REPUBLIC OF TURKEY
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

PREDICTING VENUES IN LOCATION BASED SOCIAL NETWORK

OMAR FIRAS KANAAN ALMALLAH

MSc. THESIS
DEPARTMENT OF COMPUTER ENGINEERING
PROGRAM OF COMPUTER ENGINEERING

ADVISER
ASSOC. PROF. DR. SONGÜL VARLI ALBAYRAK

İSTANBUL, 2017

REPUBLIC OF TURKEY

YILDIZ TECHNICAL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

PREDICTING VENUES IN LOCATION BASED SOCIAL NETWORK

A thesis submitted by **OMAR FIRAS KANAAN ALMALLAH** in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE** is approved by the committee on 15.06.2017 in Department of Computer Engineering, Computer Engineering Program.

Thesis Adviser

Assoc. Prof. Dr. Songül VARLI ALBAYRAK

Yıldız Technical University

Approved By the Examining Committee

Assoc. Prof. Dr. Songül VARLI ALBAYRAK

Yıldız Technical University _____

Prof. Dr. Hasan DAĞ, Member

Kadir Has University _____

Assist. Prof. Dr. Oğuz ALTUN, Member

Yıldız Technical University _____

ACKNOWLEDGEMENTS

This manuscript is the result of three years of work in the college. During all this time, I have received the support of my family, friends, and colleagues, to whom I am incaculably grateful. These lines are just the sign of my gratitude to all of them. To beging, I would like to thank my supervisor Dr. Songül Albayrak for his constant encouragement in every aspect I had to face through these years, from the stressful talks to the (sometimes confusing) reviews received, and including those about the overwhelming first classes as a teacher and the tiresome paperwork. Thanks ever so much to Dr. Songul Albayrak for giving me the opportunity to pursue this Master's with her, for her sinscere effort and continuous work during these years, and for allowing me to participate in conferences and projects of which I have learnt so much about deadlines and priorities, and how to deal with them according to a given purpose. Special thanks to my family for their support to me in all circumstances and immediately providing me with whatever I asked for.

June 2017

Omar ALMALLAH

TABLE OF CONTENTS

Page

LIST OF SYMBOLS vi

LIST OF ABBREVIATIONS vii

LIST OF FIGURES ix

LIST OF TABLES x

ABSTRACT xii

ÖZET xiii

CHAPTER 1

INTRODUCTION 1

1.1 Literature Review 1
1.2 Aim Of Study 2
1.3 Hepothesis 2

CHAPTER 2

GENERAL INFORMATION 3

2.1 Related Work 8

CHAPTER 3

DATASET PROPERTIES 13

3.1 Dataset Prepration 16
3.2 Dataset Analysis 20

CHAPTER 4

LOCATION PREDICTION METHODS 28

4.1 Generic Recommendations 28
4.1.1 Popular Places Algorithm 29
4.1.2 Popular Category Algorithm 29
4.2 Personalized Recommendations 30

4.2.1 User History Algorithm 31
4.2.2 Collabrative Filtering 31

CHAPTER 5

EXPERMINTAL RESULTS 35

5.1 Location Prediction 35
5.1.1 Recommendation System By Using Popular Places Algorithm 36
5.1.2 Recommendation System By Using Popular Category Algorithm 37
5.1.3 Recommendation System By Using User History Algorithm 38
5.1.4 Recommendation System By Using User-User CF 40
5.2 Evaluation And Test Results 42

CHAPTER 6

CONCLUSION 45

REFERENCES 46

CURRICULUM VITAE 49

LIST OF SYMBOLS

N	Number of user
M	Number of places
C	Number of check-in
C_u	Average number of check-in per user
C_p	Average number of check-in per place
$C_{i,j}$	Check-in number for users in placs
Tp	True positive
Tn	Ture negative
Fn	False negative
Fp	False positive
u	User
s	Venue
i	Interest

LIST OF ABBREVIATIONS

APR Average Precentile Ranking

CB Content-Based Filtering

CF Collaborative Filtering

GPS Global Positioning System

LBSN Location Based Social Network

POI Point of Interest

TOA Time of Arrival

LIST OF FIGURES

Page

Figure 3.1 Averge number of check-in per user.....15
Figure 3.2 Average number of check-in per place.....15
Figure 3.3 The first status of dataset.....17
Figure 3.4 Dataset in SQL after appling Excel find and replace function.....17
Figure 3.5 How to import dataset to oracle database.....18
Figure 3.6 Data preview with importing method in oracle dataset.....18
Figure 3.7 Choose columns in oracle database.....19
Figure 3.8 Defining the column in oracle database.....19
Figure 3.9 SQL functions to clean the dataset.....20
Figure 3.10 Date Time function in SQL.....20

LIST OF TABLES

Page

Table 3.1 Average properties of the dataset: Total number of user (N), places (M) and check-in (C), average number of check-in per user (Cu) and per place(CP) 14
Table 3.2 Dataset information in SQL.....21
Table 3.3 Number of check-in for each user (Ascending)..... 21
Table 3.4 Number of check-in for each user (Descending)..... 22
Table 3.5 Number of check-in for users(Ascending) 22
Table 3.6 Number of check-in for users (Descending)..... 22
Table 3.7 Number of check-in for location depend on users (Ascending)..... 23
Table 3.8 Number of check-in for location depend on users (Descending)... 23
Table 3.9 Number of check-in for location depend on place (Ascending)..... 23
Table 3.10 Number of check-in for location depend on place (Descending).....23
Table 3.11 Number of check-in for users according to the location (Ascending) 24
Table 3.12 Number of check-in for users according to the location (Descending) 24
Table 3.13 Check-ins for users from London ..... .24
Table 3.14 Check-ins for users from Austin ..... .25
Table 3.15 Check-ins for users from Dallas.....25
Table 3.16 Number of check-in for users from Dallas.....26
Table 3.17 Number of check-in for users from Austin..... .26
Table 3.18 Number of check-in for users from London..... .26
Table 3.19 Number of check-in for user from Austin with places.....26
Table 3.20 Number of check-in for user from Dallas with places.....27
Table 3.21 Number of check-in for user from London with places.....27
Table 5.1 Austin city table for the most popular places..... 36
Table 5.2 London city table for the most popular places..... 36
Table 5.3 Dallas city table for the most popular places.....37
Table 5.4 Austin city table for the most popular categories.....37
Table 5.5 London city table for the most popular categories.....38
Table 5.6 Dallas city table for the most popular categories..... .38

Table 5.7 The most popular places in Austin city according to the history of selected users……. 39
Table 5.8 The most popular places in Dallas city according to the history of selected users…… 39
Table 5.9 The most popular places in London city according to the history of selected users ……...…….....…….. 40
Table 5.10 Austin city result by using User-user CF…….…….…..…..... 40
Table 5.11 Dallas city result by using User-user CF ……..... 41
Table 5.12 London city result by using User-user CF …….…..... 41
Table 5.13 Austin city the precision, recall, accuracy and APR……..… 43
Table 5.14 Austin precision@N, recall@N and APR@N in CF ……..... 43
Table 5.15 Dallas city the precision, recall, accuracy and APR……..... 43
Table 5.16 Dallas precision@N, recall@N and APR@N in CF ……..... .. 43
Table 5.17 London city the precision, recall, accuracy and APR…….…..... 44
Table 5.18 London precision@N, recall@N and APR@N in CF ……..... 44

ABSTRACT

PREDICTING VENUES IN LOCATION BASED SOCIAL NETWORK

Omar ALMALLAH

Department of Computer Engineering
MSc. Thesis

Advisor: Assoc. Prof. Dr. Songül VARLI ALBAYRAK

The circulation of the social networks and the evolution of the mobile phone devices has led to a big usage of location based social network applications such as Foursquare, Twitter, Swarm and Zomato on mobile phone devices, which signifies a huge dataset containing a blend of information about users behaviors, social society network of each users and also information about each of venues. All this information is available in mobile location recommendation systems. These datasets are much more different from those which are used in online recommender systems; besides, they have more information and details about the users and the venues allowing to have more clear results with much more higher accuracy of the analyzing in the results.

In this paper we examine the user's behaviors and the popularity of the venue through a dataset with large check-ins from a location based social services, i.e. Foursquare, by using large scale dataset containing both user check-in and location information. Our analysis exposes across 3 different cities. The analysis of this dataset reveals a different mobility habits, preferring places and also location patterns in the user personality. The information about the user's behaviors and each of the location popularity can be used to know the recommendation systems and to predict the next move of the users, depending on the categories form which the users attend to visit, according to the history of each user's check-ins.

Keywords: Generic Recommendations, Personalized Recommendations, Location based social networks.

YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

ÖZET

LOKASYON TABANLI SOSYAL AĞLARDA YER TAHMİNİ

Omar ALMALLAH

Bilgisayar Mühendisliği Bölümü

Yüksek Lisans Tezi

Tez Danışmanı: Doç Prof. Dr. Songül VARLI ALBAYRAK

Sosyal ağların gelişimi ve cep telefonu cihazlarının evrimi, cep telefonlarında Foursquare, Twitter, Swarm ve Zomato gibi konum tabanlı sosyal ağ uygulamalarının büyük bir kullanıma açılmasını sağladı. Bu veri seti, kullanıcıların davranışları, her bir kullanıcının sosyal paylaşım ağı ve mekânların her biri hakkında bilgi içeren mobil konum öneri sisteminde mevcut olan tüm bilgileri içermektedir. Kullandığımız veri kümeleri, çevrimiçi öneri sistemlerinde kullanılanlardan çok daha farklıdır; sonuçta analizlerin doğruluk oranları çok daha yüksektir, ayrıca kullanıcı ve mekanlar hakkında daha net sonuçlara imkân veren, daha fazla bilgi ve ayrıntıya sahiptir.

Hem kullanıcı check-in hem de konum bilgilerini içeren geniş ölçekli veri kümesini kullanarak, kullanıcıların davranışlarını ve yerin popülaritesini, konum tabanlı sosyal hizmetlerden (Foursquare) büyük bir check-in veri seti aracılığıyla inceliyoruz. Analizlerimiz 3 farklı şehirde açığa çıkıyor. Veri kümesinin analizi, kullanıcının kişiliğinden sevdiği yerleri ve konum modellerini tercih eden farklı hareket ve gezme alışkanlıklarını ortaya koymaktadır. Kullanıcının davranışları ve konum popülerliklerinin her biri hakkındaki bilgiler sayesinde, tercih edilen yerler ve sistemler bilinebilir; kullanıcıların ziyaret ettikleri kategorilere ve her kullanıcının check-in geçmişine göre kullanıcıların bir sonraki hareketleri tahmin edilebilir.

Anahtar Kelimeler: Genel Tavsiyeler, Kişiselleştirilmiş Tavsiyeler, Konum tabanlı sosyal ağlar.

CHAPTER 1

INTRODUCTION

1.1 Literature Review

The evaluation of the availability of the location- acquisition technology such as GPS empowers people to share and add their location dimension to online social networks under the name of location based social network (LBSN).

