Sharif University of Technology

Masoud Tahmasbi Fard

| | |

Student ID: 402200275

EE577: Massive Data Analytics

Assignment #3- Twitter User Engagement Recommender System

February 4, 2024

Table of Contents

1 Task Description

In this assignment, we implement a Twitter User Engagement Recommender System using a dataset of Persian users' tweets. Each tweet contains multiple elements describing the tweet, including the text, user id, generator uid and tweet id (for "generated"-type tweets), quoted uid and tweet id (for "quoted"-type tweets), replied uid and tweet id (for "replied"-type tweets), retweeting uid and tweet id (for "retweeted"-type tweets), and more. Leveraging these fields, we develop a recommender system in PySpark that can recommend potentially similar users and, based on that, potentially similar tweets to engage users. The implementations utilize PySpark to analyze the large-scale tweet dataset and generate personalized user and tweet recommendations. Specifically, we extract user similarity features and feed them into a collaborative filtering algorithm to produce user-based recommendations. The goal is to recommend accounts and content that are relevant and engaging to Twitter users.

2 Algorithm Description

2.1 Theoretical Model

The recommender system is implemented using a user-based collaborative filtering algorithm, which makes recommendations by finding users with similar interests and recommending tweets related with those similar users.

To find similar users, we utilize Pearson correlation (1), a technique that measures the linear relationship between two users' tweet engagement patterns. It returns a value between -1 and 1 indicating how strongly two users' engagement tendencies align. A correlation of 1 means perfect alignment where users engage with precisely the same tweets, 0 is no correlation, and -1 is perfect misalignment.

$$\operatorname{sim}(x, y)=\frac{\sum_{s \in S_{xy}}\left(r_{xs}-\overline{r_x}\right)\left(r_{ys}-\overline{r_y}\right)}{\sqrt{\sum_{s \in S_{xy}}\left(r_{xs}-\overline{r_x}\right)^2} \sqrt{\sum_{s \in S_{xy}}\left(r_{ys}-\overline{r_y}\right)^2}} \tag{1}$$

The algorithm first calculates a Pearson similarity score between the tweet engagement profile of a given user and all other users. It selects other most similar users. Then from those similar users, the algorithm examines the tweets they engaged with that the given user has not engaged with. The highest scoring of these tweets, based on the aggregate of the similar users' engagements, are recommended to the given user.

2.2 Implementation

To implement the recommender system, a tweet scoring scheme is designed to quantify a user's engagement with each tweet. The scheme assigns points as follows:

- 3 points if the user generated the original tweet
- 2 points if the user quoted or replied to the tweet
- 1 point if the user retweeted the tweet

A scoring function encapsulates this logic, taking in default parameters for generated, quoted, reply, and retweet scores that can be customized as needed.

The implementation utilizes a map-reduce framework over the user RDD. Five target users are chosen from different communities and viewpoints. For each target user, the Pearson Correlation is calculated between them and all other users to measure tweet engagement profile similarity. Specifically, the correlation examines the subset of tweets rated by both user x and user y, comparing their assigned engagement scores r_{xs} and r_{ys}. In our Pearson correlation calculation, we omit subtracting the average rating values for r_x and r_y for two reasons. First, averaging is typically done to account for user rating biases; however, in our domain user averages do not represent systematic biases. Second, some users have no rating variance (e.g. a user only retweets), causing their average to equal every rating and produce undefined correlation values. By removing averages, we avoid invalid divisions while not needing to correct biases. The raw rating vectors still capture relative tweet engagement differences between users necessary for measuring similarity. Omitting averages retains computational stability without affecting the utility of the resulting correlations for modeling tweet engagement tendencies. As a consequence, the calculated similarity is not necessarily between -1 and 1.

The most similar users to the target are selected. The tweets those similar users engaged with that the target user did not are extracted. The tweets are scored according to:

$$tweet\ score(t) = sim(x, y) * r_{yt} \tag{2}$$

Where r_{yt} is similar user y's rating of tweet t. The top scoring tweets are recommended to the target user. This personalized recommendation pipeline allows suggesting potentially relevant tweets tailored to users based on those with historically analogous engagement tendencies.

3 Results

In this section the target user's tweets, similar users id's, and recommended tweets are shown for 5 different type of users.

3.1 User #1

```
target_user = grouped_rdd.takeSample(False,1,55)[0]
target_user
```

```
('1697206791715635200',
 [('1710302619371982848', '1736250426230493213', 1),
  ('1710302619371982848', '1736250178363883872', 1),
  ('1710302619371982848', '1736249928018764235', 1),
  ('1578454116107517962', '1735951927584682444', 2),
  ('1633911157755879425', '1736108738921033809', 1),
  ('1654409094076268546', '1736287140860829901', 1),
  ('1205723159984574464', '1736622445505741159', 2),
  ('1654179648761102337', '1734655981412573264', 1),
  ('1654179648761102337', '1734657820010991962', 1),
  ('1654179648761102337', '1734661707908120774', 1),
  ('1654179648761102337', '1734661220601352349', 1),
  ('1654179648761102337', '1734659777710498122', 1),
  ('1654179648761102337', '1734658201302577473', 1),
  ('1398658879827886085', '1734656185226424457', 2),
  ('1710302619371982848', '1734799518305034308', 2),
  ('1710302619371982848', '1734798971577536976', 1),
  ('1710302619371982848', '1734798971577536976', 2),
  ('1710302619371982848', '1734798173476950425', 1),
  ('1697206791715635200', '1734826498190717030', 3),
  ('1147519752740491266', '1734531657623859526', 1),
  ('1659343792422395907', '1734526333823176854', 1),
  ('1489683639180869634', '1734555309257982198', 1),
  ('1697206791715635200', '1734830893259239623', 3),
```

Figure 1: Targer user 1

```
[ ] target_user_tweets = parsed_rdd.flatMap(lambda x: target_user_tweets_parser(target_user,x))
    target_user_tweets.collect()

    [('وصیت نامه یکی از  شهدای بزرگوار  :  خدایا نشود که در زندگی دو دستی به درخت زندگی بچسبم و همچون میوه ایی پوسیده بیافتم خدایا تا ایمانی هست مرا برای خود قربانی کن\n#گمنام_مثل_مادر\n#لشکر_علویان https://t.co/CzBw8KoKmM',
      'generated'),
     ('مهمه\u200cزندگی پر از انتخاب هایی است که مسیر ها رو عوض می کنند، اینکه قهرمان زندگی چه شخصی باشه در عاقبت بخیری خیلی\n#گمنام_مثل_مادر\n#لشکر_علویان https://t.co/QGEkchpZl4',
      'generated'),
     ('حضور مردم قدر شناس دیار علویان در مراسم وداع با لاله\u200cها\n#وداع_با_لاله\u200cها\n#گمنام_مثل_مادر https://t.co/JWvgiokZzn',
      'generated'),
     ('دیار علویان و مردم همیشه در صحنه\n#وداع_با_لاله\u200cها\n#گمنام_مثل_مادر https://t.co/1NNoFp1AGe',
      'generated'),
     ('در پی اسم و...\n#گمنام_مثل_مادر\n#وداع_با_لاله\u200cها https://t.co/QgF5gIXki9',
      'generated'),
     ('می زنند،توهین می کنند و به آب هم میزنن\u200cسنت قلم\u200cقبلا باور نمی کردم،ولی باچشمان خودم می بینم همونایکه دم از وحدت و برادری می زنند،متأسفانه اینجا علیه مقدسات اهل همه توصیه وتأکیداتشون اطاعت نمی کنید؟\u200cتنگش!نمی دونم فازشون چیه!شمایکه خودتون رو پیرو رهبری فرزانه مان میدونید،چجور از آن',
      'generated'),
     ('و اینک رهروان مکتب حاج قاسم در بابل ،فرزندان خوشنام\n\nخسته نباشی پهلوون                 \n\nمسافر من\u200cاومدی\u200cخوش      \n\nشهید گمنام سلام\nوطن را در آغوش میکشد💔\n\n#گمنام_مثل_مادر\n#لشکر_علویان https://t.co/W8TdS4BqlM',
      'generated'),
     ('گمنام_مثل_مادر\n# ها\u200cوداع_با_لاله\n# بوی مادر میدهد\n آمدنت چه آمدنیست که عجب هوای شهر\n شهید گمنام سلام',
      'generated'),
     ('چه زیباست ..\nهستند در کهف گمنامی و ندارند نام پروای که مردانی خاطرات 🤲 #گمنام_مثل_مادر\n#لشکر_علویان https://t.co/Zj4jzA9tCG',
      'generated'),
     ('گمنام_مثل_مادر\n#لشکر_علویان\n#تو نشان از مادر داری گمنام بمان\nبخدا گمنام بودن زیباست\nگمنام بمان شهید عزیز',
      'generated'),
     ('دیار علویان این روزها مهمان دارد\nاز سال مهمانانی ویژه\u200cدفاع مقدس و حماسه های\nهستند مادر مثل گمنام# که مهمانانی \n\n#لشکر_علویان https://t.co/ioQCzwZBqC',
      'generated'),
     ('لشکر_علویان\n#گمنام_مثل_مادر\n#.شهید گمنام مایه برکت برای نظام و باعث وحدت و همدلی میشود',
      'generated'),
```

