green	
1321	2022-10-11	la equidad x cortulua	60 a 70	10	1.50	5.0	Green	
1325	2022-10-11	ferro carril x almagro	60 a 70	10	1.50	5.0	Green	
1426	2022-10-20	blooming x palmaflor	gols	10	1.50	5.0	Red	
1435	2022-10-21	river x torque	gols	10	1.50	5.0	Red	
1440	2022-10-21	brugge x challeroi	gols	10	3.40	24.0	Green	
1497	2022-10-26	tottenham x sporting	gols	10	1.50	5.0	Red	
1500	2022-10-26	univ. de chile x u. española	gols	20	1.50	10.0	Red	
1517	2022-10-27	vasco x sampaio	gols	10	1.50	5.0	Red	
1531	2022-10-28	magdeburg x heidenheim	gols	10	1.50	5.0	Red	

Para simplificar a análise, valores de "Minutos" com menos de oito ocorrências foram agrupados como "outros", exceto o período "80 a 90", que foi mantido separado. Após a padronização, o dataset foi validado e preparado para análise preditiva. A exploração inicial usou value_counts() para examinar a distribuição de "Minutos" e identificar frequências. Gráficos de barras e pizza ilustraram a proporção de resultados ("Green" e "Red") em relação aos investimentos, facilitando a comunicação visual para stakeholders. Correlações entre variáveis como investimento e ganho foram analisadas com Seaborn e Matplotlib. Finalmente, o dataset foi ajustado para modelagem preditiva, com seleção e transformação de variáveis para aprendizado de máquina, resultando em uma base de dados robusta para futuras análises de apostas.

Manipulação e Visualização de Dados

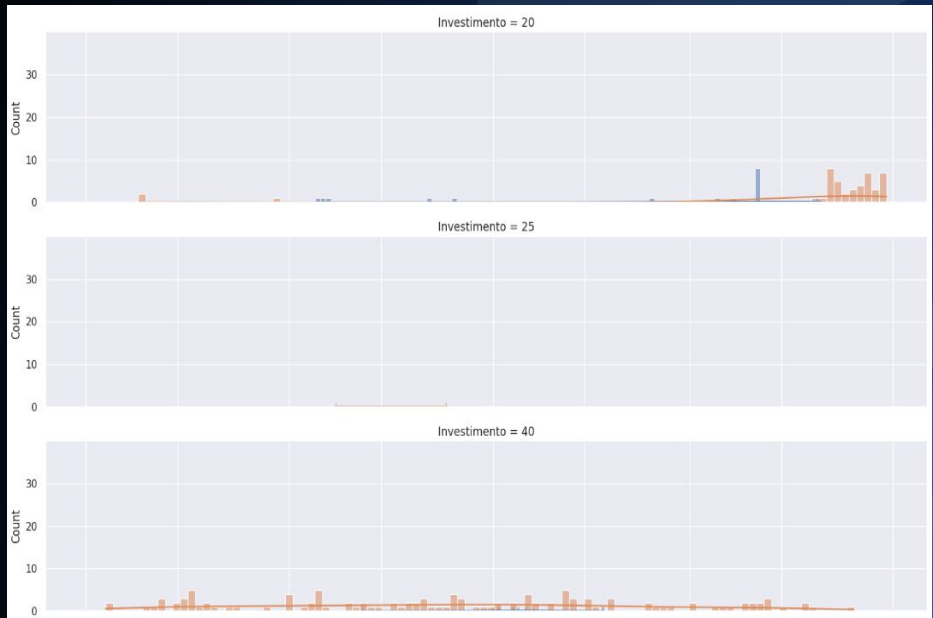


Figura 6: Grade de facetas para os investimentos nos intervalos “20”, “25” e “40”.

