T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

KONUM TABANLI SOSYAL AĞLARDA KONUM TAHMİNİ

MÜCAHİT BAYDAR

YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
BİLGİSAYAR MÜHENDİSLİĞİ PROGRAMI

DANIŞMAN
DOÇ. DR. SONGÜL ALBAYRAK

İSTANBUL, 2016

T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

KONUM TABANLI SOSYAL AĞLARDA KONUM TAHMİNİ

Mücahit BAYDAR tarafından hazırlanan tez çalışması 29.06.2016 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Tez Danışmanı

Doç. Dr. Songül ALBAYRAK

Yıldız Teknik Üniversitesi

Jüri Üyeleri

Doç. Dr. Songül ALBAYRAK

Yıldız Teknik Üniversitesi _____

Prof. Dr. Oya KALIPSIZ

Yıldız Teknik Üniversitesi _____

Prof. Dr. Hasan DAĞ

Kadir Has Üniversitesi _____

ÖNSÖZ

Bu çalışmada, konum tabanlı sosyal ağlarda konum tahmini problemi üzerinde durulmuştur. Konum tabanlı sosyal ağlar popülerliklerini her geçen gün artırırken, bu ağlar sayesinde insanların hareketlilikleri hakkında daha önce elde etmesi oldukça güç bilgilere erişmek de mümkün hale gelmiştir. Bu alanda popüler olan iki sosyal ağdan elde edilen bilgiler üzerinde kapsamlı bir analiz yapılmış ve tahmin için kullanılabilecek özellikler belirlenmiştir. Bu özellikleri bir arada kullanan ve problemi farklı açılardan ele alabilen bir yöntem önerilmiştir.

Çalışma boyunca beni yönlendiren ve yardımcı olan tez danışmanım ve hocam Doç. Dr. Songül ALBAYRAK'a,

Bu çalışmada kullandığım veri setlerini benimle paylaşan Anastasios Noulas'a,

Bugünlere gelmemde en büyük emeğe sahip, beni daima destekleyen ve yardımlarını eksik etmeyen anneme, babama ve ablama teşekkürlerimi sunuyorum.

Haziran, 2016

Mücahit BAYDAR

İÇİNDEKİLER

Sayfa

SİMGE LİSTESİ vi

KISALTMA LİSTESİ vii

ŞEKİL LİSTESİ viii

ÇİZELGE LİSTESİ ix

ÖZET x

ABSTRACT xii

BÖLÜM 1

GİRİŞ 1

1.1 Literatür Özeti 1
1.2 Tezin Amacı 13
1.3 Hipotez 13

BÖLÜM 2

VERİ ANALİZİ 14

2.1 Veri Toplanması 14
2.2 Veri Ön İşleme 15
2.3 Veri Analizi 16
2.3.1 Foursquare Analizi 17
2.3.2 Gowalla Analizi 19

BÖLÜM 3

KONUM TAHMİNİ PROBLEMİ 22

3.1 Problem Tanımı 22
3.2 Problemin Formüle Edilmesi 22
3.3 Tahmin Yöntemleri 23
3.3.1 Tekil Özellikler İle Tahmin 23

3.3.2 Önerilen Yöntem İle Tahmin 25
3.4 Test Yöntemi ve Ölçütler 26
3.5 Test Ortamı 27

BÖLÜM 4

DENEYSEL SONUÇLAR 28

4.1 Foursquare Sonuçları 28
4.2 Gowalla Sonuçları 33
4.3 Foursquare-Gowalla Karşılaştırılması 38

BÖLÜM 5

SONUÇLAR VE GELECEK ÇALIŞMALAR 41

KAYNAKLAR 43

ÖZGEÇMİŞ 45

SİMGE LİSTESİ

A_k	k kullanıcısının arkadaş seti
K	Kullanıcı seti
Ka	Kategori seti
Ka_k	k kullanıcısının ziyaret ettiği kategori seti
$Liste_A$	Tahmin listesi
$Liste_B$	Tahmin listesi
M	Mekân seti
m_t	Yer bildiriminin yapıldığı zaman aralığında mekanın popülerliğini
m_c	Yer bildiriminin yapıldığı mekan
$r_{k,ka}$	k kullanıcısının ka kategorisini tercih etme oranı
P_t^{10}	Kategoriler arası geçiş ihtimali (10 dakikadan az süre için)
P_t^*	Kategoriler arası geçiş ihtimali (zaman kısıtlması olmadan)
$Skor_{k,m}$	m mekanının skoru
$Skor_m$	m mekanının skoru
Y_k	k kullanıcısının yer bildirimi seti
{k,m}	Yer bildirimi verisi

KISALTMA LİSTESİ

API	Application Programming Interface
GPS	Global Positioning System
HMM	Hidden Markov Model
RMF	Regularized Matrix Factorization
RWR	Random Walk with Restart
TF-IDF	Term Frequency-Inverse Document Frequency

ŞEKİL LİSTESİ

Sayfa

Şekil 1. 1 Harita ve Komşuluk Matrisi 3
Şekil 2. 1 Foursquare aylara göre yer bildirimi sayısı 17
Şekil 2. 2 Foursquare şehirlere göre yer bildirimi sayısı 17
Şekil 2. 3 Foursquare şehirlere göre mekân sayısı 18
Şekil 2. 4 Foursquare şehirlere göre kullanıcı sayısı 18
Şekil 2. 5 Gowalla aylara göre yer bildirimi sayısı 19
Şekil 2. 6 Gowalla şehirlere göre yer bildirimi sayısı 20
Şekil 2. 7 Gowalla şehirlere göre mekân sayısı 20
Şekil 2. 8 Gowalla şehirlere göre kullanıcı sayısı 21
Şekil 4. 1 Foursquare Austin tahmin sonuçları 28
Şekil 4. 2 Foursquare Dallas tahmin sonuçları 29
Şekil 4. 3 Foursquare San Francisco tahmin sonuçları 30
Şekil 4. 4 Foursquare Londra tahmin sonuçları 31
Şekil 4. 5 Foursquare önerilen yöntem tahmin sonuçları 32
Şekil 4. 6 Gowalla Austin tahmin sonuçları 33
Şekil 4. 7 Gowalla Dallas tahmin sonuçları 34
Şekil 4. 8 Gowalla San Francisco tahmin sonuçları 35
Şekil 4. 9 Gowalla Stockholm tahmin sonuçları 36
Şekil 4. 10 Gowalla önerilen yöntem tahmin sonuçları 37
Şekil 4. 11 Austin önerilen yöntem tahmin sonuçları 39
Şekil 4. 12 Dallas önerilen yöntem tahmin sonuçları 39
Şekil 4. 13 San Francisco önerilen yöntem tahmin sonuçları 40

ÇİZELGE LİSTESİ

Sayfa

Çizelge 1. 1 Kategori Aktivite Geçişi 7
Çizelge 1. 2 Kategori Mekân Geçişi 7
Çizelge 4. 1 Foursquare şehirlere göre yer bildirimi, kullanıcı, mekân, ortalama kullanıcı başına düşen yer bildirimi ve ortalama mekân başına düşen yer bildirimi sayısı 32
Çizelge 4. 2 Gowalla şehirlere göre yer bildirimi, kullanıcı, mekân, ortalama kullanıcı başına düşen yer bildirimi ve ortalama mekân başına düşen yer bildirimi sayısı 37

ÖZET

KONUM TABANLI SOSYAL AĞLARDA KONUM TAHMİNİ

Mücahit BAYDAR

Bilgisayar Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

Tez Danışmanı: Doç. Dr. Songül ALBAYRAK

Mobil cihazlardaki ve kablosuz ağlardaki gelişmelere bağlı olarak son yıllarda konum tabanlı sosyal ağlar da popülerliğini artırmaktadır. Bu sosyal ağlar kullanıcıların yeni mekânlar keşfetmelerine, fotoğraf, video ve konumlarını paylaşmalarına ve arkadaşlık kurmalarına olanak sağlamaktadır. Konum tabanlı sosyal ağlar sayesinde daha önce elde etmesi oldukça zor olan büyük miktardaki kullanıcıların hareketleri ile ilgili verilere erişmek mümkün hale gelmiştir. Bu verileri kullanarak kullanıcılara başarılı öneriler sunulması hem sosyal ağ sağlayıcılarına, hem kullanıcılara, hem de sistemdeki mekânlara yarar sağlamaktadır.

Konum tabanlı sosyal ağlar sayesinde elde edilebilecek çok fazla veri olmasına karşın şu an için elde edilebilen veriler genelde oldukça az ve ham verilerdir. Bu veriler kullanıcının hangi tarihte nerede yer bildiriminde bulunduğu ve kullanıcılar arası arkadaşlık durumu bilgileridir. Bunların haricinde konum tabanlı sosyal ağlar sayesinde elde edilebilecek verilere örnek olarak; kullanıcıların mekânlara verdikleri puanlar, mekân hakkında yaptıkları yorumlar ve mekânların fiyat aralıkları gösterilebilir. Elde edilen veriler ham halde olsalar bile bu veriler işlendikten sonra birçok farklı bilgi elde edilebilir. Sadece yer bildirimi verisi kullanılarak mekânların popülerliklerine, kullanıcıların mekân ve kategori tercihlerine, mekânların popüler oldukları zaman aralıklarına ve daha birçok bilgiye ulaşılabilir.

Elde edilen veriler öncelikle analiz edilmiş ve bir ön işlemeden geçirilmiştir. Sistemi düzenli kullanmayan kullanıcı verileri elenmiştir. Benzer şekilde sadece tek bir yer bildirimine sahip mekânlar sistemden çıkarılmıştır. Bunun sebebi bu mekânların incelediğimiz özellikler ile önerilmesinin oldukça zor olmasıdır. Daha sonra yer bildirimleri yapılan konumların hangi ülke ve şehirlerde oldukları tespit edilmiştir. Yapılan incelemeler sonucu kullanıcıların büyük çoğunluğunun yer bildirimlerini aynı şehir içerisinde yaptıkları görülmüş ve öneri yapılırken bu göz önüne alınmıştır. Öneri sistemi çalıştırılırken veriler şehir bazlı incelenmiş ve buna göre başarı hesaplanmıştır.

Bu çalışmada kullanıcıların hareketleri önceden tahmin edilmeye çalışılmıştır. Yapılan işlem test için belirlenen yer bildirimini yapan kullanıcı için sistemdeki her bir mekâna puan verme ve bu puanları sıralama işlemidir. Bu sıralama sonucunda farklı liste uzunluklarında öneri yapılmış ve yer bildiriminin gerçekten bu liste içinde olup olmadığı, var ise hangi sırada var olduğu incelenmiş ve başarı hesaplanmıştır. Bunun için öncelikle elde edilen veriler teker teker kullanılmış ve başarıları incelenmiştir. Bu veriler mekân popülerliği, mekânın zaman aralıklarındaki popülerlikleri, kullanıcıların önceki mekân ve kategori tercihleri, mekânın kullanıcının evine olan uzaklığıdır. Daha sonra bu verileri bir arada kullanan bir sistem tasarlanmıştır. Bu sistemin başarısı incelenmiş ve önceki elde edilen sonuçlarla karşılaştırılmıştır.

Anahtar Kelimeler: Konum tabanlı sosyal ağlar, öneri sistemleri, yer bildirimi verisi, konum tahmini

ABSTRACT

LOCATION RECOMMENDATION IN LOCATION-BASED SOCIAL NETWORKS

Mücahit BAYDAR

Department of Computer Engineering

MSc. Thesis

Adviser: Assoc. Prof. Songül ALBAYRAK

Depending on the developments in mobile devices and wireless networks, location-based social networks have been gaining popularity in recent years. These social networks allow their users to explore new places and share their location, videos and photos and make friends. Location-based social networks also help us to get huge amount of useful information about the mobility of users which is not possible previously. By using this information, presenting good recommendations to users provide benefits to social network providers and users and venues in the system.

Even if there are lots of information can be obtained from location-based social networks, only a small part of it can be obtained which is also raw. This information is users' check-in date and places and friendship data. Rather than these information, users' ratings to venues, comments on venues and venues' price ranges can be obtained from location-based social networks. Even if the information is raw at first, after processing this information we can acquire lots of useful data. Using only the check-in data, we can get venue popularities, users' preference on venues and categories, venue popularity time ranges and much more information.

At first the data is analyzed and preprocessed. İnfrequent users and their information are removed from the dataset. The venues with only have one check-in are also removed. This process is necessary because with little information prediction is slightly hard operation. Then every check-in data is linked to a city and a country. After

examinations we realized that users tend to make check-ins mostly in the same city and we decided to use this information. Prediction system is tested city by city and results are calculated in this way.

In this work users' movements are tried to guessed. The process is giving points to the places for the selected user whose check-in is being tested according to the prediction method and ranking them with given points. After ranking places are chosen for different length of prediction list. Prediction lists are checked whether one of the places in the list is correct place. Then the order of correct place in the list is used to calculate system performance. The features that are obtained in the preprocessing and analyzing phase are used for prediction individually. These features are venue popularity, venue time range frequency, users' venue and category preferences, venue distance to the user's home and friendship. Finally a new method is proposed that is combining the individual features and performance of the new method is compared to the previous results.

Keywords: Location-based social networks, recommendation systems, check-in data, location prediction

BÖLÜM 1

GİRİŞ

Konum tabanlı sosyal ağlar araştırmacıların ilgisini uzun süredir çekmektedir. Bu alanda yapılan ilk çalışmalar sadece kullanıcıların konum bilgisini ve harita bilgisini kullanarak sonuç elde etmeye odaklanmıştır. Sonraki çalışmalarda bu bilgilere ilaveten kişisel ve genel birçok bilgi değerlendirilmiştir. Örnek olarak kullanıcıların geçmiş konum bilgileri, mekân ve kategori tercihleri, mekânların popülerlik bilgileri verilebilir. Bu yeni değerlendirmeye alınan bilgiler sayesinde sistemlerin başarısı artmıştır.

1.1 Literatür Özeti

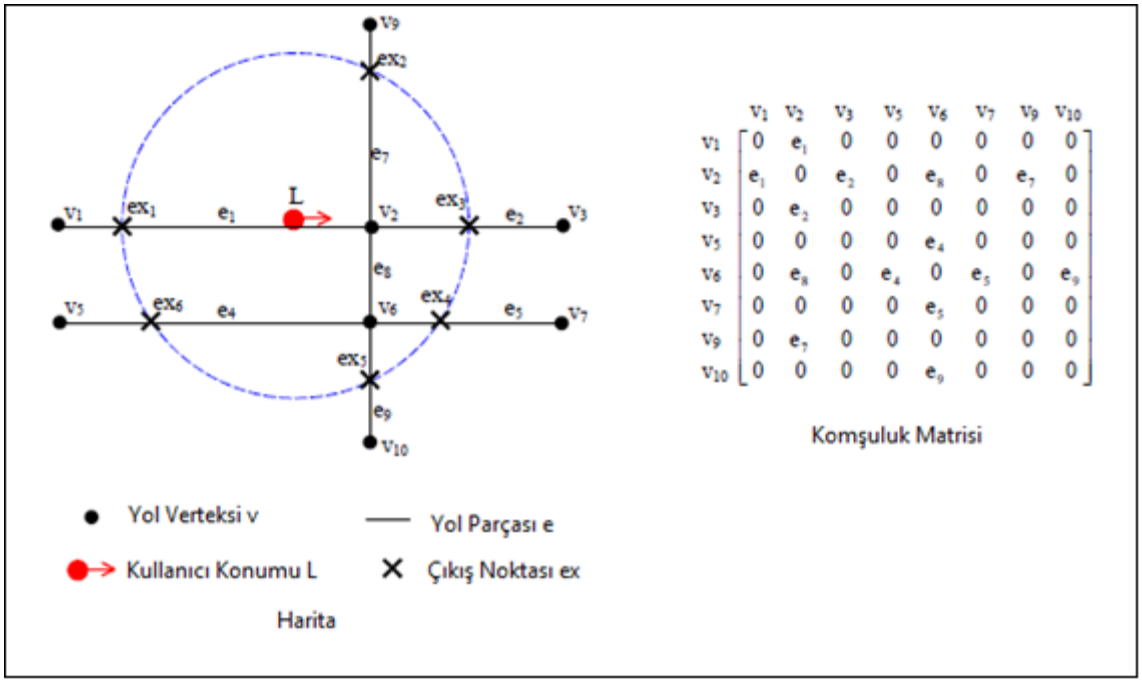
Bu bölümde konum tabanlı sosyal ağlarda konum tahmini üzerine yapılmış akademik çalışmalar incelenmiştir.