Figure 2: Targer user tweets

```
[ ] similar_users = sorted_similars_rdd.top(20)
    similar_users

    [(1.6035674514745464, '1724483603600486400'),
     (1.6035674514745464, '1632351425383522312'),
     (1.6035674514745464, '1487097029990002688'),
     (1.3416407864998738, '774470366676877312'),
     (1.3416407864998738, '1725652340043100160'),
     (1.3416407864998738, '1725280261300965378'),
     (1.3416407864998738, '1721135891308007424'),
     (1.3416407864998738, '1719916251404021760'),
     (1.3416407864998738, '1719773177503682560'),
     (1.3416407864998738, '1719696919692292096'),
     (1.3416407864998738, '1718837061237530624'),
     (1.3416407864998738, '1710302619371982848'),
     (1.3416407864998738, '1708866964020514816'),
     (1.3416407864998738, '1704606183858880513'),
     (1.3416407864998738, '1704531218174251012'),
     (1.3416407864998738, '1692155672807325699'),
     (1.3416407864998738, '1689309272180482048'),
     (1.3416407864998738, '1685928050708418560'),
     (1.3416407864998738, '1675154446966312962'),
     (1.3416407864998738, '1648401324608827392')]
```

Figure 3: Similar users and corresponding similarity score

```
[ ] recommend_tweets_ids = recomm_tweets_ids[0:20]
    recommend_tweets_ids

    [(4.810702354423639, '1734904359844982845'),
     (4.810702354423639, '1734599807044349982'),
     (4.810702354423639, '1734204787200004101'),
     (4.024922359499621, '1735389039786668038'),
     (4.024922359499621, '1734604136702935127'),
     (4.024922359499621, '1735901407872475393'),
     (4.024922359499621, '1735901649871253757'),
     (4.024922359499621, '1736285402858688631'),
     (4.024922359499621, '1736285577358573735'),
     (4.024922359499621, '1736348680079233107'),
     (4.024922359499621, '1736349049442210119'),
     (4.024922359499621, '1736349214286688597'),
     (4.024922359499621, '1736349372965655032'),
     (4.024922359499621, '1734182827418095904'),
     (4.024922359499621, '1736000149338325236'),
     (4.024922359499621, '1734227232418075035'),
     (4.024922359499621, '1734234103837327546'),
     (4.024922359499621, '1735690223877501163'),
     (4.024922359499621, '1734273745986994506'),
     (4.024922359499621, '1735714482779750596')]
```

Figure 4: Recommended tweets and corresponding tweet score

Figure 5: Recommended tweets

3.2 User #2

```
[ ] target_user = grouped_rdd.takeSample(False,1,1000)[0]
    target_user

    ('1036885525326647296',
     [('1591547142136627202', '1736619970027462949', 1),
      ('1227221292685938688', '1734241138918175049', 1),
      ('1227221292685938688', '1733918649746919740', 1),
      ('1591547142136627202', '1734596007038517749', 1),
      ('1591547142136627202', '1735659470351433916', 1),
      ('1591547142136627202', '1735310756118155394', 1),
      ('1591547142136627202', '1736442052454310094', 1),
      ('1591547142136627202', '1736076770128851061', 1),
      ('1591547142136627202', '1736096711972475174', 1)])
```

Figure 6: Targer user 2

```
[ ] target_user_tweets = parsed_rdd.flatMap(lambda x: target_user_tweets_parser(target_user,x))
    target_user_tweets.collect()

    [('در مملکتی که شورای نگهبان مصلحت اندیشی کند و مجمع تشخیص مصلحت کار فقهی!هیچ چیز درست نخواهد شد.\n#روحانی',
      'generated'),
     ('فی المثل آیا با گروهای دیگر\nصد البته نظر شما برای حقیر مهم و تعیین کننده خواهد بود\n لطفا بفرمایید از بنده انتظار دارید در انتخابات پیش رو چه کنشگری داشته باشم\n#اینبار_فرق_میکند
    و\nچه مسایلی برایتان اولویت دارد ؟\nخطوط قرمز شما چیست ؟\nاگر بله چه اشخاص ، گروه ها و جریاناتی؟\nاجماع کنیم یا خیر ؟ ....',
      'generated'),
     ('باید با کمبود\nباید تکلیفمان را با #سربازی_اجباری مشخص کنیم\nیک بار  برای همیشه باید تکلیفمان را با بی تفاوتی و در مواردی شیطنت دولت پاکستان در برقراری امنیت مرزها مشخص کنیم
    باید تکلیفمان را با مسئولین بی عرضه ای که در تهران نشسته و\n.باید با عقبه ایدئولوژیک تروریست ها در مرزهای داخلی ! مشخص کنیم\n تجهیزات و نبود آموزش نیروهای مرزبانی مشخص کنیم
    برای مرزها نسخه غلط می پیچند مشخص کنیم \n#راسک\n#ایران_تسلیت',
      'generated'),
     ('در باب گفتگو هم قرار بود دیروز ایشان تشریف بیاورند دفتر ما که ظاهرا کاری برایشان پیش آمد و از جانب ایشان موکول شد به بعد از ایام شهادت @M_Afsari2 @RezaGholami1354',
      'replied'),
     ('رای.۳\n!بحث تبادل، بحثی انحرافی است،حق پدرم تبرئه است.۲\nاند\u200cگاه دیپلمات نبوده و ازسال ۷۱ که از دادستانی استعفا میدهند، کارغیردولتی داشته\u200cپدرم هیچ.۱\n:چندنکته مهم
    وسه شده صادر ۱۴۰۱تیر۲۳بدوی، \u200cمیشود صادر تجدیدنظر حکم آتی هفته شنبه\nاست نشده نهایی هنوز ابد حبس حکم!!\n#FreeHamidNouri https://t.co/DBxSHJlQVl',
      'generated'),
     ('طبق صحیح ترین منابع -۲\n منابع شیعه و سنی معتقد و معترفند هرکس بدون امام و بیعت او بمیرد به مرگ جاهلیت مرده است فلذا جهنمی است-۱\n\n ! من امام فاطمه س را امام می دانم
    دوباره طبق منابع معتبر سنی حضرت زهرا س بدون بیعت با خلیفه اول از دنیا رفت در حالی که از او خشمگین و ناراضی -۳\n اهل سنت حضرت زهرا سلام الله علیها سیده زنان بهشت است
    سوال مهم : یا حضرت زهرا س العیاذبالله جهنمی است که نیست  یا خلیفه رسول الله ابوبکر نبوده و حضرت فاطمه س در بیعت با امام و خلیفه ای دیگر (امام علی ع) از دنیا رفته\n!بود
    است.\n#یا_زهراء',
      'generated'),
     ('برادران\n فاطمه سلام الله علیها بنیانگذار این مکتب است\n… آن مظلوم باید مقتدر باشد نه ذلیل\n:منتها یک شرط دارد\n.هیچ چیز مانند مظلومیت، رسواکننده ظلم و ظالم نیست
    و\u200cهستند این مکتب پیرو ندانند یا بدانند خواه غزه در خواهرانمان.\n\n#مقتدر_مظلوم \n#غزه_تقاوم',
      'generated'),
     ('دوما: در متون روایی و تاریخی به این مهم اشاره شده است مانند اصول کافی و من لایحضره الفقیه به سند صحیح\n اولا : تقویم جمهوری اسلامی که مبنای تشخیص امر تاریخی نیست
    سوما : وفات در زبان عربی\n.قَالَ: إِنَّ فَاطِمَةَ ع صِدِّيقَةٌ شَهِيدَةٌ وَ إِنَّ بَنَاتِ الْأَنْبِيَاءِ لَا يَطْمَثْنَ.[1]امام کاظم علیه السلام می فرمایند : فاطمه (علیهاالسلام) صدیقه ای است که شهید شده  است\n:
    رابعا: ملاحظه افرادی که تحلیل شان مانند شماست باعث شد از دهه هفتاد\n شمال مرگ و شهادت است و متعدد در اسناد تاریخی و روایی درباره افراد مختلف که شهادتشان قطعی است استفاده شده
    به بعد از لفظ شهادت استفاده شود.',
      'replied')]
```

