

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

**ÇEVİRİMİÇİ YORUMLARIN METİN MADENCİLİĞİ İLE ANALİZİ:
İSTANBUL'DAKİ ALIŞVERİŞ MERKEZLERİ ÜZERİNE BİR ÇALIŞMA**

YÜKSEK LİSANS TEZİ

Nergiz SÖNMEZ

Endüstri Mühendisliği Anabilim Dalı

Mühendislik Yönetimi Programı

HAZİRAN 2017

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

**ÇEVİRİMİÇİ YORUMLARIN METİN MADENCİLİĞİ İLE ANALİZİ:
İSTANBUL'DAKİ ALIŞVERİŞ MERKEZLERİ ÜZERİNE BİR ÇALIŞMA**

YÜKSEK LİSANS TEZİ

**Nergiz SÖNMEZ
(507141217)**

Endüstri Mühendisliği Anabilim Dalı

Mühendislik Yönetimi Programı

Tez Danışmanı: Doç. Dr. Başar ÖZTAYŞI

HAZİRAN 2017

İTÜ, Fen Bilimleri Enstitüsü'nün 507141217 numaralı Yüksek Lisans Öğrencisi Nergiz SÖNMEZ, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı “ÇEVİRİMİÇİ YORUMLARIN METİN MADENCİLİĞİ İLE ANALİZİ: İSTANBUL'DAKİ ALIŞVERİŞ MERKEZLERİ ÜZERİNE BİR ÇALIŞMA” başlıklı tezini aşağıda imzaları olan jüri önünde başarı ile sunmuştur.

Tez Danışmanı : **Doç. Dr. Başar ÖZTAYŞI**
İstanbul Teknik Üniversitesi

Jüri Üyeleri : **Prof. Dr. Selim ZAIM**
İstanbul Teknik Üniversitesi

Doç. Dr. Selçuk CEBİ
Yıldız Teknik Üniversitesi

Teslim Tarihi : 4 Mayıs 2017
Savunma Tarihi : 15 Haziran 2017





Sevgili Eşime,



ÖNSÖZ

Çalışmam boyunca beni yönlendiren ve yardımcı olan tez danışmanım Doç.Dr. Başar Öztayşi'ye değerli desteklerinden dolayı teşekkürlerimi sunarım. Ayrıca bu yolda desteklerini aldığım Öğr. Gör. Erkan Işıklı ve Araş.Gör. Sultan Ceren Öner'e de teşekkürü bir borç bilirim.

Beni her zaman desteklediği, cesaretlendirdiği ve sabırla yanımda olduğu için eşim Fatih'e ne kadar minnet duysam azdır. Bana her zaman destek olan iki aileme de yanımda oldukları için teşekkür ederim.

Nisan 2017

Nergiz Sönmez
Endüstri Mühendisi



İÇİNDEKİLER

Sayfa

ÖNSÖZ	vii
İÇİNDEKİLER	ix
KISALTMALAR	xi
SEMBOLLER	xiii
ÇİZELGE LİSTESİ	xv
ŞEKİL LİSTESİ	xvii
ÖZET	xix
SUMMARY	xxi
1. GİRİŞ	1
2. SOSYAL MEDYA GELİŞİMİ VE KULLANICI YORUMLARI	3
2.1 Web Teknolojileri ve Sosyal Medya	3
2.1.1 Web 1.0	4
2.1.2 Web 2.0	4
2.1.3 Web 3.0	5
2.1.4 Sosyal medya	5
2.1.5 Sosyal medya türleri	7
2.1.5.1 Bloglar	7
2.1.5.2 Sosyal ağlar	7
2.1.5.3 Forumlar	8
2.1.5.4 Vikiler	8
2.1.5.5 Wordpress	9
2.1.5.6 Podcast	9
2.2 Kullanıcı Yorumları	9
3. LİTERATÜR ARAŞTIRMASI	15
3.1 Alışveriş Merkezleri Üzerine Yapılan Çalışmalar	15
3.2 Foursquare Üzerine Yapılan Çalışmalar	18
3.3 Veri Madenciliği	19
3.3.1 Duygu analizi üzerine yapılan çalışmalar	19
3.3.2 Metin madenciliği-konu modelleme üzerine yapılan çalışmalar	22
4. METODOLOJİ	29
4.1 Veri Madenciliği	29
4.2 Veri Madenciliği Süreci	30
4.2.1 İş anlamak	30
4.2.2 Veriyi anlamak	30
4.2.3 Veriyi hazırlamak	31
4.2.4 Modelleme	32
4.2.5 Değerlendirme	32
4.2.6 Yayma	33
4.3 Metin Madenciliği	33
4.4 Gizli Dirichlet Tahsisi	35