Konum tabanlı sosyal ağlarda kullanıcı yer bildirimi verilerini kullanarak kişiselleştirilmiş bir öneri sistemi sunan [1] çalışmasında veriler Gowalla üzerinden elde edilmiştir. Veriler 212 bin kullanıcı ve 1,5 milyon farklı konum bilgisini içermektedir. Düzenli olarak yer bildirimi yapmayan kullanıcılar aktif sayılmamış ve elenmiş sonucunda veri tabanında 79 bin kullanıcı kalmıştır. Öneri sistemindeki amaç kullanıcılara daha önce gitmedikleri ve ilgi duyabilecekleri konumları tahmin etmektir. Gowalla üzerinden kullanıcıların mekânlara verdikleri puan bilgisine erişilemediğinden kullanıcıların mekânlara gitme sıklıklarına göre bu puan bilgisi hesaplanmıştır. Yer bildirimi verilerinden bir model oluşturmak için önce veri, kullanıcı/nokta matrisi şeklinde gösterilmiştir. Matristeki veriler kullanıcıların o noktaya olan ilgilerini temsil etmektedir. İşbirlikçi filtreleme algoritmasının amacı kullanıcıların daha önce gitmedikleri noktalara(boş olan verilere) olan ilgilerini eldeki verileri kullanarak elde

etmektir. Tahminler yapıldıktan sonra kullanıcı için en yüksek ilgiye sahip N konum belirlenebilecektir. Regularized Matrix Factorization(RMF) yönteminde kullanıcılar ve mekânlar d boyutlu bir uzayda birbirlerine eşlenmektedir. Her mekân d boyutlu bir vektöre, her kullanıcı da d boyutlu bir kullanıcı tercihine tekabül eder. Bu vektörler kullanılarak her bir mekân için kullanıcıların derecelendirme değerleri hesaplanır. Çalışmada öğrenme algoritması olarak "stochastic gradient descent" kullanılmıştır. Önce sisteme rastgele değerler atanmış daha sonra tahmin hataları hesaplanarak bu değerler güncellenmiştir. Bu yöntemin büyük bir avantajı da bütün verinin aynı anda ana hafızaya taşınmasına gerek duymamasıdır. Bunun yerine aynı anda sadece tek bir eğitim verisini işleyebilmektedir. RMF yönteminin bir diğer avantajı ise çok hızlı çalışmasıdır. Yöntem bir kez eğitildiğinde, O(1) zamanda (iki vektörün nokta çarpımı için gereken zamanda) çalışabilmektedir. Dezavantajı ise algoritmanın uygun değerlerinin belirlenmesinin zor oluşudur. Sistem performansı değerlendirilirken veriler konum tabanlı sosyal ağın en yoğun olarak kullanıldığı iki şehirden alınmıştır. Sonuçlar standart ortalama öge öneri sistemiyle karşılaştırılmıştır. Her iki şehir verileri içinde önerilmiş olan kişiselleştirilmiş işbirlikçi filtreleme yöntemi standart yöntemi geride bırakmıştır. Buradan hareketle konum tahmini için kişiselleştirilmiş işbirlikçi filtreleme yöntemlerinin bu tip veriler için uygulanabilir olduğu sonucu çıkartılmıştır.

Konum tabanlı servislerde konum tahmini üzerine yapılan ilk çalışmalardan biri olan [2] bu çalışmada kullanıcıların konumlarının anlık olarak bilindiği varsayımı yapılmıştır. Çalışmada harita tabanlı bir konum tahmin modeli önerilmektedir. Model bir veri tabanı modülü, bir rota tahmin modülü, bir konum belirleme modülü ve hata kontrol mekanizmasından oluşmaktadır. Sistem temel olarak bir mobil istemci ve sabit sunucu kullanmaktadır. Her istemcinin GPS alıcısı olduğu ve bu veriyi anlık olarak sunucu ile paylaşabildiği varsayılmıştır. Sistem çalışırken kullanıcının konumu harita üzerinde işaretlenmekte (Şekil 1. 1) ve anlık hızına ve tahmin büyüklüğüne göre belirli bir alan içerisinde kalan mekânlara gidebileceği rotalar hesaplanmaktadır. Rotalar arasında seçim yapmak için bir çizge ile modelleme yapılmakta ve olasılıkları hesaplamak içinde bir olasılık matrisi (Şekil 1. 1) kullanılmaktadır. Olasılıklar kullanıcının önceden geçtiği yollar ve kullanıcı profil bilgilerine göre hesaplanmaktadır. Çalışma bu alanda yapılacak

yeni çalışmalarda kullanılabilecek bu yöntemi sunmakta ancak yöntemin başarılarıyla ilgili herhangi bir veri paylaşmamaktadır.



Şekil 1. 1 Harita ve Komşuluk Matrisi

Konum tabanlı bir öneri sistemi oluşturan [3] çalışmasında bayes ağları kullanılarak bir model oluşturulmuştur. Model içerik kayıt toplayıcısı, tavsiye modülü ve sistem eşleştirme modüllerinden oluşmaktadır. Kullanıcı verileri mobil cihaz aracılığıyla toplanmaktadır. Kullanıcıdan isim, cinsiyet, yaş, kan grubu, aylık gelir ve tercih edilen yemek bilgileri alınmaktadır. Bunların haricinde sisteme hava durumu bilgileri, sıcaklık, gün, mevsim ve kullanıcının yeri gibi bilgilerde verilmektedir. Toplanan veriler ön işleme sonrası Bayes ağlarının parametrelerini eğitmek için kullanılmaktadır. Sisteme yeni bir istek geldiğinde her bir özellik için en yüksek ihtimal hesaplanmakta ve bu ihtimaller ışığında veri tabanında en benzer mekân bulunmaktadır. Bu çalışmada yalnızca 4 kullanıcı ve 50 mekân verisi kullanmıştır ve veriler bir haftalık bir aralıkta alınmıştır. Kullanılan verilerin azlığı başarıyı doğrudan etkilemiştir ve elde edilen sonuçların başka çalışmalarla karşılaştırılması mümkün olmamıştır.

Mobil cihaz konum verilerini kullanarak sosyal etkinlik önerisinde bulunan [4] çalışmasında veriler Airsage Inc. firmasının veri tabanından elde edilmiştir. Firma

kullanıcılarının konum bilgilerini kullanarak trafik yoğunluk hesaplaması yapmaktadır. Sosyal etkinlikleri elde etmek için Boston Globe Calendar internet sitesi incelenmiş ve 500 farklı etkinlik verisi elde edilmiştir. Daha sonra konum olarak birbirine yakın olan etkinlikler ve kullanıcılar elenmiş ve sonucunda 2519 kullanıcı ve 53 sosyal etkinlik kalmıştır. Bu çalışmada özel olarak hakkında daha önceden bir veri sahibi olmadığımız kullanıcıların hareketleri tahmin edilmeye çalışılmıştır. Böyle bir durumda tahmin yapılabilmesinin tek yolu kullanıcının konumunu ve mekânların popülerlik ve konumlarını kullanmaktır. Çalışma boyunca farklı teknikler incelenmiş ve başarıları birbirleriyle karşılaştırılmıştır. Kullanılan yöntemler sırasıyla şu şekildedir:

1- Popüler etkinlikler: Kullanıcıya en popüler etkinlik önerilir.

2- Coğrafi olarak yakın etkinlikler: Kullanıcıya en yakınındaki etkinlik önerilir.

3- Çevredeki popüler etkinlikler: Kullanıcıya evine yakın alandaki popüler etkinlikler önerilir.

4- Terim Frekansı-Ters Metin Frekansı(TF-IDF): Bazı etkinlikler genele bakıldığında popüler olmayabilir ancak küçük bir alan incelendiğinde popüler olduğu söylenebilir. Genel olarak popüler olmayan ancak bu küçük alanda popüler olan etkinliklerin önemi artırılarak bu etkinlikler önerilmeye çalışılmaktadır.

5- K-en yakın konum: Kullanıcının bulunduğu konuma en çok benzeyen k-en yakın konum hesaplanır ve buradaki etkinlikler önerilir.

6- K-en yakın etkinlik: Kullanıcının bulunduğu konumdaki etkinliklere benzeyen k-en yakın etkinlik hesaplanır ve bu etkinlikler önerilir.

Değerlendirme için gerçek listeden bir kullanıcı çıkartılmış ve bu kullanıcı için tahmin yapılmıştır. Bu yöntem bütün kullanıcılar için tekrarlanmıştır. Sonuç olarak en başarılı algoritmalar çevredeki popüler etkinlikler ve TF-IDF yöntemleri olmuştur. Hakkında daha önceden bilgi sahibi olmadığımız kullanıcılara öneri yaparken en başarısız yöntem de coğrafi olarak yakın etkinlikleri önerme olmuştur.

Konum tabanlı sosyal ağlarda konum tahmini için kullanıcıların ve mekânların sosyal ve coğrafi karakteristiklerini göz önüne alarak bir arkadaş tabanlı işbirlikçi filtreleme sistemi geliştiren [5] çalışmasında veriler Foursquare üzerinden elde edilmiştir. Var

olan geleneksel tavsiye sistemleri çok fazla hesaplama gücü ve zaman gerektirdiğinden konum tabanlı sosyal ağlara uygulanmak için uygun değildir. Çalışmada arkadaş tabanlı işbirlikçi filtreleme kullanılmış ve başarısı değerlendirilmiştir. Foursquare üzerinden elde edilen verilerde 58569 kullanıcı ve 96219 konum verisi bulunmaktadır. Verilerde önce kullanıcıların arkadaşlarıyla hangi oranda aynı mekânlarda bulundukları incelenmiştir. Kullanıcıların %4ünün arkadaşlarıyla %10dan daha fazla aynı mekânlarda bulundukları ortaya çıkmıştır. İşbirlikçi filtrelemenin kullandığı temel mantık da benzer kullanıcıların benzer tercihlere sahip olacağı yani benzer mekânları ziyaret edecekleri varsayımıdır. Yine veriler incelendiğinde kullanıcıların %96sının arkadaşlarıyla %10dan daha az mekânda beraber bulundukları ve %87.7sinin hiç ortak mekânda bulunmadıkları ortaya çıkmıştır. Bu birçok arkadaşın birbirine yakın ilgileri olmadığını ve tavsiye sistemi kurulurken bu arkadaşlıkların çok fazla bir önem arz etmeyeceğini göstermektedir. Bu nedenle hangi arkadaşlıkların birbiriyle daha fazla ortak noktası olduğunu tespit etmek ve bu arkadaşlıkları ön plana çıkartmak gerekmektedir. Birbirlerine konum olarak yakın olan arkadaşların daha fazla aynı mekânda bulundukları ve birbirlerinden uzaklaştıkça bu sayının azaldığı görülmüştür. Öncelikle arkadaş tabanlı işbirlikçi filtreleme kullanılmış ve filtre uygulanırken sadece arkadaş olan kişiler incelemeye alınmıştır. Daha sonra model biraz geliştirilerek birbirlerine konum olarak yakın olan arkadaşların daha fazla ortak mekânda bulundukları gözlemi de göz önüne alınarak coğrafi ölçülü arkadaş tabanlı işbirlikçi filtreleme modeli sunulmuştur. Bu modelde arkadaşlar coğrafi konumlarına göre farklı değerlerle ağırlıklandırılmış ve başarının artması beklenmiştir. Önerilen model random walk with restart(RWR) ve two state-of-art teknikleriyle karşılaştırılmıştır. Yapılan deneyler sonucunda önerilen model karşılaştırılan tekniklerle yakın sonuçlar elde etmiştir. Ancak bu yöntemin avantajı gerek duyduğu hesaplama gücünün diğer yöntemlere göre daha az olmasıdır. Bu sayede diğer yöntemlere göre çok daha hızlı sonuç vermektedir.

Konum tabanlı sosyal ağlarda kullanıcı hareketlerinin incelendiği [6] çalışmasında amaç insanların hareketlerinin daha iyi analiz edilmesi ve bu alanda daha başarılı tavsiye sistemlerinin geliştirilmesine katkı sağlanmasıdır. Veriler Foursquare'in yaklaşık 700 bin kullanıcısının 100 günlük süreyle yaptıkları 12 milyon yer bildirimini ve yaklaşık 3

milyon coğrafi etiketli ve kategorilenmiş mekân bilgilerini içermektedir. Çalışma insanların günlük ve haftalık düzenli olarak arkadaşlarıyla beraber gerçekleştirdikleri aktiviteleri ve bu aktiviteler arasındaki geçişleri açıklamaya çalışmaktadır. Foursquare yer bildirimlerinden insanların aktivitelerinin zamansal ve mekânsal olarak nasıl şekillendiği ve nasıl bir düzen içerdiği anlaşılabilmektedir. Yer bildirimleri hafta içi ve hafta sonu olarak incelendiğinde karşımıza iki farklı grafik çıkmaktadır. Hafta içi bildirimler üç noktada belirgin bir şekilde artmaktadır. Bunlar sabah işe giderken, akşam yemeği zamanı ve eve dönüş zamanlarıdır. Hafta sonları ise belirgin bir artış gözlenmemektedir. Hafta içi iş ve ofisle ilgili yer bildirimleri geniş bir yer tutarken bunlar yerlerini hafta sonları boş zaman aktivitelerine bırakmaktadırlar. Bunların yanında evde yapılan yer bildirimleri sürekliliğini korurken akşam 6'ya doğru ufak bir artış gözlenmektedir. Yer bildirimleri arası zaman incelendiğinde toplam yer bildirimlerinin %10'dan fazlasının 10 dakikalık bir arayla, %30'dan fazlasının 100 dakikalık arayla ve %20'lik bir kısmının da 2000 dakikalık bir arayla gerçekleştirildiği görülmektedir. Buradan hareketle ardışık yapılmış olan yer bildirimleri arasındaki zaman az ise bunların birbirleri ile alakalı ve önem arz eden yer bildirimleri olduğu ve aradaki zaman farkı arttıkça bu ilişkinin azaldığı söylenebilir. Yer bildirimleri arası mesafe incelendiğinde, birbirini takip eden yer bildirimlerinin %20sinin 1 km mesafe içerisinde, %60ının 1 ila 10 km'lik bir alan içerisinde ve %20lik bir kısmının da 10 km'den uzak mesafe içerisinde yapıldığı görülmüştür. Ardışık yer bildirimleri arası mesafe ve zamanın birbirleriyle aynı yönde etki gösterdikleri söylenebilir. Ardışık yer bildirimleri arası zaman fazla ise mesafe de fazladır çıkarımı yapılabilir. Kullanıcıların belirli aktivitelerden sonra bunları takiben başka bir aktiviteye geçip geçmediği incelenmiş ve Çizelge 1. 1'deki sonuçlar elde edilmiştir.

Çizelge 1. 1 Kategori Aktivite Geçişi

Kategori	Kategori	P_t^{10}	P_t^*
Tren	Tren İstasyonu	0,48	0,30
Terminal	Havaalanı	0,46	0,17
Kapı	Havaalanı	0,45	0,22
Fas	Lunapark	0,39	0,06
Tren İstasyonu	Tren	0,38	0,22
Araba Kiralama	Havaalanı	0,36	0,18
Uçak	Havaalanı	0,33	0,19
Tramvay	Havaalanı	0,33	0,19
Sinema	Alışveriş Merkezi	0,30	0,08
Köprü	Otoban	0,28	0,10

Kullanıcıların belirli mekânlardan sonra bunları takiben başka bir mekâna geçip geçmediği incelenmiş ve Çizelge 1. 2'deki elde edilmiştir.

Çizelge 1. 2 Kategori Mekân Geçişi

Kategori	Kategori	Uzaklık(km)	Yer bildirimleri
Tren İstasyonu	Tren İstasyonu	3,5	231
Tren İstasyonu	Teknoloji	0,16	187
Hafif Raylı	Alışveriş Merkezi	0,02	122
Tren İstasyonu	Anıt	2,0	101
Tren İstasyonu	Kongre Merkezi	0,57	97
Alışveriş Merkezi	Sinema	0,01	73
Anıt	Park	4,34	67

Çizelge 1. 2 Kategori Mekân Geçişi (devamı)

Kategori	Kategori	Uzaklık(km)	Yer bildirimleri
Alışveriş Merkezi	Elbise	0,50	64
Ofis	Ofis	0,60	61
Oyun parkı	Karaoke	0,50	59

Kullanıcıların hareketlerinin iyi analiz edilmesi ve bu analizin konum tahmin sistemlerinde kullanılması hiç şüphesiz sistemlerin başarısının artmasına yardımcı olacaktır.

İnsanlar ne zaman yeni bir mekânı ziyaret ederler sorusuna cevap arayan [7] çalışmasında konum tahmini yapmak için kişiselleştirilmiş random walk modeli önerilmiş ve diğer makine öğrenme algoritmalarına göre %5-18 daha başarılı sonuçlar elde edilmiştir. Çalışmada kullanılan veriler Foursquare ve Gowalla'dan elde edilmiştir. Foursquare verileri 5 aylık bir süreçte 925 bin kullanıcı tarafından 5 milyon farklı mekânda yapılmış olan 35 milyon yer bildirimini kapsamaktadır. Gowalla verileri 18 aylık bir süreçte 216 bin kullanıcı tarafından 1,5 milyon mekânda yapılmış olan 12 milyon yer bildirimini kapsamaktadır. Veriler incelendiğinde kullanıcıların daha önceden gitmedikleri mekânlara gitme eğiliminin yüksek olduğu görülmüştür. Kullanıcıların yaptıkları yer bildirimlerinin %60-80i daha önceden gitmedikleri mekânlarda yapılmaktadır. Bu kullanıcılara yeni, ziyaret edilmemiş mekânların tavsiyesinin ne kadar önemli olduğunu göstermektedir. Algoritma beş farklı bilgiyi kullanmaktadır. Bunlar;

1- Mekânın popülerliği: Mekânı kaç kişinin ziyaret ettiği bilgisi

2- Kategori tercihi: Kullanıcıların daha önceki ziyaretlerinde hangi kategoriyi ne kadar ziyaret ettikleri bilgisi

3- Arkadaşların tercihleri: Kullanıcının arkadaşlarının ziyaret ettikleri mekânların bilgisi

4- Eve yakın olma: Mekânların kullanıcıların evlerine olan uzaklık bilgisi

5- Kafa uyumu ve benzerlik: Benzer kullanıcıların benzer yerleri ziyaret edebilecekleri varsayımı

Random walk bağlı yapılı bir sistemde nodlar arasında gezme fikrine dayanmaktadır. Nodlar arasındaki geçiş olasılıkları eldeki bilgilere dayanarak hesaplanmaktadır. Bu olasılıklara göre sistem belirli bir gezinme sonrası bir denge konumuna ulaşacaktır. Veriler 30 günlük parçalara bölünmüş ve çapraz doğrulama tekniğiyle test yapılmıştır. Sonuçlar 4 başlık altında verilmiştir.