The location based social network has greatly reinforced and affected people's attraction to the social media over the recent years. The location based social network enables the users to have access to "Check-in" in any places and share their location with their online friends. Thus, it will decrease the gap between the real world and the virtual world by making a bridge between what the users are living in the real world and what the are sharing in their social media life. The availability of the large amount of social data on LBSNs provides a big unmatched chance to study the human behaviors and their daily activities through data analysis and what people are looking and searching for.

In this work, we focus on predicting venues by using location based social network methods. Also depending on the history of the users, popularity of the places and attending the places by their categories, in order to find the most visited places in our datasets. In other hand, find the similarity between each of the users by using user-user collaborative filtering, through three different cities Austin, Dallas and London each with different types of categories and number of users. At the end, the result of the three methods in the three cities is compared to find the highest accuracy, precision, recall and average percentile ranking (APR).

1.2 Aim of Study

The main aim of this thesis is to find the best recommender system for the users by applying three different methods which depends on the number of the check-ins that the users make in different places under the name of different categories for each of the venues.

The objectives of this thesis are as the follow-up:

- Gain knowledge about recommendation system.
- Present the dataset in different majority.
- Build a recommendation system that depends on users' behaviors.
- Predict the most popular places in three different cities.
- Apply three different methods to speculate the best for users.

1.3 Hypothesis

The idea of the recommendation system is to predict the next move of the people through the social network applications. In order to reach to many points on how to predict what is suitable for the users, also all of the current recommendation systems that suggest different types of products or services for the people using social media applications. Whereas the methods in this thesis present the prediction system so as to recommend the users to visit some places according to three points, including their histories, popularity of place and the type of categories. The whole idea of the recommendation system, i.e. predicting places for the users in the social media, is innovative; therefore, too many challenges have occurred. To have such a system or framework, it needs the combination of different branches of knowledge such as sociology and also the computer science in order to make the best speculation for the users.

Predicting the places for the users and picking the best one that makes our prediction system better for showing the best information from the dataset require us to know the type of algorithms suitable for our dataset. Aslo the prediction system based on the behaviors of the users and what they are looking for.

CHAPTER 2

GENERAL INFORMATION

The social networks is a social framework made up from individuals correlated out of one or more than certain types of mutuality, such as common interest, friendship, general information and knowledge. The service of the social network builds on the idea of reflecting the real life of people on online life through the websites and social media. Also provides the users with too many ways to explain and share their activities, events, ideas and interests so as to transfer them from the real life to online life.

The progress of the availability of the location procuration technologies such as the GPS authorize users to add and/or use the location measure of online social networks in different ways such as sharing and uploading the location form while visiting a place and making activities to the social network service, e.g. twitter, Facebook, Flickr [1], and even share their present location on the websites like Foursquare application [2].

The location can be represented by the longitude-latitude coordinates, symbols (such as office, home and shopping mall). Also the users can post their locations with comments, tags and any thing that they prefer to represent their locations and memories with. The location is a new object in the location based social networks (LBSNs) to evaluate the new relations between one user and the other users or one location and the other locations and make relations between the users and the locations. The new recommendation system scripts such as location and itinerary recommendations, enabled by using the new knowledge such as friends, social media recommendations.

The vast use of the mobile phone devices and the location-based services in the world have led to coin the new term of “online social media”, which is location based social network. The location based social network uses GPS, Web 2.0 technology and the mobile phone devices allow the users to share and add their own locations, by refering

to as “check-in”. Online location based social network site provides location-based service that allows the users to make check-in at any physical places the users visit and directly include the check-in of the location in their post. “Check-in” is an online action, which allows the users to add locus to their comments and photos through their social media applications. Not only the term check-in refer to add and share a spatial dimension to the social network. In other Word, it implies a coloration between the real world and the virtual World, by making the users share their whole life with their friends and the other users via photos, videos, texts and even symbols.

The evolution of the mobile phone has brought for big usage of the location based social network applications such as Twitter, Facebook, Google, Latitude, and foursquare. The location based service can be defined as a software service using the location information data to control features and this information service now has a large number of users who depend on it in different majorities like health, entertainments and personal life. LBSN contain services to recognize a location and give the longitude and the latitude of a person or object like restaurant or the location of friends. LBSN allows the users to track the places that they are search. There are several locating methods to identify the location of a person or an object like control plane locating and self reported position. In the control plane locating method, the device brings the location based signal from the closest cell phone toward the phone which doesn’t have GPS. This operation is slower than connecting to the GPS service, this method is based on the E-911 mandate and it is sill being used as a safety measure. The other method is the self-reported position. This method is the cheapest location based technology. It was used for tracking the players, and in the mixed reality game referred to as Uncle Roy All Around You in 2003. Also in the Augmented reality games in 2006 in order to announce the location and give a map for their position of other players. The other method is the near location based service (N_LBS), which depends on how to match the devices close to each other, such as Bluetooth low energy. This application allows users to share their information and location by the calculate the distance between each of the devices. For now, most of these application and method depend on the idea of GPS, such as local position system and indoor positions system [3].

Global Positioning System (GPS) circles the earth two times in a day in a very specific orbit and sends signal to the earth, in the same time the GPS receivers receive the signal and the trilateration calculates the user signal and give the exact location. Typically receiver of the GPS, compares and calculates the time difference in the transmitter and receiver. Then tell the GPS how far the GPS satellite is. The receiver of the GPS should reach at least three satellites to give the specific latitude and longitude of the location while more than three satellites the receiver can determine 3-D position latitude, longitude and altitude. After calculating the exact location of the user, GPS can determine the other information such as speed trip distance, bearing track, sunrise and sunset [4]. Global Positioning System has become common and popular for integrating the users' devices like computers, smart phone, tables, cars and so on. The new generation of multimedia like mobile phone can easily connect to online LBS such as Google maps and Yandex maps to help the users give and know the destination. It has also become like a definition of places that help the users find the place they are searching for [5]. The GPS structure consists of three layers which respectively include the space layer, the user layer and the control system. The technology of GPS based on TOA (time of arrival), which is the time that the signal goes from the satellite to the receiver [6].

Location based social networks, which enable the additional social network dynamics such as geo-coding and geo-tagging are types of social networks. These techique use geographic services as the main point in this system; further, users can visit any location in the world and make geo-tagging about anything that they want to show along with their photos, comments, and videos accordingly. In particular, the geographic service presents it by using these three layers of users, locations and the content layer [7][8]. Here it's sure that the users can exploit information from one layer to another, according to these layers we can calculate the similarity of the information between the users for instance, the places where most of the users can prefers. LBSNs have improved the quality of service for, firstly, the recommendation of social events, places, friends and activities, secondly, users' behaviours and community detection, and finally, personalized recommendation of social events, locations, and friends. Recently, the service of location based social network has led to the emergence of mobile devices, tables and other devices, which give the users access to the web. In this time, the users can take the advantages of the service location based social network to show the

location of place or object via using mobile phone; moreover, users can share information related to their location.

Users of location based social network application can record the location where they have visited, referred to 'check-in'. The check-in mostly consists of active users, date, time, places and accompanying people. These records of each check-in allow the users to keep track of the place they have visited [9], also these records let the user receive notifications about other contacts and where they have made check-in and their opinion about there. On the other hand, the users can check other users' comments about the places the users want to visit. These application allows to check everything about a place before setting off to. While the user is wondering notification can receive about each area passed from directly the program showing the rates and the comments of other users about a place. All of these applications also allow the user to check the place and everything about it through comments and rates before they give their own ones about that place. The other benefit is to tag other people who are with you and show it as a notification in the application. These applications allow their users to select a list of friends and create their own lists of friends as a social networking system [10]. The new generation of mobile phone or tablets have the capalability to determine any position in many different ways by using an integrated GPS chip or W-LAN traingularization. These smart generations of phone allow the user to show and have what they are looking for.

Recommendation system is a technology, deployed in different environment like movies, restaurant, cafe, theatre, and too many different kind of categories, are recommended to the users (visitor, customer, reader, and tourists). By using this kind of system they can represent any kind of categories. But in the same time it will be more expensive to solve this types of problems, because the users can ask for anything with different options they are looking for it in just one of the categories. Each application has a specific list of option for the places [11].

Recommendation systems are kind of personal systems because they depend on the history of the user and always shows the same categories that the users are interested in and always looking for. Typically, recommendation system predicts users' performances depending on history rating for the places. Also speculates what a user prefers or asks for, also recommendation system does not consider the other comments or social effects at all just looks at the user performance, and reveals the influence of the

social network links on the activities, users performance and tastes of individual [12]. Recommendations system has become important since 1990 and the people started to look for this kind of systems and applications to make faster and easier decisions by giving the best options about what they are interested in. There were plentiful attempts to create such a program, e.g. Rinner and Raubal in 2004, designed for finding hotels by showing the service and the prices of booking. These types of programs, at the beginning, needed a manually input from the users to describe what they is looking for [13]. There are too many new ways and attempts so as to develop the performance and accuracy of the social network by using new generation of smart devices and new programs, using GPS system.

One of the most popular location based social network application is Foursquare. Foursquare is a mobile application, which become too much popular as local search and discovery application which recommends the users to find the places they prefer. It also provides recommendations of the locations, much more close to the users. Foursquare helps and lets the users to search and look for different categories such as restaurants, markets, malls, sport places and so forth. Another help of this application is to look and search for another places by entering the names of the place. Foursquare also displays the personalized suggest depending on the time like in the morning and recommend a breakfast places and so on. The recommendations in this application will depend on the history of the users and what they always search and looking for, like if the user is looking always for dinner places it will directly suggest for the user new places for dinner. Through this application, the users can create and add their own POI (point of interest) in the application. They can make check-in at the area they are in, also the users can make check-in public or private as they prefer and collect there own points as a kind of entertainment. The users can seek for places around them to find what they are interested in and check the other people check-in and comment for that place. In Foursquare, users can easily add their friends as usual in social network programs. Foursquare also allows the users to share their check-in in other programs though it as a kind of activate social network. Foursquare, offer discount and badges for the users who are using it as daily social network program and this one is a kind of motivation for the users to make them make more check-in [14][15].

Exploration, reconnoitring and analysing the foursquare dataset are very useful for the users in different majority for business and customers. According to the percentages,

59% of the users in both majority business and normal customer are looking for Foursquare dataset for local business; subsequently. It allows the businessmen to know and understand where and how to reach a good number of customers for their new projects depending on the post and the check-in of the users and calculate and count which area would be good to open their new projects [16]. For the location based social network, users who always look for new location and are interested in venues also implies good reputation, that is, it's very useful and important for users to choose place. Thus, exploration and the analysis of Foursquare dataset would be very helpful to understand the foursquare costumers, also we may be interested in the popular places according to the number of the check-in.

The location based social network (LBSN) is not only for the idea of sharing and adding the location to the social network so that people can see their location, where they attend to go. It's also consists of a new framework made up from different and individual correlated by too many types of interdependency derived from the locations that the users share and add from the physical World. As well as their expression through the location tagged media content such as comment, photos, video and symbols. The interdependency contains not only the share of the location of two persons in the same physical location or have the same history of check-in, but also the knowledge such as behaviors, activities and similar common interest.