5. UYGULAMA	45
5.1.1 Türkiye’ de AVM sektörü	45
5.1.2 Mekân popülerlik analizi	47
5.1.3 İstanbul AVM’leri konu modelleme	54
5.1.3.1 Veri ön işleme	54
5.1.3.2 MALLER ile analiz	55
5.1.3.3 Konuların etiketlenmesi	60
6. SONUÇ VE ÖNERİLER	63
6.1 Sonuçlar ve Yorumlar	63
6.2 Çalışmanın Eksikleri	76
6.3 Gelecek Çalışmalara Öneriler	76
KAYNAKLAR	79
EKLER	87
ÖZGEÇMİŞ	109



KISALTMALAR

AVM	: Alışveriş Merkezi
ABD	: Amerika Birleşik Devletleri
HTML	: Hyper Text Markup Library
XHTML	: Extensible HyperText Markup Language
LBSA	: Lokasyon Bazlı Sosyal Ağlar
MP3	: MPEG-1 Audio Layer III
ANOVA	: Analysis of Variance
AHP	: Analytic Hierarchy Process
TOPSIS	: Technique for Order Preference by Similarity to Ideal Solution
CBS	: Coğrafi Bilgi Sistemi
HITS	: Hyperlink-Induced Topic Search
DDİ	: Doğal Dil İşleme
VIKOR	: Vlsekriterijumska Optimizacija I Kompromisno Resenje
SVM	: Support Vector Machine
API	: Application Programming Interface
SMO	: Sequential minimal optimization
IBM	: International Business Machines
DHL	: Dalsey, Hillblom ve Lynn
LDA	: Latent Dirichlet Allocation
Yİİ	: Yiyecek İlaç İdaresi
CRISP-DM	: Cross Industry Process for Data Mining
NLP	: Natural Language Process
ID	: Identification
VM	: Veri Madenciliği
GDT	: Gizli Dirichlet Tahsisi



SEMBOLER

α : Alfa
 β : Beta
 Γ : Gama
 θ : Teta
 η : Eta





ÇİZELGE LİSTESİ

Sayfa

Çizelge 2.1: Web Teknolojilerinin Gelişim Dönemleri.....	3
Çizelge 5.1: İstanbul'daki AVM Verileri (Verilerin Genişletilmiş Devamı Ek A)...	49
Çizelge 5.2: Örnek Düzenlenmiş Yorum Verisi.....	55
Çizelge 5.3: Output-Topic-Keys Çıktı Verisi.....	59
Çizelge 5.4: Optimize Edilen Konular.	60
Çizelge 5.5: Konuların Etiketlenmesi.....	61
Çizelge 6.1: AVM Yorum Sayılarının Konular Bazındaki Değerleri.	65
Çizelge A.1: İstanbul'daki Alışveriş Merkezleri Verileri.....	88
Çizelge C.1: Alışveriş Merkezlerinin Konu Dağılımına Etki Eden Yorum Sayıları Çizelgesi.....	102