1- Metotların performansları: Metotlar incelendiğinde en yüksek başarıyı random walk'ın yakaladığı görülmektedir. Bununla beraber popularity metodu da yüksek başarı elde etmiştir.
2- Şehirlerdeki performanslar: Daha çok yer bildirimi verisi bulunan şehirlerdeki tahminlerin daha başarılı olduğu gözlenmiştir.
3- Kullanıcı aktivitesinin etkisi: Daha çok mekân ziyaret etmiş kullanıcılar daha az yeni mekân ziyaret etme eğiliminde olmaktadırlar. Çok mekân ziyaret etmiş kullanıcılar için yapılan tahminler de başarı düşmektedir.
4- Tahmin listesinin boyutunun etkisi: Liste boyutu büyüdükçe tahmin başarısı artmaktadır.

Kullanıcıların hareketlilik özelliklerini inceleyerek elde edilen veriler üzerinden tahmin yürütmeye çalışan [8] çalışmasında eğiticili öğrenme yöntemleri denenmiştir. Veriler Foursquare'den elde edilmiş ve 5 aylık bir süreyi kapsamaktadır. 35 milyon kullanıcının yaklaşık 5 milyon farklı mekânda yaptığı yer bildirimleri üzerinde çalışılırken sadece aralarında 24 saatten az zaman olan ardışık yer bildirimlerine odaklanılmıştır. Kullanıcıların konumları tahmin edilirken tahmin listesinin boyutu 50 olarak belirlenmiştir. Veriler üç ana başlık altında incelenmiştir.

1- Kullanıcı Hareketlilik Özellikleri
    a- Önceki Ziyaretler: Kullanıcının daha önceden bir mekânı kaç defa ziyaret ettiğinden hareket ederek daha sonra tekrar ziyaret edebilme ihtimali hesaplanmaya çalışılmaktadır.

b- Kategori Tercihi: Kullanıcının yer bildirimlerinin hangi kategoriye ait oldukları araştırılmaktadır.

c- Sosyal Filtreleme: Kullanıcının arkadaşlarının bir mekânda kaç defa yer bildirimi yaptıklarına bakılmaktadır.

2- Genel Hareketlilik Özellikleri

a- Popülerlik: Bir mekânda toplamda kaç defa yer bildirimi yapıldığı hesaplanmaktadır.

b- Coğrafi Uzaklık: Kullanıcının o an bulunduğu yerden mekânlara olan uzaklıklar hesaplanmaktadır.

c- Sıra Uzaklığı: Kullanıcının bulunduğu yere yakın mekânların yoğunlukları hesaplanmaktadır. Bunun sebebi bir sonraki hamlenin sadece uzaklığa değil aynı zamanda mekânların yoğun olarak bulunmasına bağlı olarak gerçekleştiği varsayımıdır.

d- Aktivite Geçişi: Belirli mekânların birbiri ardına ziyaret edildiği göz önüne alınarak hesaplama yapılmaktadır.

3- Zamana Bağlı Özellikler

Burada kullanıcıların ziyaretlerinin zamana bağlı değişimleri incelenmektedir. Bu zaman aralıkları belirli bir saat, belirli bir gün veya belirli bir haftayı kapsayacak şekilde değişmektedir.

Özellikler ayrı ayrı değerlendirildiğinde en başarılı sonuçları geçmiş ziyaretler ve mekânın saate bağlı değerlendirmesi sağlamıştır. Özellikleri tek bir algoritmada birleştirerek M5 karar ağaçları ve linear ridge regression ile de testler yapılmıştır. M5 karar ağaçları en yüksek başarıyı sağlamıştır. Özelliklerin ayrı ayrı değerlendirilmesi yerine bir araya getirilerek bir sistem oluşturulmasının başarıyı arttırdığı görülmüştür.

Konum tabanlı sosyal ağlarda kullanıcıların bir mekânda yaptıkları ilk yer bildirimini tahmin etmeye çalışan [9] çalışmasında yeni bir tahmin algoritması önerilmiştir. Veriler Brightkite ve Gowalla'dan elde edilmiştir. Brightkite 50 binin üzerinde kullanıcının 800

bin civarı mekânda yaptığı 2,5 milyon yer bildirimi verisine, Gowalla ise 200 bin kullanıcının 1 milyonun üzerinde mekânda yaptığı 6 milyon yer bildirim verisine sahiptir. Kullanıcıların yaptıkları ziyaretlerin genellikle önceki ziyaretlerine yakın yerler olduğu varsayımında bulunulmuş ve bu incelenmiştir. Yapılan inceleme sonucunda yeni ziyaretlerin Brightkite'ta %67 Gowalla'da %81'lik bir kısmının önceki ziyaretlere 10 km'lik mesafede oldukları görülmüştür. Yine arkadaşlık etkisi incelenmiş ve kullanıcıların yaptıkları yeni ziyaretlerin Brightkite'ta %31'inin Gowalla'da ise %23'ünün daha önceden arkadaşlarının veya arkadaşının arkadaşlarının ziyaret ettikleri yerler olduğu görülmüştür. Çalışmada 5 adet algoritma test edilmiş ve 2 adet yeni algoritma önerilmiş ve sonuçlar karşılaştırılmıştır.

1- Kullanıcı Bazlı İşbirlikçi Filtreleme: Benzer kullanıcılar mekânlar üzerinde benzer tercihlere sahip olurlar fikrine dayanmaktadır. Kullanıcıların ziyaret ettikleri mekânlar ve ziyaret etme sıklıkları üzerinden profiller oluşturulmakta ve filtreleme yapılmaktadır.

2- Konum Bazlı İşbirlikçi Filtreleme: Kullanıcılar benzer mekânları ziyaret eder varsayımını kullanmaktadır. Konum profilleri oluşturulur ve karşılaştırılır.

3- Yakın Konum Komşuluğu: Kullanıcının daha önceden ziyaret ettiği yerler listelenir ve bu listedeki konumlara yakın mekânlar puanlanır.

4- Kullanıcı Bazlı İşbirlikçi Filtreleme: Kullanıcının arkadaşlarının ziyaret ettikleri yerlere göre puan hesaplanır.

5- Random Walk With Restart: Sadece kullanıcının arkadaşları değil arkadaşlarının arkadaşları da kişinin tercihleri üzerinde etkilidir. Bu önermeyi incelemek için RWR kullanılmaktadır.

6- Arkadaşlık Bazlı Sayfa Boyama Algoritması: Kullanıcıların arkadaşlarının ziyaret ettikleri mekânlar değerlendirilerek tahmin yapılmaktadır.

7- Konum-Arkadaşlık Bazlı Sayfa Boyama Algoritması: Kullanıcıların sadece arkadaşlarının değil, daha önceden aynı mekânda bulunmuş ve arkadaşı olmayan kullanıcılarında önceki ziyaretleri göz önüne alınarak bir tahmin yürütülmektedir.

Bütün algoritmalar test edilmiş ve en başarılı algoritma Konum-Arkadaşlık bazlı sayfa boyama algoritması olmuştur. Bunun sebebi sadece eski tercihlere bakmak yerine, mekân konumlarını ve sosyal etkileşimleri de hesaba katmasıdır.

Konum tahmini için Hidden Markov Model kullanan [10] çalışmasında önce kullanıcının hangi kategoriyi tercih edeceği tahmin edilmiş ardından hangi mekânı tercih edeceği tahmin edilmeye çalışılmıştır. Bu yöntem sayesinde tahmin uzayı 5,5 kat azalırken tahmin başarısı ise %13 artmıştır. Çalışmada Gowalla üzerinden 230 bin kullanıcıya ait 13 milyon yer bildirimi verisi alınmıştır. Kullanıcıların bir yıllık zamanda ortalama yer bildirimi sayısı 58 olarak hesaplanmıştır ve bu değerin üstünde yer bildirimi yapan kullanıcılar aktif kullanıcı olarak varsayılmıştır ve gerisi veri tabanından çıkarılmıştır. Mekânlar 9 farklı kategori altında incelenmiştir. Toplanan veri setinde 817 bin farklı konum bulunmaktadır. Mekânlar sıralanırken 4 farklı ölçüt kullanılmıştır.

1- Yer bildirimi sayısı: Mekânda yapılan toplam yer bildirimi sayısı
2- Kullanıcı sayısı: Mekânı ziyaret eden toplam kullanıcı sayısı
3- Kullanıcı sayısı X yer bildirimi sayısı
4- Bir kullanıcı tarafından mekânda yapılan en çok yer bildirimi sayısı

Çalışmadaki amaç kullanıcıların bir sonraki adımda hangi konuma gideceklerini tahmin etmektir. Kullanıcıların bir sonraki adımlarını doğrudan tahmin etmek oldukça zor bir işlemdir. Bu yüzden önce kategori tahmininde bulunulmuş ve tahmin uzayı azaltılmış, ardından konum tahmini yapılmıştır. Kullanıcının o anda bulunduğu konum bilgisi kullanılarak belirli bir mesafe uzaklığındaki bütün mekânlar listelenir. Testler yapılırken bu mesafe 400 metre kabul edilmiştir. HMM'den elde edilen kategori bilgisini kullanarak önce kategori tahmin edilir. Ardından sıralanmış listede bu kategoriye denk gelen ilk mekân tavsiye edilir. Eğer bu kategoriden bir mekân listede yok ise bir sonraki kategoriye geçilir. Kategori tahmini yapıldığında tahmin başarısının %13 oranında arttığı gözlenmiştir. Kullanıcıların hareketlerinin ve yer bildirimlerinin kategori seviyesinde modellendiğinde sonuçların daha başarılı olduğu söylenebilir. Kategori tahmini yapılmadığında tahmin başarısı %27 seviyesinde kalmıştır. Kategori tahmini

yapıldığında ortalama 9 mekân arasından tahmin yapılmaya, kategori tahmini yapılmadığında ise ortalama 52 mekân arasından tahmin yapılmaya çalışılmıştır. Kategori tahmini yapmak tahmin uzayını 5 kat azaltmıştır.

1.2 Tezin Amacı

Konum tahmini, konum tabanlı sosyal ağlarda önemli bir yere sahiptir. Konum tahmininin başarılı bir şekilde yapılması hem servis sağlayıcılara, hem kullanıcılara, hem de mekân sahiplerine yarar sağlamaktadır. Bu çalışmada, konum tabanlı sosyal ağlardan elde edilmiş yer bildirimi, mekân ve arkadaşlık verileri kullanılarak konum tahmini problemine bir çözüm önerisi yapılmıştır.

1.3 Hipotez

Konum tabanlı sosyal ağlardan elde edilen veriler oldukça ham durumdadır. Bu verilerin incelenmesi ve işlenmesi sonucu konum tahmini için kullanılabilecek birçok özellik elde edilebilmektedir. Bu özellikler ayrı ayrı tahmin için kullanılabileceği gibi beraber kullanılarak daha başarılı sonuçlar da elde edilebilir.

BÖLÜM 2

VERİ ANALİZİ

Bu bölümde, tez çalışması boyunca kullanılmış olan veri setinin elde edilme süreci, verinin incelenmesi, ön işleme aşaması ve kapsamlı bir analizi sunulmuştur.

2.1 Veri Toplanması

Veriler iki farklı konum tabanlı sosyal ağ olan Gowalla ve Foursquare'den elde edilmiştir. Her iki veri seti de daha önce bu alanda çalışma yapmış bir ekipten hazır olarak alınmıştır. Veriler üzerindeki çalışmayı sadece en popüler şehirlerle kısıtlayarak sistemin en yoğun kullanıldığı ve en çok kullanıcı ve mekânın bulunduğu yerlere odaklanılmıştır.

Foursquare: 2009 yılında kurulduktan sonra kısa sürede çok popüler oldu. Şu an konum tabanlı sosyal ağlar arasındaki en popüler sosyal ağ konumundadır. 45 milyondan fazla kullanıcıya ve 65 milyondan fazla mekânı barındıran sistemde bu güne kadar 8 milyardan fazla yer bildirimi yapılmış durumdadır.

Foursquare'e ait veri seti kullanıcılarının Twitter üzerinden yaptıkları yer bildirimi paylaşımları taranarak elde edilmiştir. Dolayısıyla Foursquare'deki bütün yer bildirimlerini kapsamamaktadır. Tahmini bir hesapla verilerin elde edildiği dönemde bütün sistemin yaklaşık %20-25'lik bir kısmını kapsadığı düşünülmektedir.

Foursquare verileri 4 farklı şehri kapsamaktadır. Bunlar Londra, Austin, Dallas ve San Francisco'dur. 2011 yılında 9 aylık bir dönemde 117 bin farklı kullanıcı tarafından 100 bin farklı mekânda yapılmış olan 1,4 milyon yer bildirimi verisi bulunmaktadır. Yer

bildirimi verisi şu bilgileri içermektedir; kullanıcıya ait kimlik numarası, zaman, mekâna ait kimlik numarası. Mekânlara ait veriler şu bilgileri içermektedir; mekâna ait kimlik numarası, mekânın adı, mekânın bulunduğu enlem ve boylam değerleri, kategori adı.

Gowalla: 2009 yılında kullanıma açıldıktan sonra kısa sürede oldukça büyük bir kitleye ulaşmıştır. 2011 yılında Facebook tarafından satın alındıktan sonra kullanımına son verilmiştir. Veriler genel kullanıma açık API aracılığıyla elde edilmiştir. Gowalla veri seti 2009 yılı ocak ayı ile 2010 yılı ağustos ayları arasında yaklaşık 19 aylık bir dönemi kapsamaktadır. Bu dönemde yaklaşık 100 bin kullanıcı 1,5 milyon farklı mekânda 10 milyonun üzerinde yer bildiriminde bulunmuştur. Mekânlar 283 farklı kategori altında toplanmıştır. Yer bildirimi verisi şu bilgileri içermektedir; kullanıcıya ait kimlik numarası, zaman, mekâna ait kimlik numarası. Mekânlara ait veriler şu bilgileri içermektedir; mekâna ait kimlik numarası, mekânın adı, mekânın bulunduğu enlem ve boylam değerleri, kategoriye ait kimlik numarası. Bunların haricinde kullanıcıların arkadaşlık ilişkilerini gösteren veri kümesi de bulunmaktadır.

2.2 Veri Ön İşleme

İlk etapta elde edilen veriler oldukça ham verilerdir. Ancak bununla beraber bu verileri kullanarak birçok farklı analiz yapmak ve yeni bilgiler elde etmekte mümkündür. Bu amaçla veriler üzerinde bir takım ön işleme çalışmaları yapılmıştır.

Yer bildirimi verileri kullanılarak her bir mekânın, kategorinin ve kullanıcının toplam yer bildirimi sayıları elde edilmiştir. Verilerin ait olduğu dönemde sistemi az kullanan kullanıcılar pasif kullanıcı kabul edilmiş ve bu kullanıcılar sistemden çıkarılmıştır. Bir kullanıcı eğer 9'dan az yer bildiriminde bulunduysa pasif kullanıcı olarak değerlendirilmiştir. Benzer şekilde mekânlar için de bir eleme uygulanmıştır. Eğer bir mekânda hiç yer bildirimi yoksa ya da sadece bir tane var ise bu mekânda veri setinden çıkarılmıştır. Burada amaç kullanıcılar hakkında elde edilen bilgileri kullanarak kişiselleştirilmiş bir öneri sistemi hazırlamaktır ve hakkında az bilgi olan kullanıcılar için sistem başarılı şekilde çalışmayacaktır.

Enlem ve boylam değerlerine sahip olduğumuz mekânların hangi ülke ve şehirlerde olduğunu tespit etmek için ek bir veri tabanı kullanılmış ve bu veri tabanındaki verilerle

karşılaştırma sonucu bütün mekânların hangi ülke ve şehirlerde oldukları tespit edilmiştir. Elde edilen sonuçlara göre verilerin büyük çoğunluğu birkaç ülkede toplandığı, benzer şekilde yer bildirimlerinin büyük bir kısmı az sayıda şehre dağıldığı gözlenmiştir. Sistem performansını test ederken veriler şehir şehir ayrılarak değerlendirilmiştir. Burada kullanıcıların genelde aynı şehir içerisinde yer bildiriminde bulundukları varsayımı yapılmıştır. Bu sayede hem daha küçük veri setleriyle daha hızlı sonuçlar almak mümkün olacak hem de öneri yapılırken sadece bir şehir içerisinde tahmin yürütülecek ve başarının artması beklenecektir.

