

UNIVERSIDADE PRESBITERIANA MACKENZIE

RAYAN CRHISTOFER GOMES DA SILVA, GUILHERME SOARES FROTA, BRUNO

LUIGI DOS SANTOS TOBIAS, THÁBATA CRISTINA SEIXAS BALBINO

ANÁLISE DE APOSTAS ESPORTIVAS NO FUTEBOL

SÃO PAULO

2024

SUMÁRIO

1. INTRODUÇÃO.....	3
2. GRUPO DE TRABALHO.....	3
3. PREMISSAS DO PROJETO.....	4
3.1 DEFINIÇÃO DA EMPRESA E ÁREA DE ATUAÇÃO.....	4
3.2 APRESENTAÇÃO DOS DADOS.....	5
3.3 OBJETIVOS E METAS.....	6
4. DEFINIÇÃO DO PRODUTO ANALÍTICO.....	7
4.1 BIBLIOTECAS E REPOSITÓRIO GITHUB.....	8
4.2 DEFINIÇÃO DA BASE DE DADOS E ANÁLISE EXPLORATÓRIA.....	8
4.3 TRATAMENTO DA BASE DE DADOS (PREPARAÇÃO E TREINAMENTO).....	9
4.4 BASES TEÓRICAS DOS MÉTODOS ANALÍTICOS.....	9
4.5 CÁLCULO DA ACURÁCIA.....	10
5. MANIPULAÇÃO E VISUALIZAÇÃO DOS DADOS.....	10
7. CONCLUSÕES OBTIDAS NA ANÁLISE EXPLORATÓRIA DE DADOS PYTHON.....	19
8. MODELO DE APRENDIZADO DE MÁQUINA PARA APOSTAS NO FUTEBOL.....	20
9. MÉTODO DE AVALIAÇÃO DO MODELO DE APRENDIZADO DE MÁQUINA PARA APOSTAS NO FUTEBOL.....	24
9.1 CONCLUSÕES SOBRE O MODELO DE APRENDIZADO DE MÁQUINA.....	26
10. CONCLUSÕES GERAIS.....	28
11. CRONOGRAMA.....	30
12. REPOSITÓRIO GITHUB.....	30
13. REFERÊNCIAS BIBLIOGRÁFICAS.....	30

1. INTRODUÇÃO

Nos últimos anos, o setor de apostas esportivas tem experimentado um crescimento exponencial, impulsionado pelo avanço da tecnologia e pela popularização do esporte. Esse aumento gerou um vasto volume de dados, que, se analisados de forma adequada, podem oferecer insights valiosos sobre o comportamento dos apostadores e as dinâmicas das partidas. Nesse contexto, a ciência de dados se apresenta como uma ferramenta essencial, capaz de transformar dados brutos em informações estratégicas que podem impactar a tomada de decisões nas apostas.

O presente projeto tem como objetivo desenvolver um modelo preditivo para classificar apostas em futebol como bem-sucedidas e mal sucedidas, utilizando o dataset "Analisando Apostas de Futebol". Esse conjunto de dados fornece informações relevantes, incluindo valores de investimento, odds, ganhos e resultados das apostas, permitindo uma análise aprofundada das variáveis que influenciam o sucesso nas apostas. Através da aplicação de técnicas de análise estatística preditiva e aprendizado de máquina, buscamos identificar padrões e tendências das apostas esportivas do futebol.

2. GRUPO DE TRABALHO

O grupo de trabalho é composto por quatro integrantes, cada um com responsabilidades específicas e complementares, visando à realização de uma análise detalhada sobre os momentos mais favoráveis para apostas em futebol, com base no desempenho histórico de apostas em diferentes jogos. A seguir, são descritas as funções e atribuições de cada membro:

- Membro: Rayan Christofer Gomes da Silva

Função: Banco de Dados e Análise Estatística

Responsabilidades: Rayan será responsável pela construção e gestão do banco de dados, garantindo a integridade e a organização adequada das informações coletadas sobre apostas. Além disso, aplicará métodos estatísticos avançados para explorar e interpretar os dados, identificando padrões e momentos favoráveis para apostas, além de formular conclusões fundamentadas nas análises realizadas.

- Membro: Thábata Cristina Seixas Balbino

Função: Banco de Dados e Análise Estatística

Responsabilidades: Thábata compartilhará com Rayan a responsabilidade pela montagem do banco de dados, assegurando a qualidade e a precisão das informações. Ela também participará da aplicação de técnicas estatísticas para a exploração e interpretação dos dados, colaborando na identificação de correlações e na análise dos resultados obtidos, contribuindo para a elaboração de insights relevantes sobre os melhores momentos para realizar apostas.

- Membro: Guilherme Soares Frota

Função: Visualização de Dados

Responsabilidades: Guilherme será encarregado de criar e desenvolver visualizações dos dados. Utilizando ferramentas especializadas, ele traduzirá os resultados da análise estatística em gráficos e outras representações visuais que facilitem a compreensão dos melhores momentos para apostas. Suas visualizações contribuirão para a clareza e a comunicação eficaz dos resultados a diferentes públicos, tanto acadêmicos quanto leigos.

- Membro: Bruno Luigi dos Santos Tobias

Função: Revisor e Documentador

Responsabilidades: Bruno atuará como revisor e documentador do trabalho. Sua responsabilidade inclui a revisão minuciosa do relatório final, garantindo a precisão e a clareza do conteúdo apresentado. Ele também será responsável pela documentação do processo de análise, desde a coleta de dados até a interpretação dos resultados, assegurando que o trabalho esteja bem estruturado e em conformidade com os padrões acadêmicos.

3. PREMISSAS DO PROJETO

3.1 DEFINIÇÃO DA EMPRESA E ÁREA DE ATUAÇÃO

A empresa responsável pelos dados contidos no dataset [1] é a "Game Winner", uma plataforma que atuava no mercado de apostas esportivas. Embora a Game Winner não esteja

mais em operação, os dados coletados e disponibilizados por Rodrigo Mota em parceria com Maury Lukas fornecem uma base valiosa para a análise do comportamento de apostas em futebol.

O mercado de apostas esportivas, especialmente no contexto do futebol, tem se expandido de forma significativa nos últimos anos, impulsionado pelo aumento da popularidade do esporte e pela evolução das tecnologias de informação. Esse mercado permite que os apostadores façam previsões sobre os resultados de eventos esportivos, como partidas de futebol, e apostem com base nessas previsões.

A dinâmica do mercado de apostas é composta por várias partes interessadas, incluindo casas de apostas, apostadores e organismos reguladores. As casas de apostas, que podem operar tanto online quanto fisicamente, oferecem odds (cotações) que representam a probabilidade percebida de um resultado específico. Essas odds são ajustadas com base em diversos fatores, como desempenho passado das equipes, condições do jogo, e até mesmo a movimentação de apostas pelos usuários.

Os apostadores, por sua vez, são indivíduos ou grupos que buscam identificar oportunidades para lucrar por meio de suas apostas. A análise de dados históricos e estatísticas se tornou uma prática comum entre apostadores, pois fornece uma base para decisões mais informadas. Com o advento das plataformas digitais, o acesso a dados e ferramentas analíticas se tornou mais fácil, permitindo que apostadores desenvolvam estratégias mais eficazes.

Além disso, o mercado de apostas esportivas é regulado por leis específicas que variam de país para país. Essas regulamentações são essenciais para garantir a segurança e a integridade do mercado, protegendo tanto os apostadores quanto as casas de apostas.