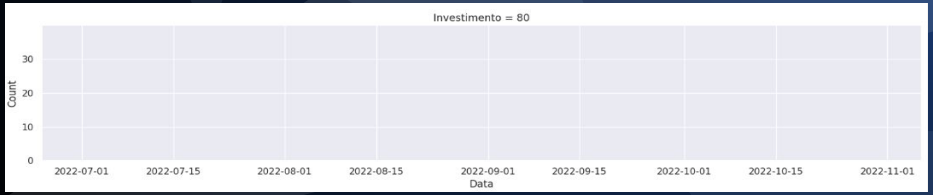


Figura 7: Grade de facetas para os investimentos no intervalo “80”.

Por fim, a análise investigou a relação entre os períodos de aposta (“Minutos”) e os resultados ao longo do tempo, identificando períodos possivelmente ligados a perdas, o que oferece insights para otimizar futuras estratégias de apostas.



Figura 8: Período de apostas e os resultados ao longo do tempo para os intervalos “10 a 20”, “20 a 30” e “60 a 70”.

Manipulação e Visualização de Dados

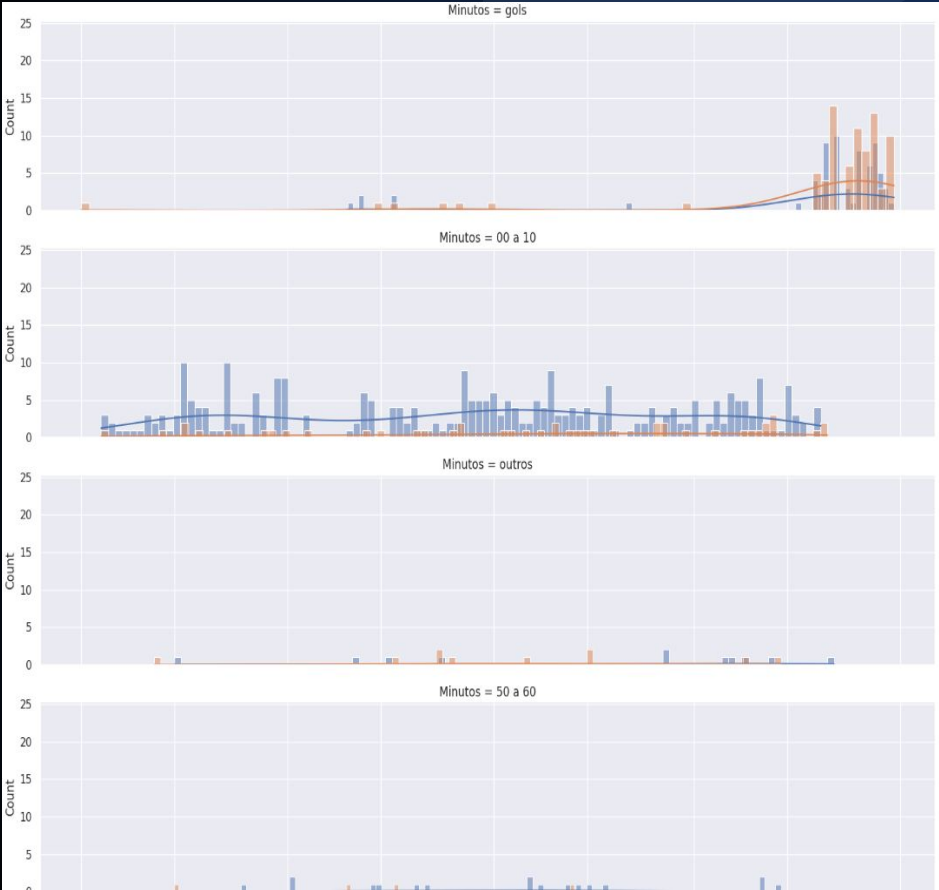


Figura 9: Período de apostas e os resultados ao longo do tempo para os intervalos “Minutos = gols”, “00 a 10”, “Minutos = outros” e “50 a 60”.

