Case Study-01 Mining Twitter

Subject- NOVAK DJOKOVIC

Submitted by- Group 7

Motivation and Background

In the Australian Open 2019 men's final, that was held on January 27th, Novak Djokovic, the 31-year-old Serbian beat Rafael Nadal who has a longstanding rivalry of 13 years with Djokovic, and got his seventh Australian Open victory. The number-one-ranked player successfully dominated the game and kept his position.

In his professional career, Rafael Nadal and Roger Federer are two of the strongest rivals to Djokovic. The Trio have dominated the game over the last decade.

We thought it would be an interesting study to find out more about Djokovic's twitter profile and particularly how these high-profile athletes are able to utilize their fan-following for a noble cause – their respective (charity) foundations.

Our Business Question was: "Given the number of twitter followers (in Millions) for each of them, how efficient are they in converting/convincing/promoting their followers to also show interest in their foundation? That is, the conversion factor of Twitter account followers to also follow their foundation"

Let's look at how they matched up!

Data Source

As per the requirement, we found Novak Djokovic's Twitter account. We got his followers, friends, and his Tweets. We collected data using key word "DjokerNole" (his twitter handle) to search related Tweets. Nole is his nick name. Because Djokovic can speak Serbian, English, French, German, and Italian, his Tweets are not all in English. This also increases the difficulty of processing data.

Twitter is our main data resource. Except Twitter, we also refer to other online news and tennis game reports.

Methodology

1. A Twitter developer Account was created to generate the OAUTH logins and Tokens. Python coding was done to perform data collection, analysis and reporting of data.

2. We used the Twitter API's (like Get Followers/ids, Get Friends/ids, Get Users/lookup etc.) to collect tweets and user account information. This provided the data records necessary for the first 3 questions listed.

3. The data obtained was then used in further analysis of tweets by finding the word count, hashtags, user mentions, followers, friends and so on. This result was then plotted in tabular format.

4. A few of the data were then saved in the JSON file.

Business Question

We are trying to provide inputs to the Novak Foundation, evaluating Nole's Social Influence Transfer Rate from two aspects, by mining the mutual followers of Novak Djokovic (@DjokerNole) and Novak Foundation (@novakfoundation) and by comparing the Social Influence Transfer Rate between same level tennis players of this decade, Rafael Nadal (@RafaelNadal) and Roger Federer (@rogerfederer). By performing this evaluation we intend to guide the next steps of Novak's foundation in decision making and efforts in charity field. The logic of this business question is followed by this process graph (fig.1). In this case, the more Social Influence Transfer Rate demonstrates how the more foundation's followers come from player's followers, which is an index for company to define whether they achieve their charity purpose among player's followers.

Based on our upper limits to mining users' data fail to reach all the followers in million level within project deadline. But it is necessary to make our assumption rational, so we set the upper limit for each follower searching by ratio of two groups, and the upper limit could be the sample of our project when assuming that the rest (testers) would have the same trend. The limits will be shown in followed table (table 1).

The relationship among player's performances, Social Influence and Foundation

	Real Followers	Rounded Ratio (top:bottom)	Limits (Sample)
DjokerNole	8.53M	170 to 1	10000
novakfoundation	50.2K		60
RafaelNadal	15.6M	754 to 1	30000
frnadal	20.7K		40
RafaelNadal	12.4M	419 to1	20000
frnadal	29.6k		50

Ratio of followers and limits' selection

Table.1

Note: The above table was used to arrive at just the Sample Size to be used for all these twitter profiles and find the correlation. This does not indicate any overlap/mutual followers.

Results

Problem 1: JSON file attached

Problem 2: Tweets and Tweet Entities

1. Top 30 words

```
+--------------------+-------+
|        Word        | Count |
+--------------------+-------+
|                    |  675  |
|     DjokerNole     |  543  |
|        RT          |  417  |
|        a           |  229  |
|        of          |  223  |
|      tennis        |  211  |
|       dad          |  209  |
|       Life         |  201  |
| dadlifeisthebestlife |  200  |
|  httpstcoZsBfPRx5Zw |  200  |
|        No          |  174  |
|        is          |  104  |
|      TCurin        |   99  |
|     hitwithme      |   93  |
|     possible       |   88  |
|       Hero         |   86  |
|       arms         |   86  |
|       legs         |   86  |
|       But          |   86  |
|     anything       |   86  |
|   MondayMotivaton  |   86  |
|  httpstcomGk3aIpn3z |   86  |
|   jelenadjokovic   |   51  |
|       the          |   50  |
|        u           |   48  |
|  novakfoundation   |   45  |
|        I           |   43  |
|       and          |   37  |
|       vrti         |   35  |
|     Prnjavoru      |   35  |
|       you          |   34  |
+--------------------+-------+
```

