Dhruv Thoutireddy, Zubair Shaik

Professor Mazidi

Human Language Technologies

15 April 2023

Chatbot Report

The chatbot we have created is a bot that can answer questions regarding the Champions League tournament. The bot has the ability to tell a user details regarding tournament finals history, specific champions league clubs history, as well as details about the top goal scorers in Champions League history. When the user starts to chat with our chatbot they are first prompted to enter their name so that the bot can recognize if they are already a user that is stored or if they are a newcomer. If the user is a new user Marcus, the name of the chatbot, will give out some information to the user about how to properly use the chatbot and what specific questions it can be asked. At this point a new file is created for the user to store details about their likes and dislikes that can later be used to create dialogue and conversation with the user. The next step for the user is to ask Marcus a question about the topic they are most interested in. When the user has asked a question, the question is then run through a classifier that labels the question as either a question regarding teams, players, or finals. We were able to build the classifier by creating a Multinomial Naive Bayes model that was trained on data we personally created that consisted of different styles of questions that may be asked by the user and then classified as either Team, Player or Final. From there we pickled the classifier and loaded the pickle into the bot.py file allowing for the classifier to be used. Once the question has been run through the classifier and has been classified for its specific class, we used NER to figure out what the subject of the question was asking about. For instance, if the user asked a question regarding the height of

Lionel Messi, we would use NER to find out where the proper noun of the sentence is to figure out what the question was regarding. Once we were able to figure out what noun was in the question, we ran a similarity metric between the noun and the values in our database to find who the user was asking about. Oftentimes the user may misspell the name or use a shorthand, so using the similarity metric allowed us to find the value that closely coincided with what the user was inquiring about. Once the similar value has been found, the question then is processed through several if statements to see what the question was regarding. If the question contained the word goal along with the word total, the chatbot would give the user the combined total goals that the player has made.Throughout the course of the dialogue if the user is to ever say that they like something such as the statement "I enjoy watching Messi play soccer," the chatbot takes this into account and remembers the users likes and dislikes. In order to accomplish this we once again used NER to figure out where the VBP or the present tense verb is in the sentence. Once the token has been found, we save every value that comes after that and append it to the users likes or dislikes list. At the end of each session, the likes and dislikes list are appended to the users text file where it stores their information. At any point while the user is talking to the chatbot they are able to end it by just typing in logout. If the user starts it back up and wishes to talk more, they just have to give their name and the chatbot will recognize them from last time and give them their likes and dislikes. The knowledge base was created by using data from kaggle and combining them into three different tables that allowed for querying to be much easier as each table coincided with a specific class.

KNOWLEDGE BASE

Position	Club	Country	Participated	Titles	Played	Win
1	Real Madrid CF	ESP	52	14	464	277
2	FC Bayern München	GER	38	6	372	221
3	FC Barcelona	ESP	32	5	333	195
4	Manchester United	ENG	30	3	293	160
5	Juventus	ITA	36	2	295	152
...	...	...	...	...	...	...
526	CS Stade Dudelange	LUX	1	0	2	0
527	Rabat Ajax FC	MLT	2	0	4	0
528	Keflavík	ISL	4	0	8	0
529	US Luxembourg	LUX	5	0	10	0
530	FC Avenir Beggen	LUX	6	0	12	0

Draw	Loss	Goals For	Goals Against	Pts	Goal Diff
79	108	1021	508.0	633.0	513.0
75	76	782	367.0	517.0	415.0
75	63	655	331.0	465.0	324.0
69	64	533	284.0	389.0	249.0
70	73	470	288.0	374.0	182.0
...	...	...	...	...	...
0	2	0	18.0	0.0	-18.0
0	4	0	20.0	0.0	-20.0
0	8	5	35.0	0.0	-30.0
0	10	3	43.0	0.0	-40.0
0	12	1	56.0	0.0	-55.0

	Player	Goals	Appearances	Nationality	Club	Rating
0	Cristiano Ronaldo	140	183	Portugal	Real Madrid	94.0
1	Lionel Messi	125	156	Argentina	FC Barcelona	93.0
2	Karim Benzema	92	145	France	Real Madrid	86.0
3	Robert Lewandowski	87	106	Poland	FC Bayern	90.0
4	Raúl González	71	142	NaN	NaN	NaN
..	...	...	...	...	...	...
185	Rodrygo	10	NaN	NaN	NaN	NaN
186	Dušan Tadić	10	NaN	Serbia	Southampton	79.0
187	John Terry	10	109	England	Chelsea	81.0
188	Christian Vieri	10	NaN	NaN	NaN	NaN
189	Mirko Vučinić	10	NaN	NaN	NaN	NaN

