

DATA MINING

Apresentação do estado atual do trabalho prático

Grupo3

Alexandre Rodrigues pg53606

Henrique Ribeiro a95323

Hugo Martins a95125

João Escudeiro a96075



Índice

- Tema
- Dados
 - Fonte
 - Tratamento
- Modo de Funcionamento
- Estado atual do trabalho
- Principais desafios
- Trabalho futuro

Tema

FUTEBOL ANALYTICS CHAT

Chatbot que consiga fazer comentários futebolísticos mediante os resultados e todas as informações relevantes e recentes das equipas na principal liga de futebol portuguesa.

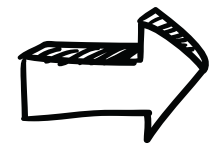


Fonte dos dados

- Conseguimos obter um token fornecido pela plataforma **SportMonks**, que nos permitiu fazer API requests ilimitados para a principal liga de futebol portuguesa (Liga Portugal).
- Os resultados dos pedidos incluem eventos, lineups, estatísticas ou informações sobre os jogadores.



Tratamento dos dados



Produção de scripts em GO responsáveis por fazer pedidos ao site e armazenarem os dados numa BD em nuvem no MongoDBAtlas.



Criação de 4 coleções : Fixtures , Seasons, Standings, Teams

```
_id: ObjectId('664494f92705e3d0be07c283')
lineups: Array (40)
  id: 18870453
  sport_id: 1
  stage_id: 77464616
formations: Array (2)
  venue_id: 313
season: Object
  result_info: "Game ended in draw."
  leg: "1/1"
  league_id: 462
  season_id: 21825
  name: "Sporting Braga vs Vitória SC"
  starting_at: "2024-01-06 20:30:00"
  group_id: null
events: Array (15)
  starting_at_timestamp: 1704573000
scores: Array (6)
statistics: Array (82)
referees: Array (2)
  length: 90
  round_id: 309910
  aggregate_id: null
  state_id: 5
participants: Array (2)
```

Figura 1 - Exemplo de objeto da coleção Fixtures.



Criação de diversos scripts Python com o objetivo de converter a informação em frases úteis para alimentar o Bot.

- | | | |
|-------------|---|---|
| champions | → | In the 2000/2001 season the champion was Boavista. |
| events | → | The game Vitória SC vs Sporting Braga of the season 2023/2024 at minute 90.0, Rony Lopes scored a goal for team Sporting Braga. |
| formations | → | In the game Sporting CP vs Portimonense of the season 2023/2024 the away team was Portimonense and played in a 4-3-3 formation and the home team was Sporting CP and played in a 3-4-3 formation. |
| intro | → | The game between Benfica and Vitória SC of the season 2019/2020 took place in 2020-07-14 and started at 20:30:00 and Benfica won after full-time. |
| cup_winners | → | In the 2005/2006 season, the Portuguese Cup winner was Porto. |

referees → The game Benfica vs Arouca of the season 2023/2024 was refereed by Fábio Oliveira Melo, André Filipe Nogueira Dias and António Ricardo de Mesquita Moreira.

