



Available online at [www.qu.edu.iq/journalcm](http://www.qu.edu.iq/journalcm)

JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS

ISSN:2521-3504(online) ISSN:2074-0204(print)



## **Prediction Future Location for Multiple account users Across Social Network**

**Alia Karim Abdul Hassan<sup>a</sup> , Bashar Saadoon<sup>b</sup> , Hadeel Satar humadde<sup>c</sup>**

<sup>a</sup> Computer science Dept.University of Technology Baghdad, Iraq, Email: [110018@uotechnology.edu.iq](mailto:110018@uotechnology.edu.iq)

<sup>b</sup> Computer science Dept.University of Technology Baghdad, Iraq, Email: [110043@uotechnology.edu.iq](mailto:110043@uotechnology.edu.iq)

<sup>c</sup> Computer science Dept.University of Technology Baghdad, Iraq, Email: [111818@student.uotechnology.edu.iq](mailto:111818@student.uotechnology.edu.iq)

### ARTICLE INFO

#### Article history:

Received: 03 \11 \ 2019

Revised form: 04\12\2019

Accepted : 09/12/2019

Available online: 23\01\2020

#### Keywords:

social network, foursquare dataset, LBSN, Gowalla dataset, frequency pattern.

### ABSTRACT

The personal data of people over social networks bring many opportunities for developing a new application. Some users create multiple accounts on social networks. Correlate user information across social networks helps to design algorithms to predict the future location of multiple accounts of a single individual. Most social media site applications need to share their sites with friends and family and this feature has made many people need to predict the site. Prediction is one of the foremost essential issues that need to be investigated for mobility management In mobile computing systems. In this paper, a proposed algorithm for future location prediction of multiple account users based on using frequent pattern mining. First, apply the frequency algorithm. The proposed algorithm experimented using "Gowalla and foursquare datasets" were presented simple method can be used to infer the user location using publicly available attributes and also the geographic information associated with allocatable friends. We find that it is possible to infer the user city with high accuracy. The proposed algorithm prediction achieved accuracy 96% in the Gowalla dataset when applying the support and confidence for each region that visited the user in a specific time period through some years or many months.

DOI : 10.29304/jqcm.2019.11.4.653

## 1 . Introduction

Today became most people use the Social Media Network (SMN) [1]. There are several applications of the social network; these applications are "Facebook, Twitter, Instagram" and more others. Users use several social network applications for the purpose that needed in the life such as use Facebook to keep communication with a friend and family members. And use Twitter to post news, Foursquare for location-based social applications. And use LinkedIn for job search [2]. For this reason many of users register in multiple social networks. The problem of multiple accounts of users is believe the person sends an image or video to another user. But this user is the same username because they have multiple accounts but in along time. this person discovers that the users have multiple accounts are same and don't another user. The benefits of multiple accounts the user that registers in the signal social network the information is incomplete but the user that registers in multiple social networks the information of profile is complete to improve online services for example information diffusion and business intelligence [3]. Prediction is similar to the classification, except that we are trying to predict the value of a numerical variable (e.g., amount of purchase): rather than a class (e.g., the purchaser or No purchaser)[4]. prediction is an important task for analyzing social networks. Much of the research in data mining social networks, is focused on using the association in order to derive interesting data about the social media network such as communities, or labeling the nodes with class labels. For example, in a social network, friend links are continuously created over time. Therefore, a natural question is to determine or predict future links in the social network. The prediction process may use either the structure of the network or the attribute-information in the different modes[5].

The organization of this work follows the related work explain in section2 and the basic definition of this technique that used in this paper explains in section3 and section4 display the structure of the proposed method finally section5 presented the results and experimental for this method.

### 1. Related work

Recent research on prediction user location across the Social Media Network (SMN) using the techniques of data mining and machine learning.