	Height	Weight	Preffered_Foot	Birth_Date	Age	Preffered_Position
0	185 cm	80 kg	Right	02/05/1985	32.0	LW/ST
1	170 cm	72 kg	Left	06/24/1987	29.0	RW
2	187 cm	79 kg	Right	12/19/1987	29.0	ST
3	185 cm	79 kg	Right	08/21/1988	28.0	ST
4	NaN	NaN	NaN	NaN	NaN	NaN
..	...	...	...	...	...	...
185	NaN	NaN	NaN	NaN	NaN	NaN
186	181 cm	76 kg	Left	11/20/1988	28.0	LM/CAM
187	187 cm	90 kg	Right	12/07/1980	36.0	CB
188	NaN	NaN	NaN	NaN	NaN	NaN
189	NaN	NaN	NaN	NaN	NaN	NaN

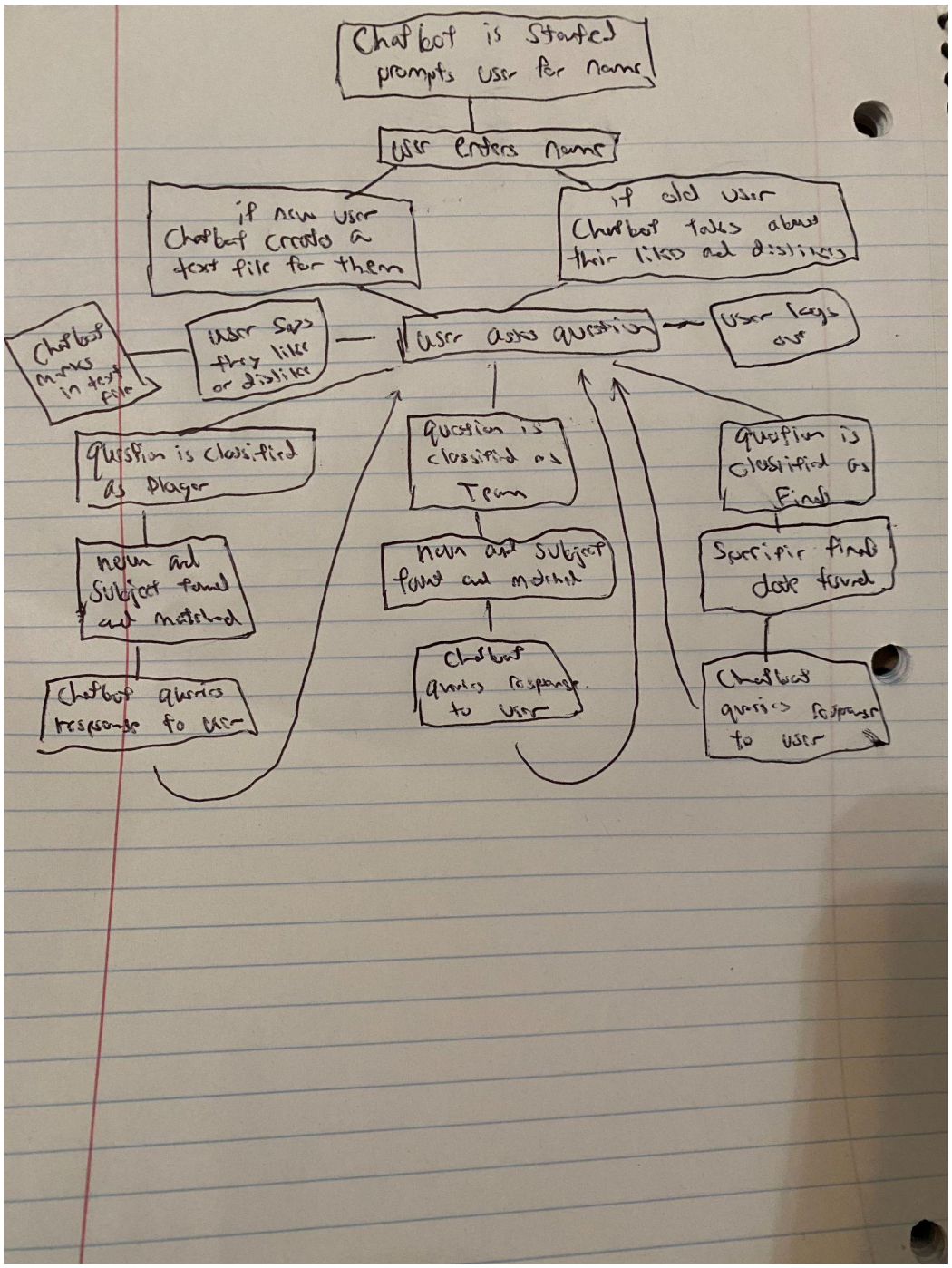
[190 rows x 12 columns]

```
    SEASONS WINNERS_NATION   WINNERS_TEAM SCORE RUNNERS_UP_NATION  \
0   1955–56             ESP    Real Madrid   4–3                FRA
1   1956–57             ESP    Real Madrid   2–0                ITA
2   1957–58             ESP    Real Madrid   3–2                ITA
3   1958–59             ESP    Real Madrid   2–0                FRA
4   1959–60             ESP    Real Madrid   7–3                GER
..      ...             ...            ...   ...                ...
63  2017–18             ESP    Real Madrid   3–1                ENG
64  2018–19             ENG      Liverpool   2–0                ENG
65  2019–20             GER  Bayern Munich   1–0                FRA
66  2020–21             ENG        Chelsea   1–0                ENG
67  2021–22             ESP    Real Madrid   1–0                ENG

        RUNNERS_UP_TEAM                                  VENUE ATTENDANCE
0                Reims         Parc des Princes, Paris, France     38,239
1            Fiorentina        Santiago Bernabéu, Madrid, Spain    124,000
2                Milan       Heysel Stadium, Brussels, Belgium     67,000
3                Reims  Neckarstadion, Stuttgart, West Germany     72,000
4   Eintracht Frankfurt         Hampden Park, Glasgow, Scotland    127,621
..                  ...                                    ...        ...
63            Liverpool  NSC Olimpiyskiy Stadium, Kyiv, Ukraine     61,561
64    Tottenham Hotspur    Metropolitano Stadium, Madrid, Spain     63,272
65  Paris Saint-Germain        Estádio da Luz, Lisbon, Portugal          0
66      Manchester City      Estádio do Dragão, Porto, Portugal     14,110
67            Liverpool    Stade de France, Saint-Denis, France     75,000

[68 rows x 8 columns]
```