2.1 Related Work

There are two main research related to our work the first one is about recommendations system and the second one is about the prediction system using the history of the user and the prediction system according to the categories.

There is a big number of researches about data mining algorithms that basis on the recommendation system idea [17]. These systems are taking the users performance and choose such as rating, commenting and tagging as an input to predict the new move for the user and to show the popularity of the venues, all these are calculating under the term of collaborative filtering. In most of new research work focuses on the dataset which colected from the web site, namely movies such as Netflix [18] and music such as Yahoo Music and also another programs which is give a huge dataset.

In [19], categories type was the main goal to model implicit users' motion pattern. In there job, they focus in two points the first of all, the prediction categories according to the user activities. Secondly, the prediction according to the location was given by the rated category distribution. The dataset used in this paper is Gowalla dataset for 13 million check-in for more than 230,000 users in 12 months, through this dataset they try to explain the predication categories idea. The used framework was mixed hidden Markov model to speculate the next categories that the user is looking for, according to this prediction model the feature was the predication were focus in the user activities instead of the prediction space for all other activities of other users. According to their experiment the result with the categories, prediction was 5.45 times smaller which means the prediction accuracy was 13.21% higher.

Location based service Foursquare; they collected and took snapshot for check-in information in two cities in UK (Cardiff and Cambridge), which almost 19,000 check-in for nearly to 3000 users. In this paper, they are exacted and discussed how users of location based social network by phone present different behavior. Depending on their places and where the users prefer to go, they observed that a considerable percentage of users used to visit a regular number of locations, as well as showing a tendency to visit social friends groups. According to their result, they found the average for each city depending on the population of each city and looking at the number of the active venues in that city; and also, they found the average in each venues and the average for users [20].

In [21], the researcher depend on analysis of users' check-in history also social interaction patterns of using network structures of Foursquare users and venues dataset, they also focus on the venues geographical information and its effects on users' behaviors. when choosing the place to better understand the sensitive factors. The dataset used in this project consist of 485381 users, 83999 venues, 1.02 million check-ins and 27 million friend links. They applied four type of model (Trust-based collaborative filtering (TCF), Interest-based collaborative filtering (ICF), Geographic Model (GEOM) and Fusion Model (FM) to analysis the dataset and understand the critical factors related to the users check-in behaviors.

In [22], they evaluate a series of ways and techniques of identifying users from their own check-in information. They applied two techniques to the analyze data according to users' check-in over time and also the frequency of visit to a specific location. The

techniques depended on users' identification to analyze the dataset (Trajectory-based Identification, Frequency-based Identification and Measuring the Complexity of the Identification Task) depended on spatio-temporal trajectory emerging for their users check-in. They applied also a hybrid way to exploit both types of information. In this research they used three real world dataset to have more high degree of accuracy in their results. They classifed more than 80% of the users' activities correctly according to their dataset as well. They collect their dataset from Bright kite, Gowalla and Foursquare dataset and it's about more than 13 millions check-ins for more than 260 thousands in almost 2 millions locations, depending on these dataset they showed the GPS information and also the frequency of visit to a specific place.

In [23], in this paper they focus on users history information to predict the next steps or the next place that he/she planning to visit, when users make a check-in by using foursquare application and showe it by using Twitter. They collected this dataset by depend on each twit will contain any information about check-in through Foursquare because of that the dataset here is about 180 millions check-in in different categories. In this paper they applied an end to end prototype included automated data pre-processing, automated model learning and automated prediction and result evaluation/visualization. At the end, this paper showe method for speculating the user future move and which place the user where looking for depending on their history. The task of successive personalized point of interest (POI) in LBSNs considered in another paper by focusing on how to solve the POI recommendation or prediction. They observed two prominent properties in the check-in sequence: personalized Markov chain and region localization. They submitted a matrix factorization method, namely FPMC- LR, to firm the personalized Markov qualifications and the centralized zone. The idea here is not just about personalized Markov chain in the check-in sequence, but also to look around the localized area. Through this project they reduced the cost by using the localized information by discard the noisy unimportant information to raise the recommendation, the dataset here contain information get collected from Foursquare and Gowalla about all the users history with the time stamp and location details [24].

In [25], they inspected how venue discovery behavior characterizes the huge check-in data from two location based social service, Foursquare and Gowalla for more than 11 cities and through they analyzed they focus on how often the people will try to visit new place and according to the dataset that get used between 60% to 80% of the users didn't

visit new places for more than 30 days, in another word it mean how many of the users were looking for new places. The dataset collected from Foursquare and Gowalla, for Foursquare dataset contain 35 millions check-in made by 925000 users across around 5 millions venues in a period of 5 months. For Gowalla dataset contain around 13 millions check-in made by 216000 users across 1.5 millions in a period of 18 months. In this project they applied more than one algorithms. Also they compare the result to discover which one has the highest percentage for the prediction for the next place. The recommendation algorithms that they used depend on different points according to the information that the dataset contains, the algorithms were visiting popular venues, attending venues by category, following friends, staying close to home, like-mindedness and similarity and walk random walk approach. They propose to calculated the average of each users in each city and compared with the other 10 cities. Also calculating the average number of check-in in each place. The calculation in this project was showing the number of population for each place and the number of visits with their averages and according to all these result they can understand the rush hour and the minimum and maximum numbers of visits for each place and each city.

According to result of averages they could showed and explained the next move of the users if it will be in new or old place and count the rush hour for both of the dataset Foursquare and Gowalla. In another project they focus on the problem of the time aware point of interest recommendation systems, which recommending to the user to visit new places at a given time according to the users interest. To explain the geographical the temporal influence in the point of interest POI recommendation systems. They suggested Geographical-Temporal influences Aware Graph (GTAG) to calculate and record the check-in also the geographic and temporal influence. In this project to make the Geographical-Temporal influences Aware Graph (GTAG) more effective and more efficient recommendation they developed the performance of it by propagate methods named Breadth-first Preference Propagation (BPP). According to this algorithm the recommendation system will returns results within at most 6 diffusion steps. The recommendation systems results were time aware because of the perception that the users looking for different places at different time. The dataset that get used were Real world dataset collected from Foursquare and Gowalla respectively, The *Foursquare* dataset contains 342,850 check-ins and The *Gowalla* dataset, contains 736,148 check-in. The Foursquare dataset it's about Singapore city, for Gowalla dataset it's for

California City. Both of the dataset showed the proposed graph-based approach out-performs state-of-the-art POI recommendation system [26].

Most of these researches have focus on the idea of predicting venues in location social networks from different points. As we can see some of these focus on the history of the users and what the user attend to go and visit, according to the history the recommendation system will give the next step and recommend to the users depends on there history. From another side some researches focus on the check-in and what the check-in contain such as tags, comment and also giving a rate to the place that the users visit. Also some papers focus on the idea of tagging friends and how this tag will affect on the other friends to make them visit this place as a recommend system. Relating to tag friend recommend system there's also some recommendation system depends on the comment and what the comment that the other users give for the places as recommendation system technique. In other hand, there's users will visit the place according to the comment that the other visitors give when the visit that place. Also there's researchers focus on the point that the user will visit new place instead of visiting the same type of places but in the same time under the same categories. Also the idea of attend to popular places according to the numbers of check-in the other users make in one of the places or depending on the rank of the place. The recommendation system idea for predicting new or the same kind of categories has focus on most of the type of speculation technique. In order to give the users what they are looking for, also showing the advantage of predicting the next move for the users, to help the users to make better decision for choosing the places.

CHAPTER 3

DATASET PROPERTIES

Location based social network has been very popular subject and the dataset has become very interesting point to attract millions of people footprint and what they look for. Our dataset had been collected from foursquare application. It shows different categories and venues that the users of this program prefer to look for. In order to study personalized location recommendation foursquare dataset one of the best option in this time to understand what the users are interested in. Users can have access to foursquare by using there own phone very easily and show their status by make check-in in any place that the users wants. Through these programs, users can write any comments about their status with the rate that want to give for this place. Also make share and tags for anyone from their friends in their contents, also foursquare allows the users to create their own list about the places they prefer best for them based on what the user always looks for.

The foursquare dataset used in our research dealing with different kind of categories in three different cities which they are London, Austin and Dallas the dataset shows check-in in too many places for different categories according to too many users. Each of cities has different number of check-in with different check-in in all the places in each of the cities for different categories and positions. Each for the cities dataset has it's own properties about the number of users, check-in and the categories. For London city dataset, it contains more than 4 millions check-in for more than 100 thousands users in more than 40 thousands different positions and places for different categories. Also it contain the longitude and latitude of each venues. For the check-in it show the date and also shows the number of the street and the building nbumber. These dataset had been taken in 2011 from forsquare application, according to these dataset there's some user made more than 200 check-in and other less these number until one check-in in different categories, date and time. According to all these information we can predict

the best places also speculate the rush hour and give the best recommendation for the others users.

For Austin city dataset, it contain more than 1.5 million check-in for more than 40 thousands users in more than 12 thousands venues and positions for different categories, also it contains the longitude and latitude of each positions and for the check-in and depicts the date & time for all the check-in and representing the number of the street and the building, these dataset had been also taken in 2011 for March and April for different users and there own check-in.

For Dallas Foursquare dataset, it contains more than 1.4 millions check-in in different places in Dallas city for more 35 thousands users made check-in in more than 15 thousands places and venues for different categories, also it has the longitude and latitude of each venues and for the check-in it explains the date and time for each check-in and also contain the number of the street and the building number, these dataset had been also taken in 2011 for March and April for different users and there own check-in.

Table 3.1 Average properties of the data set: Total number of user (N), places (M) and check-in (C), average number of check-in per user (Cu) and per place (CP)

City	**N**	**M**	**C**	**Cu**	**Cp**
London	104076	43584	4162121	11.7	97
Austin	42122	12971	1474270	7.5	115
Dallas	35593	15751	1637232	13.7	105

We calculated the number of users for each of the three cities and we represent the number of the users as (N). The highest number of users in all of the three cities is in London city comparing to the two cities. For the number of places in the three cities we calculated it, the highest city that it has the highest number of places is London city, also for check-in London city is highest one. In order to measure what our dataset contain and how to understand what the dataset has, also what kind of information could be taken as benefit form the dataset for our analyzing and predicting system.

After pointing in the analyzing focus on the average of the number of the check-in according to the number of both users and places to show the average for each of the check-in per user and the check-in per place. This analyzing show the real number of the check-in for each of the cities and it is not about the number of check-in and the user

or places number. The first analysis for the check-in per user, the highest average is for Dallas city by average 13.7. Then, we analysis and calculate the check-in number per place, the highest rate is for Austin city by average 115 in each of the places in Austin city. We show average rate for check-in number per the users and per the places in charts.