Neste projeto, a análise dos dados disponibilizados pela Game Winner visa explorar os padrões de apostas em futebol, identificando os momentos mais favoráveis para realizar apostas com base nas informações históricas. Essa abordagem permitirá uma melhor compreensão do funcionamento do mercado de apostas.

3.2 APRESENTAÇÃO DOS DADOS

O dataset "Analisando Apostas de Futebol" contém uma variedade de informações relevantes sobre as apostas realizadas em partidas de futebol no ano de 2022 da empresa "Game Winner". Abaixo, apresentamos uma descrição detalhada das colunas presentes na base de

dados, exemplificando com uma pequena amostra que ilustra a estrutura e o conteúdo dos dados.

Os dados estão disponíveis na planilha CSV denominada "**futebol.csv**", que inclui as seguintes colunas:

- **Data:** Representa a data em que o jogo ocorreu.
- **Jogo:** Nome das equipes que se enfrentam na partida, permitindo a identificação das partidas específicas.
- **Minutos:** Intervalo de minutos em que a aposta foi realizada ou algum evento específico ocorreu (por exemplo, número de gols na partida).
- **Investimento:** Valor (em R\$) apostado pelo usuário.
- **Odd:** Fator multiplicador caso o evento apostado aconteça.
- **Ganho:** Valor retornado ao apostador caso ganhe, considerando o valor investido e a Odd (descontando o valor apostado, ou seja, apenas o saldo líquido).
- **Resultado:** Indica se a aposta foi “Green” (vencedora) ou “Red” (perdedora).

A seguir, apresentamos uma pequena amostra do dataset com algumas apostas realizadas:

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 UCY 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

Figura 1: Amostra o dataset “Futebol.csv”.

A apresentação dos dados visa fornecer uma compreensão clara da estrutura e conteúdo do dataset, permitindo que os membros da equipe e interessados no projeto entendam como as informações estão organizadas e quais variáveis serão analisadas. Através dessa apresentação, serão identificados padrões e insights que auxiliarão na formulação de estratégias para determinar os momentos mais propícios para realizar apostas em partidas de futebol.

3.3 OBJETIVOS E METAS

O presente projeto tem como objetivo principal analisar as apostas realizadas em partidas de futebol, utilizando o dataset "Analisando Apostas de Futebol". Essa análise buscará identificar padrões e momentos mais favoráveis para realizar apostas, com base nos dados históricos disponibilizados. A seguir, são apresentados os objetivos e as metas específicas que guiarão o desenvolvimento do projeto.

3.3.1 OBJETIVOS GERAIS

- **Analisar os dados históricos de apostas em futebol:** Explorar e interpretar o dataset para entender as características das apostas e identificar tendências de sucesso.
- **Identificar os momentos propícios para apostas:** Desenvolver um modelo que aponte os intervalos de tempo nos jogos em que as apostas foram mais bem-sucedidas.
- **Avaliar a performance das apostas:** Investigar a relação entre o valor das odds, o investimento realizado e os resultados obtidos, com o intuito de entender como esses fatores influenciam o sucesso das apostas.

5

3.3.2 METAS ESPECÍFICAS

1. **Realizar uma análise exploratória detalhada:** Conduzir uma análise descritiva dos dados, incluindo a visualização de padrões e a identificação de correlações entre variáveis.
2. **Desenvolver um modelo preditivo:** Implementar um modelo simples que indique, com base em apostas anteriores, os momentos do jogo em que as apostas foram mais bem-sucedidas.
3. **Produzir visualizações informativas:** Criar gráficos e outras representações visuais que comuniquem de forma clara os resultados da análise e os insights obtidos.
4. **Documentar todo o processo:** Elaborar um relatório final que compile a metodologia utilizada, os resultados alcançados e as conclusões tiradas a partir da análise dos dados.
5. **Realizar uma apresentação final:** Preparar e apresentar os resultados do projeto para a banca avaliadora.

Esses objetivos e metas serão fundamentais para orientar as atividades do grupo ao longo do desenvolvimento do projeto, assegurando que a análise proposta seja abrangente e significativa, contribuindo para uma melhor compreensão do comportamento nas apostas em futebol.

4. DEFINIÇÃO DO PRODUTO ANALÍTICO

A Etapa 2 do projeto visa definir as diretrizes para a execução da análise de dados, bem como estabelecer os métodos e as ferramentas que serão utilizadas. O database "futebol.csv" será analisado por meio da linguagem Python, utilizando suas bibliotecas, métodos estatísticos e modelos de aprendizado de máquina. A seguir, são apresentados os tópicos que compõem esta seção.

4.1 BIBLIOTECAS E REPOSITÓRIO GITHUB

Para a execução colaborativa do trabalho, foram selecionadas as seguintes bibliotecas da linguagem Python:

- **Pandas:** Para manipulação e análise de dados, especialmente para operações em DataFrames.
- **NumPy:** Para operações numéricas e manipulação de arrays, facilitando cálculos estatísticos.
- **Matplotlib e Seaborn:** Para visualização de dados, permitindo a criação de gráficos e plots informativos que ajudam na interpretação dos resultados.
- **Scikit-learn:** Para aplicação de métodos de aprendizado de máquina, incluindo a divisão dos dados em conjuntos de treino e teste, e a implementação de modelos preditivos.
- **Statsmodels:** Para análise estatística mais profunda, permitindo a realização de testes estatísticos e a construção de modelos de regressão.

O repositório no GitHub será utilizado para o versionamento do código e a colaboração entre os membros da equipe. Um repositório será criado com o nome "[PROJ_APL_2-AN-LISE-DE-APOSTAS-ESPORTIVAS-NO-FUTEBOL](#)", onde todos os

scripts, notebooks Jupyter e documentos de suporte serão organizados de maneira a facilitar o acesso e a edição por todos os integrantes do grupo.

4.2 DEFINIÇÃO DA BASE DE DADOS E ANÁLISE EXPLORATÓRIA

A base de dados a ser utilizada é o arquivo "futebol.csv", que contém informações sobre apostas realizadas em diversas partidas de futebol. A análise do database será realizada por meio da linguagem Python, utilizando suas bibliotecas, além de métodos estatísticos para extrair insights significativos. A análise exploratória inicial envolverá:

- **Estatísticas descritivas:** Cálculo de métricas como média, mediana, desvio padrão, e contagens de valores únicos em colunas relevantes.
- **Distribuição dos dados:** Visualização da distribuição dos investimentos, odds e ganhos, utilizando histogramas e boxplots.
- **Correlação entre variáveis:** Avaliação da relação entre diferentes variáveis, como investimento e ganho, utilizando gráficos de dispersão e matrizes de correlação.

4.3 TRATAMENTO DA BASE DE DADOS (PREPARAÇÃO E TREINAMENTO)

O tratamento da base de dados incluirá as seguintes etapas:

6. **Remoção de entradas ausentes:** Todas as entradas com valores faltantes serão excluídas para garantir a integridade dos dados.
7. **Conversão de tipos de dados:** Garantir que as colunas estejam no formato correto (por exemplo, converter a coluna 'Data' para o tipo datetime).
8. **Normalização:** Se necessário, os dados serão normalizados para facilitar a comparação entre variáveis.
9. **Divisão dos dados:** A base de dados será dividida em conjuntos de treino e teste, permitindo a validação do modelo.

4.4 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.