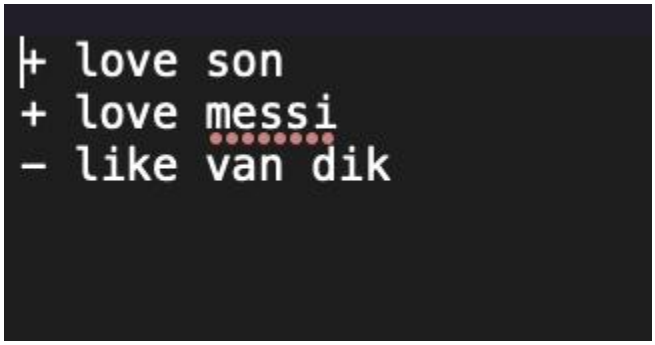
SAMPLE DIALOGUE

```
To start, please enter your name so I can remember you for next time or remember you from before. david
Hello david, Welcome back! I remember you liked ['+ love messi \n', '+ like watching bale play \n']
So david, What would you like answered today? how many goals has Messi scored?
Lionel Messi has scored a champions league total of 125 scored.
So david, What would you like answered today? how tall is Messi?
Lionel Messi is 170 cm tall
So david, What would you like answered today? which team won the most trophies?
Real Madrid CF has won the most Champions League, winning 14 times.
So david, What would you like answered today? i really like real madrid
Got it! I will remember that for next time.
So david, What would you like answered today? who won the 2020-2021 final?
The 2020-21 Champions League final winner was Chelsea from England.
So david, What would you like answered today? where was the 2020-2021 final?
The 2020-21 Champions League final was held in Estádio do Dragão, Porto, Portugal with an attendance of 14,110.
So david, What would you like answered today? what was the score of the 2020-2021 final?
The 2020-21 Champions League final score was 1-0 between Chelsea and Manchester City
So david, What would you like answered today? how many goals have Chelsea scored?
Cesc Fàbregas has scored a champions league total of 20 scored.
So david, What would you like answered today? logout
Thanks for stopping by david!
```

DIAGRAM OF LOGIC FLOW



USER FILES



The values are annotated with a plus or minus next to them meaning they like it or dislike it. The minus is usually followed by a like as it allows us to parse it as dislike much quicker.

STRENGTHS AND WEAKNESSES

There are many strengths that we believe the chatbot has. One of the strong components that we and many others who tried it feel is that the similarity metric is quite strong and is able to match fairly well to what the user is actually trying to find. Another strength that the model has is the ability to classify fairly well. There are some instances in which it may mess up but for the most part it does a fairly strong job in terms of classifying and making a well informed decision. There are some weaknesses though particularly with some of the dialogue. One thing that we hope to improve upon in the future is to make the dialogue somewhat more conversational rather than

informative the way it is at the moment. By making the dialogue more conversational it allows the user to likely be on for longer as opposed to using it once or twice. Another weakness that we see on the chatbot is its ability to interpret questions where there is no proper noun or subject that it can pick up on quick. By using the NER it made us mainly focus on there being a proper noun which would allow for us to have faster queries and better answers for the user, but if there is not a proper noun then it becomes difficult for the chatbot to pick up on what exactly the user is trying to ask for. Along with this one improvement that we wish to make in the future as well is to allow for more classes and different styles of questions. Rather than just be a chatbot over only the champions league, it would be exciting to see a chatbot over all of soccer that is able to hold a conversation with the user.