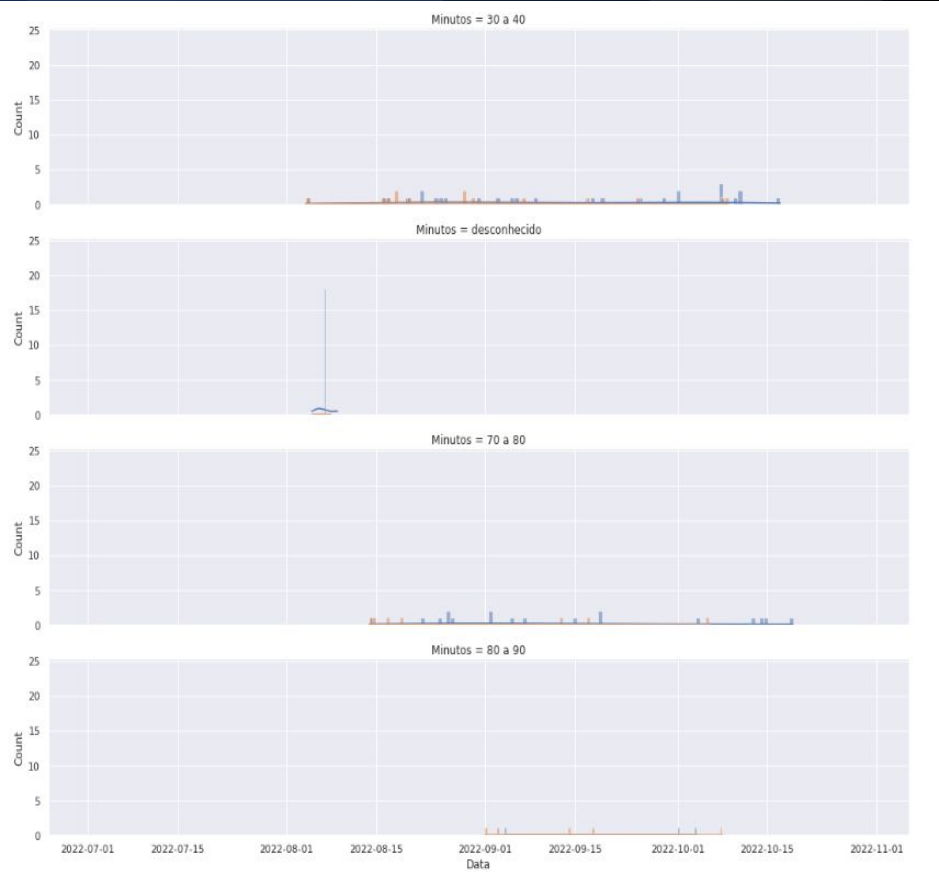


Figura 10: Período de apostas e os resultados ao longo do tempo para os intervalos “30 a 40”, “Minutos = desconhecido”, “70 a 80” e “80 a 90”.

Manipulação e Visualização de Dados

A análise detalhada revelou que apostas de valor 40 estavam majoritariamente ligadas a perdas, enquanto os ganhos se concentravam em apostas de 10, que vinham diminuindo ao longo do tempo. Apostas de 20 também mostraram perdas significativas, sugerindo possíveis falhas nos critérios de investimento.