1. Salvatore et. al. [6] (2011) defined new features for prediction depend on the properties of places visited by the user. They study the problem of designing a link prediction system for location-based social networks and describe a supervised learning framework that discovers the prediction features to predicate new associates among friends – of – friend and place of a friend. This work used the "Gowalla dataset" where 30% of new links among "place-friend" such as among users who visit the same places and show how this prediction 15 times smaller while still 66% thus proposed new prediction features. These results displayed a new position in a real-world link system on LBSN.
2. Eunjoon et. al [7] (2011) developed a model of human movement in location-based social network through the movement of user and mobility patterns become have a high degree of variation and freedom where the exhibit structure of patterns to geographic and social constraint by using cell phone location information. In this work used three large of the dataset, two of them from two online location-based social networks and one from social network applications. These datasets are: "Facebook, foursquare, Gowalla". in this research aim to understand human motion and dynamics and explain 10% to 30% of all human movement while 50% to 70% to explain periodic behavior.
3. Tatiana et.al [8] (2012) proposed a method to infer the home located in a social media network that becoming a large volume of personal information available on the social media network such as (professional associations). This study executes the large scale study in three popular social networks: google+, foursquare, Twitter where this work focused on inferring the user's home location that considers private attribute and also geographic data links with locatable friends. This approach achieved a high accuracy of around 67% for foursquare and 72% for Google+ and 82% for Twitter.
4. Ole.J. et.al [9] (2013) presented in this work test link prediction between two types of the dataset in the mobile social network from location-based social network and records of reality mining. location-based social network includes the Gowalla and Brightkite dataset and another type of these datasets called( Call Data Record CDR) and Gowalla dataset called( Location Data Record LDR) where contain location information. These features used for prediction. In this work apply different techniques on the dataset where these techniques are: decision tree, support vector machine(SVM), Naïve Bayes, logistic Regression. The experimental results of this work achieved higher accuracy both (reality mining and Gowalla dataset) in the "decision tree, logistic Regression" classifier: the maximum dataset size is 7,902 but the minimum dataset size is 349. The result of decision tree precision (0.99), recall is (0.96 ) and f- the measure is (0.97).

5. Shamila et.al. [10](2014) presented survey for frequency pattern mining to find links of patterns for data streams where in this work discover how this algorithms used to get frequent pattern (FP) of large transactional databases where algorithms these used in this work are: " Priory algorithm, Frequent Pattern (FP) Growth algorithm, Rapid Association Rule Mining (RARM), ECLAT algorithm and Associated Sensor Pattern Mining of Data Stream (ASPMS) frequent pattern mining algorithms". And explain the strengths and weaknesses of each algorithm in this study also compare these algorithms using the same database to understand the proprieties of these algorithms.

**2. Basic Definitions**

**3.1 Association rule [11]**

Association is typically used for locating tendencies within the information. That is, this technique tries to find groups of items that are commonly found together. The results can be used to determine which items are common ought together.

**Association rule** associations, patterns among items in large databases. The rules can then be used in a variety of ways. [4]

Frequent item can be used to create association rules, with the use of a degree known as the "confidence" [12]

**Support:-** The support of a rule  $A \Rightarrow B$  (where A and B are each items/events etc.) is defined as the proportion of transactions in the data set which contain the item set A as well as B.

$$\text{Support}(A \Rightarrow B) = \text{no. of transactions which contain the item set } A \& B / \text{total no. of transactions} \dots\dots\dots (1)[11]$$

**Confidence:-**Let A and B be two groups of items. The confidence  $\text{conf}(A \cup B)$  of the rule  $A \cup B$  is the conditional likelihood of  $A \cup B$  happening in a transaction, given that the transaction includes A. Therefore, the confidence  $\text{conf}(A \Rightarrow B)$  is defined as follows:

$$\text{conf}(A \Rightarrow B) = \text{sup}(A \cup B) / \text{sup}(A) \dots\dots (2)[12]$$

**3.2 Apriori algorithm[11]**

The Priory algorithm starts by calculating the supports of the specific items to create frequent 1-itemsets. The 1-itemsets are collective to generate candidate 2-itemsets, whose support is calculated. The frequent 2-itemsets are reserved. In general, the frequent item groups of length k are used to create the candidates of length (k + 1) for increasing values of k. Algorithm (1) represents the general steps of the apriori algorithm.

A support value is providing to the algorithm. First, the algorithm creates a list of contender itemsets, which contains all of the itemsets appearing within the dataset. Of the contender itemsets created, an itemset can be determined to be frequent if the number of transactions that it appears in is greater than the support value.

**Algorithm (1) : Apriori Algorithm**  
**Input:** S, support where S dataset , min- support=real  
**Output:** set of Frequent itemset.