2. Top 10 popular tweets

Count	Screen Name	Tweet ID	Text
10493	ATP_Tour	1089511083109425152	RT @ATP_Tour: The funniest video you'll watch all day 🤣 👏 @DjokerNole #AusOpen https://t.co/JMrDYNkJGQ
8312	DjokerNole	1089503231305109504	RT @DjokerNole: https://t.co/WagYOrSNF1
1877	DjokerNole	1089690634800132096	RT @DjokerNole: Wow guys we did it #TeamDjokovic #AusOpen #15 🏆 #marian #uli #gebhard #miljan #ele #edo @jelenadjokovic. There aren't enoug…
1069	DjokerNole	1094994635942842369	RT @DjokerNole: Hero 🙌🙏 "No arms. No legs. But anything is possible!" @TCurin @hitwithme #MondayMotivaton https://t.co/mGk3aIpn3z
758	DjokerNole	1093551117357572097	RT @DjokerNole: Cowaabungaaaa ⛄❄🏂🙌👍😃 https://t.co/siz5lI0J8K
753	Wimbledon	1089475900402552832	RT @Wimbledon: Unstoppable. @DjokerNole wins the #AusOpen final for his third consecutive Grand Slam title and 15th overall 🏆 https://t.co…
606	DjokerNole	1095214577082482688	RT @DjokerNole: Life of a tennis dad 🤣😂👍🎾 #dadlifeisthebestlife https://t.co/ZsBfPRx5Zw
183	novakfoundation	1092824934223618048	RT @novakfoundation: "Most importantly, you should love what you do." - @djokernole 🙌😂🎾 Who agrees?! #KeepClimbing #BelieveinTheirDreams ht…
143	denis_shapo	1089543846508281856	RT @denis_shapo: Congrats to @DjokerNole on his 7th AO and 15th slam. Truly an incredible champion!
115	MihailoTopic	1094964148117610496	RT @MihailoTopic: Traditionally we reunite and celebrate won championships! Love you so much @djokernole 🏆💙🙏🤗 Vec je postala tradicija da s…

3. Top 10 hashtags and top 10 user mentions

User Mentions	Count
DjokerNole	562
TCurin	99
hitwithme	93
jelenadjokovic	51
novakfoundation	45
pmouratoglou	26
RTS_Sport	17
ATP_Tour	17
MihailoTopic	14
BasiaID	9

Hashtag	Count
dadlifeisthebestlife	200
MondayMotivaton	86
AusOpen	15
rtsvesti	13
NDF	9
TuesdayThoughts	8
भगवान_न_भूलो	8
NittoATPFinals	4
Djokovic	4
glutenfree	3

Problem 3: Followers & Friends

(1) Friend and follower IDs

```
+--------------------+--------------------+          +--------------------+--------------------+
| Followers-ID       | Followers-ScreenName |        | Friends-ID         | Friends-ScreenName |
+--------------------+--------------------+          +--------------------+--------------------+
| 1095472421589663744 |       RosalbaVillan11 |      | click to scroll output; double click to hide   QatarTennis |
| 1095471415405334528 |          wubo62652284 |      | 439909823          |             OKSrbije |
| 1075505329050390528 |            AmnesyaC9 |        | 4105177214         |             LaverCup |
| 944915281742319616  |           BobmarkSrb |        | 182839067          |        BrainGameTennis |
| 1069933637301284864 |        Abutahir_Shafin |      | 905604214398738433 |              rodlaver |
| 1095468730379845633 |              dzenaa1 |        | 87782384           |         filipkrajinovic |
| 1095467528317489153 |        Raquel03949233 |      | 39710060           |          KAndersonATP |
| 1095467612136505352 |        Danijel67597024 |      | 759136472          |             FSSrbije |
| 305145373           |         hopeessential |       | 1161840631         |           CoSchedule |
| 1047317385680568320 |              toanhino |       | 3697308501         |          tiebreaktens |
| 57806997            |             stucas16 |        | 57471206           |         marcelomelo83 |
| 1095465192249536517 |          BPhenprapha |        | 104537355          |          olympiacosbc |
| 446100623           |             drakula40 |       | 110398012          |           matteksands |
| 1095465839707516930 |         Rajay67565207 |      | 18481641           |              LACOSTE |
| 1095465393886543872 |       armance87782096 |      | 2182094108         |            PERFECTJJIN |
| 235690555           |               brazrej |       | 724128384          |             weichen59 |
| 1095465049660030977 |         vxrtoWum17PIbnT |    | 1653548340         |             CaroleSnn |
| 1095464425836863488 |        Qrb5KHkRB7RO8qV |     | 531979487          |         Oliviahsiao0418 |
| 1095460707984920576 |        XJOhpRTQPv5pcHy |     | 17840799           |            Nic_Leigh |
| 1095294235312562177 |              OHezouani |      | 945597554          |         NadjaBraunovic |
+--------------------+--------------------+          +--------------------+--------------------+
```

(2) Compute the mutual friends within the two groups.

```
+--------------------+--------------------+
| Mutual -ID         | Mutual -ScreenName |
+--------------------+--------------------+
| 403523317          |              ZojaKu |
| 1095472421589663744 |      RosalbaVillan11 |
+--------------------+--------------------+
```

Problem 4:

A comparison is done on the foundations of the 3 ace tennis players and we see that mutual

conversion ratio (ratio between the mutual followers of foundations to player followers).