4.5 CÁLCULO DA ACURÁCIA

No contexto deste projeto, a acurácia foi escolhida como uma métrica fundamental para avaliar o desempenho inicial do modelo de recomendação de apostas esportivas. Em termos gerais, a acurácia representa a proporção de previsões corretas feitas pelo modelo em relação ao total de previsões realizadas, englobando tanto as classificações corretas de apostas bem-sucedidas ("Green") quanto as de apostas não bem-sucedidas ("Red").

O cálculo da acurácia se baseia na seguinte fórmula:

$$Acurácia = \frac{\text{Número de previsões corretas}}{\text{Total de previsões}}$$

A escolha da acurácia como métrica de avaliação reflete o interesse do projeto em obter um panorama inicial sobre a eficácia do modelo. Em um modelo de classificação binária como este, uma alta acurácia indica que o modelo está conseguindo prever a maioria dos resultados corretamente, o que é encorajador. No entanto, ao aplicá-lo no contexto de apostas esportivas, é necessário considerar limitações. A acurácia, isoladamente, não oferece uma visão detalhada de casos onde erros específicos podem impactar a decisão de apostar ou não.

Portanto, a acurácia foi acompanhada de métricas complementares como precisão e recall, que adicionam uma camada mais refinada de análise ao avaliar, por exemplo, a capacidade do modelo de evitar falsos positivos (previsões incorretas de sucesso). A precisão, especificamente, é importante para assegurar que, entre as apostas recomendadas, a maior parte tenha um potencial real de retorno positivo. Já o recall contribui para garantir que as oportunidades de apostas lucrativas não sejam descartadas desnecessariamente.

Assim, ao utilizar a acurácia como uma medida de sucesso, foi possível estabelecer um ponto de partida confiável para ajustes subsequentes do modelo, permitindo que, com o tempo, outras métricas mais robustas possam também ser aplicadas e ajustadas conforme o modelo evolui e se integra a mais variáveis relevantes.

5. MANIPULAÇÃO E VISUALIZAÇÃO DOS DADOS

Inicialmente, foram importadas as bibliotecas necessárias para manipulação e visualização dos dados, como Pandas, Seaborn e Matplotlib. O dataset foi carregado em um DataFrame, e as colunas passaram por um processo de tratamento para padronização. O caminho do arquivo CSV foi definido e carregado para `futebol_df`, e todos os espaços extras nos nomes das colunas foram removidos, garantindo uniformidade nos dados.