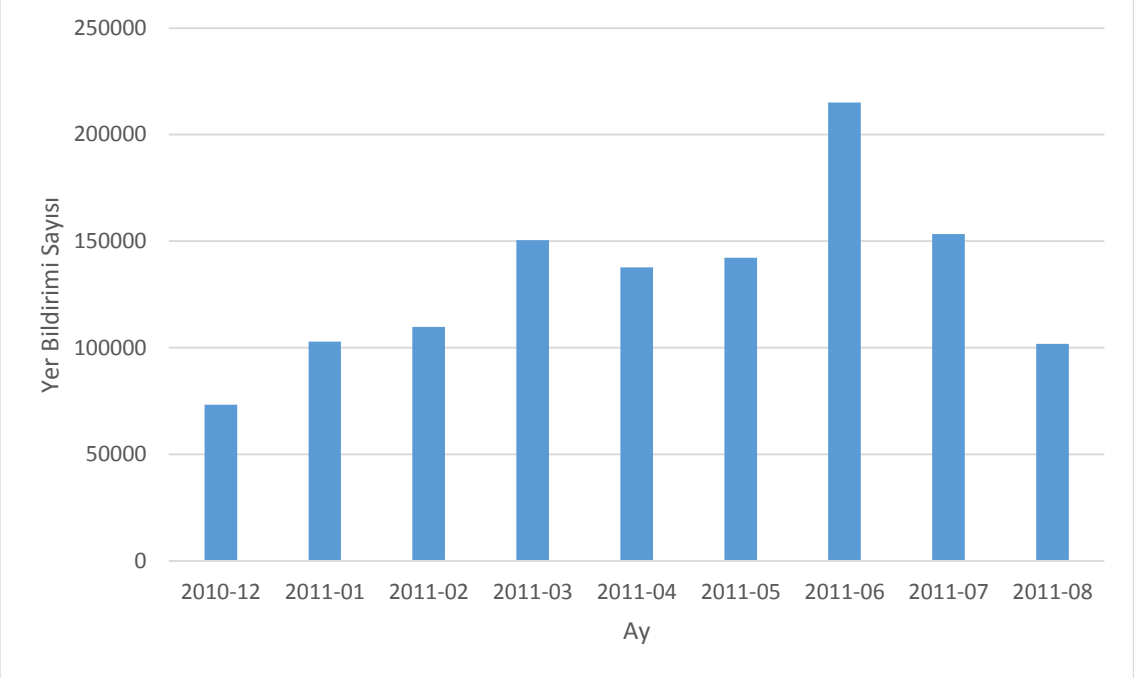
Son olarak ileride kullanılmak üzere mekânların hangi zaman aralıklarında daha popüler oldukları bilgisi de önceden hesaplanmış ve veri tabanına eklenmiştir. Gün 4 eşit zaman dilimine bölünmüştür bunlar; 00:00-06:00, 06:00-12:00, 12:00-18:00 ve 18:00-24:00'tür. Daha sonra her bir mekân için bu zaman aralıklarında ne kadar yer bildirimi yapıldığı bilgisi kullanılarak hesaplama yapılmış ve sonuç normalize edilerek her bir mekân için 4 ayrı değer veri tabanına eklenmiştir.

Son durumda mekân veri tablosu için mekâna ait kimlik numarası, kategori numarası, mekânın adı, toplam yer bildirimi sayısı, enlem ve boylam değerleri, şehir ve ülke bilgisi ve zaman aralıkları için normalize edilmiş yoğunluk değerleri bulunmaktadır.

2.3 Veri Analizi

Bu bölümde veri setleri üzerinde kapsamlı analiz yapılmış ve elde edilen sonuçlar paylaşılmıştır.

2.3.1 Foursquare Analizi



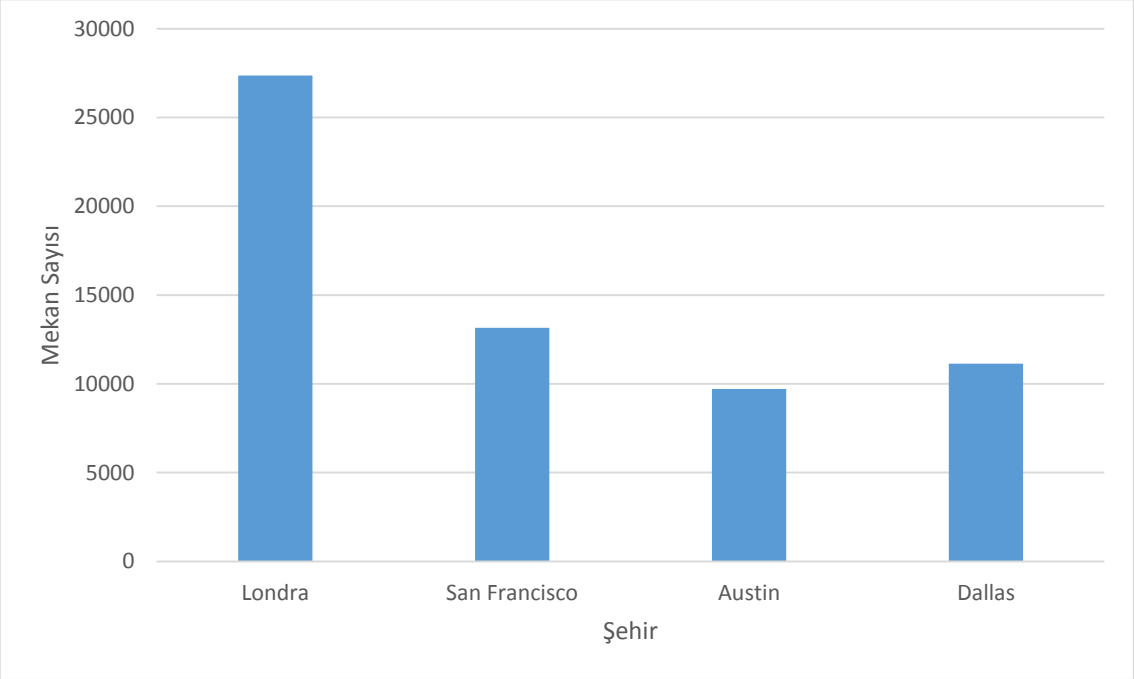
Şekil 2. 1 Foursquare aylara göre yer bildirimi sayısı

9 aylık dönem içerisinde en çok yer bildiriminin yapıldığı ay haziran olmuştur. Kış aylarında yer bildirimlerinin diğer aylara göre daha az olduğu söylenebilir. Ayrıca yer bildirimi sayısının sürekli olarak artış göstermediği de görülmektedir.



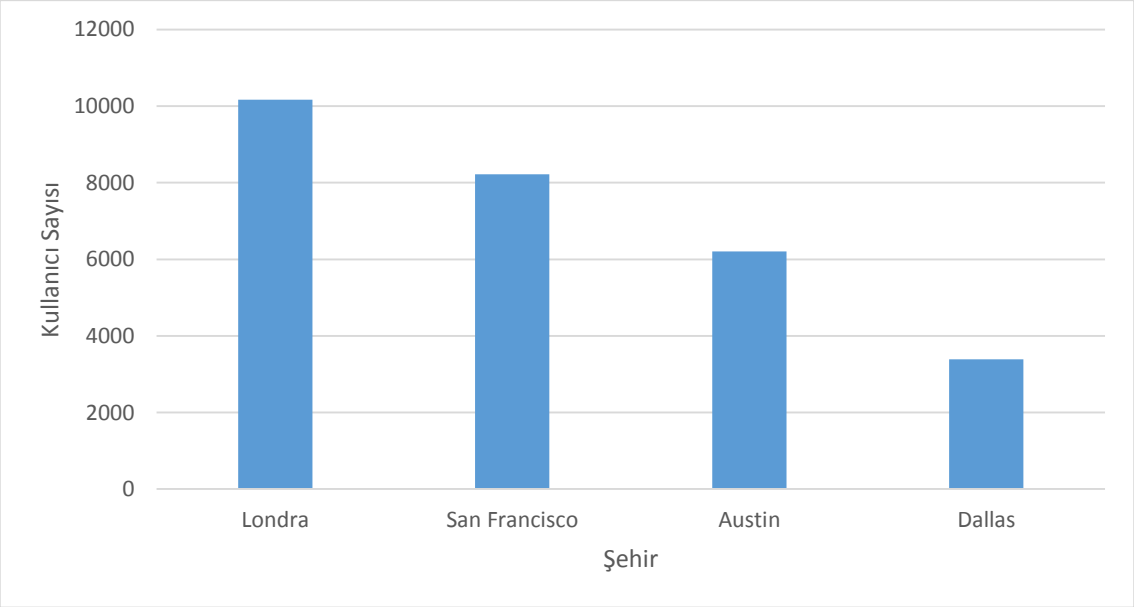
Şekil 2. 2 Foursquare şehirlere göre yer bildirimi sayısı

Aynı dönemde en çok yer bildirimi Londra şehrinde yapılmıştır. En az yer bildirimi de Dallas şehrinde görülmektedir.



Şekil 2. 3 Foursquare şehirlere göre mekan sayısı

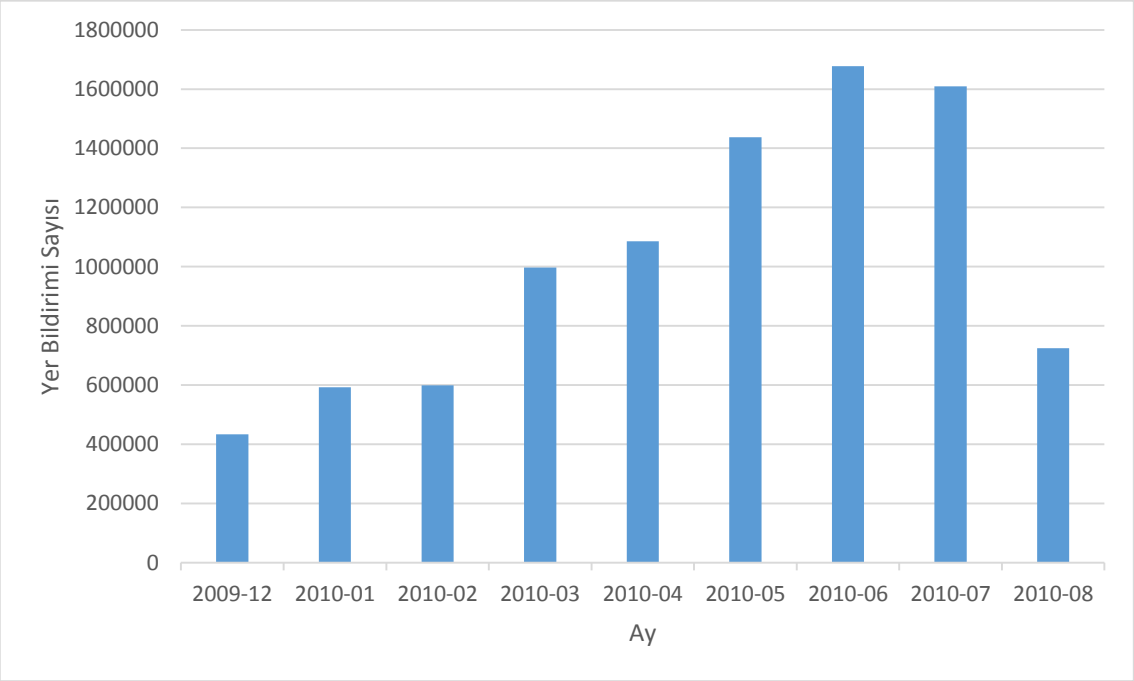
Yer bildirimi sayısı ile mekan sayısının doğrudan orantılı olmadığı grafikten görülmektedir. Dallas şehri en az yer bildirimine sahip iken mekan sayısı olarak Austin şehrini geçmiştir. Londra şehri en çok mekana sahiptir.



Şekil 2. 4 Foursquare şehirlere göre kullanıcı sayısı

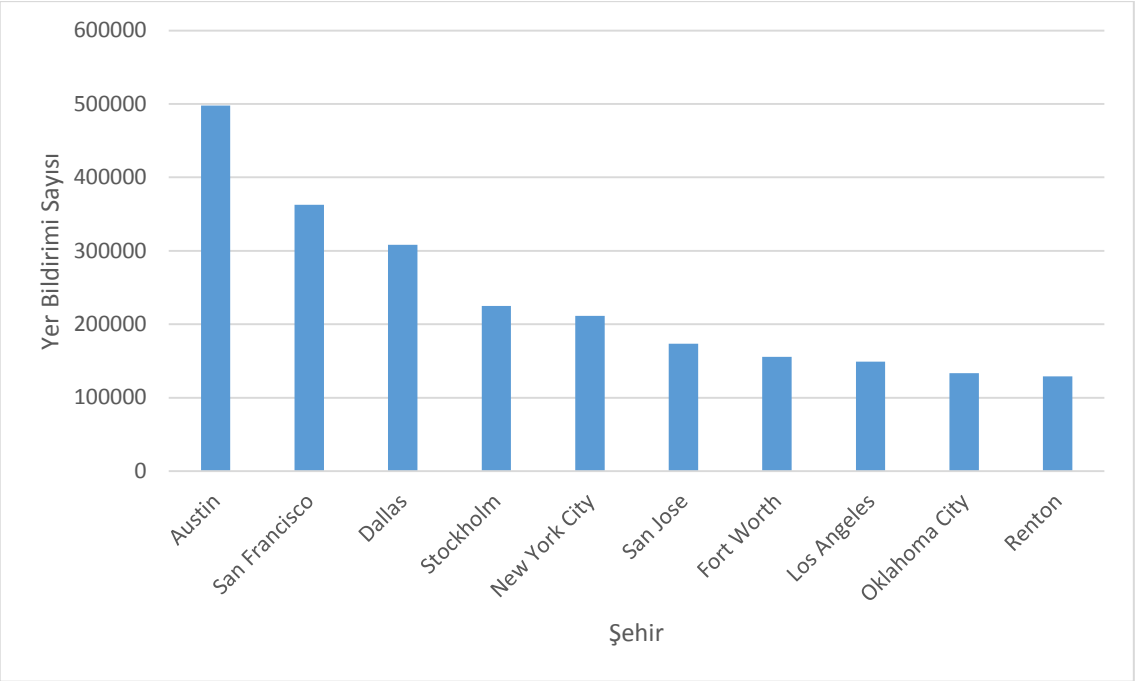
Kullanıcı sayısı ile yer bildirimi sayısı arasında bir orantı olduğu söylenebilir. En çok yer bildiriminin bulunduğu Londra'da kişi sayısı en çok iken en az yer bildiriminin bulunduğu Dallas şehrinde bu sayı en azdır.

2.3.2 Gowalla Analizi



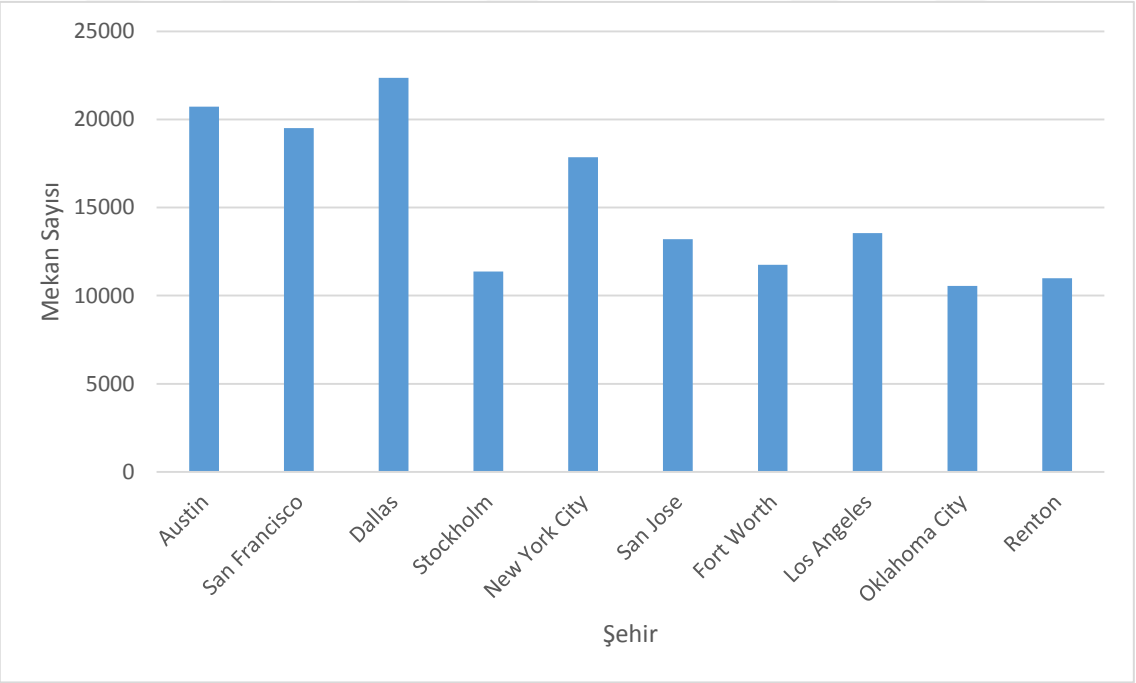
Şekil 2. 5 Gowalla aylara göre yer bildirimi sayısı

9 aylık zaman içerisinde en çok yer bildirimi haziran ayında gerçekleşmiştir. Yaz aylarında diğer aylara göre daha fazla yer bildirimi yapılmaktadır. Ağustos ayı göz önüne alınmazsa yıl içerisinde yer bildirimi sayısında sürekli bir artış olduğu söylenebilir.



