Police Bot: Enhancing Social Media Governance with Policing Bots

Milestone 2 Presentation

Group Members:

Students:

- Liam Dumbell
- Gabriel Silva
- Cody Manning

Faculty Advisor / Project Client:

- Khaled Slhoub

Computer Science Project Instructor:

- Philip Chan

Overview:

- Discussion of Task Completion:
  - Social Media API Research
  - Development API skills on target Social Media Platform
  - Development of social media account data collection system
  - Research on Bot Detection methods
  - Development of data storage solution for our project.
- Demo of data collection and storage solution
- Faculty Advisor / Project Client Feedback
- Technical Challenges Update
- Plans heading towards Milestone 3

Task Completion

Target Social Media Platform update to Reddit

- Gaining the funding for the Twitter API subscription service is unfeasible due to high costs
- YouTube, Facebook, Instagram and Reddit considered
- Reddit selected as our target platform
  - Well documented API
  - Praw 7.7.1 Python Library
  - 24¢ per 1000 API requests (Vastly cheaper than Twitter)
  - Sizeable bot presence on Reddit

Task Completion

Learning the Reddit API and starting development

- Reviewed Documentation of Reddit API
- Learned how to use the Praw 7.7.1 Python Library
  - Learned how to make API Requests
  - Learned how to interpret the responses
  - Learned how to best to categorize Reddit account data
  - Created Social Media account data collection system using Praw

Task Completion

Development of Social Media Account Data Collection System

Purpose:
Data collection of accounts that comment under posts of popular subreddits. These accounts have a chance of being bots

Functionality:

Input (What is requested from the API):

- Target Subreddit (r/)
- Target Subreddit Filter (New / Hot / Top)
- Number of Responses
- Search Depth

Response (API response):

- Username,ID,Link Karma, Comment Karma, Total Karma, Account Age,Is Verified,Total Submissions, Total Comments

Data Collection System

```
60 # Initialize the Reddit API client
61 reddit = praw.Reddit(client_id=client,
62                       client_secret=key,
63                       user_agent=user_agent)
64 #Fetch the top posts from the "programming" subreddit
65 x = input("What subreddit you want to look at?")
66 subreddit = reddit.subreddit(x)
67 top_posts = subreddit.new(limit=1)
68 depth = int(input("Depth: "))
69
70 #for post in top_posts:
71 #    print(post.title)
72     #find_info(post.id, depth)
73 top_post_ids = [post.id for post in top_posts]
74 find_info(top_post_ids, depth)
75
76 file.close()
```

```
47 def find_info(ids, depth):
48     for i in ids:
49         try:
50             submission = reddit.submission(id=i)
51             submission.comments.replace_more(limit=depth)
52             for comment in submission.comments.list():
53                 if comment.author:
54                     display_info(comment.author)
55         except Exception as e:
56             print(f"Exception {e} ocurred")
57
```

```
15 def display_info(username):
16     user = reddit.redditor(username)
17     created = datetime.datetime.fromtimestamp(user.created_utc)
18     created = created.strftime("%d/%m/%y")
19     total_comments = 0
20     total_submissions = 0
21     for comment in user.comments.new(limit=None):
22         total_comments += 1
23     for submission in user.submissions.new(limit=None):
24         total_submissions += 1
25     data = []
26     data.append(user.id)
27     data.append(user.name)
28     data.append(user.link_karma)
29     data.append(user.comment_karma)
30     data.append(created)
31     data.append(user.verified)
32     data.append(total_submissions)
33     data.append(total_comments)
34     add_to_db(data)
35     print(f"""Username: {user.name}
36 Id: {user.id}
37 Link Karma: {user.link_karma}
38 Comment Karma: {user.comment_karma}
39 Total Karma: {user.total_karma}
40 Account age: {created}
41 Is verified: {user.verified}
42 Total submissions: {total_submissions}
43 Total comments: {total_comments}
44 """)
```

Task Completion

Research on Bot Detection methods

- Academic Papers on Bot Detection Methods were supplied by our Academic Advisor
- Our Target Bot Detection Method will be chosen from this list by Milestone 3

Task Completion

Development of data storage solution for our project.

- Data compiled, stored and updated in a .csv file every time a response is received from the Reddit API
- Easy to move to a database in the future using Python code (MySQL)

Data Storage Solution

```
 9 file = open("database.csv", mode='a', newline='')
10 db = csv.writer(file)
11
12 def add_to_db(data):
13     db.writerow(data)
14
```