Figure 7: Targer user tweets

```
[ ] similar_users = sorted_similars_rdd.top(20)
    similar_users

    [(1.3416407864998738, '988684923203710976'),
     (1.3416407864998738, '952463173868818432'),
     (1.3416407864998738, '951201559597191169'),
     (1.3416407864998738, '948126072603754496'),
     (1.3416407864998738, '941220112845164544'),
     (1.3416407864998738, '919323412195102721'),
     (1.3416407864998738, '918806712995348480'),
     (1.3416407864998738, '911453224049164288'),
     (1.3416407864998738, '901777753833332736'),
     (1.3416407864998738, '880353064477233152'),
     (1.3416407864998738, '878382611730751488'),
     (1.3416407864998738, '878217414323961856'),
     (1.3416407864998738, '874242781790818305'),
     (1.3416407864998738, '855438360730578946'),
     (1.3416407864998738, '834665971075907585'),
     (1.3416407864998738, '822564412997402624'),
     (1.3416407864998738, '819266108351856640'),
     (1.3416407864998738, '818514349782405120'),
     (1.3416407864998738, '818336090998374400'),
     (1.3416407864998738, '814591408862416897')]
```

Figure 8: Similar users and corresponding similarity score

```
[ ] recommend_tweets_ids = recomm_tweets_ids[0:20]
    recommend_tweets_ids

    [(4.024922359499621, '1736200632980103386'),
     (4.024922359499621, '1736200676332417325'),
     (4.024922359499621, '1736571240553283866'),
     (4.024922359499621, '1736571274892296333'),
     (4.024922359499621, '1735849256391475619'),
     (4.024922359499621, '1735849284082303219'),
     (4.024922359499621, '1735867134490607746'),
     (4.024922359499621, '1736640572977250357'),
     (4.024922359499621, '1736663226119512083'),
     (4.024922359499621, '1735958921079112138'),
     (4.024922359499621, '1736695367943561440'),
     (4.024922359499621, '1736703321509892379'),
     (4.024922359499621, '1736722218724274318'),
     (4.024922359499621, '1735994535237325134'),
     (4.024922359499621, '1736358408054296791'),
     (4.024922359499621, '1735647971692204432'),
     (4.024922359499621, '1735648652553560540'),
     (4.024922359499621, '1736760514624016488'),
     (4.024922359499621, '1735654953446736373'),
     (4.024922359499621, '1735672954925515228')]
```

Figure 9: Recommended tweets and corresponding tweet score

```
[ ] recommended_tweets = parsed_rdd.flatMap(lambda x: recommended_tweets_parser(recommend_tweets_ids,x))
    recommended_tweets.collect()

    [('اللهم اشف کل\n\nاَللّهُمَّ صَلِّ عَلی فاطِمَةَ وَ أبیها وَ بَعْلِها وَ بَنیها وَ سِرِّ الْمُسْتَوْدَعِ فیها بِعَدَدِ ما أحاطَ بِهِ عِلْمُکَ\n\n أمَّن یُجیبُ المُضطَرَّ إذا دَعاهُ وَ یَکشِفُ السُّوءَ المریض..\n\n#اَللّهُمَّ_صَلِّ_عَلےمُحَمَّدٍوآلِ_مُحَمَّدٍ_وَعَجِّل_فَرَجَهُم',),
     ('الحمدلله الذی جعلنا من المتمسکین بولایة امیرالمومنین علی بن ابیطالب و الائمة المعصومین علیهم السلام',),
     ('اللهم اشف کل\n\nاَللّهُمَّ صَلِّ عَلی فاطِمَةَ وَ أبیها وَ بَعْلِها وَ بَنیها وَ سِرِّ الْمُسْتَوْدَعِ فیها بِعَدَدِ ما أحاطَ بِهِ عِلْمُکَ\n\n أمَّن یُجیبُ المُضطَرَّ إذا دَعاهُ وَ یَکشِفُ السُّوءَ المریض..\n\n#اَللّهُمَّ_صَلِّ_عَلےمُحَمَّدٍوآلِ_مُحَمَّدٍ_وَعَجِّل_فَرَجَهُم',),
     ('الحمدلله الذی جعلنا من المتمسکین بولایة امیرالمومنین علی بن ابیطالب و الائمة المعصومین علیهم السلام',),
     ('اللهم اشف کل\n\nاَللّهُمَّ صَلِّ عَلی فاطِمَةَ وَ أبیها وَ بَعْلِها وَ بَنیها وَ سِرِّ الْمُسْتَوْدَعِ فیها بِعَدَدِ ما أحاطَ بِهِ عِلْمُکَ\n\n أمَّن یُجیبُ المُضطَرَّ إذا دَعاهُ وَ یَکشِفُ السُّوءَ المریض..\n\n#اَللّهُمَّ_صَلِّ_عَلےمُحَمَّدٍوآلِ_مُحَمَّدٍ_وَعَجِّل_فَرَجَهُم',),
     ('الحمدلله الذی جعلنا من المتمسکین بولایة امیرالمومنین علی بن ابیطالب و الائمة المعصومین علیهم السلام',),
     ('کارکنان اب سروش هنوز نرفتن سرکار بفهمن اب شون خرابه کار نمیکنه؟؟',),
     ('چند وقت پیش تو عروسی دوستم بعد چند سال رفقای دوران دانشجویی رو دیدم هر کدوم یکی دوتا بچه نوجوون داشتن و من هنوز مجرد بودم همونایی که موقع ازدواج با پز دادن شوهرشونو تو چشم ما به منی که مجرد بودم میگفتن خوش به حالت مجردی و\u200cمیکردن...\n\nواقعا باور دارم مجرد بودن من از لطف خداس',),
     ('از ۴۵ دیقه نیم ساعت قبل اذان دیگه هیچ اداره ی خراب شده ای جواب تلفن نمیده توف توی اون نمازی که شما لجن ها میخونید و کار مردم رو انجام نمیدید بعد نمازم برن ناهار و چای ساعت یک و نیم هم تعطیل میکنن میرن خونه',),
     ('هرکی بود الان کلی ازتون پول میگرفت بعد میزاشت فالو ش کنید\n\nخیلی خوش به حالتونه که بدون پول فالور مئید https://t.co/XrD6qzuMJ9',),
     ('یه آدمایی تو دایرکت میگن خانم موسوی اجازه بدید یه جلسه با هم حضوری گفتگو کنیم آشنا بشیم که ادم نمیدونه چی بگه اخه جی جیه تو به من میخوره که میخوای منو ببینی چه حرف مشترکی با من داری اخه\nتوییتر هم شده بچه بازی',),
     ('اطلاعیه مراسم عزاداری شهادت حضرت زهرا (علیها\u200cالله\u200cسلام)\n\n👤سخنران: \nاستاد #رائفی_پور\n\n👥مداحان:\nکربلایی حسین\nحاج محمدرضا طاهری\nطاهری\n\n🗓زمان: دوشنبه ۲۷ آذر، ساعت ۱۹\n\n▪️مکان: نازی آباد، خیابان بهمنیار، کوچه ۲۴، مهدیه امام جواد\u200c(علیه\u200cالسلام) https://t.co/AoiofXXhbI',),
     ('نمیدونم چطوری یسری خانما به عکس گند با ده قلم ارایش همراه با حجاب\n\nدو دیقه عکس پروفایل عوض میکنم بی خبر بی صدا،خودم معذب میشم دوباره به عکس میزارم چهره م معلوم نباشه من که این کاربرا رو بلاک میکنم\n\nزیبای چفیه و چادر!!!!! میزارید کف تایم لاین بعد خودتونم جزئیات چهره رو شرح میدید',),
     ('تو سنی ها مگه به چهلم اعتقاد دارن اصلا\nچه خبره\nکویت برای مرگ امیر کویت\u200cچهل روز عزای عمومی !!!!',),
     ('نکته میترسید جاهای دیگه کسی نیاد برنامه تون بخوره\nشما مذهبی ها، شرق تهران تو اون کوچه پس کوچه ها و خیابونای تنگ و مساجد کوچیک برنامه اجرا نکنید خدا ازتون قبول نمیکنه؟؟؟؟\nچیه اخه کلا همه چی شرق تهران\n\nوالا بقیه مردم تهرانم به اهل بیت علاقه دارن مجلس خوب باشه میرن\n\nتو نوقتون؟؟',),
     ('عه دیروز ۶ سال شد\n\nبرای تک تک لایک ها و ریت ها و منشن ها باید جواب پس بدیم\nخدا عاقبتمونو ختم به خیر کنه\n\n #MyXAnniversary https://t.co/kgIcP3LCKf',),
     ('اللهم اشف کل\n\nاَللّهُمَّ صَلِّ عَلی فاطِمَةَ وَ أبیها وَ بَعْلِها وَ بَنیها وَ سِرِّ الْمُسْتَوْدَعِ فیها بِعَدَدِ ما أحاطَ بِهِ عِلْمُکَ\n\n أمَّن یُجیبُ المُضطَرَّ إذا دَعاهُ وَ یَکشِفُ السُّوءَ المریض..\n\n#اَللّهُمَّ_صَلِّ_عَلےمُحَمَّدٍوآلِ_مُحَمَّدٍ_وَعَجِّل_فَرَجَهُم',),
     ('عه جشنواره فیلم حقیقت شروع شده چرا انقد کم تبلیغ میکنن خبرش رو ندیدم جایی',),
```