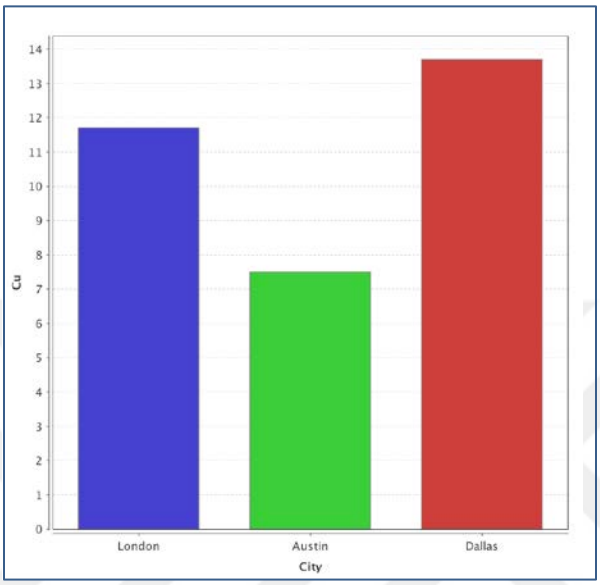

Figure 3.1 Average number of check-in per user

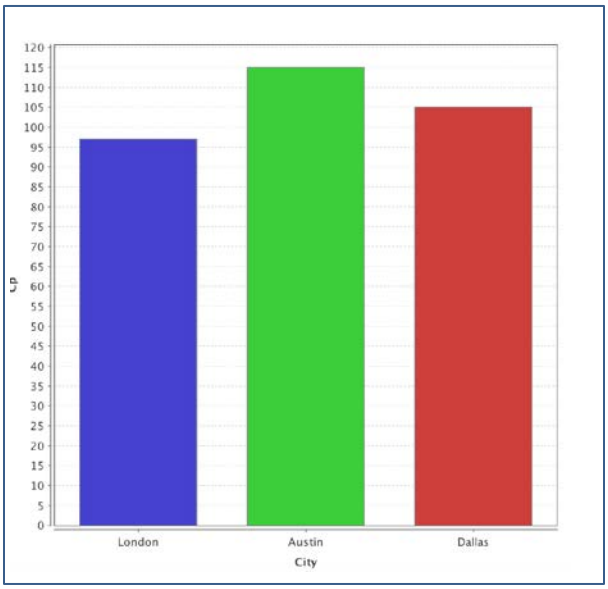

Figure 3.2 Average number of check-in per place

In both of the Figures, we show charts explaining the average number of the check-in according to the number of users and the number places in each of the three cities. The first chart explains the average number of check-in per user in the three cities, the highest average is for Dallas city. The second chart depicts the check-in number per place in the three cities in our dataset, the highest rate is for Austin city for the average number of check-in per places.

3.1 Dataset Prepration

The main idea of the data mining is dealing with the dataset to show how to analyze the dataset to make more clear and more useful. In this thesis, we deal with big dataset which collected from Foursquare program. Which it showing a number of check-in made by the users of Foursquare program in three different cities London, Dallas and Austin. Each of these cities has two separate dataset the first one for the user check-in information which is contain the user identity, date and time of the check-in. Also all the number of check-in for each user because some of the users have more than 200 check-in and in the other hand there are some which has just one check-in. The second onefor the location informationwhich conatin longitude, latitude, street number and categoriry for each of the places. In this dataset was over than 8 million check-in for more than 180 thousands of users in different kind of categories which is has too many different places. This kind of big dataset it's the mean idea of data mining subject and how to prepare these data to deal with it and how to prepare the dataset to take the benefit from analyzing it.

The first status of the dataset was like a text file and contain too many symbols in such as point, question marks, gamma and etc.

```
@$493501236*;;;;*[('4ad8f7fbf964a5208e1621e3', datetime.datetime(2011,
('4ae33705f964a520c69121e3', datetime.datetime(2011, 4, 9, 23, 35, 38))
76628764*;;;;*[('440da323f964a52092301fe3', datetime.datetime(2011, 3,
('4a4e1452f964a52016ae1fe3', datetime.datetime(2011, 3, 28, 15, 40, 21)
datetime.datetime(2011, 3, 31, 9, 26, 37)), ('4bb52afeef159c74446274f7'
27, 11)), ('440da2cbf964a52091301fe3', datetime.datetime(2011, 3, 31, 9
('4d7550abba682d43ba873976', datetime.datetime(2011, 3, 31, 19, 4, 48))
datetime.datetime(2011, 4, 4, 18, 51, 50)), ('4d7a66c27451236ae632e1fe'
7, 28)), ('4c4ba6efc668e21e99fa55fa', datetime.datetime(2011, 4, 5, 18,
('4b3bec1bf964a520727e25e3', datetime.datetime(2011, 4, 8, 16, 1, 41)),
datetime.datetime(2011, 4, 8, 16, 17, 1))]
20554299*;;;;*[('43ed0a2ff964a5204b2f1fe3', datetime.datetime(2011, 4,
14172960*;;;;*[('42accc80f964a5204c251fe3', datetime.datetime(2011, 4,
4176161*;;;;*[('440da2cbf964a52091301fe3', datetime.datetime(2011, 4, 2
('440da2cbf964a52091301fe3', datetime.datetime(2011, 4, 3, 18, 49, 35))
```

Figure 3.3 The first status of our dataset (Random status)

61097042,	4b6196e7f964a5200e192ae3	2011, 3, 31, 11, 58, 45
106702719,	4acf8c21f964a5206dd420e3	2011, 4, 1, 10, 32, 50
16517630,	4af00bcff964a52081da21e3	2011, 4, 12, 18, 24, 22
14372858,	4b5e726cf964a520268f29e3	2011, 4, 3, 16, 42, 53
49647700,	4ad7a112f964a520050d21e3	2011, 3, 27, 20, 5, 22
44171727,	4bd73095304fce72c63133ab	2011, 3, 28, 15, 9, 59
6331022,	4b8baecdf964a520d6a632e3	2011, 3, 29, 18, 0, 7
17869997,	4b998372f964a5208d8035e3	2011, 4, 3, 2, 10, 51

Figure 3.4 Dataset in SQL after appling Excel find and replace function

Here, we apply (find and replace) function in excel to take all the symbols from the dataset and change with gamma or make it empty. After that, we applied (text to column) option to separate the dataset to columns.

After we finish the preparation in the excel program, the dataset of each city is imported to oracle program. In order to make the dataset more clear and also calculate the number of the check-in in each city and merge the both dataset of each city the user information and the location information with each other to have much more usable dataset.

The main idea of using oracle database program in this thesis is that the dataset that we are dealing with it a big dataset that the Excel Microsoft office can't deal with this much of dataset. Oracle database program allow to the users to deal with the big dataset much more easily by using function, which it is written, by using SQL functions to clean and analyzing the dataset as it's shown in the next figure.

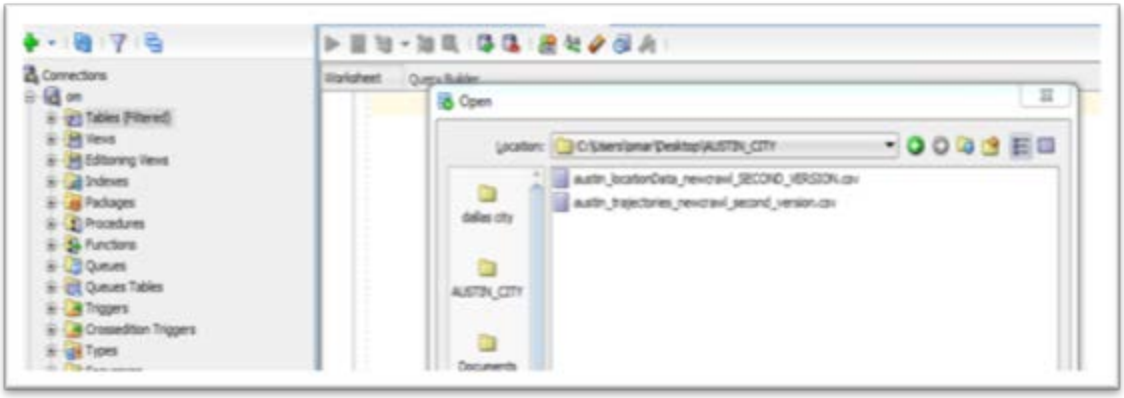

Figure 3.5 How to import the dataset to oracle database

Import the dataset to the oracle first of all to deal with the data much more easily we saved the data type under. CSV which allow to us to Import and change the dataset as we wanted much faster and easily in the same time. Importing the dataset to oracle database allow the users to clean and separate the data by using import methods in oracle to take symbols from the dataset.

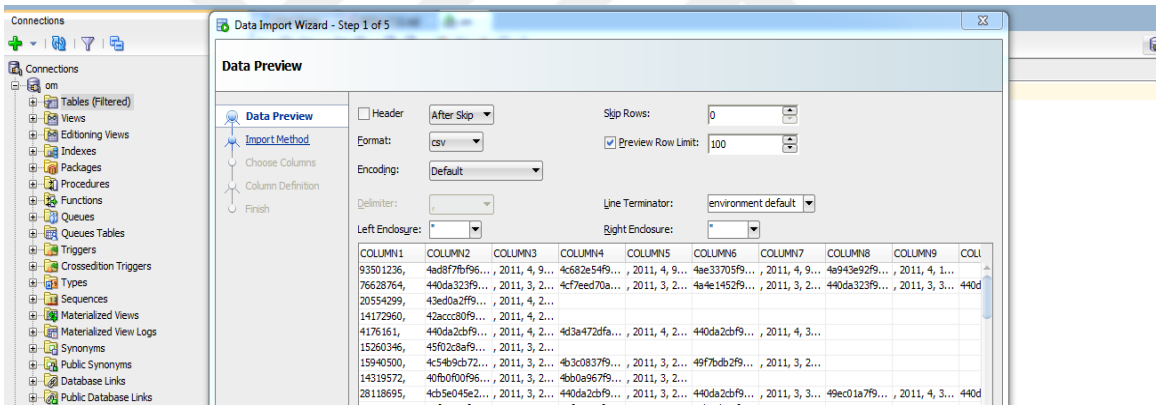

Figure 3.6 Data preview with importing method in oracle dataset

Data preview allows the user to show the dataset as he/she looking for, by separating the dataset as the user prefer and also show the header in the dataset. Data preview allow users to skip the rows from the dataset also select the type of the format that the user will deal with.

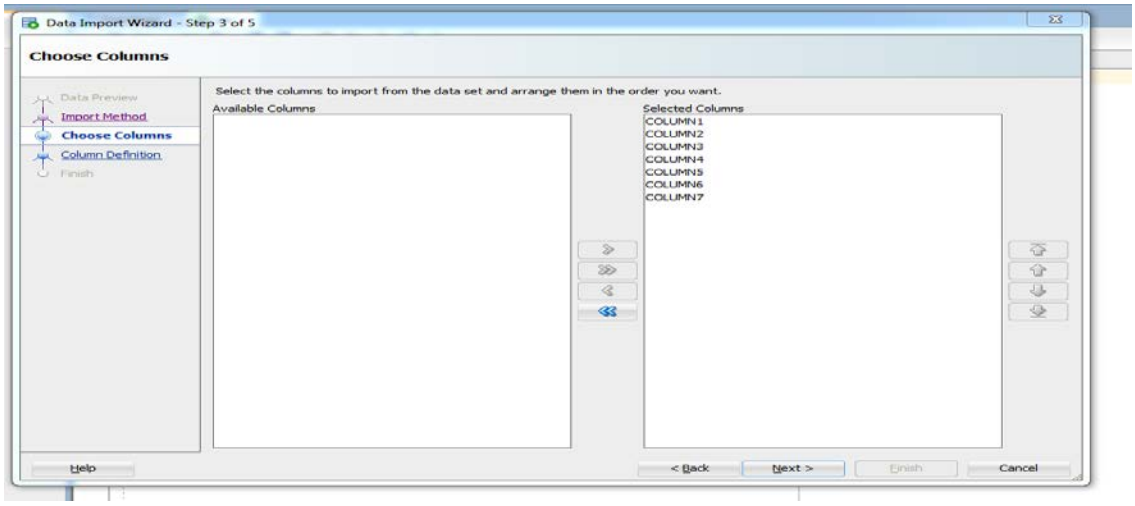

Figure 3.7 Choose columns in oracle database

Choosing the columns option in oracle allow the users to select only the columns that the users are looking for and ignore all the other columns.