	A	B	C	D	E	F	G	H
1	ID	NAME	LINK_KARMA	COMMENT_KARMA	ACC_AGE	VERIFIED	SUBMISSIONS	COMMENTS
2	3nwce0yr	del1ro	1	1019	26/04/19	TRUE	0	166
3	5tmty0ji	busdriverbuddha	66629	93973	01/03/20	TRUE	575	1000
4	5kk1wsur	jddddddddddd	299	76304	09/02/20	TRUE	38	990
5	ncg62	Icecoldkilluh	48	945	07/05/15	TRUE	6	83
6	iux4j	StormyWaters20	7628	99164	14/10/14	TRUE	508	1000
7	dttz15i9	stupsnon	1	9236	10/08/21	TRUE	3	1000
8	hm3yz45y	SchwiftyMcpoop	8	3706	15/12/21	TRUE	9	828
9	6csnaw5o	Material-Resour	722	9039	02/11/20	TRUE	0	2
10	wgpu9	georgehank2nd	29	7579	17/03/16	TRUE	15	1000
11	m4k3pmwqp	I_hate_networki	1	327	19/10/23	TRUE	1	40
12	7tbc4	metaphorm	1453	94257	26/05/12	TRUE	107	1000
13	wq9bs	diabolical_diarrh	1717	36182	29/03/16	TRUE	138	998
14	hgw4dum30	hotdog20041	145	836	12/08/23	TRUE	5	138
15	3sn4u	MattsFace	1868	12850	28/12/09	TRUE	347	999
16	d0scw	azizfcb	24311	8693	05/09/13	TRUE	218	1000
17	4itnkguo	0-Joker-0	207	996	04/09/19	TRUE	5	310
18	7hlla	wineblood	1307	85010	18/04/12	TRUE	66	848
19	1zong	ludflu	337	4110	19/06/07	TRUE	8	291
20	ba4so2fz	mtgtfo	1	27906	01/04/21	TRUE	0	1000
21	8xt4t55r	Dat_Dapper_Ow	183	6410	18/11/20	TRUE	10	999
22	bbenkqo1s	Healey_Dell	1	1485	15/05/23	TRUE	0	244
23	72w8k07v	Johan_Viisas	179	12900	05/07/20	TRUE	15	637
24	kcto96fk	Kalad1nBrood	29	456	18/07/22	TRUE	6	90
25	35pcp	BossOfTheGam	3082	51777	29/05/08	TRUE	323	1000
26	mqisxb0i	diegoquezadac2	1	2	17/01/23	TRUE	1	8
27	6atcj	virtualadept	92	25439	23/11/11	TRUE	18	1000
28	jvcu97pyi	spacebird4321	1	733	16/09/23	TRUE	2	307
29	11x1tx	losangeleskings	1168	17640	05/10/16	TRUE	65	1000
30	8xt4t55r	Dat_Dapper_Ow	183	6410	18/11/20	TRUE	10	999
31	364ad	BigGrayBeast	3150	89179	13/06/08	TRUE	240	996
32	c7tmbw5r5	Dr4gonflyaway	14	1542	28/05/23	TRUE	17	331
33	pidfbt9q	Bobmarleysjoint	1	3782	02/07/22	TRUE	0	109
34	11l5l2zb	jk_zhukov	7	4388	14/03/18	TRUE	5	615

Moving to a Database (Future)

```python
import mysql.connector
def batch_execute_ddl(conn, ddl_file_path): # connection
    cursor = conn.cursor()
    ddl_file = open(ddl_file_path)
    sql = ddl_file.read()

    for result in cursor.execute(sql, multi=True): # remove multi if you're executing 1 statement
        if result.with_rows:
            print(f"Rows returned: {result.statement}")
            print(result.fetchall())
        else:
            print(f"Number of rows affected by statement {result.statement}: {result.rowcount}")
    conn.close()
    ddl_file.close()
```

```python
import mysql.connector
def main():
    host = 'botdb.cdxq5ejycv4r.us-east-1.rds.amazonaws.com'
    user = 'Liabell'
    password = open("credentials.txt").read() # contains password
    conn = mysql.connector.connect(host=host,
                                   user=user,
                                   password=password)
    ddl_file_path = "bot_database.sql"

    batch_execute_ddl(conn, ddl_file_path)

main()
```

```python
import csv
def download_account(conn):
    cursor = conn.cursor()
    count = 0

    with open("data/imdb_ddl/database.sql", "r", encoding="UTF-8") as file:
        for line in file:
            cursor.execute(line)
        cursor.execute("SELECT COUNT(*) AS NumRowsInserted FROM accounts")
        count += cursor.fetchone()[0]
    conn.commit()
    print(count, "rows inserted for table actors")
```

Data Collection System DEMO

https://www.youtube.com/watch?v=2sVS3xRCGxo

Faculty Advisor / Project Client Feedback

- Reddit was agreed to be a good social media platform to focus on for the time being
- Our current progress with the Reddit API was deemed satisfactory
- Discussed what needed to be focused on for Milestone 3
- A Progress Evaluation document was also provided to our client that overviewed our contributions, developments, plans and feedback for Milestone 2 which was signed.

Milestone 2

Task	Completion	Cody	Gabriel	Liam	To Do
Research as many social media APIs as possible (with the possibility of switching from twitter if it becomes unfeasible)	100%	33%	33%	33%	
Gain a rudimentary understanding of the API and environment of whatever new social media platform we choose	60%	20%	20%	20%	Keep expanding knowledge of our new social media platform of choice.
Develop a system to collect basic data on social media accounts	100%	25%	50%	25%	
Research known bot detection methods	40%	20%	10%	10%	This is an ongoing process in our project, we need to break down and choose a detection method as our jumping off point
Research and potentially find a way to store the data we collect	50%	10%	10%	30%	For the moment, we just store the data in a standard csv file, a DB would be more efficient for larger datasets, so we should consider it

Milestone 3

Task	Cody	Gabriel	Liam
Improve data collection system	50%	25%	25%
Research and decide on a single starting bot detection method	33%	33%	33%
Start implementation of chosen bot detection method	25%	25%	50%
Create a working demo of rote bot detection (with the data collection integration)	25%	50%	25%

Technical Challenges Update

Progress on Resolving Challenges

- Resolved Twitter API subscription cost issue by switching target to Reddit
- Gained rudimentary experience working with the Reddit API
- Gained rudimentary experience working with the Reddit Virtual Environments
- Expanded experience working with and coding Bots
- Gained rudimentary experience working with the Praw Python Library
- Expanded HTML knowledge

Technical Challenges that require attention going forward

- We need to significantly deepen our understanding of various bot detection methods
- Gain more experience working with the Reddit API and Praw Library

Moving Towards Milestone 3:

- Continue Development for Data Collection and Storage systems
- Work on Efficiency of our Systems
- Conduct in depth research on various Bot Detection methods
- Choose a Target Bot Detection method
- Develop and implement a Target Bot Detection method into our framework
- Create a working Bot Detection System within our framework

This concludes our presentation, Thank You