---

Require:  $S \neq \phi, 0 \leq \text{min- support} \leq 1$   
 Step 1: get frequent item sets  
           freq set []  $\leftarrow$  null  
 Step 2: for all itemsets i in S do  
           if support  $\geq$  min- support then  
           freq set []  $\leftarrow$  i  
           End if  
 Step3: End for

**3.3 Frequency Pattern[13]**

Frequent Pattern Mining (FPM) is one of the maximum intensively investigated problems in terms of computational and algorithmic development.

During the primary few a long time of inquiring about in this region, the essential center of work was to discover FPM calculations with superior computational effectiveness. It to begin with populates all length-one visit designs in a visit design data-store, FP. At that point, it creates a candidate design and computes its back within the database. In the event that the bolster of the candidate design is rise to or higher than the least support threshold the design is put away in FP. The method proceeds until all the visit designs from the database are found.

### 3. Dataset[14]

In this study used two dataset for location based social network (LBSN) are:gowalla dataset and foursquare dataset.

#### A. Gowalla Dataset

is a location-based social network dataset. It is available where the user shared a location by check-in with a friend the classes of problems defines a location and social relations of users are existing this dataset is discontinued. Gowalla dataset collected by cho.et.al, it includes 6,442,890 /check-in and 196,591 /users from 1,280,969 locations defined from February 2009 - October 2010.Gowalla dataset provides API directly access publically available data. Each check-in in this dataset contains user identification and location identifier and timestamp and GPS coordinates that refers to venues itself and consider more shared among many users and lower identifiable. This thesis used the "Gowalla dataset" that consists of (35) attributes and (1000) state.

#### B. Foursquare Dataset

It is a location-based social network dataset. The foursquare dataset is one and widely and popular dataset contain 30- million users and 3 billion check-ins. it is collected by Gao et.al. it is a collection of 2,073,740 -check-in and 18,107 users from 43,063 -a location from August 2010 to November 2011. In the foursquare dataset the social ties collected directly from this dataset also this dataset contain a user identifier and place identifier and timestamp and GPS organizes where the information of location in this dataset is sense much more unique than others dataset. This means less unique pieces of information are more shared among users that help when exploited to identify users where GPS location of user performed more uniques than GPS coordinate of the venue itself. The foursquare dataset consists of (273) states.

## 4. Proposed Method (*Prediction for future location*)

In this section a proposed algorithm to predicte for future location based on Association rule and Apriori algorithm Frequency Pattern will be described.

### 5.1 Frequent Pattern Mining (*FPM*)

The first step in this part Frequent Pattern Mining that consider an analytical operation for location that finds frequent patterns, associations from data sets found in various kinds of databases such as relational databases, transactional databases, this operation aims to find the redundancy of item (region) through specific period that enables us to predict the occurrence of a specific item based on the occurrence of other items in the transaction. Frequent patterns are patterns which appear frequently of the region that users are visited in a dataset. A *frequent itemset* is one which is made up of one from these forms. The steps of (Frequent Pattern Mining (FPM)):

1. Determine the specific time period that users are visited.
2. Compute the frequency for each region.
3. Sorted the frequently from higher to lower frequencies.

algorithm (2) display the general steps of prediction for the future location where the input of this algorithm is a dataset (A or B) and the username that wants to know his future location the step1 take variable (matches) default value for matches variable is "0". In this variable put the probability of region that visited username in specific times.in step 2 using for loop to apply all record in dataset(A) after then in step3 checking if username equal to the record in dataset if yes then select the city name for this username if no go-to for loop to make another record after then determine the specific time and in step5 compute the frequent region for this username and return the output sorted( matches )and max(matches) that represent frequent pattern of region and predict region.

Figure (1) represents the structure of a proposed system that represent same e steps of the algorithm(2) where start with a request from user enter username and search in the dataset after then determine the specific time and compute the frequent pattern of region, predict region as the output of the proposed system.