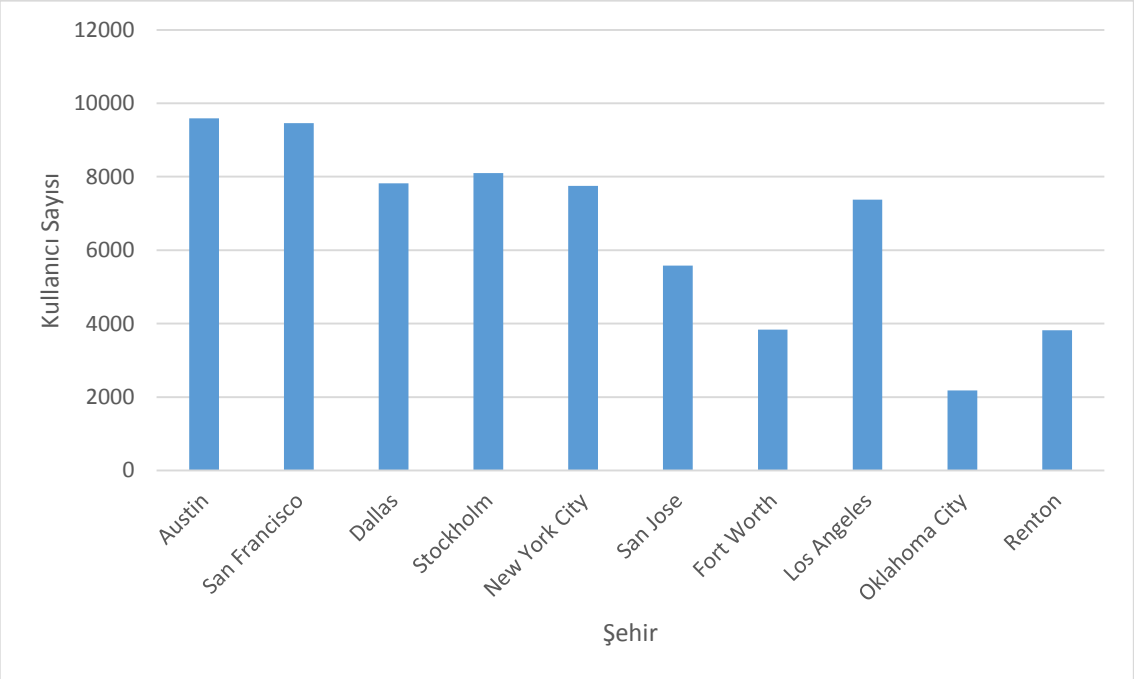
Şekil 2. 6 Gowalla şehirlere göre yer bildirimi sayısı

Şekil 2. 6'da Gowalla veri tabanında bulunan yer bildirimi açısından en popüler 10 şehir verilmiştir. Stockholm şehri hariç bütün şehirler Amerika Birleşik Devletleri'nde bulunmaktadır. En popüler üç şehir Austin, San Francisco ve Dallas'tır.



Şekil 2. 7 Gowalla şehirlere göre mekân sayısı

Şekil 2. 7'e bakıldığında şehirlerde yer bildirimi sayısı ile mekân sayısı arasında doğrudan bir orantı olmadığı söylenebilir. En popüler şehir Austin iken en çok mekân Dallas şehrinde bulunmaktadır. Yine Stockholm şehri en popüler dördüncü şehir iken mekân sayısı açısından sekizinci sıradadır.



Şekil 2. 8 Gowalla şehirlere göre kullanıcı sayısı

Şekil 2. 8'e bakıldığında şehirlerde yer bildirimi sayısı ile kullanıcı sayısı arasında doğrudan bir orantı olmadığı söylenebilir. Bununla beraber kullanıcı sayısı açısından en yoğun 5 şehrin aynı zamanda en popüler 5 şehir olduğu da görülmektedir.

BÖLÜM 3

KONUM TAHMİNİ PROBLEMİ

Bu bölümde konum tabanlı sosyal ağlarda konum tahmini problemi ayrıntılı olarak incelenmiş ve tahmin yöntemleri açıklanmıştır.

3.1 Problem Tanımı

Konum-tabanlı sosyal ağlar kullanıcılarının konumlarını paylaşmalarına olanak sağlamakta ve bu konum bilgisine zaman bilgisini de ekleyerek bu bilgileri saklamaktadırlar. Bu saklanan bilgiye yer bildirimi denmektedir. Biz de çalışmamızda bu yer bildirimi verilerini kullanarak kullanıcıların bir sonraki yer bildirimlerini nerede yapacaklarını tahmin etmeye çalışmaktayız. Yer bildirimi verilerine ek olarak mekân ve arkadaşlık verilerine de sahibiz.

Çalıştığımız veri tabanlarında çok fazla mekân bulunduğundan dolayı tek tahminde doğru mekânı bulabilme olasılığımız oldukça düşüktür. Bu problemi aşabilmek için tek tahmin yapmak yerine bir tahmin listesi hazırlayıp, listeden herhangi bir mekânın gerçekten ziyaret edilecek mekân olup olmadığına bakılmaktadır. Çalışmamız sırasında farklı uzunluklarda listeler hazırlanarak tahminler yapılacak ve başarı ölçülecektir.

3.2 Problemin Formüle Edilmesi

Veri setimizde kullanıcılardan oluşan bir K seti ve mekânlardan oluşan M seti bulunmaktadır. Her bir yer bildirimi y, {k,m} şeklinde bir ikili ile temsil edilmektedir. k yer bildirimini kimin yaptığını m ise yer bildiriminin nerede yapıldığını göstermektedir.

Y, yer bildirimi kümesini temsil ederken Y_k ise k kullanıcısının yapmış olduğu yer bildirimi kümesini temsil etmektedir. Ka_k k kullanıcısının ziyaret ettiği kategorileri göstermektedir. k kullanıcısının arkadaşlarının kümesi A_k ile tanımlanmıştır. Konumu tahmin edilmeye çalışılan bir k kullanıcısı için sistemdeki her mekân için bir skor hesaplanmakta ve mekânlar bu skorlara göre sıralanmaktadır. Her bir mekâna verilen skor $Skor_m$ veya $Skor_{k,m}$ ile gösterilmiştir. Daha sonra istenilen liste uzunluğuna göre en yüksek puana sahip mekânlar seçilmektedir.

3.3 Tahmin Yöntemleri

Tahmin yöntemleri olarak öncelikle veri ön işleme ve veri analizi aşamalarında konum tahmini için kullanılabileceğine karar verdiğimiz özellikleri seçtik. Daha sonra bu tekil özellikleri bir arada kullanarak yeni bir yöntem önererek sistemin tahmin başarısını arttırmayı amaçladık. Bu bölümde tekil özellikleri ve önerilen yeni yöntemi ayrıntılı olarak açıklanacaktır.

3.3.1 Tekil Özellikler İle Tahmin

- Mekân Popülerliği: Tahmin yöntemi olarak kullanılan ilk özellik mekân popülerliğidir. Mekân popülerliği bir mekânda yapılan toplam yer bildirimi sayısı olarak tanımlanmıştır. Bu özellikte kullanıcılar popüler olan mekânlara gitme eğilimdedirler varsayımında bulunulmuştur.

$$Skor_m = \sum \{k, m\} \mid \forall k \in K \tag{3.1}$$

- Eve Olan Uzaklık: Veri tabanımızda kullanıcıların evlerine dair bir bilgi bulunmamaktadır. Bu sebeple bir kullanıcının evi en çok ziyaret ettiği mekândır varsayımında bulunulmuştur. Daha sonra her bir mekâna, kullanıcının evine olan uzaklığına göre ters orantı kurarak bir puan verilmiştir. Bu özellikte kullanıcılar evlerine yakın olan mekânlara gitme eğilimde olurlar varsayımında bulunulmuştur.

$$Skor_{k,m} = \frac{1}{\mathrm{dist}\{k_{ev}, \mathrm{m}\}} \tag{3.2}$$

- Yakın Popüler Mekânlar: Bu özellik için kullanıcılar evlerine yakın ve popüler olan mekânları tercih ederler varsayımında bulunulmuştur. Kullanıcıların evlerine en yakın 1000 mekâna popülerliklerine göre skor verilmiştir.

$$Skor_{k,m} = \left(\sum \{k,m\} \mid \forall k \in K\right) \wedge m \in yakın(k_{ev}, 1000) \quad (3.3)$$

- Önceki Ziyaretler: Mekânlara kullanıcının o mekâna yapmış olduğu ziyaret sayısı skor olarak atanmıştır. Buradaki varsayım kullanıcılar daha önceden ziyaret ettikleri mekânlara gitme eğilimde olurlardır.

$$Skor_{k,m} = \sum \{k,m\} \in Y_k \quad (3.4)$$

- Arkadaşlık: Kullanıcının arkadaşlarının bir mekânda yapmış oldukları toplam yer bildirimi sayısı o mekâna skor olarak atanmıştır. Bu özellik için kullanıcılar arkadaşlarının bulundukları mekâna gitme eğilimindedirler varsayımı yapılmıştır.

$$Skor_{k,m} = \sum \{f,m\} \mid \forall f \in A_k \quad (3.5)$$

- Kategori Tercihi: Kullanıcıların kategori tercihleri tahmin için kullanılmıştır. Kullanıcıların önceki ziyaretlerinde hangi oranda hangi kategoriyi tercih ettikleri tespit edilmiş ve bu oranlarda tercih ettikleri kategorilerden en popüler mekânlar seçilmiştir. Bu özellikte kullanıcılar mekânları seçerken belirli bir kategori tercihinde bulunmaktadırlar varsayımı yapılmıştır. Mekânlara skor verilirken mekânların popülerlikleri göz önüne alınmıştır. $r_{k,ka}$ kullanıcının ka kategorisini tercih etme oranı olarak gösterilmiştir.

$$Skor_{k,m} = \sum_{i=1}^{n} \{k,m\} \mid \forall k \in K \wedge m \in List_i \wedge List_i \subset ka_i$$

$$size(List_i) \propto r_{k,ka_i} \quad (3.6)$$

- Mekân Zaman Aralığı Yoğunluğu: Bu özellikte gün 4 eşit parçaya bölünmüş ve mekânların bu zaman aralıklarındaki yoğunlukları hesaplanmıştır. Bu özellik kendi başına tahmin yöntemi olarak kullanılmamış, daha sonra diğer yöntemlerle birleştirilerek kullanılmıştır. Bu özellik için mekânların yoğunlukları gün içerisinde değişmektedir varsayımı yapılmıştır.

3.3.2 Önerilen Yöntem İle Tahmin

Tekil özelliklerle testler yapıldıktan sonra ilk sonuçlar elde edilmiştir. Bu ilk sonuçlar ışığında hangi özelliklerin ne şekilde kullanılması gerektiği hakkında bir fikir sahibi olunmuştur. Mekân popülerliği ve mekân zaman aralığı yoğunluğunun beraber kullanılmasıyla, sadece mekân popülerliğini kullanmaya kıyasla daha başarılı sonuçlar alındığı gözlenmiştir. Bu iki özelliğe ek olarak kategori tercihini de kullanmanın başarıyı daha da arttırdığı görülmüştür. Bunlara ek olarak tekil özellikler içerisinde en başarılısı önceki ziyaretler olmuştur.

Aday mekân kümesini azaltmak ve gerçek konum-tabanlı sosyal ağları daha iyi taklit edebilmek için yer bildiriminin yapıldığı konumun bir başlangıç noktası olarak kullanılmasına karar verilmiştir. Daha sonra bu başlangıç noktası etrafındaki en yakın 1000 mekân tespit edilmiştir. Bu sayede başlangıç noktası etrafında yaklaşık olarak 3-5 kilometrelik bir dairesel alan kapsanmıştır. Bu noktadaki varsayımımız kullanıcının mekân önerisi için telefonunu kullandığı anda bu 3-5 kilometrelik alan dışarısında bir mekânı tercih etmeyeceği yönündedir. Böyle bir daraltma işlemini gerçek konum-tabanlı sosyal ağlarda kullanmaktadırlar. Mekânlara skor vereceğimiz aşamada sadece bu alan içerisinde kalan mekânlara skor atanacaktır.

Son olarak tahmin listesini iki eşit parçaya ayırıp her iki liste içinde farklı yöntemler ile aday mekânlar seçilmesine karar verilmiştir. İlk parça için aday mekânlar önceki ziyaretler özelliği kullanılarak seçilmiştir. İkinci parça için aday mekânlar kategori tercihi, mekân popülerliği ve mekân zaman aralığı yoğunluğu kullanılarak belirlenmiştir. Mekân zaman aralığı yoğunluğu kullanılırken yer bildiriminin yapıldığı zaman aralığı dikkate alınmıştır.

ListeA ve ListeB olmak üzere tahmin listesinin yarısı uzunluğunda iki listemiz olduğu varsayılmaktadır. m_t yer bildiriminin yapıldığı zaman aralığında mekanın popülerliğini temsil etmektedir. m_c yer bildiriminin yapıldığı mekanı göstermektedir. Yöntem şu şekilde formüle edilebilir;

$$Skor_{k,m1} = \sum_{i=1}^{n} \{k, m_z\} \mid \forall k \in K \ \wedge m \in List_{A_i} \wedge \ List_{A_i} \subset ka_i \ \wedge m \in close(m_c, 1000)$$

$$Skor_{k,m2} = \sum \{k, m\} \mid \exists k \in K \ \wedge \ m \in close(m_c, 1000)$$

$$Liste_A = rank(norm(Skor_{k,m1}))$$
$$Liste_B = rank\left(norm(Skor_{k,m2})\right)$$
$$Liste = Liste_A + Liste_B$$
$$rank(Liste) \quad (3.7)$$

3.4 Test Yöntemi ve Ölçütler

Sistem test edilirken birini dışarıda bırak çapraz doğrulama yöntemi kullanılmıştır. Veri setinden tahmin edilmeye çalışılan yer bildirimi çıkarılmış ve geri kalanı ile sistem eğitilmiştir. Daha sonra yer bildirimi tahmin edilmeye çalışılmış ve sisteme geri eklenmiştir. Bu şekilde bütün veri seti test edilmiştir. Testler yapılırken her bir şehir için sistem performansı ayrı ayrı test edilmiştir. Burada kullanıcılar büyük oranda aynı şehirde yer bildiriminde bulunmaktadırlar varsayımı kullanılmıştır. Tahmin listeleri 10'dan 100'e kadar 10 farklı değer almışlardır.

Sonuçlar değerlendirilirken üç farklı ölçüt kullanılmıştır. İsabet, tahmin listesinden herhangi bir mekan, tahmin edilmeye mekan ise 1 diğer durumlarda 0 değerini almaktadır. Tahmin listesini T, tahmin edilmeye çalışılan mekanı m kabul edecek olursak, şu şekilde formüle edilebilir;

$$\text{İ}sabet = \begin{cases} 1, & m \in T \\ 0, & m \notin T \end{cases} \quad (3.8)$$

Kesinlik yapılan tahminin doğruluğunu göstermektedir. Eğer yapılan tahmin doğruysa kaçıncı sıradaki tahminin doğru olduğuna göre bir puan almaktadır. Eğer ilk sıradaki tahmin doğruysa 1 olmakta ve başarılı tahmin listede aşağılara indikçe bu değer 0'a yaklaşmaktadır. Şu şekilde formüle edilebilir;

$$Kesinlik = \frac{|N| - k + 1}{|N|} \quad (3.9)$$

Yüzdelik başarı ilk iki ölçütün birleşimidir. Yapılan tahmin başarılı değil ise 0, başarılı ise kesinlik değerini almaktadır. Şu şekilde formüle edilebilir;

$$Yüzdelik\ Başarı = \begin{cases} \frac{|N| - k + 1}{|N|}, & m \in T \\ 0, & m \notin T \end{cases} \quad (3.10)$$

3.5 Test Ortamı

Testler Windows 8.1 işletim sisteminde gerçekleştirilmiştir. Veri tabanı olarak PostgreSQL kullanılmıştır. Programın kodları Java platformunda yazılmıştır.