ŞEKİL LİSTESİ

Sayfa

Şekil 2.1: İşletmelerin İyi veya Kötü Olduğunu Öğrenmek İçin Çevrimiçi Yorum Okuyor Musunuz?	11
Şekil 2.2: Çevrimiçi Yorumlara Nereden Ulaşırsınız?	11
Şekil 2.3: Hangi Alanda Çevrimiçi Yorumları Okursunuz?	12
Şekil 4.1: Klasik Metin Madenciliği.....	34
Şekil 4.2: Gizli Dirichlet Tahsisi Levha Gösterimi	38
Şekil 4.3: Yumuşatılmış Levha Gösterimi	39
Şekil 5.1: DHC Restlet Eklenti Mekan Bilgileri Çekme Ekranı	47
Şekil 5.2: Locations Dosyası İçerik Görünümü.....	48
Şekil 5.3: Veri Çekme Program Görünümü	48
Şekil 5.4: Veri Çekme Programı Sonuç Ekranı	48
Şekil 5.5: AVM Popülerliği (Frekansı)	50
Şekil 5.6: Puan Durumu Genel Görünümü.....	51
Şekil 5.7: Puan Durumu AVM Dağılımı	52
Şekil 5.8: Yorumların AVM Bazlı Yüzdelik Dağılımı.....	53
Şekil 5.9: Düzenlenmemiş Veri Seti.....	54
Şekil 5.10: Komut Ekranından Mallet Çağırımı.....	56
Şekil 5.11: Verilerin Mallet Uygulamasına Aktarımı	57
Şekil 5.12: Örnek Veri Eğitim Ekran Görüntüsü.....	58
Şekil 6.1: Tüm Yorumlar Bazında Konu Dağılım Grafiği	63
Şekil 6.2: En Yüksek Yorum Sahip İlk 10 AVM Konu Dağılımları Grafiği	64
Şekil 6.3: Cevahir AVM Konu Dağılım Grafiği	66
Şekil 6.4: Forum İstanbul AVM Konu Dağılım Grafiği	67
Şekil 6.5: İstinye Park AVM Konu Dağılım Grafiği	68
Şekil 6.6: Viaport Outlet AVM Konu Dağılım Grafiği	69
Şekil 6.7: Marmara Park AVM Konu Dağılım Grafiği	70
Şekil 6.8: Zorlu Center AVM Konu Dağılım Grafiği.....	71
Şekil 6.9: Akasya AVM Konu Dağılım Grafiği	72
Şekil 6.10: Vialand AVM Konu Dağılım Grafiği	73
Şekil 6.11: Mall of İstanbul AVM Konu Dağılım Grafiği	74
Şekil 6.12: Marmara Forum AVM Konu Dağılım Grafiği.....	75
Şekil B.1: 11.-20. Sıradaki AVM'ler İçin Konu Dağılım Grafiği	93
Şekil B.2: 21.-30. Sıradaki AVM'ler İçin Konu Dağılım Grafiği	94
Şekil B.3: 31.-40. Sıradaki AVM'ler İçin Konu Dağılım Grafiği	95
Şekil B.4: 41.-50. Sıradaki AVM'ler İçin Konu Dağılım Grafiği	96
Şekil B.5: 51.-60. Sıradaki AVM'ler İçin Konu Dağılım Grafiği	97
Şekil B.6: 61.-70. Sıradaki AVM'ler İçin Konu Dağılım Grafiği	98
Şekil B.7: 71.-80. Sıradaki AVM'ler İçin Konu Dağılım Grafiği	99
Şekil B.8: 81.-90. Sıradaki AVM'ler İçin Konu Dağılım Grafiği	100
Şekil B.9: 91.-94. Sıradaki AVM'ler İçin Konu Dağılım Grafiği	101