	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 × 7 columns

Figura 2: dataset “Futebol.csv” carregado em um 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

```

Figura 3: Informações do dataframe.

Especificamente, a coluna "Data" foi convertida para o tipo datetime, permitindo análises temporais mais precisas e facilitando futuras ordenações ou filtros por data. A coluna "Jogo" foi padronizada com remoção de espaços em branco e conversão para letras minúsculas, permitindo uma identificação mais eficiente de duplicatas.

Para preencher os valores nulos na coluna "Minutos", foi adicionada a categoria "desconhecido", representando períodos não registrados. Termos inconsistentes ou duplicados, como "aos", "ao" e variações de "múltipla" foram ajustados na coluna "Minutos", mantendo uma categorização uniforme e padronizada. Em seguida, identificaram-se registros duplicados, que foram consolidados ao somar os valores das colunas "Investimento" e "Ganho" e ao eliminar as linhas repetidas, evitando redundâncias e preservando a integridade dos dados financeiros.

	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	botafogo 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 sampao	gols	10	1.50	5.0	Red
1531	2022-10-28	magdeburg x heidenheim	gols	10	1.50	5.0	Red

Figura 4: processo de padronização e limpeza dos dados.

Os valores de "Minutos" com ocorrência inferior a oito foram agrupados sob a categoria "outros", para simplificação e organização da análise, com exceção do período "80 a 90", que foi mantido em separado por se tratar de uma fase importante na análise de escanteios. Por fim, o dataset final foi verificado para garantir a consistência e preparado para futuras etapas de análise preditiva e estatística.

Após a padronização e limpeza dos dados, o próximo passo envolveu a análise exploratória para entender melhor as características do dataset. O uso de funções como `value_counts()` permitiu verificar a distribuição dos valores na coluna "Minutos", identificando quais períodos eram mais frequentes e quais apresentaram menor ocorrência. Essa análise foi fundamental para informar as decisões do agrupamento e assegurar que categorias relevantes fossem mantidas.

A visualização gráfica foi empregada para ilustrar as distribuições e relações entre as variáveis. Gráficos de barras e pizza foram criados para demonstrar a proporção de resultados ("Green" e "Red") em relação aos investimentos, permitindo uma interpretação visual clara do desempenho financeiro. Essa abordagem não apenas facilitou a identificação de padrões, mas também possibilitou a comunicação dos resultados de maneira acessível para stakeholders.

Além disso, a análise estatística foi conduzida para explorar correlações entre as variáveis, como o investimento e o ganho, bem como entre diferentes jogos e suas respectivas odds. O uso de bibliotecas como Seaborn e Matplotlib permitiu a criação de gráficos que ajudam a visualizar essas relações, enriquecendo a compreensão dos dados.

Por fim, após a exploração inicial e visualização, o dataset foi preparado para a modelagem preditiva. As variáveis foram selecionadas com base em sua relevância e a transformação dos dados foi realizada para garantir que estivessem em formatos apropriados para algoritmos de aprendizado de máquina. Essa etapa preparou o caminho para a construção de modelos que poderiam prever resultados futuros em apostas esportivas, utilizando técnicas de aprendizado supervisionado e não supervisionado.