Figure 10: Recommended tweets

3.3 User #3

```
[ ] target_user = grouped_rdd.takeSample(False,1,2001)[0]
    target_user

    ('1574096104450412544',
     [('1232300332497002496', '1735364117530112342', 1),
      ('1249146360584114177', '1735407610952593460', 1),
      ('1577715113679020032', '1735408119700717955', 1),
      ('1454479575015165958', '1735368209698541690', 1),
      ('226928166', '1734521930948972600', 1),
      ('1049384664', '1734603377471062075', 1)])
```

Figure 11: Targer user 3

```
[ ] target_user_tweets = parsed_rdd.flatMap(lambda x: target_user_tweets_parser(target_user,x))
    target_user_tweets.collect()

    [('او او او\nصدای مردم ایران علیه دیکتاتور مثل همیشه جهانی شد و حتی به ارتش اسرائیل هم رسید💥\n\n#صادق_بوقی https://t.co/lFZN0C2hG3',
      'generated'),
     ('این چیزها رو نکشید بچه\u200cها 😂 https://t.co/nMb1Q5j61c', 'generated'),
     ("ملت اسپانیا می\u200cتواند بهترین الگوی میهن\u200cپرستان ایرانی در راه خیزش ملی باشد. اسپانیایی\u200cها برای احیاء هویت ملی از اسلام اموی، فاشیسم، جمهوری\u200c،
    تجزیه\u200cطلب\u200cها و در نهایت از کمونیست\u200cهای چپ\u200cگرا عبور کردند. ایرانیان نیز این توان را دارند. در این مسیر همه چیز به 'سوگند قاهره' برمی\u200cگردد.
    https://t.co/2AiM1hQWiC",
      'generated'),
     ('13 سال حبس برای #مهسا_یزدانی مادر عزیز جاویدنام #محمدجواد_زاهدی به جرم دادخواهی . \n\nنکته اصلی اینه که با بایکوت کامل مازندران توسط شبکه های خبری و حقوق بشری،دادگاه
    انقلاب سازی طی یک سال گذشته سنگین ترین احکام رو به راحتی وعدم ترس از افکارعمومی صادر کرد\nهیچوقت از مازندران نگذرید https://t.co/W5NOpA3K9V',
      'generated'),
     ('🔴لشکرکشی به #اسلو نه تنها مُوَلد محبوبیت (اصلاحطلبان) نگردید،بلکه جامعه ی آزادیخواه هویت دار و سلبریتی زده و مخالف #واقعی جمهوری اسلامی را مصمم تر در ابراز ناخشنودی از  این
    جایزه ی فرمایشی کرد تا رشته ی تیم فائزه،بینه شود.\n\nکاری از @Radiomashrooteh \nبا تشکر از @SaghiZvr https://t.co/aqSv9aGyWa',
      'generated'),
     ('چهره چپ ها و بی وطن ها روز به روز برای مردم نمایان تر می شود.\nپارسال در کامنت های گلشیفته قلب مردم مشاهده می شد، امروز خشم مردم!\n\nجا داره به این جمله پرویز ثابتی
    (در اوج محبوبیت چپ در ایران) دوباره نگاهی کنیم. https://t.co/w5N7s6L6VZ',
      'generated')]
```

Figure 12: Targer user tweets

```
[ ] similar_users = sorted_similars_rdd.top(20)
    similar_users

    [(1.3416407864998738, '945011311364198400'),
     (1.3416407864998738, '874899555892047873'),
     (1.3416407864998738, '635905500'),
     (1.3416407864998738, '2862756077'),
     (1.3416407864998738, '2290250094'),
     (1.3416407864998738, '1691135810152681493'),
     (1.3416407864998738, '1685283609635377152'),
     (1.3416407864998738, '1661997428285874177'),
     (1.3416407864998738, '1650891665077334017'),
     (1.3416407864998738, '1643665017366097920'),
     (1.3416407864998738, '1638852367868207106'),
     (1.3416407864998738, '1638667917972602880'),
     (1.3416407864998738, '1637515179394056193'),
     (1.3416407864998738, '1594224521950674946'),
     (1.3416407864998738, '1593036005455511552'),
     (1.3416407864998738, '1593027443861987328'),
     (1.3416407864998738, '1587417286842908672'),
     (1.3416407864998738, '1580960286978236416'),
     (1.3416407864998738, '1580447698200805378'),
     (1.3416407864998738, '1578814191674314762')]
```

Figure 13: Similar users and corresponding similarity score

```
[ ] recommend_tweets_ids = recomm_tweets_ids[0:20]
    recommend_tweets_ids

    [(4.024922359499621, '1734638622920298899'),
     (4.024922359499621, '1736829679644868705'),
     (4.024922359499621, '1735674240941035753'),
     (4.024922359499621, '1735703356704727448'),
     (4.024922359499621, '1734303030772683235'),
     (4.024922359499621, '1735743796418150653'),
     (4.024922359499621, '1735431247524999614'),
     (4.024922359499621, '1736551527198658808'),
     (4.024922359499621, '1735489509817532581'),
     (4.024922359499621, '1735490187638034513'),
     (4.024922359499621, '1735491111907393642'),
     (4.024922359499621, '1735492658921923071'),
     (4.024922359499621, '1735493682059452843'),
     (4.024922359499621, '1734512533652857019'),
     (4.024922359499621, '1734908880897929434'),
     (4.024922359499621, '1734665362262671482'),
     (4.024922359499621, '1734665362262671482'),
     (4.024922359499621, '1735713588088607109'),
     (4.024922359499621, '1736168545619231053'),
     (4.024922359499621, '1736169799418331389')]
```