ÇEVİRİMİÇİ YORUMLARIN METİN MADENCİLİĞİ İLE ANALİZİ: İSTANBUL'DAKİ ALIŞVERİŞ MERKEZLERİ ÜZERİNE BİR ÇALIŞMA

ÖZET

Dünyada olduğu gibi Türkiye’de de hızlı gelişen Alışveriş merkezleri kentsel hayatın ayrılmaz bir parçası haline gelmekte ve sayıları giderek artmaktadır. Alışveriş Merkezleri, teknolojinin neredeyse zorunluluk haline getirdiği çevrimiçi platformlarda yerlerini almaya başlamıştır. Böylelikle son kullanıcılara iletmek istedikleri mesajı, bilgilendirmeyi bu yolla yapmaya başlamışlar ve kullanıcılarla daha yakın iletişime geçmişlerdir. Bu platformların en bilinenlerinden biri ve mekân değerlendirme konusunda kullanıcılar üzerinde en çok etkisi olan ağ Foursquare'dir. Burada işletme sahipleri uygulama üzerinden kendi sayfalarını oluşturmakta ve insanlar ne düşündüklerini özgürce söyleyebilmekte, mekânı değerlendirip diğer insanları yönlendirebilmektedir. Yapılan araştırmalarda da günümüzde insanların firmaların söylediklerinden çok diğer kullanıcıların söylediklerine inandıkları ortaya çıkmıştır. Böyle bir ortamda kullanıcıların bıraktıkları yorumlar çok değerli hale gelmektedir.

Bu çalışmada en etkili sosyal medya platformlarından biri olan Foursquare üzerinde İstanbul’daki alışveriş merkezlerine bırakılan yorumlar ile veriler analiz edilerek mekân popülerlikleri analizi ve alışveriş merkezleri için metin madenciliği yöntemlerinden biri olan konu modelleme yapılmıştır. Youmlar Foursquare API kullanılarak bir program aracılığıyla elde edilmiştir. Yorumların yanında toplam check-in sayısı, tekil kullanıcı sayısı, puan bilgileri gibi veriler de alınmıştır. Çalışmada MALLET yazılımı kullanılarak Konu Modelleme yöntemlerinden biri olan Gizli Dirichlet Tahsisi yapılmıştır. Bu alanda Türkiye’de yapılan çalışmaların çok çok az olduğu görülmüştür. Bu yöntemle elde edilen konular, bir etiketleme işlemine tabii tutulmuş, ardından her yorumun hangi konuya yakınlığı olduğu bilgisi kullanılarak her bir alışveriş merkezinin en çok konuşulan konuları ortaya çıkarılmıştır. Bu çıktılara bakıldığında 6 konu içerisinde en çok Genel AVM Atmosferi hakkında yorumların bırakıldığı, ardından sırasıyla Otopark, Kitle-Yoğunluk, Yiyecek-İçecek Hizmetleri, AVM Konumu ve Eğlence-Ek Hizmetler konularının onu takip ettiği bulunmuştur. Bu

alışmanın ıktıları ile İstanbul’da hayatını sürdüren alışveriş merkezlerinin insanların gözünde nasıl algılandığı ortaya ıkarılmıştır. alışmanın amacı her bir alışveriş merkezinde son kullanıcıların oluşturduğu içeriklerin temalarının neler olduğunu ıkarmak, alışveriş merkezlerinde yapılacak öneri sistemleri için bilgi sağlamak ve alışveriş merkezleri yöneticilerine konuşuldukları konular ile stratejilerinde yol gösterici olmaktır. Ayrıca yapılan metin madenciliği alışması ile Türke dili için yaşanan zorluklar ortaya ıkmış ve gelecek alışmalar için öneriler sunulmuştur.



ANALYSIS OF ONLINE REVIEWS WITH TEXT MINING: A STUDY ON SHOPPING CENTERS IN ISTANBUL

SUMMARY

The consumption phenomenon that has become increasingly common in today's societies has caused many forms to enter into our lives. The most known of these formations is, of course, shopping centers. As it is in the world, fast developing shopping centers in Turkey are becoming an integral part of urban life and the numbers are increasing. Although Turkey has a 30-year-old acquaintance with shopping centers, this period is more than 60 years in countries that have completed development like the United States.