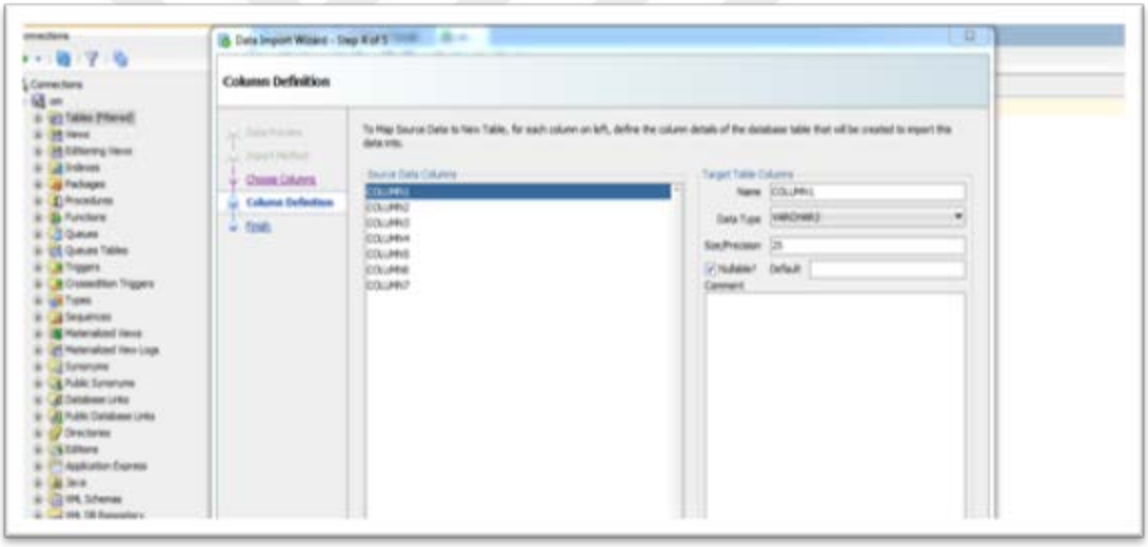

Figure 3.8 Defining the column in oracle database

Oracle allows the users to select the type of the columns of the dataset and also select the type of the column and give the specific size of each column.

After all these steps, shown by using oracle database tools to import the dataset to the program and choose only the columns, we want to give the name and the type of each columns. We apply SQL functions and formula which is dealing with the dataset to merge the two tables of each city for our dataset. Also write the date and time function for each check-in in the dataset, also show the check-in of each users separately form

the other users. SQL allow the users to find the result about each point in the dataset that the user is dealing with by write the quarry for each step of the codes.

```
alter TABLE DALLAS_SECOND rename  COLUMN COLUMN3 to DATE_TIME;


UPDATE  DALLAS_SECOND
   SET DALLAS_SECOND.DATE_TIME =REPLACE(DALLAS_SECOND.DATE_TIME, ' ','')


   SELECT * FROM NEW_CRAWL_DALLAS


UPDATE  DATE_DALLAS3
   SET DATE_DALLAS3.DATE_TIME =REPLACE(DATE_DALLAS3.DATE_TIME, ' ','')


UPDATE  NEW_CRAWL_DALLAS
   SET NEW_CRAWL_DALLAS.CATEGORY =REPLACE(NEW_CRAWL_DALLAS.CATEGORY, '"','')
```

Figure 3.9 SQL functions to clean the dataset

By using SQL, the users can update their dataset by cleaning it through removing anything from the dataset or to add anything to the dataset like column, rows and change the type of the dataset also the names of the columns.

```
CREATE TABLE DALLAS_DATE_CHANGE AS
SELECT T.*, (CASE WHEN SUBSTR(T.DATE_TIME, LENGTH(DATE_TIME),1)=',' THEN  SUBSTR(T.DATE_TIME,2,LENGTH(DATE_TIME)-2)
ELSE SUBSTR(T.DATE_TIME,2,LENGTH(DATE_TIME)) END) AS DATE2
FROM DALLAS_SECOND T

create table DALLAS_DATE_TIME AS
SELECT T.*,TO_CHAR(TO_DATE(T.DATE2,'YYYY-MM-DD hh24:mi:ss'),'YYYY-MM-DD hh24:mi:ss') AS DATE4 FROM DALLAS_DATE_CHANGE T

 SELECT * FROM DALLAS_DATE_TIME
SELECT * FROM DALLAS_SECOND
SELECT T.*,REGEXP_COUNT(DATE3,',') FROM DATE_DALLAS7 T WHERE REGEXP_COUNT(T.DATE3,',') <> 2


 SELECT * FROM DALLAS_DATE_TIME
SELECT * FROM NEW_CRAWL_DALLAS
```

Figure 3.10 Date Time function in SQL

In each dataset, dealing with the prediction function the most important point for the predict is the date and time function by using SQL function and formula for the date and time. We got the formula that we used to calculate the next move for the users, in another word we predict the dataset of us.

3.2 Dataset Analysis

According to cities, users make check-in by categories in Foursquare. In this thesis three cities had been used: London, Austin and Dallas. Importance of categories affect exact specific popular places which users made check-in and get clustered in some places. For instance, if food category is the most famous category then people create a cluster point in all food places. Furthermore, sometimes people's behaviors depend on other people's decisions. Popularity of place has a prominent side effect on people's choices. Also, sometimes an individual person travels more than one place, which analyses should been prepared not only check-ins but also number of users and locations. In this thesis,

analyses of users, locations and categories are found out according to cities. Data analyses and predictions has a great influence recommendation in new places.

Table 3.2 Dataset information in SQL

USER ID	LOCATIOID	DATE_TIME	LATITUDE	LONGTIUDE	CATEGORY	PLACE
93501236	4ad8f7fbf964a5208e1621e3	2011-04-09 17:30:05	30.270779999999998	-97.817279999999997	Mexican	Tres Amigos
76628764	440da323f964a52092301fe3	2011-03-28 12:57:05	30.263577000000002	-97.740089999999995	Convention Center	Austin Convention Center
20554299	43ed0a2ff964a5204b2f1fe3	2011-04-02 23:14:40	30.257452000000001	-97.807464999999993	Mall	Barton Creek Square Mall

In order to represents the dataset after applying all analyzing steps in the dataset preparation so as to have a clean and clear dataset that we can analysis it more easily.

In this part of the analysis, we focus on how to explain our dataset and show too many calculation from our dataset to explain the behaviors of the users and what the users is going for. Also it shows the number of check-in of the users in the three cities and in which places and calculate the average number of the check-in for all the users in the three cities.

Table 3.3 Number of check-in for each user (Ascending)

USERID	CITY	NO.OF CHCK-IN
36150551	DALLAS	607
111659789	DALLAS	538
23172319	DALLAS	479
74079602	AUSTIN	478

Table 3.4 Number of check-in for each user (Descending)

USERID	CITY	NO.OF CHCK-IN
107042277	LONDON	1
17376218	LONDON	1
621172315	DALLAS	2
74279381	AUSTIN	2

Both of the tables above show the number of check-in of the users in the three cities in two ways descending and ascending to represent the number of check-ins that the users had made in the three cities.

Table 3.5 Number of check-in for users (Ascending)

CATEGORY	CITY	PLACE	NO.OF CHECK-IN
Playground	AUSTIN	Foursquare court Pepsi MAX Lot	1724
Train Station	LONDON	London Victoria Railway Station VI	1673
Hotel	AUSTIN	The Driskill	1633
Museum	LONDON	Tower of London	1590

Table 3.6 Number of check-in for users (Descending)

CATEGORY	CITY	PLACE	NO.OF CHECK-IN
Bakery	LONDON	Treacle	1
Home	DALLAS	Georgian Apartment - Dallas	1
Cemetery	AUSTIN	Travis County International Cemetery	1
Home	AUSTIN	Chamboard Condos	2

In order to present the number of check-in in the three cities in different categories and which places in this categories, also we show the dataset in ascending and descending way according to the categories.

Table 3.7 Number of check-in for location depend on users (Ascending)

USERID	CITY	CATEGORY	NO.OF CHECK-IN
51138891	LONDON	Train Station	247
208854807	DALLAS	uncategorized	232
42168168	DALLAS	Home	226

Table 3.8 Number of check-in for location depend on users (Descending)

USERID	CITY	CATEGORY	NO.OF CHECK-IN
53138711	LONDON	Train Station	1
218754862	DALLAS	uncategorized	2
431632145	AUSTIN	Home	2

Also depending on the number of check-in in location for users, we analyze the number of check-in according to the cities and the categories.

Table 3.9 Number of check-in for location depend on place (Ascending)

PLACE	CITY	NO.OF CHECK-IN
RVIP Lounge Karaoke RV	AUSTIN	914
Waitrose	LONDON	911
Cedar Door	AUSTIN	903

Table 3.10 Number of check-in for location depend on place (Descending)

PLACE	CITY	NO.OF CHECK-IN
La Ciabatta	LONDON	1
Woodland Hills Pool	AUSTIN	1
Child's Play	DALLAS	1

In the tables above, we analyze our dataset according to the places and numbers of the check-in in these places, explaining the dataset and represent it in ascending and descending to show both side of the dataset form the top and the down of the dataset.

Table 3.11 Number of check-in for users according to the location (Ascending)

LOCATIONID	CITY	PLACE	NO.OF CHECK-IN
4ac518cef964a520f8a520e3	LONDON	Tower Bridge	770
4a4509d9f964a520b4a71fe3	DALLAS	House of Blues	768
40b13b00f964a52083f31ee3	AUSTIN	Club de Ville	755

Table 3.12 Number of check-in for users according to the location(Descending)

LOCATIONID	CITY	PLACE	NO.OF CHECK-IN
4d19e653cc216ea870a080d3	LONDON	Hackett	1
4d5184919ffc236a3ae923a7	DALLAS	La Paloma Taco-Taco	1
4c92b7b52bf7b60c3a32650a	AUSTIN	Darque Tan	1

According to the number of check-in in the location, we count the total number of check-in for the location and we show the name of the location as a place.