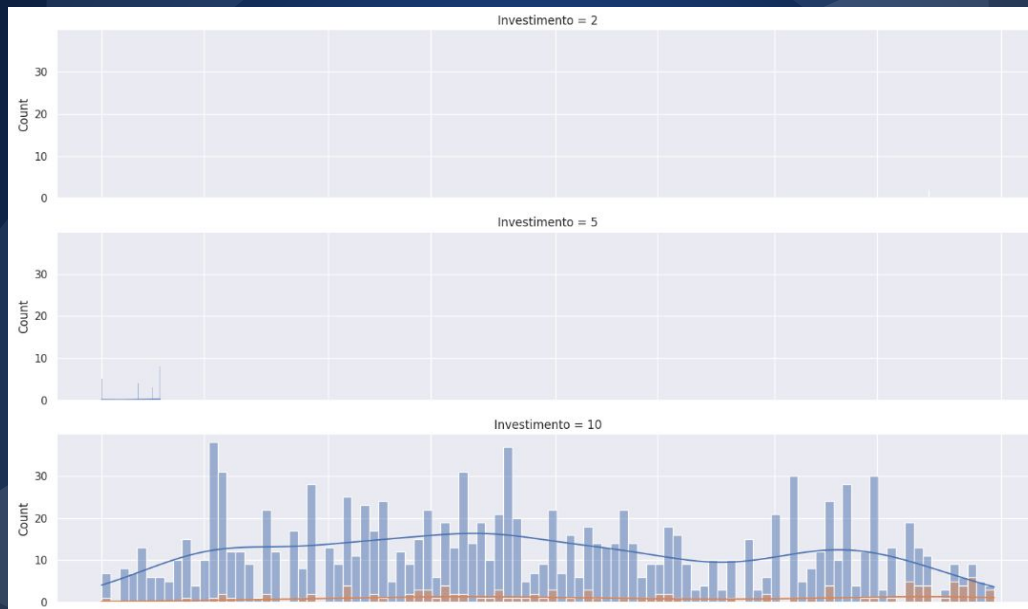


Figura 5: Grade de facetas para os investimentos nos intervalos “2”, “5” e “10”.

Análise Exploratória de Dados com Python

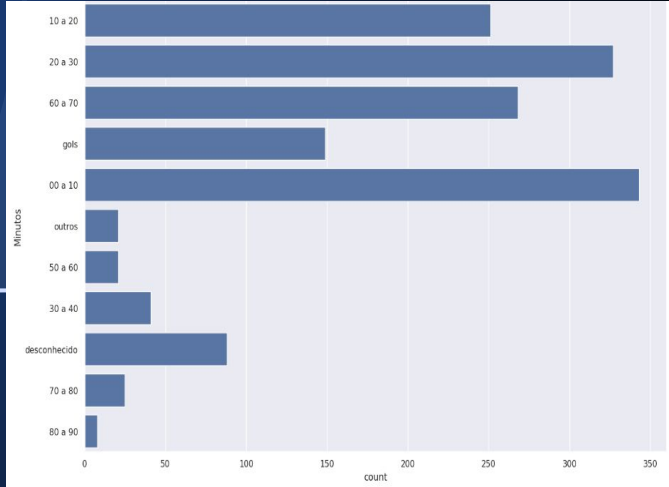
A etapa de Análise Exploratória de Dados (AED) [3] foi iniciada com o objetivo de entender melhor a distribuição dos valores de aposta no conjunto de dados. Ao analisar a coluna "Investimento", observou-se que aproximadamente 86% das apostas eram no valor de 10, evidenciando uma concentração significativa nesse montante.

Figura 1: Distribuição da proporção das apostas em valores definidos.



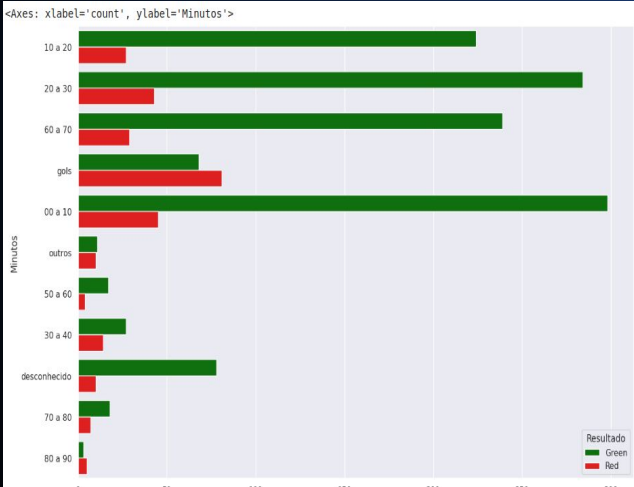
Em seguida, foi realizada uma análise da distribuição dos critérios de "Minutos" utilizando um gráfico de barras, permitindo uma visualização clara da frequência de ocorrências para cada período de aposta.

Figura 2: Gráfico sobre a representação da frequência de ocorrências para cada período de aposta.