O resultado desse processo de limpeza, exploração e preparação é um conjunto de dados robusto e confiável, pronto para análises mais profundas e aplicações práticas em cenários de apostas esportivas.

6. 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.

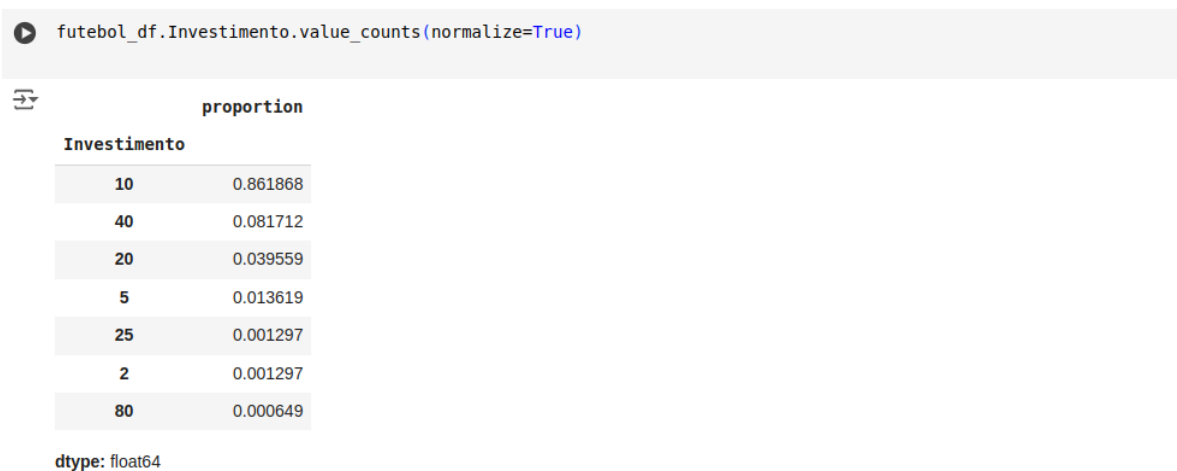


Figura 5: Distribuição da proporção das apostas em em valores definidos.

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.

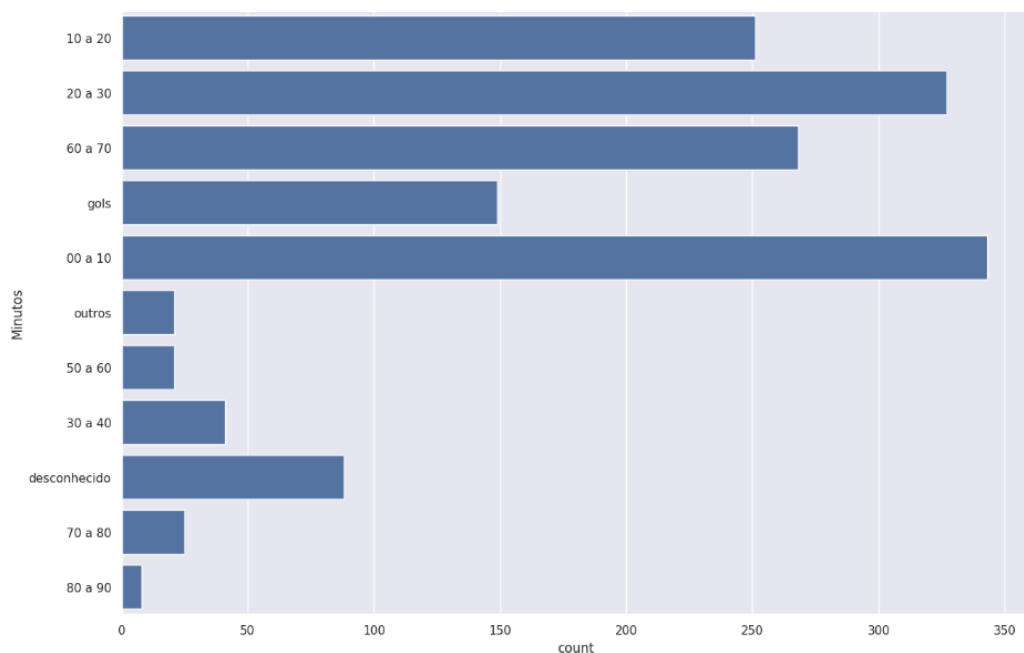


Figura 6: Gráfico sobre a representação da frequência de ocorrências para cada período de aposta.

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").

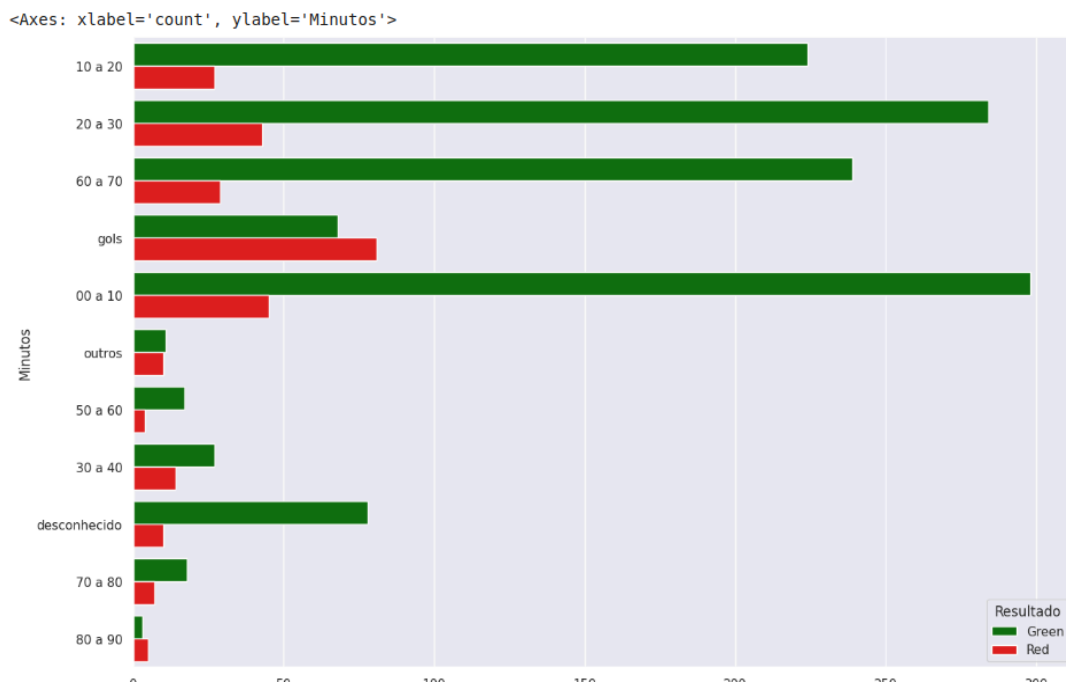


Figura 7: Resultados positivos (“Green”) e negativos (“Red”) nos intervalos de tempo definidos.

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.

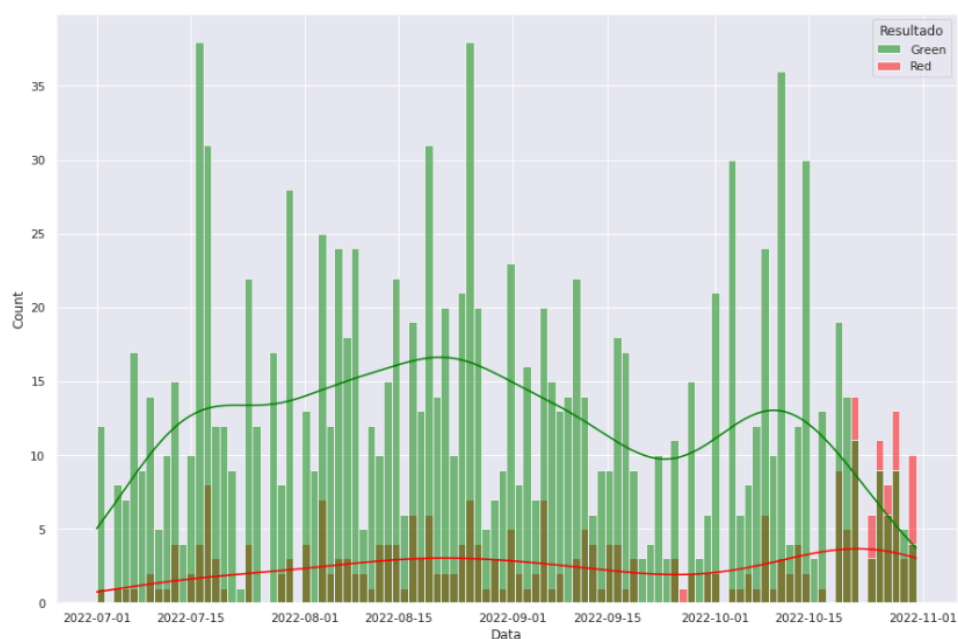


Figura 8: Histograma dos resultados ao longo do tempo

A investigação se aprofundou ao verificar quais valores de aposta estavam associados a resultados "Green" ou "Red". Utilizando uma grade de facetas, foi possível visualizar a performance dos resultados conforme os valores de investimento. Os dados mostraram que os investimentos de 40 estavam quase todos associados a perdas, enquanto os ganhos eram predominantemente provenientes de apostas no valor de 10, que também estavam se tornando menos frequentes com o tempo. Além disso, observou-se que apostas de 20 estavam apresentando perdas significativas, levantando questões sobre a eficácia dos critérios utilizados.

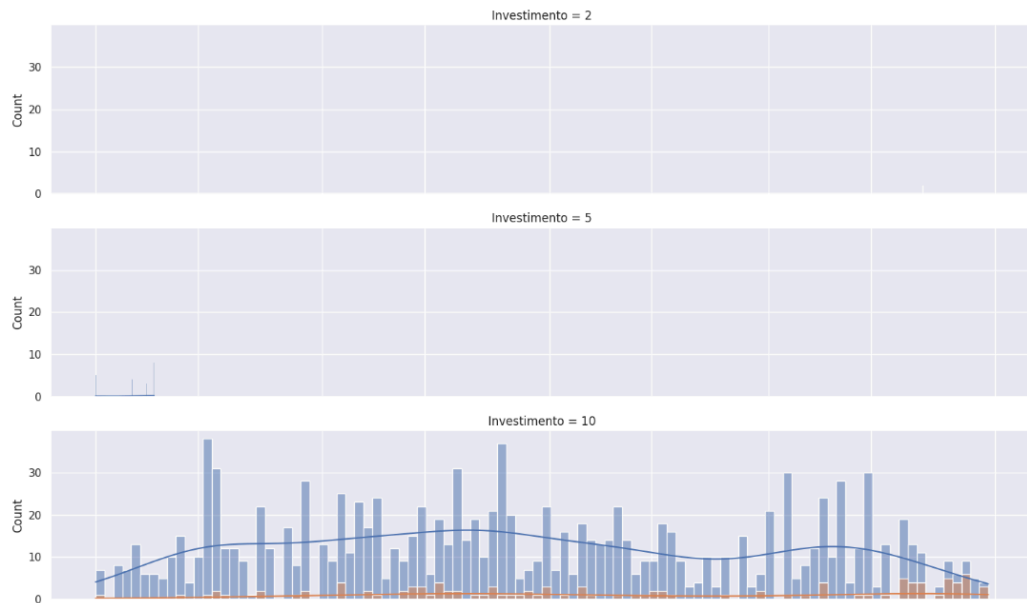


Figura 9: Grade de facetas para os investimentos nos intervalos “2”, “5” e “10”.

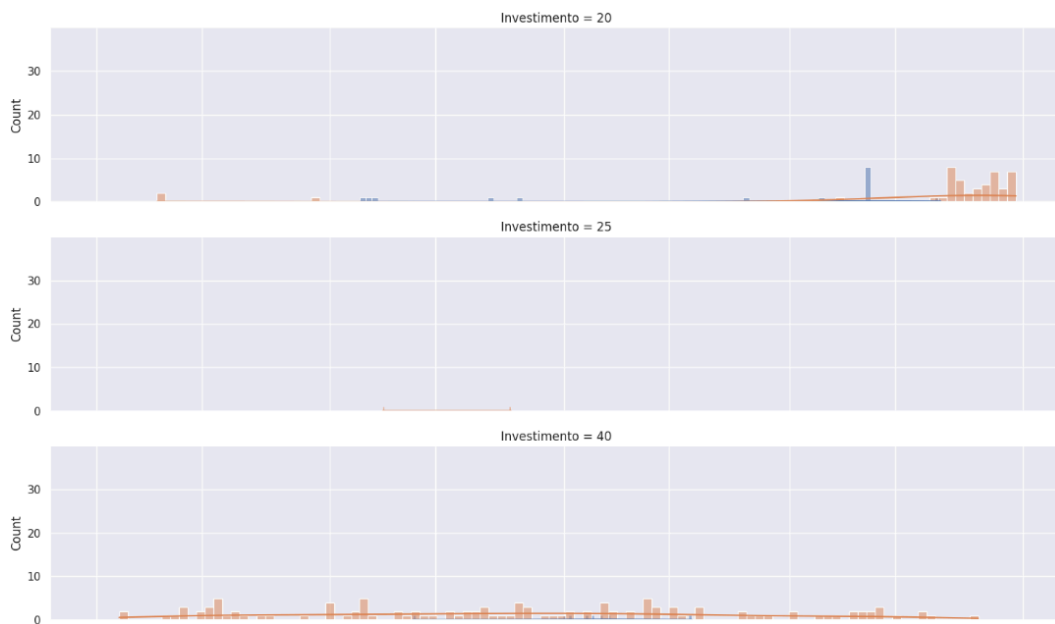


Figura 10: Grade de facetas para os investimentos nos intervalos “20”, “25” e “40”.

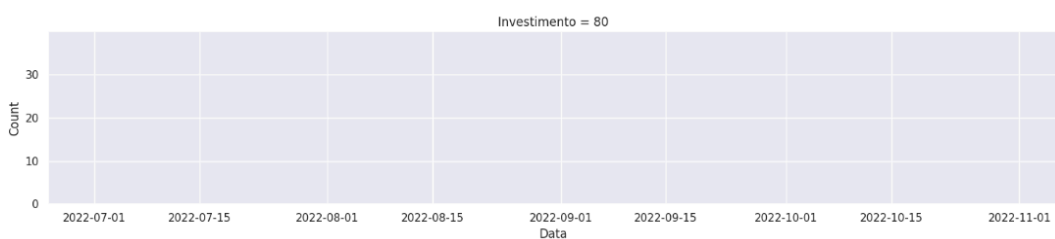


Figura 11: Grade de facetas para os investimentos no intervalo “80”.

Por fim, a análise prosseguiu para explorar a relação entre os períodos de aposta ("Minutos") e os resultados ao longo do tempo, novamente utilizando uma grade de facetas. Essa abordagem permitiu uma compreensão mais clara sobre quais períodos poderiam estar contribuindo para os resultados negativos, estabelecendo uma base sólida para decisões futuras e estratégias de apostas.

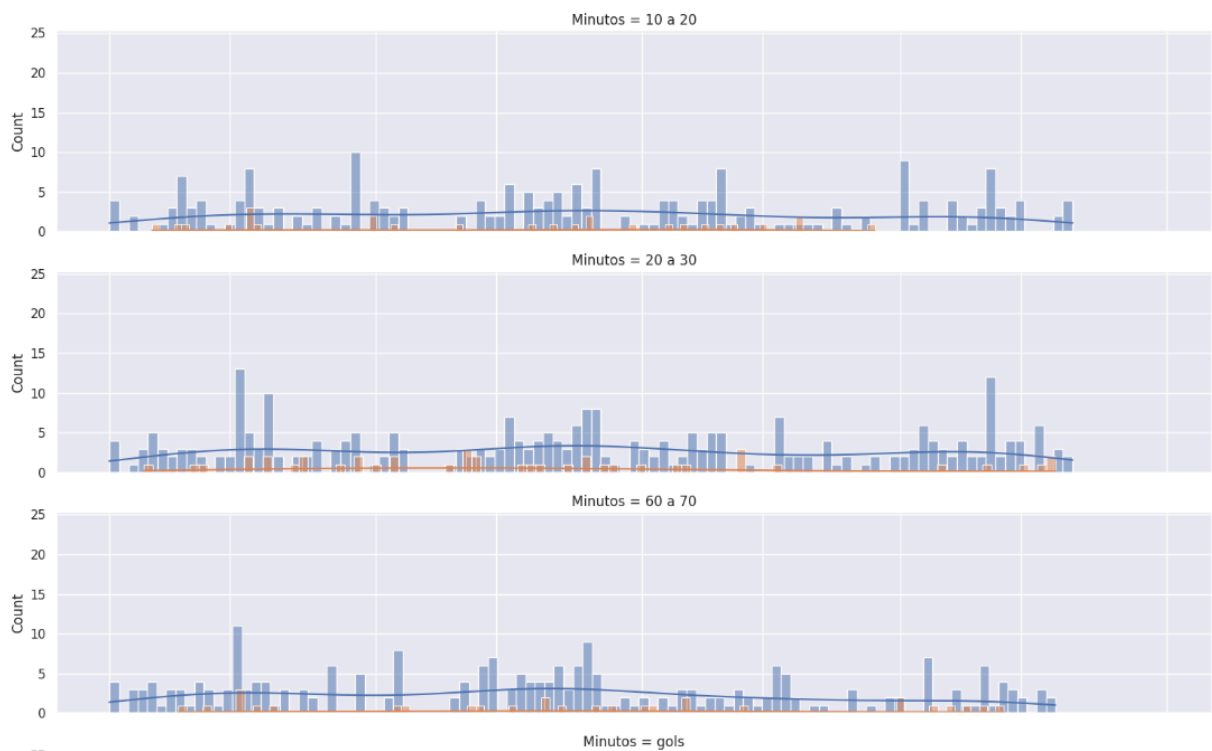


Figura 12: Período de apostas e os resultados ao longo do tempo para os intervalos “10 a 20”, “20 a 30” e “60 a 70”.

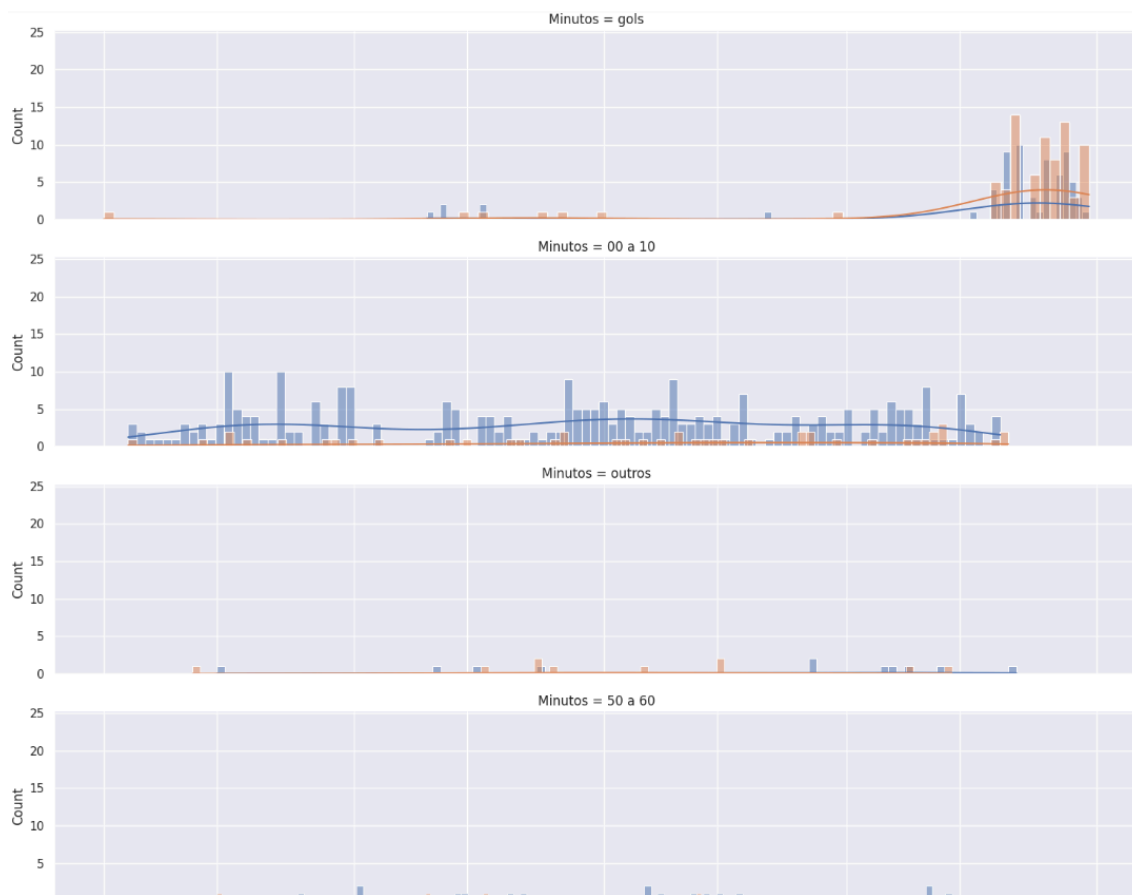


Figura 13: Período de apostas e os resultados ao longo do tempo para os intervalos “Minutos = gols”, “00 a 10”, “Minutos = outros” e “50 a 60”.



Figura 14: 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”.

7. CONCLUSÕES OBTIDAS NA ANÁLISE EXPLORATÓRIA DE DADOS PYTHON

A análise realizada neste trabalho revelou insights valiosos sobre o desempenho das apostas esportivas, particularmente em relação à distribuição dos valores apostados, a frequência das apostas em diferentes intervalos de tempo, e o sucesso ou fracasso associado a essas apostas.

A distribuição dos valores das apostas mostrou uma clara predominância das apostas de **10**, que representaram **86%** do total apostado. Essa escolha sugere uma estratégia conservadora, onde os apostadores preferem minimizar os riscos associados a apostas mais elevadas. A presença de valores mais altos, como **40**, embora menos frequentes, não deve ser ignorada, pois podem indicar momentos de confiança ou risco calculado.