Table 3.13 Check-ins for users from London

USERID	LOCATION ID	DATE_TIME	CATEGORY	PLACE	CITY
9627492	4b632d3ff964a52057682ae3	2011-03-29 19:44:44	Train Station	Hendon Railway Station HEN	LONDON
9627492	4ad9cbdbf964a520ee1a21e3	2011-03-26 18:00:59	Pub	The Old Crown	LONDON
9627492	4aeefc63f964a52002d521e3	2011-04-22 07:41:16	Bar	The Bridge Bar and Eating House	LONDON

Table 3.14 Check-ins for users from Austin

USERID	LOCATIONID	DATE_TIME	CATEGORY	PLACE	CITY
14172960	49bbcaeff964a520ed531fe3	2011-03-14 04:56:02	Karaoke	RVIP Lounge Karaoke RV	AUSTIN
14172960	4d1c114e1356224bb39e0cbe	2010-12-30 04:59:00	Home	Alison	AUSTIN
14172960	4abaa233f964a520318220e3	2011-04-21 00:44:00	Bowling Alley	The Highball	AUSTIN

Table 3.15 Check-ins for users from Dallas

USERID	LOCATIONID	DATE_TIME	CATEGORY	PLACE	CITY
15940500	4d5325bb3a9df04d0b71ec05	2011-03-29 17:59:27	BBQ	Luckie's Smokehouse	DALLAS
15940500	4bd5f8766f649521b10d70ec	2011-03-08 21:24:33	Sandwiches	Texadelphia	DALLAS
15940500	4acf765af964a520dbd320e3	2010-12-22 21:37:33	CafA	The Dream Cafe	DALLAS

In the previous tables, we took a sample of users from the three cities where we have to show their check-in information and location and places the users tend to visit. The check-in information for each of the users contains all the information about the users' check-in, such as date and time also the location of that place that user visit.

Table 3.16 Number of check-in for users from Dallas

USERID	CITY	NO.OF CHECK-IN
14464426	DALLAS	92
41593502	DALLAS	46
24229358	DALLAS	552

Table 3.17 Number of check-in for users from Austin

USERID	CITY	NO.OF CHECK-IN
22410724	AUSTIN	70
14185185	AUSTIN	420
63479402	AUSTIN	35

Table 3.18 Number of check-in for users from London

USERID	CITY	NO.OF CHECK-IN
41321300	LONDON	320
254569172	LONDON	40
68940126	LONDON	520

Here we show a number of users form Dallas, Austin and London that made a different number of check-in in each of the three cities, each of these users made a number of check-in in different positions and categories. To show the number of check-in for each of the user's according to the cities not just according to the type of places that the user prefer to visit.

Table 3.19 Number of check-in for user from Austin with places

USERID	PLACE	NO. OF CHECK-IN	CITY
93501236	Mt. Bonnell	6	AUSTIN
93501236	Wild About Music	1	AUSTIN
93501236	Shady Grove	2	AUSTIN

Table 3.20 Number of check-in for user from Dallas with places

USERID	PLACE	NO. OF CHECK-IN	CITY
32960932	Texadelphia	2	DALLAS
32960932	Charlie Palmer at The Joule	2	DALLAS
32960932	Benihana	12	DALLAS

Table 3.21 Number of check-in for user from London with places

USERID	PLACE	NO. OF CHECK-IN	CITY
16517630	The Heights at St Georges Hotel	1	LONDON
16517630	The Ship	3	LONDON
16517630	Frank's CafA & Campari Bar	2	LONDON

The previous tables show different users who had made a different number of check-ins in some places where they tend to visit in all of the three cities of our dataset. According the users in our dataset, most of the users make check-in in too many places. In the same time some of the places has too many check-in number from the same users while others it’s has just one check-in from user and too many check-in from other. The idea is to show the number of check-in for each of the users while the next step for the user we will focus on it on the next chapter.

CHAPTER 4

LOCATION PREDICTION METHODS

Location based system network can improve the services on Generic recommendation of social activities, events, friends and location, the other one is the personalized recommendations of activities, events, friends and users mobility. All of the recommendation system algorithms depend on the dataset in different ways, like following friends algorithms depends on what the friends prefer and like. The algorithms which depending on the history of users advise the users where to visit.

4.1 Generic Recommendations

Generic recommendations contain the same recommendations list (activities, event, location and so on) for all users. This type of recommender system is the most easy and simple kind of recommender system which is depend on the idea of counting and calculating frequencies of appearance and occurrences of some given information and dimensions. For instance, the simple recommender system can count just the check-in and give rank for each place in different categories and according to this the users can look for new place that it recommended for them. For the location recommender, it would count the frequency for each location and count the check-in and then, it can recommend the highest locations saving and sorting these places in decreasing or increasing order of their points and scores and select the famous and popular location. Also an interesting location can be defined as cultural location such as the most popular places in any city and also the commonly frequented famous area such as restaurants, shopping mall, sport courts and so on. On theother hand, the activity recommendations also concerned, an activity recommender can tool up the users with the most famous and popular activities that may take venues at a given location like shopping food places. Such a system that provide a simple activity and location recommender was Gowalla.com web site. And this system was showing the

target that the users looking for. For example the users will enter what he/she looking for such as coffee and also will select the city and then the system will advice to the users which places is the best for the thing that the users are searching for it. All this depending on the score of the places from the top to the end of the list. In this thesis we use two methods, which depends on the number of check-in for the venues [27].

4.1.1 Popular Places Algorithm

Visiting the places according to their popularity is one of the most known techniques. It is also not a personal recommendation system way but it is based on the historical popularity of the places. The algoritim also counts the number of the check-ins of users in one of the places and then according to the number of the check-ins the place it will have baseline rank from the users.

$$\hat{r_k} = \sum_{i \in U} C_{ik} \tag{4.1}$$

The popularity of the place score $\hat{r_k}$ of place k, also C is represent the number of check-in, and i is represent the users. This method supposes that the probability of checking in is proportional to how many of the users have made check-in in one of the places. In other word the users will make check-in in at the most famous places, the people will look to the most popular venues in the application according to the number of check-in form the users and how much rank that the place took from the users.

4.1.2 Popular Category Algorithm

Attend to places according to their categories is another type of recommending system. Because in order to find the most popular category firstly calculate the number of the check-in for each of the places. Also find these places under the name of which category, also find the popularity of the category based on thr number of check-in made by users in venues of each of their category in different majority. Then, populate a list of the categories depend on their places.

$$\hat{r_y} = \sum_{i \in U} C_{iy} \tag{4.2}$$

The popularity of the place score $\hat{r_y}$ of place y, also C is represent the number of check-in, and i is represent the users.

4.2 Personalized Recommendations

The personalized recommender system depends on the check-in histories of the usersand the users behaviors. Then, compare and count the history of users with each other and find the correlation of the performance for the users and suggest to the users new places, events and activities. In particular, the personalized recommender take the advantage of the time that some of the user has visited a place and give a rate or comment on that place and predict for the user unvisited place similar to the history of that user [27].

There are three approaches in the context of the recommender system as it follows: collaborative filtering (CF), content-based filtering (CB) and hybrid methods.

Collaborative filtering (CF) is a technique used by recommendation system, collaborative filtering have two different senses firstly is the narrow sense and second is the general sense [28]. In the narrow sense, collaborative filtering is a method which automatically predict the next move or place for the users depending on the performance and the information that collected from many users. The idea of the collaborative filtering is that if one person like x has the same opinion with another user (person) like y about one place it mean that person x can take the opinion and the idea of the person y about another place and so on. In other hand, the general sense of the collaborative filtering is a process for information and data using different techniques involving collaboration through different data and information also viewpoint and so on [29]. The applications of collaborative filtering generally depend on the idea of big dataset in their analysis in both of the senses general and narrow. The collaborative filtering has been applied to too many big datasets in different majority such as financial dataset, environmental, web applications and social media application. CF recommend those places, events, activities in the city or country that have been rated with a high rate from different users and according to the statistics measurement it will recommend to the users the most high rated places.

Content-based filtering (CB) also referred to as cognitive filtering, the recommending idea in the content based filtering depend on the comparison between the users history and profile and the content of the items or the places and according to this information after comparing the items content and the users favorite places it

will recommend to the users new places, event and activities. The content of each item represented as the comment and the description or terms normally the words that occur in the file or document.in the other side, the users profile is showed by the same terms and built up the analyzing of the content of items which the users had been search and looked for before [27].

4.2.1 User History Algorithm

The personal recommendation system depends on the users histories. And which places that the users attend to visit also how many check-in that the users made in one of the places, according to the number of check-in the system can predict which type of places that the users prefer to visit. Recommendations are computed on the proposition of the history of the users and show the most places that the user attends to visit.

4.2.2 Collaborative Filtering

Collaborative filtering also known as social filtering, filters information by using the other user's opinion and recommend. It is based on the idea of users who agreed in their estimation about an item in the past are probably will agree in the future about the same item [17]. The fundamental assumption of the collaborative filtering is that if users such as X and Y rate an item N similarly, or they have similar behaviors and interest, then the system will recommend to each of them similarly [30]. The types of collaborative filtering (CF) algorithm have been proposed in the recent time. It can define two major types of CF algorithm [31], memory based algorithm and model based algorithm. Memory based algorithm can be motivated by notice the other people comment and recommend for place or movie etc., in other word recommendations from like-minded friends. These algorithm apply nearest-neighbor like scheme to speculate the user's ratings based on the similar ratings given by like-minded users. In the model-based collaborative filtering first learn the descriptive model of users' performance and after it use it to predict ratings.

Collaborative filtering is an algorithm build on three idea: users have similar behaviors and habits, their habits and performance are stable, then we can speculate their next chose according to their past performance. The first step in the collaborative filtering is to take out the user's history profile which can be shown as ratings or check-in matrix with each of the rate or check-in that the user given to an

item or place. The matrix here consist of table and each row of these table represent a user, also each of the columns represent the item or the place [17]. The second step is to calculate the similarity between each of the users according to their interest, also find their nearest neighbors.

There are several types of collaborative filtering algorithm such as user to user collaborative filtering and item to item collaborative filtering. User to user collaborative filtering algorithm, here we try to find similar person (look alike people) and offer products or place that the person had been chosen it in the past. Also if there is another had been in the same place or like the same product. In this algorithm the process is taking too much time, because it need to compare each of the users with each other and with themselves in the same time. In the second one, we are trying to find items which is looking a like for the user that the user prefer to go or to have. The process in this algorithm is taking less than the first one, because we don't need to compare each of the users. Also we don't need to check all the score or check-in between users, just we need fixed number of place-place or product-product look alike matrix is fixed all the time.

User-User collaborative filtering, also known as K-NN collaborative filtering. It produce first in GroupLens Usenet article recommender [32], the BellCore video recommender [33] used the user-to-user collaborative filtering.

User-User CF is an easy algorithm to interpret it, it's all depend on find the other user who have similar history in their behaviors to visit and attend to do the same kind of habit. User-user collaborative filtering looks for other users who have high agreement with one of the users on an item or place that they both had been rated, then these users who had been rated the place or the item will be weighted by their acceptation to the user that we want to calculate the prediction and the similarity for him/her.