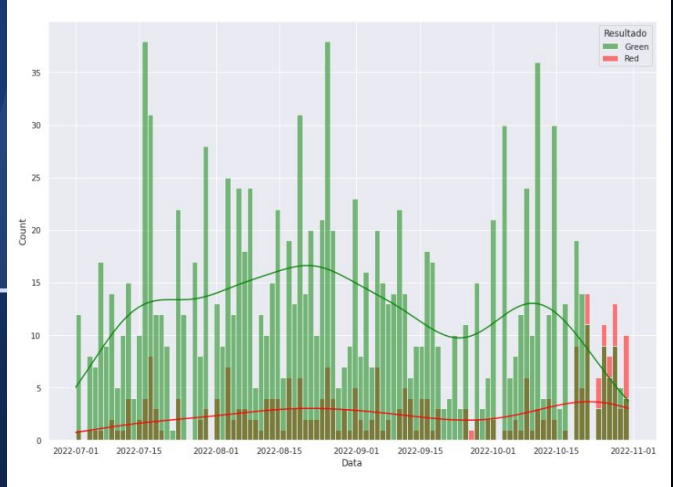
Para aprofundar a análise, as ocorrências de apostas foram subdivididas com base nos resultados obtidos, utilizando cores para distinguir entre resultados positivos ("Green") e negativos ("Red").

Figura 3: Resultados positivos ("Green") e negativos ("Red") nos intervalos de tempo definidos.



A visualização dos resultados ao longo do tempo foi feita através de um histograma, que indicou uma predominância de resultados positivos na maior parte do período analisado. Contudo, notou-se uma recente diminuição nos resultados positivos, acompanhada de um aumento nos resultados negativos.

Figura 4: Histograma dos resultados ao longo do tempo



Conclusões Obtidas na Análise Exploratória de Dados Python



A análise revelou insights importantes sobre o desempenho das apostas esportivas, destacando a predominância de apostas de 10 (86% do total), refletindo uma abordagem conservadora. Apostas maiores, como as de 40, embora raras, indicam momentos de maior confiança. A frequência de apostas foi concentrada em intervalos populares, com categorias menos frequentes agrupadas como "outros", promovendo diversificação. Observou-se um melhor desempenho em apostas de escanteios (resultados "Green"), enquanto apostas em gols tiveram mais perdas (resultados "Red"), indicando uma estratégia de aposta em gols menos eficaz. Um histograma mostrou que, apesar dos ganhos iniciais, a taxa de sucesso caiu nas apostas recentes, especialmente em gols, sugerindo impacto de mudanças no mercado ou na estratégia. Em resumo, a análise da distribuição de valores, da frequência de apostas e do desempenho ao longo do tempo evidencia a importância de ajustar estratégias com base nos dados, maximizando a chance de sucesso futuro nas apostas.

Modelo de Aprendizado de Máquina Para Apostas no Futebol

Este modelo de aprendizado de máquina, baseado na técnica de Regressão Logística, tem como objetivo prever a probabilidade de sucesso ("Green") ou insucesso ("Red") em apostas de futebol. A Regressão Logística é amplamente utilizada para classificação binária, buscando prever uma de duas saídas com base em variáveis de entrada. No projeto, as saídas representam o resultado das apostas, utilizando informações históricas como o minuto do jogo, a odd (cotação) e o valor do investimento para fazer as previsões.

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, recall_score, f1_score, r2_score
import os
```

Figura 1: Bibliotecas utilizadas no modelo de aprendizado de máquina para aposta esportiva



Modelo de Aprendizado de Máquina Para Apostas no Futebol

O processo de modelagem inicia-se com o pré-processamento dos dados de apostas históricas, essencial para garantir que as variáveis estejam no formato adequado. Cada aposta inclui informações como:

Intervalo de minuto do jogo (quando a aposta foi feita),

Odd da aposta (valor oferecido pela casa de apostas),

Valor investido,

Resultado da aposta, classificado como "Green" (sucesso) ou "Red" (insucesso).

Durante o pré-processamento, a coluna do minuto do evento é convertida em valores numéricos, substituindo os intervalos de minutos por uma média para facilitar a análise. Textos irrelevantes, como "Escanteio" ou "Gols", são tratados ou removidos.