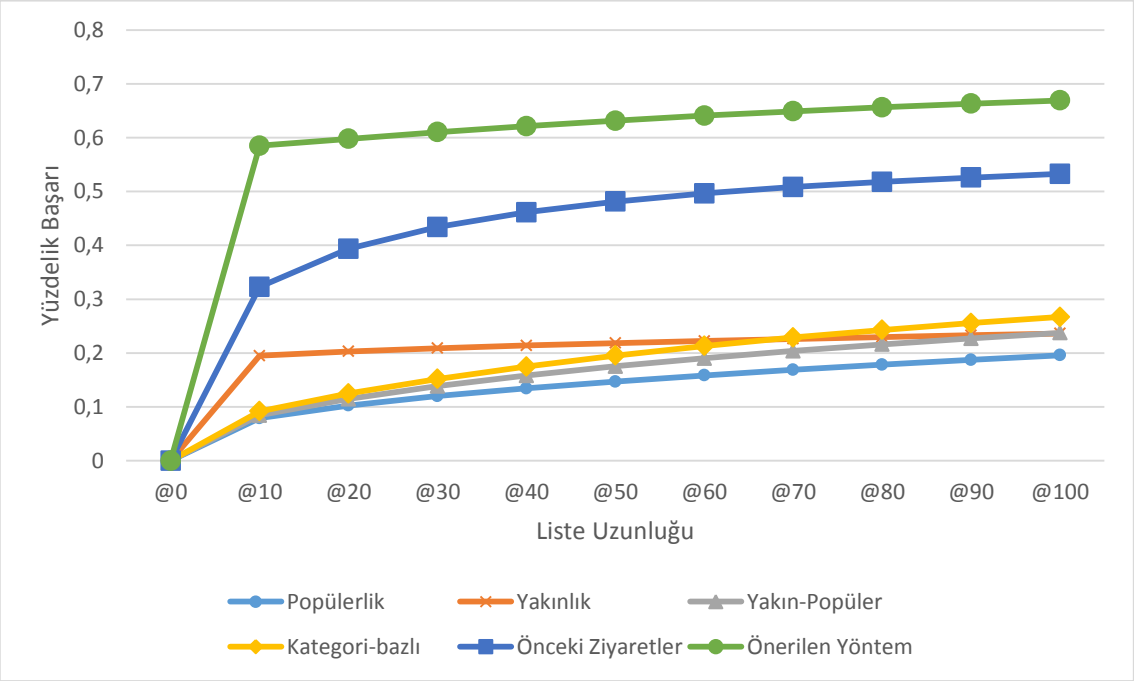
BÖLÜM 4

DENEYSEL SONUÇLAR

Bu bölümde sistem Foursquare ve Gowalla için farklı liste uzunluklarında test edilmiş ve elde edilen sonuçlar paylaşılmıştır. Sonuçlar yüzdelik başarı ölçütünde gösterilmiştir.

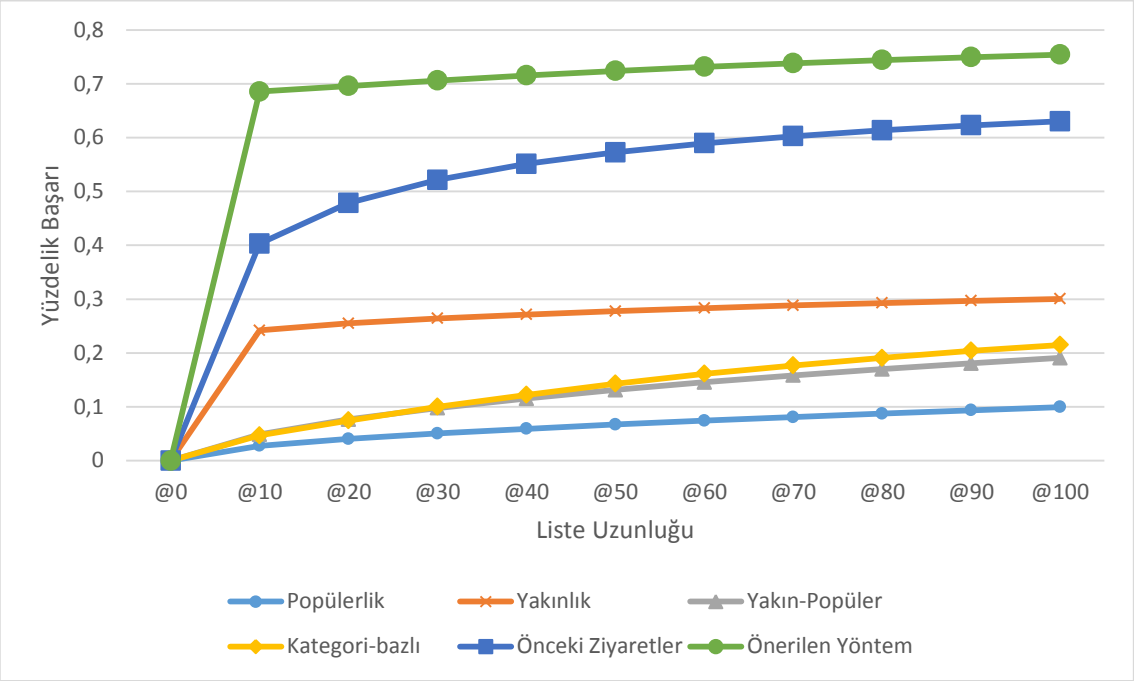
4.1 Foursquare Sonuçları

Sistem her bir tekil özellik ve önerilen yöntem için ayrı ayrı test edilmiştir. Testler şehir bazlı olarak yapılmıştır.



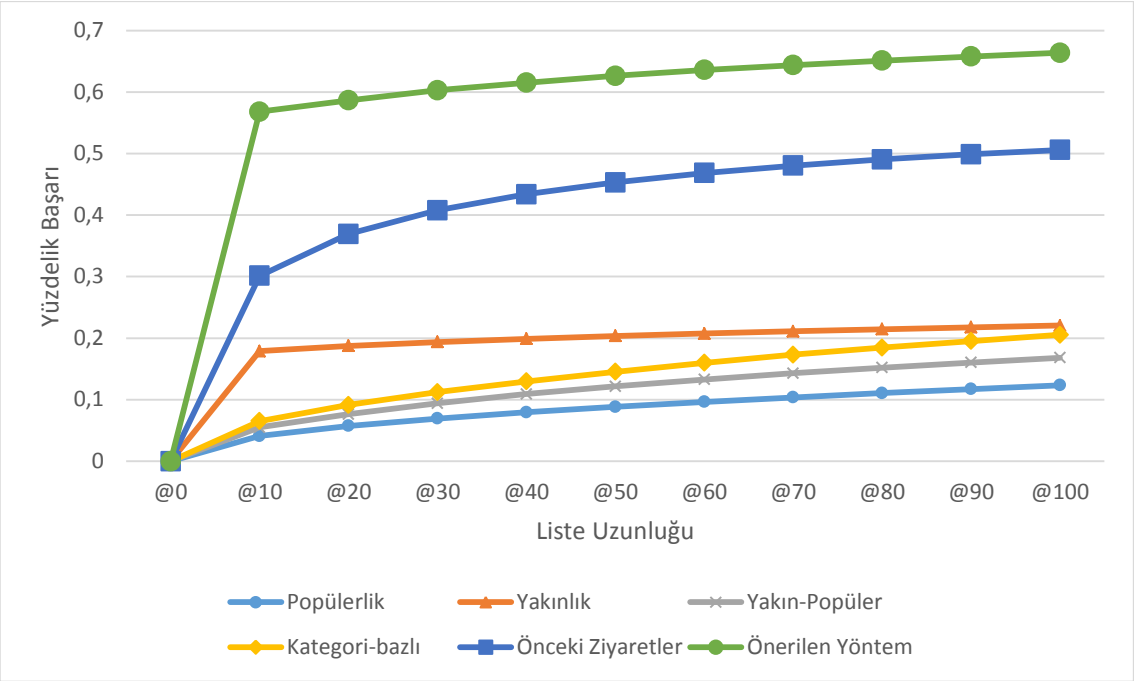
Şekil 4. 1 Foursquare Austin tahmin sonuçları

Austin şehri için en başarılı sonuçları önerdiğimiz yöntem sağlamıştır. Bütün liste uzunluklarında en yüksek yüzdelik başarıyı vermiştir. İkinci en başarılı yöntem önceki ziyaretler olmuştur. Önceki ziyaretler diğer tekil özelliklere göre çok daha başarılı olmuştur. Yakınlık özelliği 50 liste uzunluğuna kadar üçüncü en başarılı yöntem olmuş, liste uzunluğu 50'yi geçtiğinde kategori-bazlı özelliği tarafından geçilmiştir. En başarısız yöntem bütün liste uzunlukları için popülerlik özelliği olmuştur.



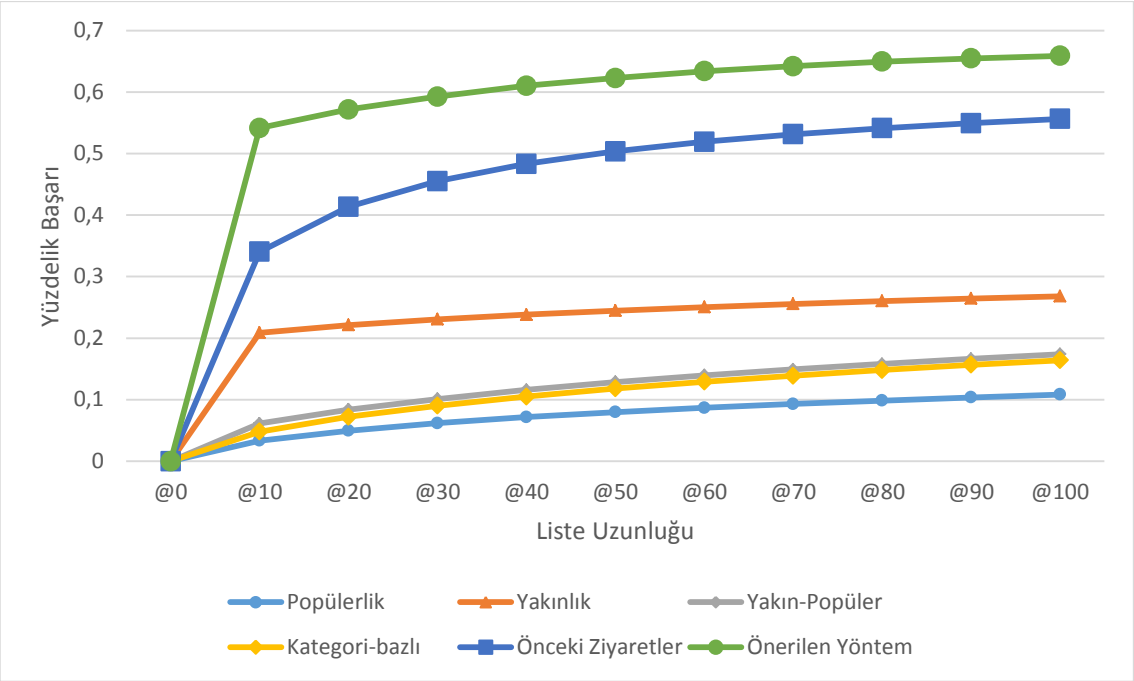
Şekil 4. 2 Foursquare Dallas tahmin sonuçları

Dallas şehrinde en yüksek yüzdelik başarıyı bütün liste uzunluklarında önerdiğimiz yöntem vermiştir. İkinci en başarılı yöntem önceki ziyaretler olmuştur. Önceki ziyaretler diğer tekil özelliklere göre çok daha başarılı olmuştur. Yakınlık özelliği üçüncü en başarılı yöntem olmuş ve önceki ziyaretler hariç diğer tekil özellikleri geride bırakmıştır. En başarısız yöntem bütün liste uzunlukları için popülerlik özelliği olmuştur. Kategori-bazlı ve yakın-popüler özellikleri de birbirine yakın performans göstermiş ve en başarısız ikinci yöntem olmuşlardır.



Şekil 4. 3 Foursquare San Francisco tahmin sonuçları

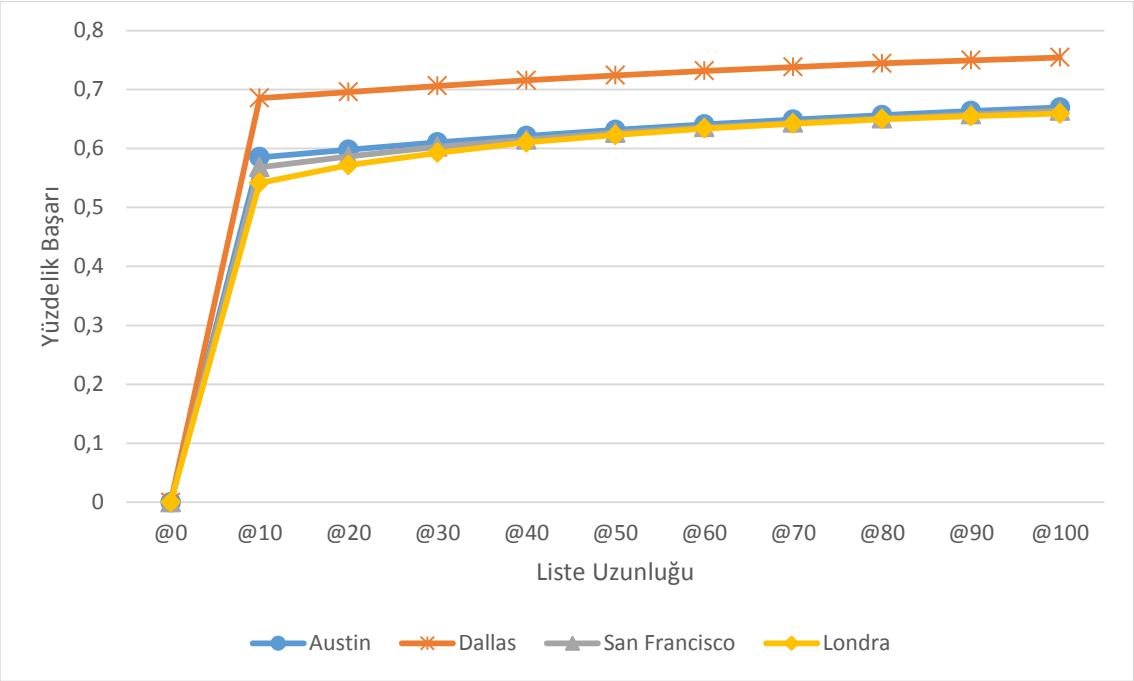
San Francisco şehri için en başarılı sonuçları önerdiğimiz yöntem sağlamıştır. İkinci en başarılı yöntem önceki ziyaretler olmuştur. Önceki ziyaretler diğer tekil özelliklere göre çok daha başarılı olmuştur. Yakınlık özelliği üçüncü en başarılı yöntem olmuş ve liste uzunluğu arttıkça başarısı çok az bir artış göstermiştir. En başarısız yöntem bütün liste uzunlukları için popülerlik özelliği olmuştur. Yakın-popüler en başarısız ikinci ve kategori bazlı en başarısız üçüncü yöntem olmuşlardır. Tahmin yöntemlerinin başarı sırası liste uzunlukları değişirken değişmemiştir.



Şekil 4. 4 Foursquare Londra tahmin sonuçları

Londra şehri için en başarılı sonuçları önerdiğimiz yöntem sağlamıştır. Bütün liste uzunluklarında en yüksek yüzdelik başarıyı vermiştir. İkinci en başarılı yöntem önceki ziyaretler olmuş ve önerilen yönteme yakın bir performans göstermiştir. Önceki ziyaretler diğer tekil özelliklere göre çok daha başarılı olmuştur. Yakınlık özelliği üçüncü en başarılı yöntem olmuştur. Yakın-popüler ve kategori bazlı özellikleri en başarısız ikinci yöntem olmuşlardır. En başarısız yöntem popülerlik özelliği olmuştur.

Önerdiğimiz yöntem bütün şehirler için diğer yöntemleri geride bırakmıştır. Benzer şekilde bütün şehirlerde en başarılı ikinci yöntem önceki ziyaretler olmuştur. Yakınlık özelliği genel olarak en başarılı üçüncü yöntem olurken popülerlik bütün şehirlerde en başarısız sonuçları vermiştir. Önerdiğimiz yöntem diğer yöntemlerden ortalama 0,42 puan daha başarılı sonuçlar vermiştir. Bununla beraber en başarılı ikinci yöntem olan önceki ziyaretlerden ortalama 0,16 puan daha başarılı olmuştur.



Şekil 4. 5 Foursquare önerilen yöntem tahmin sonuçları

Önerdiğimiz yöntemin başarısı şehirden şehre değişmektedir. En başarılı sonuçları Dallas şehrinde elde etmiştir. Diğer üç şehir için elde edilen sonuçlar birbirine yakındır.