#### Algorithm (2): Prediction for Future Location.

**Input: Multiple account of user, time period.**

**Output: Frequent pattern of region, predict region**

Step1: initialization matches = 0, List=[ ] // list to put all region

Step2: for each account in multiple user identification  
Get (the specific time period)

Step 3: select city name from dataset A

Step 4:List. append (city name)

Step 5 if the List [matches] =List [matches+1] then  
Matches = matches +1

Step 6: output =sorted( List)

Step 7: for i 1 to length (list)

    Compute support List [i] using equation(1)

    Compute confidence List [i] using equation(2)

Step 8: max= max(support), max (confidence)

Step 9: return output, max

For example to execute the proposed algorithm(2) on gowalla dataset for time period [2009-2010] to user. Table (1) and Figure (2) display the frequency pattern for cityname = "New York". where observed that "New York" region represent a higher redundancy in 2009 and 2010 therefore this region consider item in this method for both instances.

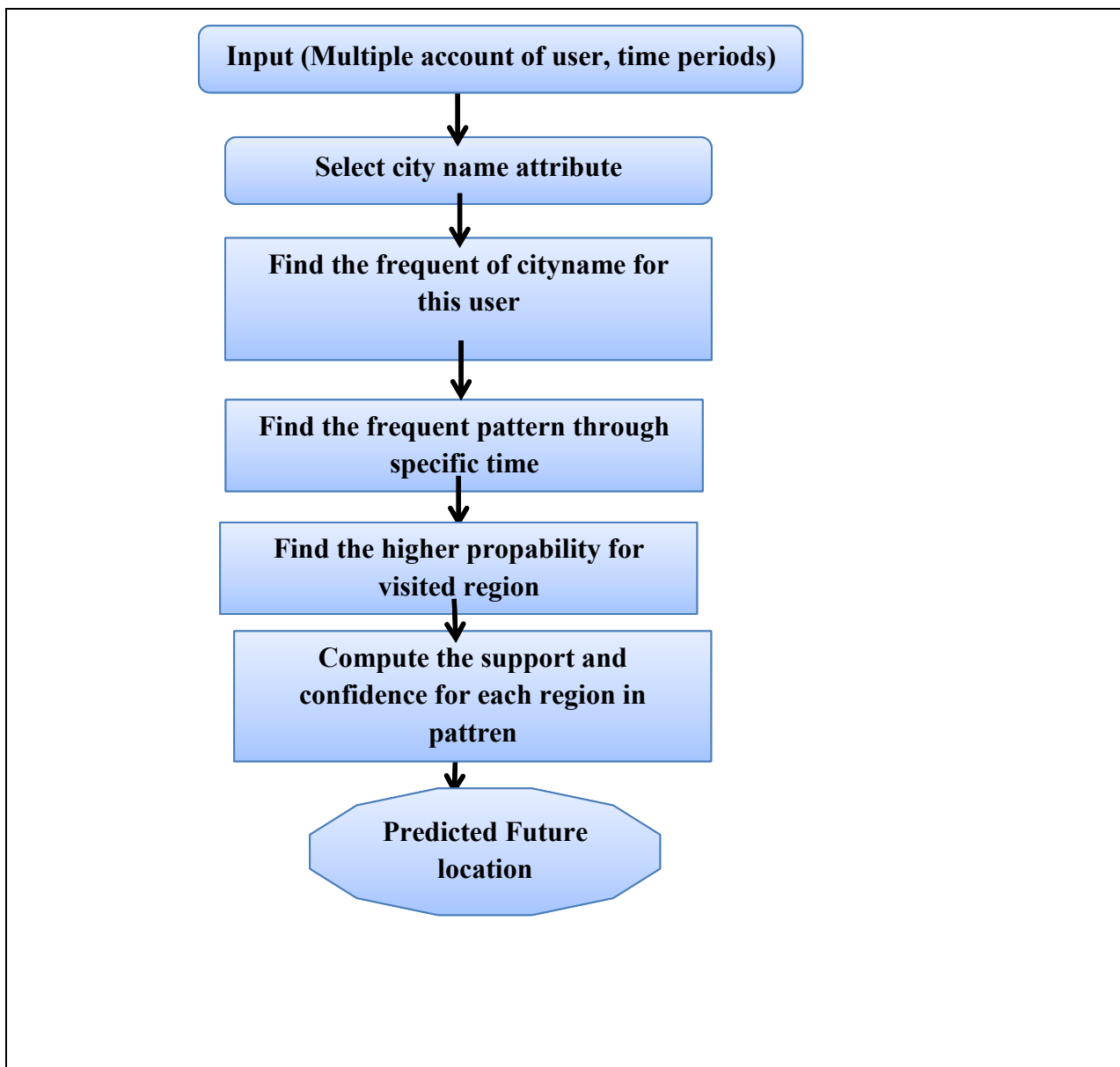
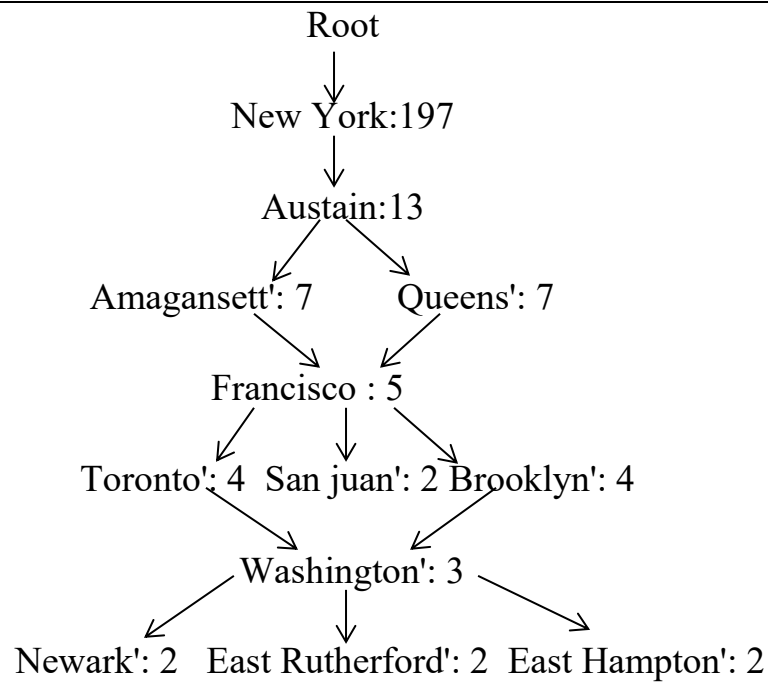
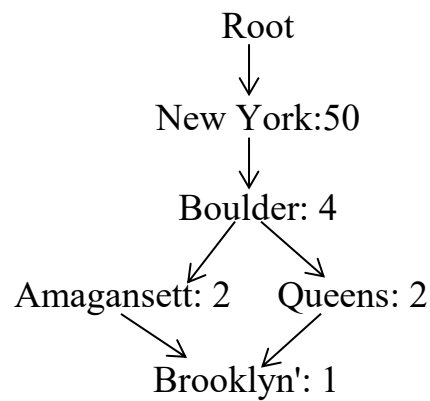