A análise da frequência das apostas nos intervalos definidos indicou uma concentração significativa nas faixas de minutos mais populares, enquanto os intervalos com menos de oito

ocorrências foram agrupados sob a categoria "outros". Essa categorização pode facilitar a identificação de tendências, mas também destaca a necessidade de diversificação nas apostas para evitar a saturação de determinados períodos.

Os dados indicaram que os intervalos de tempo em que as apostas em escanteios foram realizadas resultaram em um desempenho superior, refletindo um maior número de resultados "Green". Em contraste, as apostas em gols mostraram uma tendência de perda, evidenciada pelo aumento nos resultados "Red". Isso sugere que a mudança de estratégia para apostas em gols não foi eficaz, ressaltando a importância de avaliar a eficácia de diferentes critérios de aposta ao longo do tempo.

A visualização dos resultados em um histograma ao longo do tempo permitiu observar a evolução do desempenho das apostas. As análises mostraram que, embora a maioria das apostas tenha resultado em ganhos inicialmente, houve uma tendência de queda na taxa de sucesso nas apostas recentes, especialmente nas apostas em gols. Isso sugere que mudanças nas condições do mercado ou na estratégia de apostas podem ter impactado negativamente os resultados.

Em suma, este trabalho demonstrou a importância de uma análise detalhada das apostas esportivas, com ênfase na distribuição de valores, na frequência de apostas e na avaliação do desempenho ao longo do tempo. A capacidade de identificar padrões e adaptar estratégias com base nos resultados é essencial para maximizar as chances de sucesso nas apostas. A análise contínua e a disposição para aprender com os dados obtidos podem proporcionar uma base sólida para decisões informadas e bem-sucedidas no futuro.

8. MODELO DE APRENDIZADO DE MÁQUINA PARA APOSTAS NO FUTEBOL

Este modelo de aprendizado de máquina [4], desenvolvido com a técnica de **Regressão Logística**, visa prever a probabilidade de sucesso (chamado de "Green") ou insucesso (chamado de "Red") em apostas de futebol. A Regressão Logística é um método estatístico amplamente empregado para resolver problemas de classificação binária, onde o objetivo é prever a ocorrência de uma de duas possíveis saídas com base nas variáveis de entrada. Neste

projeto, essas saídas representam o sucesso ou insucesso das apostas, e o modelo utiliza informações históricas como o minuto do jogo, a odd (valor da cotação) e o valor do investimento para gerar 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 15: Bibliotecas utilizadas no modelo de aprendizado de máquina para aposta esportiva

O processo começa com o pré-processamento dos dados de apostas históricas, que é crucial para garantir que as variáveis estejam no formato adequado para o modelo. Cada aposta no conjunto de dados inclui informações específicas, como:

- **Intervalo de minuto do jogo** (em que 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 referente ao minuto do evento é convertida para valores numéricos, com intervalos de minutos sendo substituídos por uma média para facilitar a análise. Qualquer texto irrelevante para o cálculo, como "Escanteio" ou "Gols", é adequadamente tratado ou removido.