Çizelge 4. 1 Foursquare şehirlere göre yer bildirimi, kullanıcı, mekân, ortalama kullanıcı başına düşen yer bildirimi ve ortalama mekân başına düşen yer bildirimi sayısı

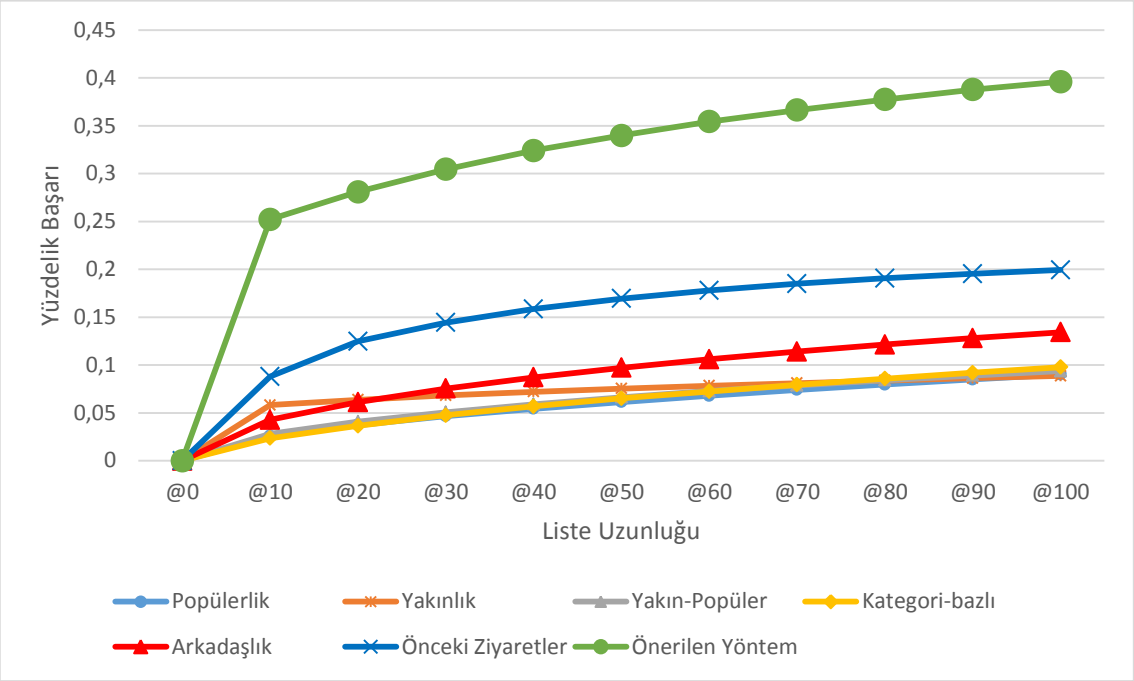
	YER BİLDİRİMİ SAYISI	KULLANICI SAYISI	MEKÂN SAYISI	ORTALAMA KULLANICI BAŞINA DÜŞEN YER BİLDİRİMİ SAYISI	ORTALAMA MEKÂN BAŞINA DÜŞEN YER BİLDİRİMİ SAYISI
DALLAS	196356	3392	11129	57,88	17,64
AUSTİN	262466	6207	9708	42,28	27,03
SAN FRANCİSCO	339447	8218	13150	41,30	25,81
LONDRA	444689	10165	27353	43,74	16,25

Çizelge 4. 1'deki verilere bakılacak olduğunda kullanıcı başına düşen ortalama yer bildirimi sayısı ile önerilen yöntemin başarısı arasında doğrudan bir ilişki olduğu söylenebilir. Dallas şehrinde ortalama kişi başına düşen yer bildirimi sayısı diğer

şehirlere oranla daha fazladır. Diğer üç şehir için ise bu rakamlar birbirine oldukça yakındır. Bununla beraber ortalama mekân başına düşen yer bildirimi sayısı ile yöntem başarısı arasında doğrudan bir ilişki olmadığı da söylenebilir.

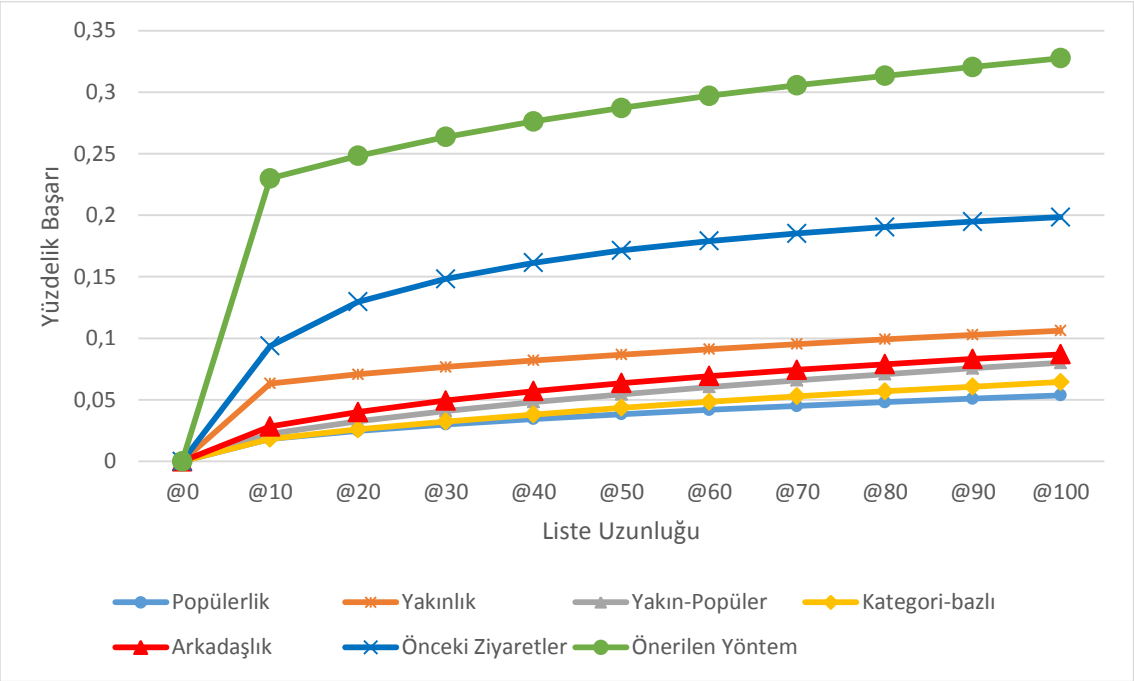
4.2 Gowalla Sonuçları

Sistem her bir tekil özellik ve önerilen yöntem için ayrı ayrı test edilmiştir. Testler şehir bazlı olarak yapılmıştır.



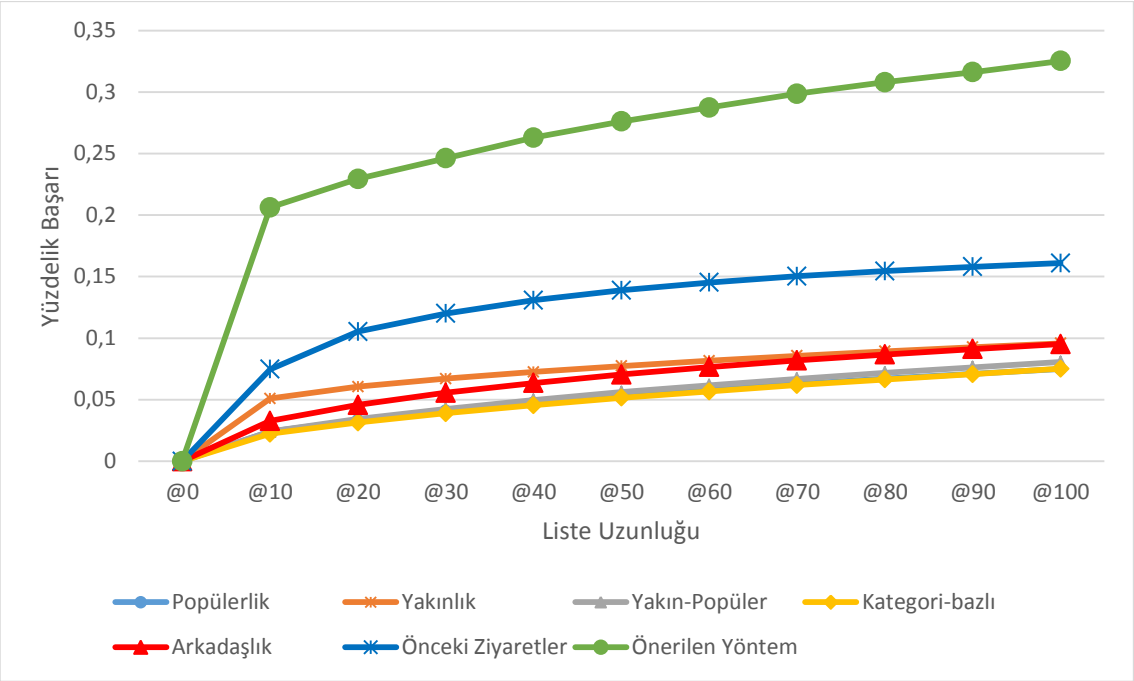
Şekil 4. 6 Gowalla Austin tahmin sonuçları

Austin şehri için en başarılı sonuçları önerdiğimiz yöntem sağlamıştır. Bütün liste uzunluklarında en yüksek yüzdelik başarıyı vermiştir. İkinci en başarılı yöntem önceki ziyaretler olmuştur. Tahmin için arkadaşlık özelliğini kullanan yöntem en başarılı üçüncü yöntem olmuştur. 10 liste uzunluğu için yakınlığın gerisinde kaldıysa da diğer liste uzunluklarında yakınlığa göre daha başarılıdır. Yakınlık özelliği 50 liste uzunluğuna kadar dördüncü en başarılı yöntem olmuş, liste uzunluğu 50'yi geçtiğinde başarısı pek artmamış ve diğer başarısız yöntemlere yakın bir performans ortaya koymuştur. En başarısız yöntemler popülerlik, yakın-popüler ve kategori-bazlı özellikleri olmuştur.



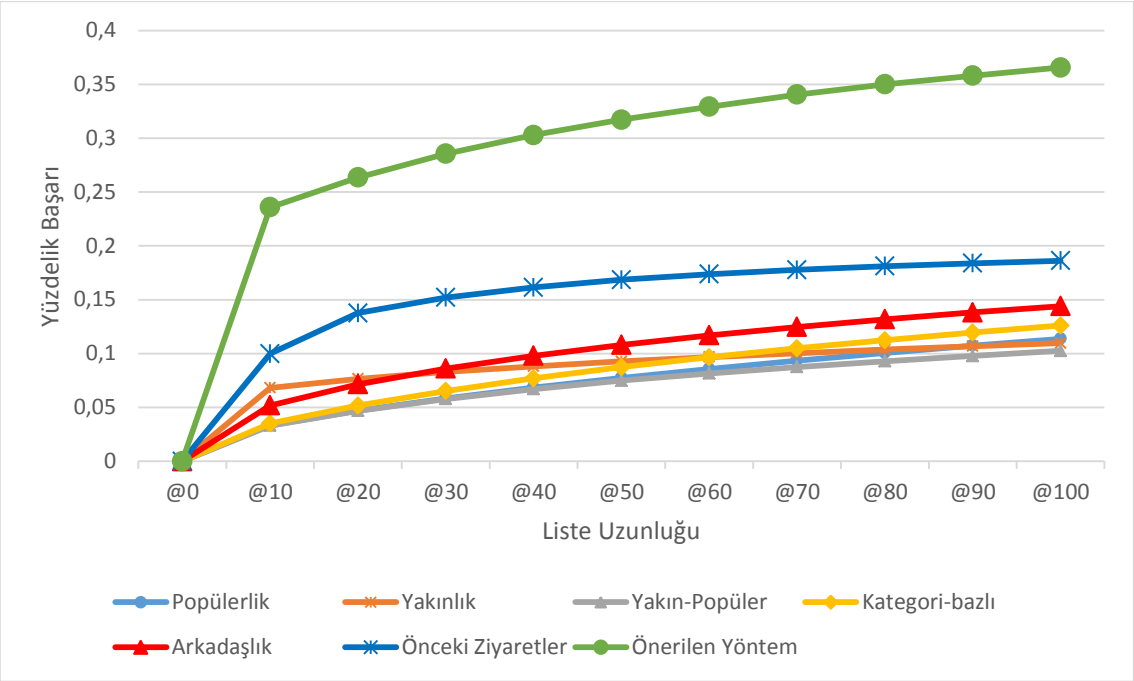
Şekil 4. 7 Gowalla Dallas tahmin sonuçları

Dallas şehri için bütün liste uzunluklarında en başarılı sonuçları önerdiğimiz yöntem sağlamıştır. İkinci en iyi performansı önceki ziyaretler göstermiştir ve diğer tekil özelliklere göre oldukça başarılı olmuştur. Yakınlık özelliği en başarılı üçüncü yöntem olmuştur. Arkadaşlık özelliği en başarılı dördüncü yöntem olurken popülerlik özelliği bütün liste uzunluklarında en kötü performansı göstermiştir. Kategori-bazlı ve yakın-popüler özellikleri en başarısız ikinci ve üçüncü yöntemler olmuşlardır.



Şekil 4. 8 Gowalla San Francisco tahmin sonuçları

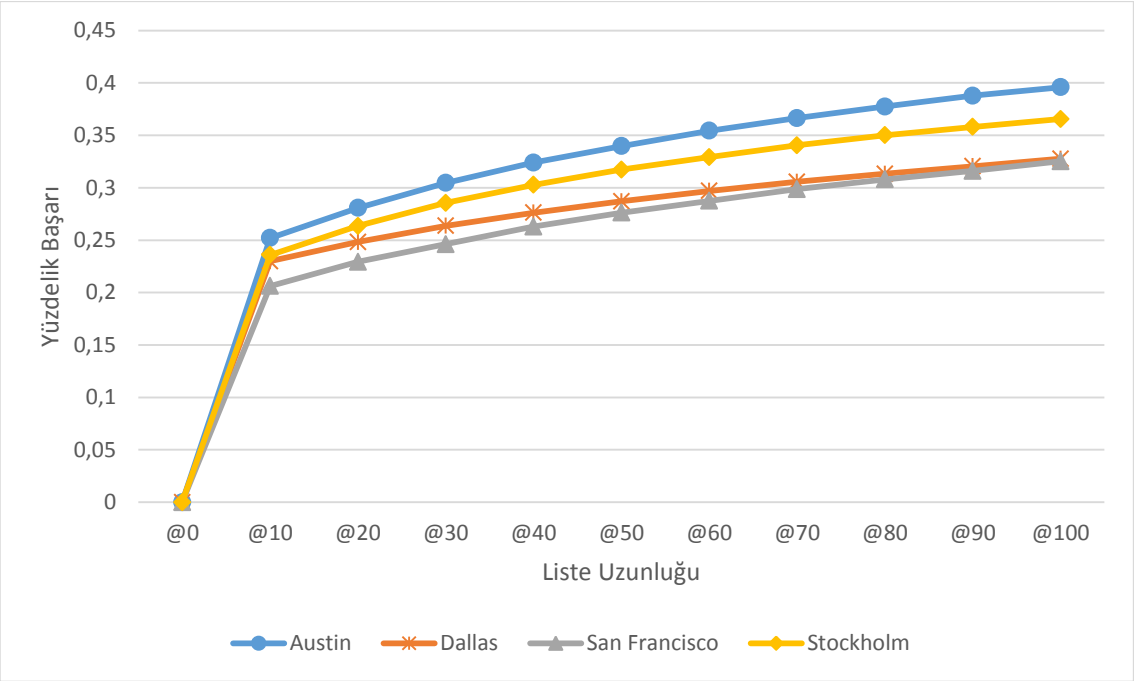
San Francisco şehri için en başarılı sonuçları önerdiğimiz yöntem sağlamıştır. İkinci en iyi performansı önceki ziyaretler göstermiştir ve diğer tekil özelliklere göre oldukça başarılı olmuştur. Yakınlık özelliği en başarılı üçüncü yöntem olmuştur. Arkadaşlık özelliği en başarılı dördüncü yöntem olurken liste uzunluğu arttıkça yakınlık özelliğine yakın bir performans göstermiştir. Kategori-bazlı, yakın-popüler ve popülerlik özellikleri birbirine yakın sonuçlar vermiş ve en başarısız özellikler olmuşlardır.



Şekil 4. 9 Gowalla Stockholm tahmin sonuçları

Stockholm şehri için en başarılı sonuçları önerdiğimiz yöntem sağlamıştır. Bütün liste uzunluklarında en yüksek yüzdelik başarıyı vermiştir. Önceki ziyaretler en iyi ikinci performansı göstermiştir. Yakınlık özelliği 30 liste uzunluğuna kadar en başarılı üçüncü özellik olurken 30'dan uzun listeler için arkadaşlık daha başarılı olmuştur. Kategori-bazlı, yakın-popüler ve popülerlik özellikleri birbirine yakın sonuçlar vermiş ve en başarısız özellikler olmuşlardır.

Önerdiğimiz yöntem bütün şehirler için diğer yöntemleri geride bırakmıştır. Benzer şekilde bütün şehirlerde en başarılı ikinci yöntem önceki ziyaretler olmuştur. Yakınlık özelliği ve arkadaşlık özelliği genel olarak en başarılı üçüncü yöntem olurken popülerlik bütün şehirlerde en başarısız sonuçları vermiştir. Önerilen yöntem diğer yöntemlerden ortalama 0,20 puan daha başarılı sonuçlar vermiştir. Bununla beraber en başarılı ikinci yöntem olan önceki ziyaretlerden ortalama 0,13 puan daha başarılı olmuştur.



Şekil 4. 10 Gowalla önerilen yöntem tahmin sonuçları

Önerdiğimiz yöntemin başarısı şehirden şehre değişmektedir. En başarılı sonuçları Austin şehrinde elde etmiştir. İkinci en başarılı şehir Stockholm olurken San Francisco ve Dallas birbirine yakın performans göstermiştir.