Figure (1) the structure of proposed system



**a.**



**b.**

**Figure (2) A. FPM in 2010 B. FPM in 2009**

**Table (1) Frequency pattern in 2010 and 2009**

Item	Year	Frequency pattern
New York	2010	New York, Austain, Amagansett, Queens Francisco, Toronto, Washington, Newark, East Rutherford, East Hampton.
New York	2009	New York, Boulder, Amagansett, Queens, Brooklyn

The frequency pattern is found then find the support and confidence is computed for each region by applying the a priori algorithm where take min support (0.1,0.2,0.3,0.4.... Etc.).

Where in 2010 the support and confidence are:

The support of: San Juan: 0.3

The support of: Philadelphia: 0.1

The support of: Newark: 0.1

The support of: New York: 3.0

The support of: Flushing: 0.1

The support of: Queens: 0.2

The confidence of: San Juan: 50 %

The confidence of: San Juan: 50 %

The confidence of: New York: 95 %

The confidence of: Queens: 50 %

And in 2009 the support and confidence are:

The support of: Boulder: 0.8

The support of: Brooklyn: 0.2

The support of: New York: 4.2

The support of: Amagansett: 0.4

The support of: Queens: 0.4

The confidence of: Boulder: 25 %

The confidence of: Brooklyn: 100 %

The confidence of: New York: 95 %

## 5. Implementation Experimental Results

this system to discover future location for the suspect users to follow the location during specific time period to discover current location where used in this tools two algorithm (association rule and frequency pattern). Firstly apply frequency pattern algorithm and obtained the sequential pattern of location through specific time period after then apply association rule algorithm to know the location that suspect user visited continuously and know the support and confidence for each region visited from the user suspectd. In Figure (3) Result of predication for future location.



```

Python 2.7.14 Shell
File Edit Shell Debug Options Window Help
fred-wilson 2010-03-20T13:00:02 Newark
{'u': 1, u'Austin, TX': 1, u'Arlington': 1, u'East Rutherford': 2, u'Philadelphia': 1, u'Newark': 2, u'New york': 2, u'Flushing': 1, u'San Francisco': 5, u'Brooklyn': 4, u'Queens': 7, u'Toronto': 4, u'Mississauga': 1, u'East Elmhurst': 1, u'Austin': 12, u'Amagansett': 7, u'Lockhart': 1, u'San juan': 2, u'East Hampton': 2, u'San Juan': 2, u'Washington': 3, u'New York': 195, u'Secaucus': 1}
(257, u'Newark')
fred-wilson 2009-08-05T20:39:14 Boulder
fred-wilson 2009-08-05T13:08:01 Boulder
fred-wilson 2009-08-05T02:57:59 Boulder
fred-wilson 2009-08-04T23:45:50 Boulder
fred-wilson 2009-08-04T12:32:11 New York
fred-wilson 2009-08-04T02:23:06 New York
fred-wilson 2009-08-04T01:21:21 New York
fred-wilson 2009-08-03T20:37:02 Queens
fred-wilson 2009-08-03T17:23:57 New York
fred-wilson 2009-08-02T12:50:30 Amagansett
fred-wilson 2009-08-01T16:57:52 Amagansett
fred-wilson 2009-07-31T00:36:15 New York
fred-wilson 2009-07-30T15:11:07 Brooklyn
fred-wilson 2009-07-30T12:21:59 New York
fred-wilson 2009-07-29T16:32:33 New York
fred-wilson 2009-07-28T23:09:50 New York
fred-wilson 2009-07-28T17:50:56 New York
fred-wilson 2009-07-19T19:49:44 Queens
fred-wilson 2009-07-18T22:56:34 New York
fred-wilson 2009-07-18T20:41:31 New York
fred-wilson 2009-07-18T15:56:38 New York
fred-wilson 2009-07-18T13:10:58 New York
fred-wilson 2009-07-17T17:52:03 New York
fred-wilson 2009-07-17T12:22:14 New York
fred-wilson 2009-07-16T02:24:08 New York
fred-wilson 2009-07-15T11:56:28 New York
fred-wilson 2009-07-15T01:36:44 New York
fred-wilson 2009-07-14T17:26:01 New York
fred-wilson 2009-07-14T12:00:17 New York
fred-wilson 2009-07-13T18:02:29 New York
{'u'Boulder': 4, u'Brooklyn': 1, u'New York': 21, u'Amagansett': 2, u'Queens': 2}
(29, u'New York')
>>>
Ln: 297 Col: 4

```

Figure (3) Result of predication for future location

#### For example:

The user that username "Fred Wilson" visited the region in 2010" are:

#### in 2010

fred-wilson 2010-03-20T13:00:02 Newark

```
[u': 1, u'Austin, TX': 1, u'Arlington': 1, u'East Rutherford': 2, u'Philadelphia': 1, u'Newark': 2, u'New york': 2,
u'Flushing': 1, u'San Francisco': 5, u'Brooklyn': 4, u'Queens': 7, u'Toronto': 4, u'Mississauga': 1, u'East Elmhurst': 1,
u'Austin': 12, u'Amagansett': 7, u'Lockhart': 1, u'San juan': 2, u'East Hampton': 2, u'San Juan': 2, u'Washington': 3,
u'New York': 195, u'Secaucus': 1]
```

After applying the frequency pattern algorithm obtain this result in (2010) where most regions visited is "New York" is 195

#### And in 2009

```
[Boulder: 4, Brooklyn: 1, New York: 21, Amagansett: 2, Queens: 2]
```

observed the higher number of regions that visited is ('New York': 21).