```
# Verifique se a coluna 'Minutos' existe após a leitura
if 'Minutos' not in data.columns:
    raise KeyError("A coluna 'Minutos' não foi encontrada no DataFrame.")

# Função para converter a coluna 'Minutos'
def convert_minutos(minuto_str):
    minuto_str = str(minuto_str).strip() # Remover espaços em branco
    if 'a' in minuto_str: # Se for um intervalo
        partes = minuto_str.split(' ao ')
        try:
            return (int(partes[0]) + int(partes[1])) / 2 # Média do intervalo
        except ValueError:
            return np.nan # Retornar NaN para entradas não numéricas
    else:
        try:
            return float(minuto_str) # Se for um número simples
        except ValueError:
            return np.nan # Retornar NaN para entradas não numéricas

# Converter a coluna 'Minutos' para valores numéricos
data['Minutos'] = data['Minutos'].apply(convert_minutos)

# Remover linhas com valores NaN em 'Minutos'
data.dropna(subset=['Minutos'], inplace=True)

# Verifique novamente se a coluna 'Minutos' existe após a conversão
if 'Minutos' not in data.columns:
    raise KeyError("A coluna 'Minutos' não está disponível após a conversão.")
```

Figura 2: Tratamento dos dados para alimentar o modelo



Modelo de Aprendizizado de Máquina Para Apostas no Futebol

Para medir a qualidade das previsões geradas pelo modelo, utilizamos métricas de avaliação comuns em aprendizado de máquina, como:

- **Acurácia:** que indica a proporção de previsões corretas feitas pelo modelo.
- **Recall:** que avalia a capacidade do modelo de identificar corretamente as apostas "Green".
- **F1-Score:** que combina precisão e recall para fornecer uma métrica balanceada, útil em contextos onde é importante ponderar acertos e erros.
- **R²:** que ajuda a entender a variância explicada pelo modelo com base nas variáveis preditoras.

```
# Calcular métricas de avaliação
accuracy = accuracy_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"Acurácia do modelo: {accuracy:.2f}")
print(f"Recall: {recall:.2f}")
print(f"F1-Score: {f1:.2f}")
print(f"R2: {r2:.2f}")
```

Figura 3: Métricas de avaliação

Essas métricas ajudam a avaliar a performance do modelo, oferecendo uma visão clara de sua capacidade de fazer previsões precisas. Na prática, o modelo solicita informações do usuário, como o minuto da aposta, a odd e o valor a ser investido. Com esses dados, ele prevê o potencial de sucesso da aposta e retorna uma recomendação, indicando "Apostar!" ou "Não apostar", com base na probabilidade identificada.

```
Insira as informações para previsão:
Intervalo (Ex: '10 ao 20'): 10 ao 20
Odd: 1.90
Valor do investimento: 10
<ipython-input-3-6e41f720d041>:118: FutureWarning:
  user_records = pd.concat([user_records, pd.
Recomendação: Apostar!
```

Figura 4: Exemplo de saída do modelo de aprendizado

O sistema permite que o usuário compare sua aposta com casos semelhantes no histórico, usando uma margem de 25% para mais ou para menos nos valores de minuto, odd e investimento. Se não forem encontradas apostas semelhantes, o sistema informa que não há dados suficientes para uma recomendação precisa.