Figure 14: Recommended tweets and corresponding tweet score

```
[ ] recommended_tweets = parsed_rdd.flatMap(lambda x: recommended_tweets_parser(recommend_tweets_ids,x))
    recommended_tweets.collect()

    [('یعنی هموطن بدون ک ایناn\و قشر دوم چیولای بی وطنn\اول عرزشیا ک خب تکلیفشون مشخصهn\😂 جالبه عکسایی که ترند شده با عکس های شاهنشاه کون دو قشر رو بشدت سوزونده
    پس چشاتو باز کن و بعد تصمیم بگیرn\ عقایدشون به عرزشیا خیلی نزدیکتر از اون چیزیه ک فکرشو کنی',),
     ('جدی شدن داستان مهسا اصلاح طلبا به فکر یه خمینی زنانه افتادن و بهترین گزینه هم مسیح بود و خب شروع کردن بولد کردنش و شناسوندن مسیح به u200c\بعد-۱n\:ماجرای نوبل از دید من
    اما بعد این ک جو اروم شد و آب از آسیاب افتادn\مردم و بعدم با اوردنش کنار شاهزاده تقریبا به چیزی ک میخواستن رسیدن',),
     ('بمدت همه چی ساکن شد و بعد دوباره اومدن مسیح بولد کنن و چی بهتر از جایزه نوبل که یهو کارما زد به کمر مسیح و رفیقاش و دستش رو شد و همونجاها بود ک امریکا گفتش دیگه-۲
    چه کنیم چه نکنیم به فکر نرگس محمدی افتادن u200c\نیازی به حمایت امنیتی ما نداری و خب اینجا دست اقایون موند تو حنا ک',),
     ('کمتر از ی ماه قبل از نوبل ، سینا اومد یکn\ج کنیم ؟n\ ! اما خود گلشیفته هم از یاد رفتهn\ !چطوری بشناسونیم؟کی بهتر از گلشیفتهn\!اما خب نرگس محمدی رو کی میشناسه ؟ هیشکی-۳
    !ساعت برنامه شو اختصاص داد ب گلشیفته و اونجا شد ک خودمم حتی تا حدی  با حرفاش همدردی کردمو خلاصه تئاتر جدید شروع شد',),
     ('اما این احمقا اینجارو نخوندن هر کیم میخوای باشی با هر سرگذشتی اما اگه یه قدم ضدملت ایران یا فرهنگn\!بولد شدن گلشیفته استارتش خورد و طبق طبع چپ ایرونی جو عالم برش داش-۴
    !مردم ایران یا جوون هایی ک برای ایران زندونی شدن و خون دادن برداری کامل تو دل ملت آگاه و ملی گرای ایران میمیری',),
     ('و در نهایت پایان نوبل کاملا خلاف نمایشی که از روی ۵۷ تو ذهن یوکشون چیده  بودن رقم خورد-۵:)))',),
     ('اگه الان مجیدرضا زنده بود قطعا شماهارو با قاتل های خودش به یک چشمn\!چهل سالگی شو ندید ولی برای وطنش جنگید و اسمش تا ابد به عنوان قهرمان توی تاریخ ایران ثبت شد
    میدیدn\؟؟؟ فکر کردین میتونین رو خون بچهامون راه برید و ما مردم لال میمونیم https://t.co/pF8ymQwm6D',),
     ('مومن هوا موندس درسته بیرونn\u200d♀️🤦 😂ی عرزشی راننده تاکسیه پنجره رو میبنده هیچی قفلشم زده نتونم باز کنمn\😂این عرزشیا علاوه بر مغز و شعور  واقعا قلب و ریه هم ندارن
    الودس ولی دلیل نمیشه یک ساعت و نیم تو هوای مونده بدون تهویه نفس بکشم😂🤦\u200d♀️😭',),
     ('n\#KingRezaPahlavi\n#جاویدشاه\n تا ابدn\ جانم فدای پادشاهم #رضاشاه_دومn\ تنها راه نجات ملت ایرانn\ تا دل و جانمان صیقل داده شود n \ هرروز این مطلب را تکرار کنیم
    https://t.co/acHNDZ0v8n',),
     ('چه خوبششش https://t.co/weYHaUxOEd',),
     ('چه خوبششش https://t.co/weYHaUxOEd',),
     ('خانواده\u200cهای دادخواه یک هشتگ #مرگ_بر_اصلاحات زدند، #فرقه_جنون زنجیر پاره کرد.',),
     ('مرگ_بر_سه_فاسد_ملا_چپی_مجاهد  تا ابد  #پاینده_ایران_یکپارچه_جاوید_شاهنشاه_رضا_شاه_دوم نترسید از ساسیند شدن، اکانتی که برای مبارزه با #ایرانستیزان ازش\n#  عاشق این شعار م
    دلداگانش\u200cو\u200cفتنه\u200c۵۷\u200cبر\u200cاستفاده نشه بهتره نباشه. #مرگ',),
     ('مجاهدین ب چه روزی افتادن:) https://t.co/tJfE0GRbEz',),
     ('واین نسل آگاه ، با دل و جان ،از خودگذشتگیn\ n\شاهمان #رضاشاه_دومn\ ما سربازان ایران آماده ایم برایn\n\« نسل ما راه خودش را پیدا کرده»n\ نسل شیک پاسارگادی چه زیبا گفتن
    کنیم n\ما_ملت_کبیریم_ایران_رو_پس_میگیریم# جون n\n\جاویدشاه# تا ابد @PahlaviReza \n#KingRezaPahlavi \n#پهلوی_امید_ملت https://t.co/cGt9cx2GKQ',),
     ('را دوست دارند در نقش شاه #ایران ببینند. رای گیریهای مختلف مجازی، آمار گیریهای حتی اصلاحطلبان صادراتی و داخلی مثل عمار ملکی وn\ #KingRezaPahlavi \nاکثریت #ایرانیان
    بلومبرگ فارسی و حتی علی تجزیه، همه پاسخهاشون یکسان بوده. اکثریت مطلق مردم خواستار برقراری مجدد پادشاهی #پهلوی هستند',),
     ('اصلاحطلبهای پفیوز صادراتی، گرگهای در لباس میش، شماها که پارسال ما و پرچم کشورمون رو مسخره میکردید. الان که دیدید در اقلیت محضید بفکر مصادره پرچم و اعتراضات ما افتاده اید؟
    با این پرچم آرتش آریامهری چه خواهید کرد؟ برای ناکام گذاشتن این حیله اتان که مثل قرآن بر نیزه کردنست،  https://t.co/q1UOPYdtUr',),
```

Figure 15: Recommended tweets

3.4 User #4

```
target_user = grouped_rdd.takeSample(False,1,777)[0]
target_user
```

```
('4702453171',
 [('4702453171', '1734001719539810691', 3),
  ('4702453171', '1734471154646049090', 3),
  ('4702453171', '1734471452450005121', 3),
  ('1653467502972682270', '1734079495177461916', 2),
  ('25710094', '1734490763394798040', 2),
  ('60171166', '1736513161631588535', 2),
  ('17532913', '1734073641908981882', 2),
  ('4702453171', '1733813004280459569', 2),
  ('4702453171', '1734086631366311958', 3),
  ('4702453171', '1734086631366311958', 2),
  ('17532913', '1734087759646871761', 2),
  ('4702453171', '1734096096526291087', 3),
  ('4702453171', '1734099408688488815', 3),
  ('4702453171', '1734099878626705917', 3),
  ('4702453171', '1734100866527817896', 3),
  ('4702453171', '1735916939380212209', 3),
  ('4702453171', '1734103110279102546', 3),
  ('4702453171', '1726253167547924649', 2),
  ('15090494', '1734104725874713032', 2),
  ('360461731', '1734106135055958309', 2),
  ('1271067471760568322', '1734105477301117383', 2),
  ('4702453171', '1734108334402228647', 3),
  ('767126934', '1734108466426380588', 2),
  ('302678816', '1736273645201182727', 2),
  ('4702453171', '1736290273133490433', 3),
  ('302678816', '1736289193905594796', 2),
  ('767126934', '1734118849220374966', 2),
  ('712752292194357249', '1735382946834518426', 2),
  ('767126934', '1734119784206172308', 2),
```