From these result from 2009 and 2010 predicate the suspect user found in the New York region.

```

Python 2.7.14 Shell
File Edit Shell Debug Options Window Help
Python 2.7.14 (v2.7.14:84471935ed, Sep 16 2017, 20:19:30) [MSC v.1500 32 bit (Intel)] on win32
Type "copyright", "credits" or "license()" for more information.
>>>
===== RESTART: C:\Python27\frequencytime.py =====
the predication in the month 2,3/2010
{'u'San Juan': 2, u'San Juan': 2, u'Philadelphia': 1, u'Newark': 1, u'New York': 21, u'Flushing': 1, u'Queens': 2}
7
21
Value : [2, 2, 1, 1, 21, 1, 2]
the support of : San Juan : 0.2857142857142857
the support of : San Juan : 0.2857142857142857
the support of : Philadelphia : 0.14285714285714285
the support of : Newark : 0.14285714285714285
the support of : New York : 3.0
the support of : Flushing : 0.14285714285714285
the support of : Queens : 0.2857142857142857
the confidence of : San Juan : 50 %
the confidence of : San Juan : 50 %
the confidence of : New York : 95 %
the confidence of : Queens : 50 %
Value : 21
['u'Queens', u'San Juan', u'San Juan', u'New York', u'Philadelphia', u'Newark', u'Flushing']
(29, u'New York')
the predication in the month 8,9/2010
{'u''': 1, u'East Hampton': 2, u'East Rutherford': 1, u'Brooklyn': 2, u'New york': 2, u'San Francisco': 5, u'New York': 80}
10
Value : [1, 2, 1, 2, 2, 5, 80, 1, 7, 2]
the support of : East Hampton : 0.2
the support of : East Rutherford : 0.1
the support of : Brooklyn : 0.2
the support of : New york : 0.2
the support of : San Francisco : 0.5
the support of : New York : 8.0
the support of : Secaucus : 0.1
the support of : Amagansett : 0.7
the support of : Queens : 0.2
the confidence of : San Francisco : 20 %
the confidence of : New York : 87 %
the confidence of : Amagansett : 28 %
(102, u'New York')

```

A.

```

the predication in the month 7.8/2009
{'u'Boulder': 4, u'Brooklyn': 1, u'New York': 21, u'Amagansett': 2, u'Queens': 2}
5
Value : [4, 1, 21, 2, 2]
the support of : Boulder : 0.8
the support of : Brooklyn : 0.2
the support of : New York : 4.2
the support of : Amagansett : 0.4
the support of : Queens : 0.4
the confidence of : Boulder : 25 %
the confidence of : Brooklyn : 100 %
the confidence of : New York : 95 %
['u'Boulder', u'New York', u'Queens', u'Amagansett', u'Brooklyn']
(29, u'New York')
the predication in 2010
{'u'Toronto': 4, u'Austin, TX': 1, u'Arlington': 1, u'Mississauga': 1, u'East Rutherford': 1, u'Washington': 1, u'East Elmhurst': 1, u'Austin': 12, u'Lockhart': 1, u'Queens': 3}
(63, u'New York')
my predication in 2010
Count of all characters in GeeksforGeeks is :

```

B.

**Figure (4) A. Result of association rule in 2010 B. Result of association rule in 2009**

In Figure (4a,b) Apply the association rule algorithm to find the result for each region that suspect users visited where divide the year into three parts to predicate for region through specific month and predication " where are you found in future location?" such as:

User name(fred-wilson ) visited in three part in 2010 and 2009 are:

the predication in the month 2,3/2010

{u'San Juan': 2, u'San Juan': 2, u'Philadelphia': 1, u'Newark': 1, u'New York': 21, u'Flushing': 1, u'Queens': 2}

the predication in the month 8,9/2010

{u''': 1, u'East Hampton': 2, u'East Rutherford': 1, u'Brooklyn': 2, u'New york': 2, u'San Francisco': 5, u'New York': 80, u'Secaucus': 1, u'Amagansett': 7, u'Queens': 2}

the predication in the month 7.8/2009

{u'Boulder': 4, u'Brooklyn': 1, u'New York': 21, u'Amagansett': 2, u'Queens': 2}

Where observed the predicate of lastly 2010 are:

the predication in 2010

{u'Toronto': 4, u'Austin, TX': 1, u'Arlington': 1, u'Mississauga': 1, u'East Rutherford': 1, u'Washington': 1, u'East Elmhurst': 1, u'Brooklyn': 1, u'New York': 37, u'Austin': 12, u'Lockhart': 1, u'Queens': 3}.