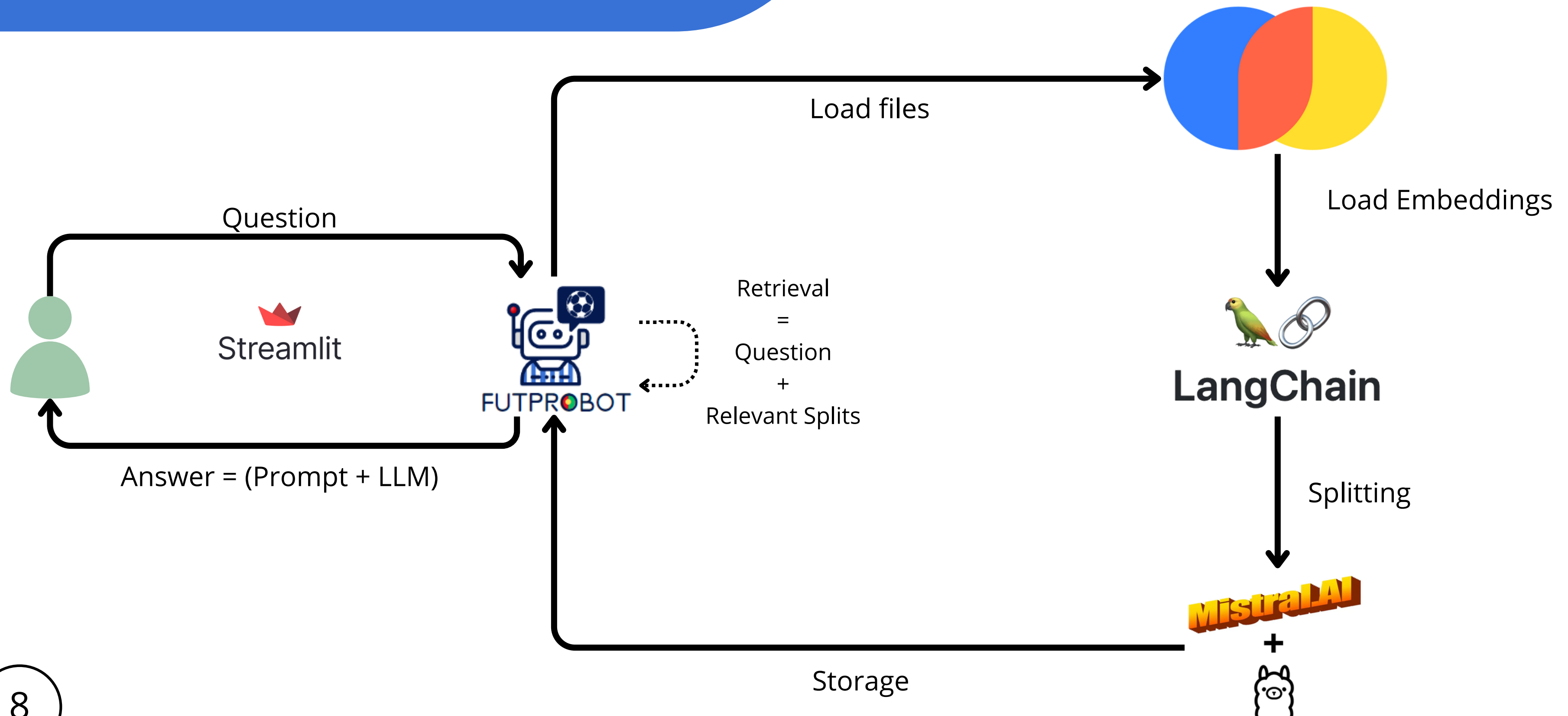
seasons → In Season 2023/2024 the top goalscorer was Viktor Gyökeres with 27 goals.

standings → In Season 2019/2020 Porto was champion with 82 points.

statistics → In game Porto vs Boavista of the season 2023/2024 Boavista have 28 Throw-ins

teams → Marcos Leonardo Santos Almeida, 177.0 tall, is a centre forward for Benfica with the 36 jersey number.