Çizelge 4. 2 Gowalla şehirlere göre yer bildirimi, kullanıcı, mekân, ortalama kullanıcı başına düşen yer bildirimi ve ortalama mekân başına düşen yer bildirimi sayısı

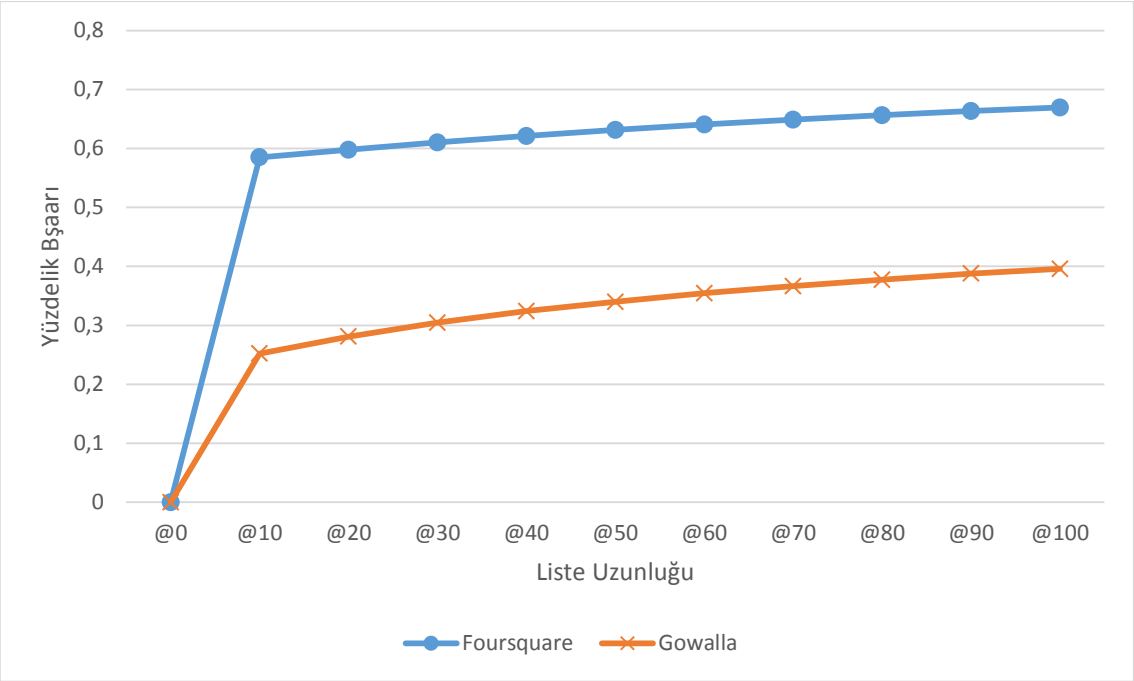
	YER BİLDİRİMİ SAYISI	KULLANICI SAYISI	MEKÂN SAYISI	ORTALAMA KULLANICI BAŞINA DÜŞEN YER BİLDİRİMİ SAYISI	ORTALAMA MEKÂN BAŞINA DÜŞEN YER BİLDİRİMİ SAYISI
DALLAS	281055	3270	22234	85,94	12,64
AUSTİN	468573	6434	20694	72,82	22,64
SAN FRANCİSCO	316673	4511	19471	70,20	16,26
STOCKHOLM	200369	3877	11354	51,68	17,64

Çizelge 4. 2'e bakıldığında önerilen yöntemin başarısı ile ortalama kullanıcı başına düşen yer bildirimi sayısı arasında doğrudan bir orantı olmadığı söylenebilir.

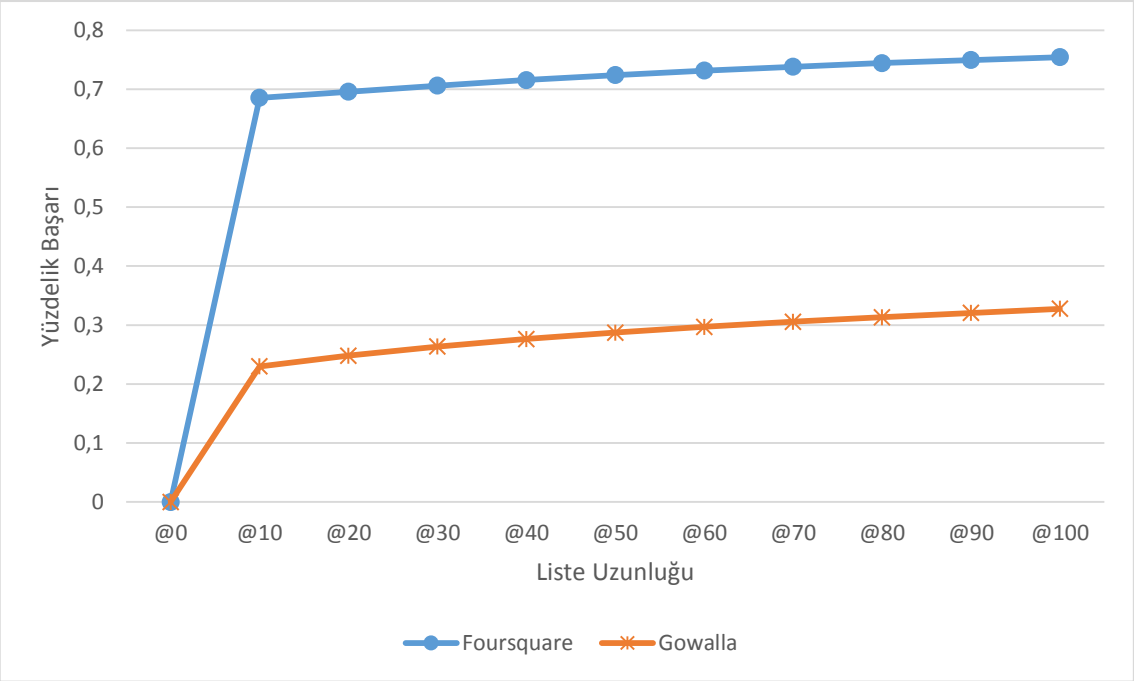
Foursquare'de olduğu gibi ortalama mekân başına düşen yer bildirimi sayısı ile önerilen yöntemin başarısı arasında da bir orantı bulunmamaktadır.

4.3 Foursquare-Gowalla Karşılaştırılması

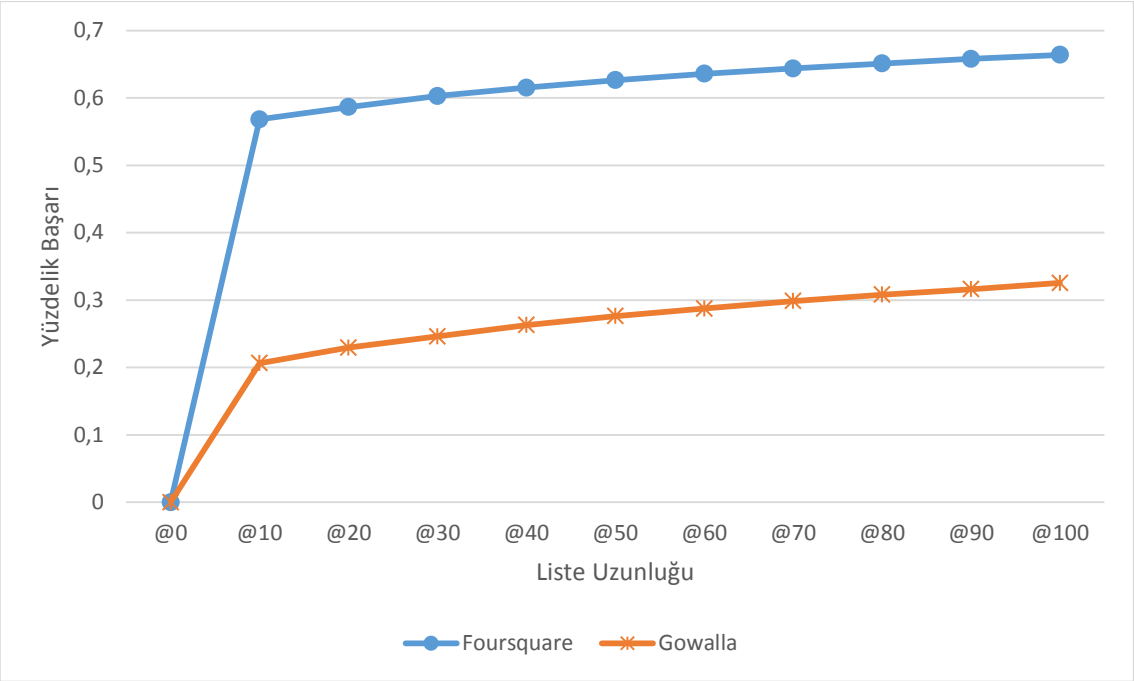
Austin, Dallas ve San Francisco şehirleri her iki tabanında bulunan şehirlerdir. Önerdiğimiz yöntemin tahmin sonuçları her iki veri tabanı içinde elde edildikten sonra karşılaştırılmıştır. Her üç şehir için de Foursquare veri tabanında elde edilen sonuçlar Gowalla'ya göre çok daha iyi bir başarı göstermiştir. Austin için önerilen yöntemin başarısı Şekil 4. 11'de, Dallas için Şekil 4. 12'de ve San Francisco için Şekil 4. 13'de verilmiştir. Austin şehri sonuçların birbirine en yakın olduğu şehir olmuştur. Dallas şehrinde ise sonuçlar arasındaki fark en fazladır. Dallas şehrinde Foursquare veri tabanı içini önerdiğimiz yöntemin yüzdelik başarısı Gowalla'ya oranla iki katından daha fazladır. Gowalla veri tabanı Dallas ve Austin şehirlerinde Foursquare'e göre daha fazla yer bildirimine sahip olmasına karşın tahmin başarısı daha düşük çıkmıştır. Yine üç şehir için de kullanıcı başına düşen ortalama yer bildirimi sayısı Gowalla için daha yüksektir. Sadece Foursquare veri tabanında mekân başına düşen yer bildirimi sayısı Gowalla'dan daha fazladır. Bu noktada başarının doğrudan yer bildirimi sayısı veya kullanıcı başına düşen ortalama yer bildirimi sayısıyla orantılı olduğu söylenemez. Tahmin başarısını etkileyen başka etkenler olabilir. Örneğin; Foursquare veri tabanına ait yer bildirimleri Gowalla'ya göre daha yenidir. Bu da sistemin daha kararlı işlediğini ve daha akla yatkın sonuçlar ürettiğini gösterebilir. Foursquare'deki kullanıcılar Gowalla'dakilere oranla daha fazla aynı mekânda yer bildirimi yapma eğilimdedirler. Bu da önerilen sistemin başarısını doğrudan etkilemektedir. Foursquare'de kullanıcıların kategori tercihlerinin de daha belirgin olarak ortaya çıktığını söyleyebiliriz. Kategori-bazlı tahmin özelliği Foursquare'de daha başarılı sonuçlar elde etmiştir. Popülerliğin de yine Foursquare'de Gowalla'ya göre daha etkili bir tercih olduğu söylenebilir. Bunun bir sebebi de Foursquare'de mekân başına düşen yer bildirimi sayısının fazla olmasıdır. Gowalla'da ise birçok mekânda çok az sayıda yer bildirimi bulunmaktadır. Austin, Dallas ve San Francisco için yüzdelik başarı sonuçları aşağıda verilmiştir.



Şekil 4. 11 Austin önerilen yöntem tahmin sonuçları



Şekil 4. 12 Dallas önerilen yöntem tahmin sonuçları



Şekil 4. 13 San Francisco önerilen yöntem tahmin sonuçları

BÖLÜM 5

SONUÇLAR VE GELECEK ÇALIŞMALAR

Bu çalışmada konum tabanlı sosyal ağlarda konum tahmini problemi üzerinde durulmuştur. Öncelikle bu alandaki popüler iki sosyal ağ olan Foursquare ve Gowalla'dan büyük miktarda veri elde edilmiştir. Daha sonra elde edilen veriler kapsamlı bir şekilde incelenmiş ve bir ön işlemeden geçirilmiştir. Bu aşamada konum tahmini için kullanılabilecek özellikler belirlenmiş ve bu özellikler ayrı ayrı test edilerek başarıları ölçülmüştür. Bir sonraki aşamada konum tahmini problemi için çıkarılmış özellikleri bir arada kullanan bir yöntem önerilmiş ve test edilmiştir. Son olarak bütün tahmin yöntemleri için elde edilen sonuçlar bir arada verilerek başarıları karşılaştırılmıştır.

Sonuç olarak konum tahmini probleminin, konum tabanlı sosyal ağlardan elde edilen verilerin incelenmesi sonucu çıkarılan özellikler ile çözülebileceği görülmüştür. Bununla beraber çıkarılan tahmin özelliklerinin bir arada kullanılmasıyla tahmin başarısının daha da artırılabileceği gösterilmiştir.

Bu alanda yapılacak gelecek çalışmalarda öncelikle sosyal ağlardan elde edilecek bilgi türlerinin ve miktarının artırılması gerekmektedir. Çalışmalarda kullanılacak bilgi miktarının artırılması elde edilen başarıları da arttıracaktır. Şu an için elde edilebilen veri türleri oldukça sınırlıdır. Bunlar yer bildirimi verisi, mekân verisi ve arkadaşlık verileridir. Bu bilgilerin haricinde elde edilebilecek veri türlerine örnek olarak mekânların puanları, fiyat aralıkları, olanakları, kullanıcıların yorumları ve puanlamaları gösterilebilir. Yeni elde edilen bilgilerin tahmin amaçlı kullanılması tahmin başarısını da

artıracaktır. Bunun nedeni problemi birçok farklı açıdan ele alma imkânı elde edilmesidir.

Konum tabanlı sosyal ağlar kullanıcı sayısını ve popülerliklerini artırdıkça hem sunacakları bilgi miktarı artacaktır hem de daha kararlı bir yapıya ulaşacaklardır. Bu nedenle daha güncel verilerle çalışmak tek başına başarıyı artırmaya yardımcı olabilecektir.

Şu ana kadar yapılan çalışmalarda işbirlikçi filtreleme ve eğiticili öğrenme yöntemleri denenmiş ve bu probleme uygulanabileceği ispatlanmıştır. Elde edilecek yeni ve farklı türden verilerle, problemi her açıdan ele alabilecek yöntemlerin de kullanılmasıyla konum tahmini için çok daha başarılı sonuçlar elde edilebilecektir.

KAYNAKLAR

[1] Berjani, B. ve Strufe, T., (2011). "A recommendation system for spots in location-based online social networks". Proceedings of the 4th Workshop on Social Network Systems. ACM, 4.

[2] Karimi, H.A. ve Liu, X., (2003). "A predictive location model for location-based services". Proceedings of the 11th ACM international symposium on Advances in geographic information systems. ACM, 126-133.

[3] Park, M.-H. Hong, J.-H. ve Cho, S.-B., (2007). "Location-based recommendation system using bayesian user's preference model in mobile devices". International Conference on Ubiquitous Intelligence and Computing. Springer, 1130-1139.

[4] Quercia, D. Lathia, N. Calabrese, F. Di Lorenzo, G. ve Crowcroft, J., (2010). "Recommending social events from mobile phone location data". 2010 IEEE International Conference on Data Mining. IEEE, 971-976.

[5] Ye, M. Yin, P. ve Lee, W.-C., (2010). "Location recommendation for location-based social networks". Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems. ACM, 458-461.

[6] Noulas, A. Scellato, S. Mascolo, C. ve Pontil, M., (2011). "An Empirical Study of Geographic User Activity Patterns in Foursquare", ICwSM, 11: 70-573.

[7] Noulas, A. Scellato, S. Lathia, N. ve Mascolo, C., (2012). "A random walk around the city: New venue recommendation in location-based social networks". Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom). IEEE, 144-153.

[8] Noulas, A. Scellato, S. Lathia, N. ve Mascolo, C., (2012). "Mining user mobility features for next place prediction in location-based services". 2012 IEEE 12th International Conference on Data Mining. IEEE, 1038-1043.

[9] Wang, H. Terrovitis, M. ve Mamoulis, N., (2013). "Location recommendation in location-based social networks using user check-in data". Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, 374-383.

[10] Ye, J. Zhu, Z. ve Cheng, H., (2013). "What's your next move: User activity prediction in location-based social networks". Proceedings of the SIAM International Conference on Data Mining. SIAM.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı	:Mücahit BAYDAR
Doğum Tarihi ve Yeri	:19.08.1991/Bursa
Yabancı Dili	:İngilizce
E-posta	:mucahitbaydar@gmail.com

ÖĞRENİM DURUMU

Derece	**Alan**	**Okul/Üniversite**	**Mezuniyet Yılı**
Lisans	Bilgisayar Mühendisliği	TOBB Ekonomi ve Teknoloji Üniversitesi	2014
Lise	Sayısal	Bursa Ulubatlı Hasan Anadolu Lisesi	2009

İŞ TECRÜBESİ

Yıl	**Firma/Kurum**	**Görevi**
2014	AYESAŞ	Yarı Zamanlı Yazılım Mühendisi

YAYINLARI

Bildiri

1. 6th WORLD CONFERENCE on INNOVATION and COMPUTER SCIENCE (INSODE-2016), Antalya **Location Prediction in Location-based Social Networks**