Also additional to the rating matrix **R**, the user-user CF requires the similarity function ***s: UxU,*** calculating the similarity between two users and a method to use the similarity and rating in order to make the prediction.

In order to generate predictions or recommendations for a user u, user-user collaborative filtering uses s to compute a neighborhood $N \subseteq U$ of neighbors of u.

When the N has been calculated, the system immediately calculate the ratings of the users in N to generate predictions for user u's preference for an item i. This almost happen by computing the weighted average of the neighboring user's ratings i using the similarity as the weight.

$$P_{u,i} = \frac{\sum_{\bar{u} \in N} s(u,\bar{u})(r_{\bar{u},i} - \bar{r}_{\bar{u},i})}{\sum_{\bar{u} \in N} |s(u,\bar{u})|} \tag{4.3}$$

Subtracting the user's rating from each other, because the scale of the rating is different from one to the other.

The definitive design idea to implement user-user collaborative filtering is by choosing similarity function. There are several types of similarity functions have been proposed in the recommendation system. In our research, we have been applied three different types of similarity functions.

Pearson Correlation, this method is calculating the statistical correlation between two user's similar ratings to find their similarity [34]. Both of the GroupLens and BellCore had been used this method [32] [33]. Pearson correlation is calculating as follow:

$$s(u,v) = \frac{\sum_{i \in I_u \cap I_v} (r_{u,i} - \overline{r_u})(r_{v,i} - \overline{r_v})}{\sqrt{\sum_{i \in I_u \cap I_v} (r_{u,i} - \overline{r_u})^2} \sqrt{\sum_{i \in I_u \cap I_v} (r_{v,i} - \overline{r_v})^2}} \tag{4.4}$$

In Pearson correlation there is a kind of suffering when the system is computing the high similarity between users with a few ratings in common. This point can be decreased by putting a threshold on the number of the co-rated place or item, also measure the similarity when the number of the co-rated place or item been less than the threshold [35]. This threshold will represent a value r_z which is neither like or dislike, with an absolute reference. This is called as Constrained Pearson correlation [36], and it's computing as follow:

$$s(u,v) = \frac{\sum_{i \in I_u \cap I_v} (r_{u,i} - r_z)(r_{v,i} - r_z)}{\sqrt{\sum_{i \in I_u \cap I_v} (r_{u,i} - r_z)^2} \sqrt{\sum_{i \in I_u \cap I_v} (r_{v,i} - r_z)^2}} \tag{4.5}$$

Cosine similarity is one of the most popular type of similarity function [37], which is calculating the similarity between two user's common ratings vector, and it's known as the simplest and the fastest type of the similarity functions.

$$s(u,v) = \frac{\vec{u}.\vec{v}}{\|\vec{u}\|.\|\vec{v}\|} \tag{4.6}$$

Jaccard similarity index (also known as Jaccard similarity coefficient) is calculating the similarity between two set of data by compare the contain of them, the range of Jaccard similarity start from 0% to 100%, also this type of similarity sensitive to the small value and may give an erroneous results, especially with the small sample in the data set [38].

$$s(u,v) = \frac{u \cap v}{u \cup v} \tag{4.7}$$

All these recommendation system is for all the types of categories and places, also for what the users is looking for such as restaurant, clubs, sport salon and so on. It's offering what the users are looking for in different ways depending on the idea of each of the recommendation system. In our thesis we applied three different algorithms of personal recommendation system and generic recommendation system. We focus on which algorithm will provide the best suggestion for the user. The algorithms focus on the history of the users, recommendation system according to the categories and the recommending according to the popularity of the places. The first one is visiting popular places, it is about visit places depending on the number of check-in on that place that the other user made in that place and according to that the recommendation system will recommend the user to visit that place. The second algorithm is attending places by categories, here the user attend new places according to the category of that place and which category that the user attend to visit depending on that the recommendation system will recommend a place to the user. The third algorithm is attending places by user history, in this algorithm the mean idea is what the user is attend to visit which place and which type of venues that the user prefer to go to.

CHAPTER 5

EXPERIMENTAL STUDIES

5.1 Location Prediction

Location based social system network has been improved in the recommendation system both in generic recommendation and personalized recommendation. We focus on our job for the recommendation system about the personalized recommendation of social activities, friends, events, and locations. In the personalized recommendation, the main idea is to understand the users' behaviors and what they use to do and where they prefer to visit. Most of the recommendation systems depends on the dataset and what the dataset contain and according to that the next step of the prediction system can show. The personalized recommendation system depends on users' behaviors and what the they use to do such as the check-in history of each user, the popularity of the places user visited and the number of the check-ins made by a user's friends in the same places.

We applied some technical methods on our dataset for the prediction system, based on history of the users and the number of the check-ins made by users in the three cities mentioned eirlier. It also depended on categories so as to find the most popular one for the three cities. On the other hand, predicting next place, where the user would like to go. Based on the popularity of the place in our dataset to know the most popular city amongst Austin, Dallas and London by taking and counting the most high rating location and according to that we predict the popular city.

5.1.1 Recommendation System By Using Popular Places Algorithm

Visiting popular places is a non-personalized baseline on the rank of the place also on the number of check-ins made by users in that place. The idea of the visiting known places is based on the name of the place famous for a specific thing. It also depends on the rank of the place given from the users and the number of check-ins made by them in the venue.

Table 5.1 Austin city table for the most popular places

PLACE	CITY	CATEGORY	NO.OF CHECK-IN
Austin convention center	Austin	Convention Center	11,219
Austin bergstrom international	Austin	Airport	9,356
Starbucks	Austin	Coffee Shop	3,170
Seaholm power plant	Austin	Concert Hall	2,768

Table 5.2 London city table for the most popular places

PLACE	CITY	CATEGORY	NO.OFCHECK-IN
Starbucks	London	Coffee Shop	8,220
Terminal 5	London	Terminal	4,290
Apple Store	London	Electronic	3,301
LondonWaterloo Railway station WAT	London	Train Station	3,278

Table.5.3 Dallas city table for the most popular places

PLACE	**CITY**	**CATEGORY**	**NO.OFCHECK-IN**
Starbucks	Dallas	Coffee Shop	4,109
NorthPark Center	Dallas	Mall	1,158
Kroger	Dallas	Grocery Store	1,146
AMC Theatre	Dallas	Cineplex	3,278

5.1.2 Recommendation System By Using Popular Categories

Attended venues by the categories of that places, the prediction system in this method depends on the popularity of the category, on other word, the type of the category that most of the users prefer to visit. In all of the three cities there is too many types of categories which it contain too many places.

Table 5.4 Austin city table for the most popular categories

CATEGORY	CITY	NO.OF LOCATION
Bar	Austin	18,839
Hotel	Austin	15,218
Mexican	Austin	11,428
Airport	Austin	9,388

Table 5.5 London city table for the most popular categories

CATEGORY	CITY	NO.OF LOCATION
Pub	London	36,544
Train Station	London	35,040
Coffee Shop	London	16,328
Bar	London	16,089

Table 5.6 Dallas city table for the most popular categories

CATEGORY	CITY	NO.OF LOCATION
Home	Dallas	9,756
Mexican	Dallas	7,462
American	Dallas	6,923
Bar	Dallas	6,109

For the popular category algorithm, we calculate the number of check-in that users made under the name of the category to find which category was the most popular on each of the three cities.

5.1.3 Recommendation System By Using User History

This method depends on the history of each user and giving the next move depending on what that user prefers to visit and tends to go. Each user has a type of venues prefferred to visit according to the program that speculates the next place for the user based on their histories. We applied this method to our dataset for each user in the three cities and predict the next move for each user depending on their history and the number of check-in that the user made in his/her favourite venues.

Table 5.7 The most popular places in Austin city according to the history of selected users

PLACE	**USERID**	**CITY**	**CATEGORY**	**CHECK-IN OFSAMPLE USER**
Rista Bar and crill	34713233	Austin	Burgers	201
Parside at lake creek	57861921	Austin	Home	187
Browning Hangar	30935019	Austin	Sculpture	159

Table 5.8 The most popular places in Dallas city according to the history of selected users

PLACE	**USERID**	**CITY**	**CATEGORY**	**CHECK-IN OFSAMPLE USER**
Andy's House	208854807	Dallas	Home	231
Amy And Spencer	125855370	Dallas	Home	196
AAA Texas	29010688	Dallas	Oter_Travel	163

Table 5.9 The most popular places in London city according to the history of selected users

PLACE	USERID	CITY	CATEGORY	CHECK-IN OFSAMPLE USER
Paddingtion Station PAD	20656631	London	Train Station	245
TFL Bus 100	40846787	London	Bus	178
London Liverpool street Railway	148302559	London	Train_stationl	169

In the previous tables the prediction system for each city depends on the check-ins of the users in all the three cities. According to the history of the users on Foursquare program to predict the next place where a user would go.

5.1.4 Recommendation System By Using User-User Collaborative Filtering

User-User CF, it's all depend on find the other user who have similar history in their behaviors to visit and attend to do the same kind of habit. User-user CF looks for other users who have high agreement with one of the users on an item or place that they both had been rated.

Table 5.10 Austin city result by using User-user CF

PLACE	CITY	CHECK-IN OF SAMPLE USER
Pimp Palace	Austin	18
YNN Austin	Austin	17
Campus Advantage	Austin	17

Table 5.11 Dallas city result by using User-user CF

PLACE	**CITY**	**CHECK-IN OF SAMPLE USER**
Thundercatsnyy Buy And Sell and thundercatsnyyare21	Dallas	32
The Office Grill	Dallas	27
Trinity Mills Station DART Rail \\/ A-train	Dallas	25

Table 5.12 London city result by using User-user CF

PLACE	**CITY**	**CHECK-IN OF SAMPLE USER**
Starbucks	London	41
Apple Store	London	32
TFL Bus 100	London	29

5.2 Evaluation and Test Results

We have evaluate the recommendation system algorithms results and compare it across the predictor system. In this section, we describe our methodology and the metrics, used here, to evaluate the recommendation system quality. In this section we separate the dataset in to train and test in order to calculate accuracy of the results. We filter the dataset from zero and unknown check-in $C_{i,j}$ value to have more clear result. We use three metrics to calculate and find the quality of these recommendation system results we have reached. We calculate the precision and the recall for each of the three methods we analyzed our dataset on for all of the three cities to major the quality of our prediction and analysis system to find the best result as shown in Tables.

The user will visit the top-N places as the result on the tables. The measurement of the recall and the precision depends on true positives (tp), false positives (fp), and false negatives (fn):

$$Precision = \frac{tp}{tp + fp}\ , Recall = \frac{tp}{tp + fn} \qquad (5.1)$$

The user also will visit the top-N places and will not attempt to visit the places, which is under the rank of N in the precision @N and recall @N.