	Real Followers	Rounded Ratio (top:bottom)	Limits (Sample)	Mutual Followers	Social Influence Transfer Rate

DjokerNole	8.53M	170 to 1	10000	16	0.160%
novakfoundation	50.2K		60		
RafaelNadal	15.6M	754 to 1	30000	18	0.060%
frnadal	20.7K		40		
RafaelNadal	12.4M	419 to1	20000	10	0.050%
frnadal	29.6K		50		

```
Fetched 5000 total followers ids for DjokerNole
Fetched 10000 total followers ids for DjokerNole

Converted Followers for Novak:  16
Percentage Conversion: 0.16

Fetched 5000 total followers ids for novakfoundation
```

```
Fetched 5000 total followers ids for RafaelNadal
Fetched 10000 total followers ids for RafaelNadal
Fetched 15000 total followers ids for RafaelNadal
Fetched 20000 total followers ids for RafaelNadal
Fetched 25000 total followers ids for RafaelNadal
Fetched 30000 total followers ids for RafaelNadal

Converted Followers for Nadal:  18
Percentage Conversion: 0.06

Fetched 5000 total followers ids for frnadal
```

```
Fetched 5000 total followers ids for rogerfederer
Fetched 10000 total followers ids for rogerfederer
Fetched 15000 total followers ids for rogerfederer
Fetched 20000 total followers ids for rogerfederer

Converted Followers for Federer:  8
Percentage Conversion: 0.04

Fetched 5000 total followers ids for rogerfedererfdn
```

Conclusions

We can conclude from the business question that, by percentage convertion, though the personal followers of Novak are far less when it compared to other ace players like Rafael Nadal and Roger Federer, the mutual followers of Novak and his foundation are more. According to our social influence transfer rate definition that is mentioned in methodology part, Novak Djokovic's followers have been influenced by his efforts on charity, as well as the transfer rate is slightly higher than other world-level tennis players. On behalf of his company, we still expect he has greater influence on his followers, especially who hasn't followed his foundation yet. The solution for this expectation and or new target could be:

1. To publish more foundation information and updated new tweets in @DjokerNole twitter handle.

2. To discuss with Nadal & Federer's business teams, in term of to start a joint project which would gather more followers to concentrate charity business.

3. To explore how our foundation or content of tweets could attract more audience via followers' survey and non-followers' survey.

4. Despite these, they could partner with other brand partners like 'Lacoste' by organizing giveaways/ selling customized products publicizing the foundation etc.

Limitations

1. The sample fails to be representative

2. Authorization for searching limit in per 15 minutes fails to meet the target in an efficient way

3. Each run exports different result with randomly. (as per the latest tweets and the user Id's are also scanned and displayed in a random fashion)

4. Initially, we tried using 'Getfollowers' as a list, but this has a twitter developer rate limit of 20 users/call and 15 calls/15 minute window. So we had to change our methodology to 'Get Followers.ID' to fetch 5000 records in one call.

References

Skiver, K. (2019, January 26). Rafael Nadal vs. Novak Djokovic: Australian Open 2019 men's final, TV channel, live stream, watch online. Retrieved February1st, 2019 from https://www.cbssports.com/tennis/news/rafael-nadal-vs-novak-djokovic-australian-open-2019-mens-final-tv-channel-live-stream-watch-online/

Skiver, K. (2019, January 25). Australian Open 2019 results: Novak Djokovic dominates Lucas Pouille to advance to third straight Grand Slam final. Retrieved February1st, 2019 from https://www.cbssports.com/tennis/news/australian-open-2019-results-novak-djokovic-dominates-lucas-pouille-to-advance-to-third-straight-grand-slam-final/

Wikipedia Novak Djokovic. Retrieved February1st, 2019 from https://en.wikipedia.org/wiki/Novak_Djokovic

Wikipedia Djokovic–Federer rivalry. Retrieved February1st, 2019 from https://en.wikipedia.org/wiki/Djokovic–Federer_rivalry

Blair, A. (2019, January 26). Australian Open 2019 | Novak Djokovic talks rivalry with Rafael Nadal.Retrieved February1st, 2019 from https://www.news.com.au/sport/tennis/australian-open-2019-novak-djokovic-talks-rivalry-with-rafael-nadal/news-story/b9535c8fa8e736ffc435a8d9e14ddf9b

Novak Djokovic's seventh Champion in Australia Open. (2019, January 29). Retrieved February1st, 2019 from http://www.chinanews.com/ty/2019/01-29/8742248.shtml

Nadal:Djokovic is unstoppable when in his top condition. (2017, December 6). Retrieved February1st, 2019 from http://sports.sina.com.cn/tennis/atp/2017-12-06/doc-ifyphxwa8073062.shtml?cre=sportspagepc&mod=f&loc=3&r=9&doct=0&rfunc=100&tj=none

Novak Djokovic Foundation Raises $100K Through Season of Giving Campaign. (2019, January 11). Retrieved February1st, 2019 from https://www.tennisworldusa.org/tennis/news/Novak_Djokovic/64957/novak-djokovic-foundation-raises-100k-through-season-of-giving-campaign-/