```
Resultados das apostas semelhantes que acabaram em Green:
ID: 1, Jogo: Ayacucho x Cantolao , Minutos: 15.0, Odd: 1.5, Investimento: 10, Resultado: Green
ID: 5, Jogo: Cordoba x Juniors , Minutos: 15.0, Odd: 1.57, Investimento: 10, Resultado: Green
ID: 13, Jogo: Braga x Botafogo , Minutos: 15.0, Odd: 1.44, Investimento: 10, Resultado: Green
ID: 19, Jogo: Galaxy x Montreal , Minutos: 15.0, Odd: 1.44, Investimento: 10, Resultado: Green
ID: 46, Jogo: Asane x Kafum , Minutos: 15.0, Odd: 1.5, Investimento: 10, Resultado: Green
ID: 50, Jogo: Aksur x Atyraur , Minutos: 15.0, Odd: 1.66, Investimento: 10, Resultado: Green
ID: 52, Jogo: Chicago x Belgrano , Minutos: 15.0, Odd: 1.61, Investimento: 10, Resultado: Green
ID: 57, Jogo: Dynamo x Borussia , Minutos: 15.0, Odd: 1.53, Investimento: 10, Resultado: Green
```

Figura 5: Modelo de aprendizado retorna apostas positivas (Green) semelhantes ao do usuário

Modelo de Aprendizado de Máquina Para Apostas no Futebol

Resultados das apostas semelhantes que acabaram em Red:

ID: 191, Jogo: Guillermo x Chacarita , Minutos: 15.0, Odd: 1.66, Investimento: 10, Resultado: Red

ID: 688, Jogo: Platense x Talleres, Minutos: 15.0, Odd: 1.5, Investimento: 10, Resultado: Red

Figura 6: Modelo de aprendizado retorna apostas negativas (Red) semelhantes ao do usuário.

Adicionalmente, todos os registros de apostas do usuário são salvos em um banco de dados, criando um histórico que possibilita a evolução contínua do modelo. Com o tempo, isso aprimora a capacidade de previsão do sistema e permite recomendações mais assertivas.

Deseja inserir mais informações? (s/n): n

Registros do usuário:

	Intervalo	Odd	Investimento	Recomendacao
0	10 ao 20	1.9	10.0	Apostar

Figura 7: Banco de dados armazena informações do usuário

Esse modelo é uma ferramenta útil e prática, que combina métodos de aprendizado supervisionado e análise estatística para auxiliar usuários na tomada de decisões em apostas de futebol, integrando a análise de dados históricos para fornecer previsões robustas e embasadas.

Método de Avaliação do Modelo de Aprendizado de Máquina Para Apostas no Futebol

O método de avaliação do modelo de aprendizado de máquina adota uma abordagem abrangente para medir seu desempenho na previsão de apostas. Ele utiliza uma matriz de confusão para analisar classificações corretas e incorretas nas classes positivas e negativas, além de métricas complementares como acurácia, precisão, recall, F1-score e coeficiente de determinação (R^2).

```
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, r2_score, confusion_matrix

def avaliar_modelo(modelo, X_test, y_test):
    """
    Avalia o desempenho do modelo de aprendizado de máquina usando várias métricas.

    Parâmetros:
    modelo (objeto): Modelo treinado de aprendizado de máquina.
    X_test (DataFrame): Conjunto de dados de teste com recursos.
    y_test (Series): Valores reais do conjunto de dados de teste.

    Retorno:
    dict: Dicionário com as métricas de avaliação do modelo.
    """
```

Figura 1: Parâmetros para avaliação de desempenho do modelo de aprendizado de máquina

Primeiramente, a matriz de confusão é utilizada para comparar as previsões do modelo com os valores reais, classificando os resultados em Verdadeiros Positivos, Verdadeiros Negativos, Falsos Positivos e Falsos Negativos, o que ajuda a identificar padrões de erro. Uma alta quantidade de Falsos Positivos pode indicar um otimismo excessivo do modelo, enquanto muitos Falsos Negativos sugerem uma aversão a recomendações. As métricas de acurácia, precisão, recall e F1-score oferecem uma análise quantitativa do desempenho do modelo em identificar apostas positivas, enquanto o coeficiente de determinação (R^2) mede a variância explicada pelo modelo, sendo mais próximo de 1 um indicativo de melhor ajuste aos dados.