Three formulas to calculate the prediction system quality are precision and the recall we applied on it to calculate the quality of prediction system. Also we applied the average percentile rank (APR) in order to understand which rank reflects users' interest, that the places ranked as higher than the others, in other word get visited and number of check-in more than the others. In order to calculate the average rank, we first define the interest as i that a user u has in the venue s, the highest number of check-in, also we define the $rank_{u,s}$ as the percentile ranking of the venue s for the user u in the ranked list of venue. To show the list of the places, if the $rank_{u,s} = 0$ then the places appear first in the list, also if the $rank_{u,s} = 1$ it mean it will show the last in the list. Average Percentile Ranking (APR):

$$\overline{APR} = \frac{\sum_{u \in U} \sum_{s \in L} i_{u,s} \times rank_{u,s}}{\sum_{u \in U} \sum_{s \in L} i_{u,s}} \qquad (5.2)$$

Table 5.13 Austin city the precision@N, recall@N and APR@N

Method / Austin city	**APR@30**	**Precision@30**	**Recall@30**
Popular places	0.691	0.057	0.065
Attending by categories	0.283	0.016	0.073
User history	0.216	0.1056	0.016

Table 5.14 Austin city the precision@N, recall@N and APR@N in CF

User-user CF/ Austin	**APR@30**	**Precision@30**	**Recall@30**
Cosine Similarity	0.569	0.062	0.065
Jaccard Similarity	0.431	0.048	0.052
Pearson Correlation	0.555	0.067	0.068

Table 5.15 Dallas city the precision@N, recall@N and APR@N

Method / Dallas city	**APR@30**	**Precision@30**	**Recall@30**
Popular places	0.672	0.052	0.061
Attending by categories	0.21	0.013	0.069
User history	0.16	0.057	0.091

Table 5.16 Dallas city the precision@N, recall@N and APR@N in CF

User-user CF/ Dallas	**APR@30**	**Precision@30**	**Recall@30**
Cosine Similarity	0.732	0.062	0.082
Jaccard Similarity	0.559	0.053	0.054
Pearson Correlation	0.691	0.061	0.075

Table 5.17 London city the precision@N, recall@N and APR@N

Method / London city	**APR@30**	**Precision@30**	**Recall@30**
Popular places	0.624	0.048	0.056
Attending by categories	0.207	0.025	0.012
User history	0.214	0.012	0.024

Table 5.18 London city the precision@N, recall@N and APR@N in CF

User-user CF/ London	**APR@30**	**Precision@30**	**Recall@30**
Cosine Similarity	0.765	0.074	0.081
Jaccard Similarity	0.585	0.049	0.062
Pearson Correlation	0.675	0.062	0.065

After we applied the Four-prediction idea according to the popularity of the places, history of the users and predicting the next move according to the categories for the places. We calculate the accuracy of each prediction system ways for the three cities to understand which of the prediction system ways will recommend to the users with more mobility for choosing the next places. According to our dataset information and what we applied on it. We found that the highest average percentile rank went to the most popular places moethod. It mean that the system will recommend to the users according to the popularity of the place, in another word according to the number of check-in for the place also the rank that the place took in the application. Each of these prediction system ways has the benefit of it and a different accuracy for each of the three ways. It’s depend also on the users idea for what the users are looking for and which type of categories that the users attend to go and therefore the places under that categories. Also, each of these results depends on the type of the dataset and what the dataset contains.

CHAPTER 6

CONCLUSION

In this thesis, we have investigated a recommendation system for location based social network. We focus in our work on the generic recommendation system and personalized recommendation system to improve the prediction system. In the generic recommendation system, we use two method (popular place algorithm and popular cateory algorithm) to predict the next step of the users through their check-in in the three cities that we have in our dataset. Also we calculate average precentile ranking, precision and recall for each of the two method.

In the personalized recommendation system, we use two method to predict the next step of the users depend on user's check-in. The two method that we applied depends on users behaviors and what the user prefer to select. User history algorithm and user-user collaborative filtering, applied to find the most popular places in the three cities depending on user behaviors. Also we calculate averge precentile ranking, precision and recall for each of three similarity methods that we applied to find which one get the highest rank from the other. According to our dataset the best type of the similarity method were cosine similarity method. For the user history method, we depend on the history of each user in order to predict the next move for each of the users.

In the terms of the future work, we seek to evaluate our recommendation system and how we predict the next move. By applying a recommendation system depen on the comment of the users. By checking the comment if it's positive or nigative. Also count these comment that give from the users as a point for the place or the item.

REFERENCES

[1] Flicker Dataset, http://www.flickr.com, 3 October 2016.

[2] Foursquare Dataset, https://foursquare.com , 3 September 2016.

[3] E. Martin, O. Vinyals, G. Friedland and R. Bajcsy, (2010). "Precise Indoor Localization Using Smart Phones", ACM Multimedia 2010, 787–790.

[4] Gps Explenation, http://www8.garmin.com/aboutGPS/ , 24 March 2017.

[5] Jiang, B. and Yao, X. B. (2006). "Location-based services and GIS in perspective. Computers, Environment and Urban Systems". 30(6): 712-725.

[6] I. A. Getting, (1993). "The Global Positioning System", IEEE Spectrum, 30.

[7] Needleman, Rafe; Claire Cane Miller: Adrianne Jeffries (2010). "Reporters' Roundtable: Checking in with Facebook and Foursquare", CNET.

[8] Daniele Quercia, et al., (2010)."Recommending Social Events from Mobile Phone Location Data", ICDM.

[9] M. C. Gonzalez, C. A. Hidalgo, and A.L.Barabasi, (2008)."Understanding individual human mobility patterns," Nature, .453(10): 779–782.

[10] C.Y. Chow, J. Bao, and M. F. Mokbel., (2010) "Towards Location-based Social Networking Services", In ACM SIGSPATIAL-LBSN.

[11] Recommendation System Intoduction: https://medium.com/recombee-blog/recommender-systems-explained-d98e8221f468#.133l6fgqy, 12 July 2016.

[12] A. Seth and J. Zhang, (2008)."A Social Network Based Approach to Personalized Recommendation of Participatory Media Content", In Int. Conf. on Weblogs and Social Media (ICWSM).

[13] Cartwright, W.Crampton, J. Gartner, G. Miller, S.Mitchell, K. Siekierska, E. and Wood, J. (2001). "Geospatial Information Visualization User Interface Issues. Cartography and Geographic Information Science", 28(1): 45-60.

[14] The full list of foursquare badges: http://www.4squarebadges.com/foursquare-badge-list/, 15 June 2016.

[15] Foursquare Badges Level up to Encourage Exploration for The Foursquare Badge Update: http://sproutsocial.com/insights/foursquare-badge-update/, 20 July 2015.

[16] How Important is Foursquare to your Business to Check Business Life: http://conversationalmarketinglabs.com/blog/2011/11/social-media-2/how-important-is-foursquare-to-your-business/, 11 November 2011.

[17] G. Adomavicius and A. Tuzhilin, (2005). “Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions”. IEEE TKDE, 17(6): 734–749.

[18] Y. Koren, (2009).”Collaborative Filtering with Temporal Dynamics. In Proceedings of KDD” , 14(9):89–97.

[19] Jihang Ye_ Zhe Zhu_ Hong Cheng (2013), “What’s Your Next Move: User Activity Prediction in Location-based Social Networks”, SIAM.

[20] G.B. Colombo, M.J. Chorley, M. J. Williams and S.M. Allen, R.M, (2011). “You are Where You Eat: Foursquare Check-in as Indicators of Human Mobility and Behavior” IEEE.

[21] Group 42, Yiying Cheng, Yangru Fang and Yongqing Yuan, (2012).“Recommendation System for Location based Social Network”, CS224W.

[22] Luca Rossi and Mirco Musolesi. (2014). “It’s the Way you Check-in:Identifying Users in Location-Based Social Networks” IEEE.

[23] Rui Zhang, Bob Price, Maurice Chu and Alan Walendowski, (2012). “Location-based Predictions for Personalized Contextual Services using Social Network Data”, PARC.

[24] Chen Cheng, Haiqin Yang, Michael R. Lyu and Irwin King, (2012). “Where You Like to Go Next: Successive Point-of-Interest Recommendation”, Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence.

[25] Anastasios Noulas, Salvatore Scellato, Neal Lathia and Cecilia Mascolo, (2012) “A Random Walk Around the City: New Venue Recommendation in Location-Based Social Networks”, IEEE.

[26] Quan Yuan and Gao Cong, (2014). “Graph-based Point-of-Interest Recommendation with Geographical and Temporal Influences”, CIKM.

[27] Symeonidis, Panagiotis, Ntempos, Dimitrios, Manolopoulos and Yannis, (2014). “p. Symeonides et al. Recommender Systems for Location Based Social Network Springersbriefs for Electrical Engineering” , Springer, 97.

[28] Francesco Ricci and Lior Rokach and Bracha Shapira, (2011) “Introduction to Recommender Systems Handbook”, Springer, 1-35.

[29] Terveen Loren and Hill Will (2001). "Beyond Recommender Systems: Helping People Help Each Other" IEEE.

[30] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins, (2001). “Eigentaste: a Constant Time Collaborative Filtering Algorithm,” Information Retrieval, .4(2): 133–151.

[31] J.S. Breese, D. Heckerman, and C. Kadie, (1998). “Empirical Analysis of Predictive Algorithms for Collaborative Filtering”, ACM.

[32] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, (1994). “GroupLens: an Open Architecture for Collaborative Filtering of Netnews”, CSCW.

[33] Hill, L. Stead, M. Rosenstein, and G. Furnas, (1995). "Recommending and Evaluating Choices in a Virtual Community of Use," in ACM CHI (95): 194-201.

[34] M. Sarwar, G. Karypis, J. A. Konstan, and J. Reidl, (2001). "Item-Based Collaborative Fltering Recommendation Algorithms".

[35] J. Herlocker, J. A. Konstan, and J. Riedl, (2002). "An Empirical Analysis of Design Choices in Neighborhood-Based Collaborative Fltering Algorithms" Information Retrieval, .5(4): 287–310.

[36] U. Shardanand and P. Maes, (1995). "Social information filtering: Algorithms for automating "word of mouth"," in ACM CHI, (95): 210–217.

[37] G. Karypis, (2001). "Evaluation of Item-Based Top-N Recommendation Algorithms", in ACM CIKM, 01: 247–254.

[38] H. R. Lasker. (1979)."Light Dependent Activity Patterns Among Reef Corals: Montastrea Cavernosa", Biological Bulletin, 156:196–211.

CURRICULUM VITAE

PERSONAL INFORMATION

Name Surname : Omar ALMALLAH

Date of birth and place :17.08.1991 , Mosul Iraq

Foreign Languages :Turkish, English , French and Mother language is Arabic

E-mail : omarfiras991@gmail.com

EDUCATION

Degree	**Department**	**University**	**Date of Graduation**
Undergraduate	Computer Engineering	Technical College of Mosul	07.07.2013
High School	Applied Sciences	Alsharqiya High School	25.06.2009

PUBLISHMENTS

Conference Papers

1. Omar F.Almallah and Songül Albayrak, (2017).” Predicting venues inLocation Based Social Network ”, Seventh International Conference on Computer Science, Engineering and Application (CCSEA 2017).

2. Omar F.Almallah and Songül Albayrak, (2017).” Predicting venues in Location Based Social Network”, International Conference on Progress in Applied Science (ICPAS’17).