After applying the confidence and support equation to find the most region that visited suspect user where observe the result of Figure(4 A,B) in table(2).

**Table(2) Support and Confidence of region**

Region	Year	Support	Confidence
East Hampton	2010	0.2	
East Rutherford	2010	0.1	
Brooklyn	2010	0.2	
New york	2010	0.2	
San Francisco	2010	0.5	20 %
New York	2010	8.0	87 %
Secaucus	2010	0.1	
Amagansett	2010	0.7	28 %
Queens	2010	0.2	
Boulder	2009	0.8	25 %
Brooklyn	2009	0.2	10 %
New York	2009	4.2	95 %
Amagansett	2009	0.4	
Queens	2009	0.4	

In table (2) observed the support and confidence for each region that visited suspect users in 2010 and 2009 where higher confidence in " New York" region in 2010 was 87% with support 8.0 an 95% in 2009 with support 4.2 where found the confidence just for higher support from the value of confidence using to predicate the future location for suspect users.

## 6. Conclusion

In this paper, the proposed method by represented two algorithm (Frequency Pattern and Association rule) where used Frequency Pattern mining algorithm to find the series of region sequentially in specific times. And used Association rule to find the support and Confidence for each region that visited by user who want discover his future location and from apply two algorithm to find the higher probability region .the accuracy of proposed method different from one user to another where the higher accuracy is 96% in gowalla dataset.

## References

- [1] M. Dangore, G. Gholap, A. Kutal, and N. Takawale, "Identifying Unknown and Identical Users in Cross-Application Social Media Networks," vol. 5, no. 10, pp. 400–402, 2016.
- [2] Y. Li, Y. O. U. Peng, W. Ji, Z. Zhang, Q. Xu, and S. Member, "User Identification Based on Display Names Across Online Social Networks," vol. 5, 2017.
- [3] B. Guruprasath, B. Dinesh, R. K. Raman, and S. M. Riswan, "Unique User Identification across Multiple Social Networks," vol. 5, no. Iii, pp. 720–723, 2017.
- [4] G. Shmueli, P. C. Bruce, M. L. Stephens, and N. R. Patel, "DATA MINING FOR BUSINESS ANALYTICS Concepts, Techniques, and Applications with JMP Pro"2017
- [5] Aggarwal, Charu C. "An introduction to social network data analytics." Social network data analytics. Springer, Boston, MA, 1-15. 2011
- [6] Scellato, Salvatore, Anastasios Noulas, and Cecilia Mascolo. "Exploiting place features in link prediction on location-based social networks." Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2011.
- [7] Eunjoon Cho, Seth A. Myers, Jure Leskovec, "Friendship and Mobility: User Movement In Location-Based Social Networks" August 21–24, 2011..
- [8] Pontes, T., Magno, G., Vasconcelos, M., Gupta, A., Almeida, J., Kumaraguru, P., & Almeida, V." Beware of what you share: Inferring home location in social networks". IEEE 12th International Conference on Data Mining Workshops (pp. 571-578). IEEE. 2012.
- [9] O. J. Mengshoel, R. Desai, A. Chen, B. Tran, and M. Field, "Will We Connect Again ? Machine Learning for Link Prediction in Mobile Social Networks."
- [10] S. Nasreen, M. Awais, K. Shehzad, and U. Naem, "Frequent Pattern Mining Algorithms for Finding Associated Frequent Patterns for Data Streams : A Survey," *Procedia - Procedia Comput. Sci.*, vol. 37, pp. 109–116, 2014.
- [11] Lai, Kenneth, and Narciso Cerpa. "Support vs. confidence in association rule algorithms."

- In Proceedings of the OPTIMA Conference, Curicó, pp. 1-14. 2001..
- [12] Charu C. Aggarwal "Data Mining and T. Textbook" Yorktown Heights  
New York USA,2015 .
- [13] C. C. Aggarwal, M. A. Bhuiyan, and M. Al Hasan, "Frequent Pattern Mining Algorithms : A Survey,"  
2014.
- [14] L. Rossi, M. J. Williams, C. Stich, and M. Musolesi, "Privacy and the City: User Identification and  
Location Semantics in Location-Based Social Networks," 2015.