```
# Calcular as métricas
acuracia = accuracy_score(y_test, y_pred)
precisao = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
matriz_confusao = confusion_matrix(y_test, y_pred)

# Exibir as métricas
print(f"Acurácia: {acuracia:.2f}")
print(f"Precisão: {precisao:.2f}")
print(f"Recall: {recall:.2f}")
print(f"F1-Score: {f1:.2f}")
print(f" $R^2$ : {r2:.2f}")
print(f"Matriz de Confusão:\n{matriz_confusao}")
```

Figura 2: Métricas para avaliação de desempenho do modelo de aprendizado de máquina

Método de Avaliação do Modelo de Aprendizado de Máquina Para Apostas no Futebol

Conclusões sobre o modelo de aprendizado de máquina

As métricas combinadas oferecem uma análise detalhada do desempenho do modelo, orientando ajustes necessários. Um recall alto e precisão baixa podem indicar que o modelo captura muitos casos positivos, mas com uma alta taxa de falsos positivos, sugerindo a necessidade de ajustes de balanceamento. Assim, o método proporciona uma avaliação robusta que detalha o comportamento do modelo, facilitando o refinamento do algoritmo para alcançar recomendações de apostas mais confiáveis e eficazes.

```
Acurácia: 0.98
Precisão: 0.97
Recall: 1.00
F1-Score: 0.99
R²: 0.74
Matriz de Confusão:
[[ 10   3]
 [  0 115]]
```

Figura 3: Avaliação do modelo de aprendizado de máquina

O modelo de aprendizado de máquina demonstrou um desempenho excepcional na previsão de apostas bem-sucedidas, alcançando uma acurácia de 0.98, uma precisão de 0.97 e um recall perfeito de 1.00. O F1-score de 0.99 reforça sua eficácia em equilibrar a identificação correta de apostas positivas e a minimização de falsos positivos e negativos. No entanto, o R^2 de 0.74 indica que ainda existe uma parte significativa da variabilidade dos resultados que não é explicada pelo modelo, sugerindo oportunidades de aprimoramento ao incluir mais variáveis ou dados adicionais.

A matriz de confusão evidenciou uma predominância de verdadeiros negativos e positivos, mas também indicou que algumas apostas foram incorretamente classificadas como bem-sucedidas. Em conclusão, enquanto o modelo é altamente eficaz na previsão de apostas, a continuidade na análise e ajustes, junto à exploração de novas variáveis, é essencial para aumentar ainda mais sua precisão e capacidade de explicar as variações nos resultados das apostas. A implementação dessas melhorias pode proporcionar recomendações ainda mais assertivas e valiosas para os usuários.

Conclusões Gerais



O projeto de análise e modelagem preditiva de apostas esportivas em futebol revelou importantes conclusões sobre a eficácia de uma abordagem metodológica. A manipulação de dados foi essencial, incluindo limpeza, remoção de duplicatas e categorização de intervalos de minutos, o que garantiu uma base de dados consistente. A análise exploratória indicou uma concentração de apostas em valores de 10, revelando tendências conservadoras e padrões temporais relevantes.

O modelo de Regressão Logística alcançou 98% de acurácia e nenhum falso negativo, evidenciando sua robustez ao prever apostas bem-sucedidas, com as variáveis escolhidas sendo pertinentes. A análise da matriz de confusão e outras métricas indicaram que, embora o modelo seja eficaz, há espaço para melhorar a explicação da variância (R^2 de 0,74), possivelmente ajustando a segmentação dos dados.

Embora o projeto tenha estabelecido uma base analítica funcional e oferecido insights iniciais, é necessário incorporar dados mais complexos, como histórico de jogos e desempenho das equipes, para aumentar a precisão das previsões e a aplicabilidade prática. Assim, o trabalho cumpriu seu objetivo de fundamentar estratégias de apostas, demonstrando o valor da integração entre análise estatística e aprendizado de máquina, e abre caminho para um sistema mais robusto e confiável no futuro.

Acesso ao Projeto e todos os arquivos no GitHub:

GITHUB. Disponível em:

https://github.com/RayanCrhistofer/PROJ_APL_2-AN-LISE-DE-APOSTAS-ESPORTIVAS-NO-FUTEBOL.git

https://colab.research.google.com/drive/148Z0fZCqWdgsLJpVyST4i33wiQZU_wgf?usp=sharing

<https://colab.research.google.com/drive/1Ux5LWEK7Wgj524v3XkCJb0M29Q076a3d?usp=sharing>