Figure 16: Targer user 4

```
[ ] target_user_tweets = parsed_rdd.flatMap(lambda x: target_user_tweets_parser(target_user,x))
    target_user_tweets.collect()
```

```
[(':)) ولی من با ۳۰۰ لیر میتونم بخرمش\nحالت عادی ۱۰۰ دلاره\nمفته واقعا\n.رو برای استور ترکیه ببینید #Blinkist شما فقط قیمت یک اپ مثل
https://t.co/snE0lCFuKD',
  'generated'),
 ('زده ۲۵ گیگ اینترنت با ۳ ماه یوتیوب پرمیوم و دیتای نامحدود برای واتساپ, یوتیوب, اسپاتیفای. همه اینا برای ۳ ماه اول ۳۷۰ n\n\.آقا چرا بسته های اینترنتی این کفار اینقدر خوبن آخه
لیر :/\n\nچرا تو ایران همچین چیزایی نداریم 😑\n#Vodafone https://t.co/xoSKCGtk1k',
  'generated'),
 ('قدم بعدی حتما خرید یک روتر\n\n😁جوشی اندرویدم که واقعا وضعیت وخیمی داره رو بعدها میدم تعمیر ولی تا اون موقع به عنوان روتر دارم استفاده میکنم ازش و خیلی خوب شده
واقعیه',
  'generated'),
 ('برای گوگل\nولی چرا سیستم پرداخت اپل اینقدرر بی دردسر تر از گوگله؟\n\n.کلی هم پایبند به قوانین مالی آمریکا هستند\nاپل هم بزرگه\nگوگل بزرگه\n\nمن درک نمیکنم این قضیه رو
همین موردم باعث میشه بازم بخوام این گوشیو نگه\n\n.باید کلی سند و مدرک و آیپی رزیدنشال جور کنی، ولی اپل میگه فقط پولو بده بهم و به روش پرداخت تو اون کشور داشته باش و تمام
دارم.',
  'generated'),
 ('@Nikolaona کل بوک\u200cمارک\u200cبزنم داخل فولدرشم مجبورم دیگه حالا ولی ها، :))',
  'replied'),
 ('@Nikolaona کل بوک\u200cمارک\u200cبزنم داخل فولدرشم مجبورم دیگه حالا ولی ها، :))',
  'replied'),
 ('امیدوارم مدت زیادی\n\nگوشیم خیلی خوب باطری نگه میداره و منی که از یک گوشی با اسکرین تایم نهایت ۲ ساعت میام، اینی که کل روز شارژ نگه میداره برام خیلی خوشحال کنندس
رو همینطوری بمونه',
  'generated'),
 ('ملت از دید آزادی اینترنت و گرون بودنش میگن که خب هر دو\n\nاون از دید اقتصادی میگه آقا وضعیت درآمدی خرابه\n.ولی هر بار که مدیرعامل شاتل توییت میزنه بهش حمله میکنن
دیگه در این حد آزادی نداریم. اینا دست حکومته و نه یک شخص و مدیریتش :))))',
  'generated'),
 ('ملت از دید آزادی اینترنت و گرون بودنش میگن که خب هر\n\nاون از دید اقتصادی میگه آقا وضعیت درآمدی خرابه\n.ولی هر بار که مدیرعامل شاتل موبایل توییت میزنه بهش حمله میکنن
دیگه در این حد آزادی نداریم. اینا دست حکومته و نه یک شخص و مدیریتش. دو :))))',
  'generated'),
 ('سال با یه گوشی سر کردم که اسکرین یک ۵\n:))) فوبیای باطری بد دارم\n.یکی بیاد یک توضیح عملی و درست حسابی بده من چطوری این آیفون رو نگه دارم که باطریش به فنا نره
۲ ساعت میداد, نیم.\nاین الان یه ۷ ساعتی میده خیلی خرکیفم 😂',
  'generated'),
```

Figure 17: Targer user tweets

```
[ ]  similar_users = sorted_similars_rdd.top(20)
     similar_users

     [(1.386750490563073, '855095592'),
      (1.386750490563073, '2564151662'),
      (1.386750490563073, '1572479734335815684'),
      (1.386750490563073, '1206642366830002179'),
      (1.386750490563073, '1152293110212263937'),
      (1.386750490563073, '1151820549653446656'),
      (1.386750490563073, '1141720339'),
      (1.3416407864998738, '941203650818560000'),
      (1.3416407864998738, '927327889'),
      (1.3416407864998738, '880550727680184320'),
      (1.3416407864998738, '860039158370316288'),
      (1.3416407864998738, '818518646968422400'),
      (1.3416407864998738, '801430424'),
      (1.3416407864998738, '75530244'),
      (1.3416407864998738, '720019748780699649'),
      (1.3416407864998738, '712037416345989123'),
      (1.3416407864998738, '36924087'),
      (1.3416407864998738, '36024428'),
      (1.3416407864998738, '30186494'),
      (1.3416407864998738, '2421831930')]
```

Figure 18: Similar users and corresponding similarity score

```
[ ]  recommend_tweets_ids = recomm_tweets_ids[0:20]
     recommend_tweets_ids

     [(4.160251471689219, '1736313883789312092'),
      (4.160251471689219, '1734302996685582511'),
      (4.160251471689219, '1736543598210908218'),
      (4.160251471689219, '1736191078921826671'),
      (4.160251471689219, '1736194209722352050'),
      (4.160251471689219, '1736196909767745970'),
      (4.160251471689219, '1735911029438083239'),
      (4.160251471689219, '1734100975072207178'),
      (4.160251471689219, '1736660798875881627'),
      (4.160251471689219, '1734851686357782537'),
      (4.160251471689219, '1734118177137717730'),
      (4.160251471689219, '1735933230006558790'),
      (4.160251471689219, '1734597062904140077'),
      (4.160251471689219, '1734599210136154571'),
      (4.160251471689219, '1736693297865806316'),
      (4.160251471689219, '1736693537289228516'),
      (4.160251471689219, '1735961908656689410'),
      (4.160251471689219, '1736321041838879073'),
      (4.160251471689219, '1734158029661077888'),
      (4.160251471689219, '1736336769023385988')]
```