```

# 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 16: Tratamento dos dados para alimentar o modelo

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 17: Métricas de avaliação

Essas métricas auxiliam na avaliação da performance do modelo, fornecendo uma visão clara sobre sua capacidade de fazer previsões assertivas.

Na prática, o modelo solicita ao usuário informações como o minuto da aposta, odd e valor a ser investido. Com esses dados, ele faz uma previsão sobre o potencial sucesso da aposta e retorna uma recomendação ao usuário, dizendo "Apostar!" ou "Não apostar" conforme a probabilidade de sucesso 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: FutureWar
    user_records = pd.concat([user_records, pd.
Recomendação: Apostar!

```

Figura 18: Exemplo de saída do modelo de aprendizado de máquina

O sistema ainda oferece ao usuário a possibilidade de comparar sua aposta com casos semelhantes no histórico, identificados com 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 19: Modelo de aprendizado retorna apostas positivas (Green) semelhantes ao do usuário

24

```
Resultados das apostas semelhantes que acabaram em Red:  
ID: 191, Jogo: Guilherme 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 20: Modelo de aprendizado retorna apostas negativas (Red) semelhantes ao do usuário.

Adicionalmente, todos os registros de apostas feitas pelo usuário são salvos em um banco de dados, criando um histórico de apostas e permitindo uma evolução contínua do modelo. Isso permite que, com o tempo, o sistema possa aperfeiçoar sua capacidade de previsão e fornecer 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 21: 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.

9. MÉTODO DE AVALIAÇÃO DO MODELO DE APRENDIZADO DE MÁQUINA PARA APOSTAS NO FUTEBOL

O método de avaliação construído para o modelo de aprendizado de máquina utiliza uma abordagem completa para mensurar seu desempenho em prever o sucesso ou o fracasso de apostas. Ele integra uma análise baseada na matriz de confusão, que detalha como o modelo

performa em termos de classificações corretas e incorretas para as classes positivas e negativas, além de outras métricas de avaliação complementares como a acurácia, precisão, recall, F1-score e o 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 22: Parâmetros para avaliação de desempenho do modelo de aprendizado de máquina

Primeiramente, a **matriz de confusão** é usada para mapear as previsões realizadas pelo modelo contra os valores reais, dividindo os resultados entre Verdadeiros Positivos (previsões corretas de sucesso), Verdadeiros Negativos (previsões corretas de fracasso), Falsos Positivos (previsões incorretas de sucesso) e Falsos Negativos (previsões incorretas de fracasso). Esse mapeamento ajuda a identificar padrões de erro no modelo. Por exemplo, uma quantidade elevada de Falsos Positivos pode indicar que o modelo é excessivamente otimista ao recomendar apostas, enquanto muitos Falsos Negativos sugerem uma tendência a evitar recomendações mesmo em cenários promissores.

Além disso, as métricas de **acurácia** (proporção total de previsões corretas), **precisão** (proporção de Verdadeiros Positivos sobre todas as predições positivas), **recall** (proporção de Verdadeiros Positivos sobre todos os casos reais positivos), e **F1-score** (média harmônica da precisão e recall) fornecem uma visão quantitativa e equilibrada sobre o desempenho geral e a eficácia em identificar apostas positivas. Já o **R^2** , utilizado como medida de variância explicada, demonstra o quanto das variações observadas no resultado das apostas é capturado pelo modelo. Quanto mais próximo de 1 o R^2 , maior a capacidade do modelo em ajustar-se 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"R²: {r2:.2f}")
print(f"Matriz de Confusão:\n{matriz_confusao}")

```

Figura 23: Métricas para avaliação de desempenho do modelo de aprendizado de máquina

Essas métricas, combinadas, permitem uma análise aprofundada do desempenho do modelo e ajudam a orientar possíveis ajustes. Se, por exemplo, o modelo apresenta um recall muito alto com precisão baixa, ele pode estar capturando muitos casos positivos à custa de uma alta taxa de falsos positivos, exigindo ajustes de balanceamento para maximizar a eficácia das previsões. O método, portanto, fornece uma avaliação robusta, detalhando o comportamento do modelo e facilitando o refinamento do algoritmo para atingir resultados mais confiáveis e ajustados ao objetivo de recomendação de apostas assertivas.