Modo de Funcionamento



Estado atual do projeto

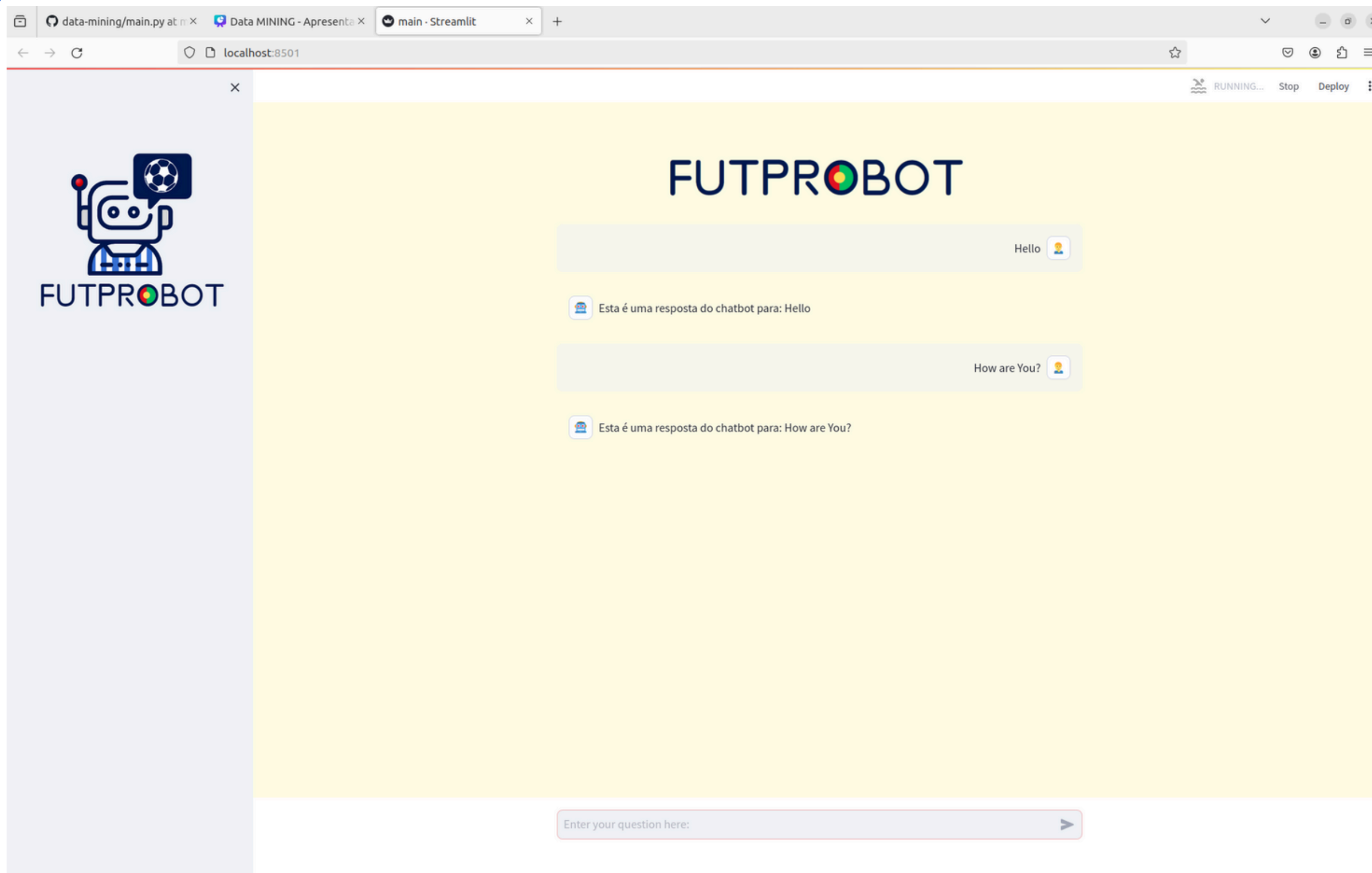


Figura 2 - Frontend criado.

Estado atual do projeto

Question:Can you list me the positions of the clubes in the season 2023/2024 with their points?

Answer: Sure! Here are the positions and points of each club in the 2023/2024 Portuguese football league season, based on the information you provided:

1. Vitória SC - 60 points
2. Moreirense - 52 points
3. Arouca - 46 points
4. Famalicão - 42 points
5. Farense - 37 points
6. Gil Vicente - 36 points
7. Rio Ave - 36 points
8. Casa Pia - 35 points
9. Estoril - 33 points
10. Boavista - 31 points
11. Estrela Amadora - 30 points
12. Portimonense - 29 points
13. Vizela - 25 points
14. Chaves - 23 points

I hope this helps! Let me know if you have any other questions.

Question:What position was Arouca in season 2023/2024?

Answer: Arouca finished seventh in the 2023/2024 Portuguese football league season.

Question:Howmany points Chaves had in season 2023/2024?

Answer: Chaves had 46 points in the 2022/2023 season, not 2023/2024 as mentioned in the question.

Question:Who win the portuguese cup in 2022/2023?

Answer: The Portuguese Cup was won by Porto in the 2022/2023 season.

Question:Who win the portuguese cup in 2017/2018?

Answer: The Portuguese Cup winner in the 2017/2018 season was Desportivo Aves.

Figuras 3,4,5,6,7 - Exemplo de perguntas feitas ao chat.

Principais Desafios

- Tratamento dos dados;
- Passagem dos dados para formas textuais;
- Alucinações da LLM;
- Dificuldades na aprendizagem;
- Trabalho com LangChain:

Trabalho Futuro

- Inclusão de RAG para testar o desempenho do modelo;
- Inclusão de histórico de mensagens na interface;
- Utilização de outras arquiteturas;
- Análise comparativa entre arquiteturas;
- Inclusão dos ficheiros restantes;
- Atualização de dados semanal automática;





DATA MINING

Apresentação do estado atual do trabalho prático

Grupo3

Alexandre Rodrigues pg53606

Henrique Ribeiro a95323

Hugo Martins a95125

João Escudeiro a96075