Figure 19: Recommended tweets and corresponding tweet score

```
[ ] recommended_tweets = parsed_rdd.flatMap(lambda x: recommended_tweets_parser(recommend_tweets_ids,x))
    recommended_tweets.collect()

[('آخیشششش\nفی شبکه بیتکوین یه نفسی کشید :))',),
 ('توکن $SYMM خیلی چیز جالبیه\nروی هردو پروتکل BRC20 و CBRC20 همزمان و بصورت دوگانه اینسکرایب میشه.\n\nاولین هست در نوع خودش!',),
 ('بخدا این انصاف نیست تایمی که ما باید بخوابیم، آمریکاییا بیدارن همه چیو پامپ میکنن\n\nخب من الان ساعت ۵ صبح با چه حالی بلند بشم اثر بریزم تو ولت؟😑',),
 ('چنان حجم پولی تزریق شده توی اکوسیستم بیتکوین که باور نمیکنید!!\nعجیب و غریب!\n\nبعد ما میریم تسک لینیا میزنیم:))',),
 ('متاسفانه تراکنشهای خودم روی $UTXO تایید نشد و invalid بودن\nولی اگر مینت کردید الان براحتی میتونید ۱۸۰۰ دلار بفروشید :))',),
 ('چه عجب بالاخره یه کندل قرمز هم دیدیم :)\n\nولکنش به میدلاین بود (قبلا چارتشو تو کانال گذاشته بودم)\nببینیم کف، قیمت رو حمایت میکنه یا باید بریم پایینتر…\n\n#bitcoin https://t.co/97we1i01xC',),
 ('اینسکریپشن روی هر بلاکچینی بجز بیتکوین،\nارزش خرید و یا نگهداری نداره!!',),
 ('حواستون به #SRC721 باشه\nپروتکل توکنهای NFT روی SRC20\n\nخبرهای خوبی براش در راهه…',),
 ('دامیننس تتر جای خیلی حساسی هست!\n#USDT.D https://t.co/UBB1eBubL3',),
 ('در یکماه گذشته چه پروژه هایی معرفی کردم؟\n\nتوکنها:\n#AVAX 3.5x\n#ATOM 1.6x\n$TIA 1.7x\n\nروی شبکه بیتکوین:\n\n$STAMP 35x\n#SRC20 48x\n$UTXO 14x\n$SATO 2.5x\n\nاینم که جدیده ولی از همین الان داره میترکونه:\n$COM 2.5x\n\nتونستید از این پروژه ها سود بگیرید؟\nبیاید بهم بگید خوشحال بشم\nاگه استفاده کردید و سود خوبی داشتید، دونیت هم میتونید انجام بدید.(به دایرکت توییتر مراجعه کنید) ❤️🌹',),
 ('این ترند صعودیه که یکسال هست دامیننس رو نگه داشته.\nاگر به پایین شکسته بشه، آلتسیزنی در پیش داریم که بیا و ببین\n\n:)) جوری که تمام رشدهایی که تا الان دیدین در مقابلش اصلا به حساب نمیان! https://t.co/p8StFdHubZ',),
 ('این ترند صعودیه که یکسال هست دامیننس رو نگه داشته.\nاگر به پایین شکسته بشه، آلتسیزنی در پیش داریم که بیا و ببین\n\n:)) جوری که تمام رشدهایی که تا الان دیدین در مقابلش اصلا به حساب نمیان!\n#bitcoinDominance https://t.co/uHTbfNbm3Q',),
 ('قرار بر این بود پست هوشمند بشه ولی الان تبدیل شده به رقیب مخابرات توی قتل بودن.\nتاخیر در تحویل\nبسته های له و لورده\nعدم آپدیت وضعیت مرسوله\nبسته\u200cای که تحویل نشده رو پستچی محترم میزنه تحویل گردیده!',),
 ('حالا اگه مثل کره شمالی اینترنت و ارتباطات خارجی رو قطع کرده بودند، این مشکلات هک پیش نمیومد\n\nهمش تقصیر شما مردمه که نمیزارید کارشونو بکنند دیگه! :))',),
 ('حالا اگه مثل کره شمالی اینترنت و ارتباطات خارجی رو کاملا قطع کرده بودند، این مشکلات هک پیش نمیومد\n\nهمش تقصیر شما مردمه که نمیزارید کارشونو بکنند دیگه! :))',),
 ('حسرت امروز؟\n\nچرا پانک بیتمپ قشنگم رو زود فروختم 😞\n@BitmapPunk',),
 ('ترند جدید مارکت نه NFT بود و نه لایه ۲\nبلکه لایه های انگل وار\ninscription روی تمام بلاکچینها بودن!',),
 ('اعتراض مدنی فقط به قضیه #صادق_بوقی تعلق میگیره؟\nمالیات سنگین، سن بازنشستگی ناجوانمردانه، تورم و گرونی، برده روسیه شدن و تاراج و حراج کردن سرمایه های ملی، مخالفت نداشت؟!\n\nبهتر نبود این انرژی رو صرف اعتراض به چیزهای مهمتری میکردیم؟',),
 ('شبکه بیتکوین زاییده!!\n\nدر حال ریزشه و قیمتها بسیار کم شده و معامله ها بخاطر فی بالا،\n\nبازی دو سر باخت!',),
 ('یادش بخیر یه دوره\u200cای هرکی یه وبلاگ داشت و توش شعر و متن\u200cهای عاشقانه منتشر می\u200cکرد',)]
```

Figure 20: Recommended tweets

3.5 User #5

```
[ ] target_user = grouped_rdd.takeSample(False,1,91)[0]
    target_user

('1693973461306109952',
 [('1574904470835798041', '1734442494039826873', 2),
  ('1574904470835798041', '1734883099647914204', 2),
  ('1574904470835798041', '1735771172472492519', 2),
  ('1693973461306109952', '1736853602969280923', 3),
  ('1693973461306109952', '1736855151917334935', 3),
  ('1693973461306109952', '1735069630308266029', 3),
  ('1693973461306109952', '1735401606160785610', 3),
  ('1693973461306109952', '1735071498032058504', 3),
  ('1693973461306109952', '1735071498032058504', 2),
  ('1574904470835798041', '1735403820757426674', 2)])
```

Figure 21: Targer user 5

```
[ ] target_user_tweets = parsed_rdd.flatMap(lambda x: target_user_tweets_parser(target_user,x))
    target_user_tweets.collect()

    [('افشین_خماند#\n ما دیوانه شده ها\nو این دنیای مودب جای «ما» نیست\nشما» شده اید»\nحتی «تو» نیستی\n تو « دیگر»',
      'generated'),
     ('گلم قرمز ، بنفش و اکنون\n...از آن روز به بعد شدم باغبان باغچه ی کوچکم با یک گل\n...تمام سیاره ها را گشتم تا به جایی رسیدم که خالی از هر کسی بود و آنگاه دلم آرام گریست
    آبی را در خود دارد...',
      'generated'),
     ('نه! هیچی، چیزی نفهمیدم، فقط خوابیدم.\n!!...امروز خواستم از عشق بگم، دیدم عشقی نیست! خواستم از عاشقونه ها حرف بزنم دیدم چیزی نیست، یادم اومد از دل تنگ بنویسم که فهمیدم
    همین...',
      'generated'),
     ('اگه گوشات هوشیار شد، تیزش کردی و صداش رو نشنیدی،\nاگه چشات صدات کرد، سرتو بلند کردی ندیدیش، باختی\nاگه دستت بی وزن شد، حرکتش دادی و نتونستی لمسش کنی، باختی
    باختی\nاگه دلت تنگ شد، موبایلت رو ورداشتی، زنگش زدی، جوابت رو نداد، باختی\nو امشب من بازنده ترینم...',
      'generated'),
     ('گفت یعنی چی!؟ گفتم دوس داشتن یا هست یا\n...خیلی زود جواب دادم دوس داشتن که اندازه نداره رفیق\nیه روز یکی از آدمای اطرافم که دنیاش خیلی رنگیه بهم گفت: چقدر دوسم داری!؟
    اندازه نداره، نیست.\nآهان گفت و زد داد یهو. میشی متوجه داشتی دوس رو کسی وقت هر گفتم !!!میگی چی نفهمیدم باز گفت...',
      'generated'),
     ('عاقبت روزی به خانه ی ما خواهد رسید؟\nاما مگر قانون این نبود که هر آنچه دیر می آید\n شاید خیلی دیر\nاز بوی تنت فهمیدم',
      'generated')]
```

Figure 22: Targer user tweets

```
[ ] similar_users = sorted_similars_rdd.top(20)
    similar_users

    [(1.0000000000000002, '1693973461306109952'),
     (1.0, '1574904470835798041'),
     (0, '999996903185100800'),
     (0, '99999538'),
     (0, '999995055090929664'),
     (0, '999988080819015680'),
     (0, '999970696435523584'),
     (0, '999964377922134016'),
     (0, '999963219425931266'),
     (0, '999959604556697602'),
     (0, '999956594686349317'),
     (0, '999953030060675072'),
     (0, '999952212783783939'),
     (0, '999951354620727296'),
     (0, '999951276'),
     (0, '999928626987569152'),
     (0, '999925604957937665'),
     (0, '999924334402531328'),
     (0, '999922321816080386'),
     (0, '999918484485128195')]
```