```

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

```

Figura 24: Avaliação do modelo de aprendizado de máquina

9.1 CONCLUSÕES SOBRE O MODELO DE APRENDIZADO DE MÁQUINA

Os resultados obtidos na avaliação do modelo de aprendizado de máquina apresentam um desempenho altamente satisfatório em relação à previsão de apostas bem-sucedidas. A **acurácia** de 0.98 indica que o modelo acertou 98% das previsões, refletindo sua capacidade de classificar corretamente a maioria das instâncias. A **precisão** de 0.97 sugere que, das previsões positivas feitas pelo modelo, 97% foram realmente corretas, o que demonstra uma forte habilidade em evitar falsos positivos.

O **recall** de 1.00 é particularmente notável, pois significa que o modelo conseguiu identificar todos os casos positivos sem falhas, ou seja, não houve nenhum falso negativo. Isso é crucial em um contexto de recomendação de apostas, onde perder uma oportunidade de aposta bem-sucedida pode ter consequências financeiras significativas.

O **F1-score** de 0.99, que é a média harmônica entre precisão e recall, reforça a ideia de que o modelo não apenas é preciso, mas também é eficaz na captura de todas as apostas que devem ser recomendadas. Esse equilíbrio entre precisão e recall é fundamental em situações onde tanto falsos positivos quanto falsos negativos têm implicações importantes.

Entretanto, o **R²** de 0.74 sugere que, apesar do desempenho excelente em termos de classificação, há espaço para melhorar a explicação da variância nos resultados das apostas. Embora o modelo capture uma boa parte das variações, ainda há 26% de variabilidade que não é explicada, indicando que outros fatores podem influenciar os resultados das apostas e que talvez mais características ou dados adicionais possam ser integrados para aprimorar a capacidade preditiva.

A **matriz de confusão** revela que o modelo cometeu 3 falsos positivos e 115 verdadeiros negativos, enquanto identificou corretamente 10 verdadeiros positivos e não teve falsos negativos. Essa distribuição reforça a eficácia do modelo em prever apostas bem-sucedidas, ao mesmo tempo em que mostra que algumas apostas foram indevidamente classificadas como bem-sucedidas.

Em suma, o modelo demonstrou um desempenho excepcional em prever apostas com alta precisão e capacidade de identificação, mas o ajuste contínuo e a exploração de variáveis

adicionais podem contribuir para melhorar ainda mais sua eficácia na explicação das variações nos resultados.

10. CONCLUSÕES GERAIS

O trabalho desenvolvido neste projeto de análise e modelagem preditiva de apostas esportivas em futebol revelou conclusões valiosas, destacando o impacto de uma abordagem detalhada e metodológica no entendimento e aprimoramento de estratégias de aposta. Ao longo do processo, diversos aspectos foram cuidadosamente analisados, com destaque para a manipulação de dados, a análise exploratória, a modelagem preditiva e a avaliação do desempenho do modelo de aprendizado de máquina.

Inicialmente, na manipulação e padronização dos dados, a limpeza e padronização dos dados foram etapas fundamentais para garantir a consistência e qualidade da base de dados. A remoção de duplicatas e a categorização de intervalos de minutos menos frequentes permitiram uma organização mais clara e uma base sólida para a análise. Essa preparação é essencial em qualquer projeto de análise de dados, pois estabelece uma fundação precisa e confiável.

Na segunda etapa, de Análise Exploratória dos Dados, a análise exploratória demonstrou que o valor das apostas tendia a se concentrar em 10, indicando uma preferência por estratégias conservadoras. Além disso, visualizações gráficas mostraram a distribuição dos resultados ("Green" e "Red") ao longo dos intervalos de tempo, permitindo uma identificação de padrões temporais importantes, como a vantagem de apostar em escanteios versus gols. A divisão dos intervalos de minutos e o agrupamento por categorias específicas também foram úteis para simplificar a análise e revelar tendências que embasaram a construção do modelo.

No Modelo de Aprendizado de Máquina, o modelo de Regressão Logística desenvolvido apresentou um desempenho notável, com altas métricas de acurácia, precisão e recall. A taxa de 98% de acurácia e a ausência de falsos negativos foram indicadores de que o modelo é robusto e confiável ao prever apostas bem-sucedidas. Esse desempenho sugere que a combinação das variáveis escolhidas (intervalo de minutos, odds e valor de investimento) é relevante para o contexto de apostas e que a técnica de classificação binária com Regressão Logística é adequada para o problema.

Em relação ao Método de Avaliação, a análise da matriz de confusão e das métricas de desempenho forneceu insights adicionais sobre o comportamento do modelo. Embora a explicação da variância dos dados ainda possa ser aprimorada (como indicado pelo R^2 de 0,74), o modelo provou ser eficaz na identificação de oportunidades de aposta. Os 25% de variância não explicada podem estar relacionados a características não capturadas devido ao agrupamento dos dados em intervalos de 25%, sugerindo que ajustes na segmentação dos dados poderiam melhorar a capacidade do modelo de captar nuances nos resultados.

Em suma, o projeto de análise e modelagem preditiva de apostas esportivas em futebol apresentou um modelo inicial, simples e primitivo, que ilustra os primeiros passos na construção de um sistema preditivo. A abordagem inicial foi eficaz na criação de uma base analítica e de um modelo de aprendizado de máquina funcional, ainda que simplificado. Com isso, foi possível visualizar o potencial de um sistema que oferece recomendações baseadas em dados (de uma amostra muito pequena) indicando tendências básicas para uma estratégia de apostas.

No entanto, para transformar esse modelo em uma ferramenta realmente útil e confiável, é necessário incorporar dados mais complexos e variados, como o histórico de jogos dos times, média de gols feitos e sofridos, desempenho em diferentes condições (como jogos em casa e fora), e outros indicadores estatísticos detalhados. Esses fatores adicionais dariam ao modelo uma compreensão mais ampla das variáveis que impactam o resultado de uma partida, aumentando a precisão das previsões e sua aplicabilidade prática no contexto das apostas.

Portanto, o projeto cumpriu seu objetivo de construir uma base analítica para as apostas esportivas, oferecendo insights práticos iniciais para futuras estratégias de aposta. As técnicas aplicadas demonstraram o valor de integrar análise estatística, aprendizado de máquina e um tratamento cuidadoso dos dados. Esse ponto de partida tem potencial de evolução contínua e abre portas para um sistema mais robusto, estruturado e confiável, ao atender as exigências e variabilidades do contexto esportivo real.

11 CRONOGRAMA

Cronograma de Atividades - Projeto Aplicado
--

ATIVIDADE	DATA PREVISTA DE CONCLUSÃO	RESPONSÁVEL	OBSERVAÇÃO	STATUS
Definir tema do Projeto	03/09/2024	EQUIPE	Tema alterado por conflito de versões.	Concluído
Criar documento e Github	03/09/2024	EQUIPE		Concluído
Revisar documento	05/09/2024	EQUIPE		Concluído
Data limite de entrega da primeira parte	05/09/2024	RAYAN	Etapa 1 e 2 com base no novo tema.	Concluído
Entrega da 3 etapa	02/11/2024	RAYAN	Etapa 3 finalizado em equipe.	Concluído
Entrega da 4 etapa	23/11/2024	RAYAN		Concluído

12. REPOSITÓRIO GITHUB

Todos os arquivos e dados, como o dataset, análise de dados exploratória com Python (Google Colab), Modelo de aprendizado de máquina (Google Colab), Storytelling [5] e a Apresentação [6] utilizados neste trabalho serão armazenados no [GitHub](#) [2].

13. REFERÊNCIAS BIBLIOGRÁFICAS

[1] LUKAS, Maury. Sports Dataset: Analisando Apostas de Futebol. Disponível em: <https://www.kaggle.com/datasets/maurylukas/sports/data?select=futebol.csv>. Acesso em: 03 out. 2024.

[2] GITHUB. Disponível em: https://github.com/RayanCrhistofer/PROJ_APL_2-AN-LISE-DE-APOSTAS-ESPORTIVAS-NO-FUTEBOL.git

[3] Análise exploratória de dados com Python. Google Colab. Disponível em: https://colab.research.google.com/drive/148Z0fZCqWdgsLJpVyST4i33wiQZU_wgf?usp=sharing

[4] Modelo de aprendizado de máquina para apostas no futebol. Google Colab. Disponível em: <https://colab.research.google.com/drive/1Ux5LWEK7Wgj524v3XkCJb0M29Q076a3d?usp=sharing>

[5] Storytelling: Apresentação - Análise de Apostas Esportivas no Futebol. Disponível em: <https://docs.google.com/presentation/d/1sitiZMHR3u8VQR6lstjY0OOiyrma0FnZ/edit?usp=sharing&ouid=116953145659391990083&rtpof=true&sd=true>

[6] SILVA, CRHISTOFER. Projeto aplicado II - Análise de apostas esportivas no futebol. Youtube, 23 nov. 2024. Disponível em: <https://youtu.be/fX8nmwvAUNQ>. Acesso em: 23 nov. 2024.