

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	1
2 GRUPO DE TRABALHO	2
3 PREMISSAS DO PROJETO	3
3.1 DEFINIÇÃO DA EMPRESA E ÁREA DE ATUAÇÃO	3
3.2 APRESENTAÇÃO DOS DADOS	4
3.3 OBJETIVOS E METAS	5
3.3.1 OBJETIVOS GERAIS	5
3.3.2 METAS ESPECÍFICAS	6
4 DEFINIÇÃO DO PRODUTO ANALÍTICO	6
4.1 BIBLIOTECAS E REPOSITÓRIO GITHUB	7
4.2 DEFINIÇÃO DA BASE DE DADOS E ANÁLISE EXPLORATÓRIA	7
4.3 TRATAMENTO DA BASE DE DADOS (PREPARAÇÃO E TREINAMENTO)	8
4.4 BASES TEÓRICAS DOS MÉTODOS ANALÍTICOS	8
4.5 CÁLCULO DE ACURÁCIA	9
5 CRONOGRAMA	11
6 REPOSITÓRIO GITHUB	12
7 REFERÊNCIAS BIBLIOGRÁFICAS	13

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 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 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

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.

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 "Analise_Apostas_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

A acurácia dos modelos será avaliada em termos percentuais, utilizando a seguinte fórmula:

$$Acuracia (\%) = \left(\frac{\text{Numero de previsoes corretas}}{\text{total de resultados reais}} \right) \times 100$$

Onde:

- **Número de Previsões Corretas:** Refere-se à soma dos verdadeiros positivos (casos em que o modelo previu corretamente uma classe positiva) e verdadeiros negativos (casos em que o modelo previu corretamente uma classe negativa).
- **Total de Resultados Reais:** É o número total de observações na base de dados.

A apresentação da acurácia em termos de porcentagem não apenas permite uma avaliação mais intuitiva do desempenho do modelo, mas também facilita a comparação entre diferentes abordagens analíticas. Adicionalmente, serão calculadas métricas complementares, como a taxa de erro, que indica a porcentagem de previsões incorretas em relação ao total de resultados reais:

$$Taxa de erro (\%) = \left(\frac{\text{Numero de previsoes Incorretas}}{\text{Total de resultados reais}} \right) \times 100$$

A apresentação das taxas de acertos e erros em termos de porcentagem não apenas proporciona uma avaliação mais intuitiva do desempenho do modelo, mas também facilita a comparação entre diferentes conjuntos de dados e abordagens analíticas. Além disso, ao calcular as taxas de acertos e erros para os resultados reais, será possível analisar a eficácia do modelo em prever os resultados de apostas.

Taxa de Acertos (Resultados Reais)

$$Taxa de acertos (Resultados Reais) (\%) = \left(\frac{\text{Total de ganhos}}{\text{Total de resultados}} \right) \times 100$$

Taxa de Erros (Resultados Reais)

$$Taxa\ de\ erros\ (Resultados\ Reais)\ (\%) = \left(\frac{Total\ de\ perdidos}{Total\ de\ resultados} \right) \times 100$$

Para complementar a análise das taxas de acertos e erros, será calculado o desvio padrão, que fornecerá uma medida da dispersão das previsões em relação à média. O desvio padrão é calculado pela fórmula:

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

Onde:

- xi representa cada um dos valores das taxas de acertos e erros.
- \bar{x} é a média das taxas de acertos e erros.
- N é o número total de observações.

O cálculo do desvio padrão permitirá avaliar a consistência das previsões, possibilitando identificar variações significativas nas taxas de acertos e erros. Essa informação é valiosa para compreender a eficácia do modelo e para identificar áreas que podem ser aprimoradas nas futuras análises e previsões.

Assim, a análise combinada da acurácia, taxas de acertos e erros, juntamente com o desvio padrão, proporcionará uma visão abrangente do desempenho do modelo, fundamentando as decisões estratégicas em relação às apostas analisadas.

5 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
Definir próximos passos	12/09/2024	EQUIPE	Aguardar observação do professor referente ao tema.	Em andamento

6 REPOSITÓRIO GITHUB

Todos os arquivos e dados utilizados neste trabalho serão armazenados no GitHub [2].

7 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