Figure 23: Similar users and corresponding similarity score

```
[ ] recommend_tweets_ids = recomm_tweets_ids[0:20]
    recommend_tweets_ids

    [(3.0000000000000001, '1736853602969280923'),
     (3.0000000000000001, '1736855151917334935'),
     (3.0000000000000001, '1735069630308266029'),
     (3.0000000000000001, '1735401606160785610'),
     (3.0000000000000001, '1735071498032058504'),
     (3.0, '1734042842803110258'),
     (3.0, '1736261559083491431'),
     (3.0, '1735326293720228136'),
     (3.0, '1736124195354890571'),
     (3.0, '1735771172472492519'),
     (3.0, '1735406823128125835'),
     (3.0, '1735410144727478743'),
     (3.0, '1735422143100244168'),
     (2.0000000000000004, '1734442494039826873'),
     (2.0000000000000004, '1734883099647914204'),
     (2.0000000000000004, '1735771172472492519'),
     (2.0000000000000004, '1735071498032058504'),
     (2.0000000000000004, '1735403820757426674'),
     (0, '1735767625915859133'),
     (0, '1735872482198860152')]
```

Figure 24: Recommended tweets and corresponding tweet score

```
[ ] recommended_tweets = parsed_rdd.flatMap(lambda x: recommended_tweets_parser(recommend_tweets_ids,x))
    recommended_tweets.collect()

    [('تو\nآنجایی\nو \n"آنجا" \nنمی دانند \nکه\nچقدر\nخوشبخت\nاست!',),
     ('این شریفی\u200cمقدم که مجری محبوب آخونده و تو همه شبکه\u200cها هست(منم بلاک کرده) امروز علنا گوشی آیفون ۱۳پرومکس تو برنامه زنده گرفته دستش\nاینا دیگه تعارف رو کنار گذاشتن و به ریش ملت میخندن',),
     ('بخیر ترین صبح قطعا اونی که \nکنار کسی که دوسش داری شروع میشه…',),
     ('آنجا که تو نیستی، کجا خواهد بود…',),
     ('تو «دیگر» \nنیستی «تو» حتی\nشده اید «شما»\nو این دنیای مودب جای «ما» نیست\nما دیوانه شده ها \n#افشین_خماند',),
     ('...تمام سیاره ها را گشتم تا به جایی رسیدیم که خالی از هر کسی بود و آنگاه دلم آرام گریست\n...از آن روز به بعد شدم باغبان باغچه ی کوچکم با یک گل\nگلم قرمز ، بنفش و اکنون...آبی را در خود دارد',),
     ('گمان میکنی پایان است،\n…و خدا همه چیز را درست میکند',),
     ('امروز خواستم از عشق بگم، دیدم عشقی نیست! خواستم از عاشقونه ها حرف بزنم دیدم چیزی نیست، یادم اومد از دل تنگ بنویسم که فهمیدم...!!!\nنه! هیچی، چیزی نفهمیدم، فقط خوابیدم. همین...',),
     ('اگه دستت بی وزن شد، حرکتش دادی و تونستی لمسش کنی، باختی\nاگه چشات صدات کرد، سرتو بلند کردی ندیدیش، باختی\nاگه گوشات هوشیار شد، تیزش کردی و صداش رو نشنیدی، باختی\nاگه دلت تنگ شد، موبایلت رو ورداشتی، زنگش زدی، جوابت رو نداد، باختی\nو امشب من بازنده ترینم...',),
     ('یه روز یکی از آدمای اطرافم که دنیاش خیلی رنگیه بهم گفت: چقدر دوسم داری!؟\nخیلی زود جواب دادم دوس داشتن که اندازه نداره رفیق\n...گفت یعنی چی!؟ گفتم دوس داشتن یا هست یا نیست، اندازه نداره.\nگفت باز نفهمیدم چی میگی!!! گفتم هر وقت کسی رو دوس داشتی متوجه میشی. یهو داد زد و گفت آهان...',),
     ('@Orwellian2017k استخدام پیمانی پلیس تنها جا دست خیلیها برای رسیدن به حداقلهاس، فاسد هستن اما اکثرا شر نیستن،\nبجز واحدهای نوپو، اطلاعات، یگان ویژه و امثالشون و اشخاص سرسپرده.\nاهداف آسانی اند اما تاثیری بر هسته اصلی سرکوب(سیاه/سرسپرده/اتش به اختیار) نداره؛\nاز قدیم گفتن مرگ جزئی از این شغله.',),
     ('از بوی تنت فهمیدم\nخیلی دیر رسیدم\nشاید خیلی دیر\nاما مگر قانون این نبود که هر آنچه دیر می آید\n…عاقبت روزی به خانه ی ما خواهد رسید؟',),
     ('بغلم کن همه جا\n…شهر حسودی بکند',),
     ('ای دیدن تو دین من\nای روی تو ایمان من\nای هست تو پنهان شده\n…در هستی پنهان من',),
     ('کار می کنم و به تو فکر میکنم\nبه خانه برمی گردم وبه تو فکر میکنم\nبه آن زن تلفن میکنم و درعین حال به تو فکر میکنم\nبه او می گویم حالت چطور است؟ و به تو فکر میکنم\nکجا برویم؟ وبه تو فکر میکنم\nبه او لبخند می زنم، چشمانم را می بندم و به تو فکر می کنم...',)]
```

Figure 25: Recommended tweets

4 Future works

Here are some additional metrics and features that could be incorporated to improve user similarity analysis and tweet recommendations:

- **Hashtag usage**

  tracking which hashtags users frequently use can help identify shared topical interests. Correlating common hashtags can boost similarity.

- **Tweet entity extraction**

  extracting mentioned users, links, media, locations etc. can uncover more semantic similarities between users' tweets.

- **Follower/following overlap**

  users sharing a lot of followers/followings may have closer alignment of interests. The social graph offers useful signals.

- **Tweet timestamp analysis**

  users tweeting at similar times of day or days of week may have linguistic/interest similarities.

- **Tweet language modeling**

  topic models and distributional semantic vectors derived from tweet text can reveal similarities.

- **Profile metadata**

  shared descriptors in user profiles like location, website, bio keywords could indicate similarities.

- **Tweet propagation patterns**

  how tweets spread can indicate influence and interest similarities between original posters.

5 Conclusion

The results verify the proper functionality of the implemented recommender system, which effectively recommends tweets highly similar to a user's related tweets. The system was tested on various users with differing tweet volume, viewpoints, and tweet subjects. In all cases, accurate recommendations were generated, with only a negligible amount of uncorrelated tweets. The implementations were demonstrated for 5 sample users due to resource constraints, but could be scaled up to all users given a more powerful system configuration.

A key advantage of this system is that it can operate in both online and offline settings. When new tweets or users are added, only the affected user recommendation lists need to be efficiently updated while all other users remain unchanged. This allows incorporating new data with minimal computation by incrementally updating tweet scores as needed, rather than in real-time.

In summary, the personalized tweet recommendation engine developed provides relevant engagement suggestions by modeling user similarity based on tweet activity patterns. The map-reduce algorithm combined with Pearson correlation user profiling effectively produces customized recommendations. The system's scalability and updatability makes it well-suited for real-world usage on Twitter's firehose of ever-growing data.

Is the implemented model online or offline?

Implemented collaborative filtering algorithm is naturally more amenable to online updates:

- User tweet engagement profiles are directly compared via Pearson correlation without a full graph precompute
- New tweets or users can have their profiles appended to the system and correlated against existing ones
- Only correlations involving the new data would need recomputation

So as new tweets come in:

- Their engagements would update related user profiles

- Those user profile changes would incrementally adjust correlations and recommendations

This localized adjustment to impacted user profiles and correlations enables easy assimilation of streaming data.

Additionally, optimization strategies like approximate similarity search techniques could accelerate inbox recommendation generation over growing user bases.

The intrinsic design of profiling users, correlating their behaviors, and making recommendations avoids an upfront graph construction allowing innate online capabilities as data evolves.