Shopping Centers in social media platforms in today's internet age has become almost a necessity in terms of not being behind other competitors. The opportunities offered by the Shopping Centers to people are increasing day by day and the shopping centers have become able to make individual contacts with the customers through the announcement of their activities, news and competitions through these platforms. In the same way, customers can access shopping centers more easily.

Companies in modern shopping centers are able to determine the profile of potential consumers, to determine the needs and desires of consumers in order to protect and / or increase their market share under intense competition conditions, briefly to obtain information about the purchasing behavior of existing and potential consumers and to decide according to the information obtained on marketing strategies Is required [1].

With the rapid development of the Internet, people are advancing parallel to this. With Web 2.0, "user generated content" has started to grow, and this has become one of the most important issues for both users and companies today. When people now like a product or service, when they do not like it, they can hear more voices compared to the old one, and an interpretation that has been taken, the view becomes accessible to a large audience. In this case, we can call it digital marketing by mouth. As a result, companies are now paying close attention to finding these meetings where people are expressing their opinions.

Fourquare, a social media platform, is at a point where places around the globe can create their own pages and affect people's preferences. Foursquare maintains its ongoing popularity on mobile platforms for many years as a social network that can identify the locations in the environment by identifying the location. Thanks to Foursquare, the location can be shared across all social networks in use. Founded in 2009, Foursquare has a serious user base. This application has the feature of not only notifying the place but also marking and saving it in order not to forget the going place, and it is in almost everyone's mobile phone. You can check the pages of these places by giving users points, checking in and commenting via Swarm which is another application connected with Foursquare. It is possible. Foursquare comments can be used as feedback to help other users choose places to visit. For example, someone who recommends a hotel or does not like the service of a restaurant guides others. Interpretations provide the basis for attracting future visitors as well as providing valuable feedback to business owners as well as the link between users and real businesses. Unlike check-ins only seen by a user's friends, comments are visible to everyone and can therefore greatly impact online information sharing and business marketing.

Data mining is crucial in understanding large, complex, and information-rich data. Significant information from this data; Problem solving, decision-making, and so on. The widely used definition of text mining in data mining is "the extraction of qualified information hidden in the data in a non-specific format, the formatting of the irregular data". Objectives such as classification and clustering of text mining texts, subject extraction from texts, generation of class particles, emotional analysis, text summarization, entity relationship modeling

In this study, one of the most effective social media platforms, Foursquare, was analyzed the spatialities left to the shopping centers in Istanbul and analyzed the popularity of the spaces and modeled one of the methods of text mining for shopping centers. Reviews has been obtained through a program using Foursquare API. In addition to comments, data such as total number of check-ins, number of unique users, and point information are also included. In the study, Latent Dirichlet Allocation which is one of the text mining methods was made using MALLET software. It has been seen that there is a one thesis work done in Turkey in this area. Experiments were conducted for different numbers in order to find the best topic, and it was determined that the

most appropriate result was for 6 subjects. The subjects obtained by this method have been subjected to a labeling process and the General Shopping Mall Atmosphere (Genel AVM Atmosferi), Food and Beverage Services (Yiyecek-İçecek Hizmetleri), Parking Lot (Otopark), Mass-Density (Kitle-Yoğunluk), Entertainment-Additional Services (Eğlence-Ek Hizmetler) and Shopping Mall Location (AVM Konumu) labels have been determined through these six tagging keys. Then, using the knowledge that each interpretation is closest to the subject, the most talked-about topics of each shopping center were revealed. When we looked at this output, it was found that the most commented about the General Shopping Mall Atmosphere in 6 subjects followed by Parking Lot, Mass-Density, Food & Beverage Services, Shopping Mall Location and Entertainment-Additional Services. With the outputs of this study, it was revealed how the shopping centers that keep their lives in Istanbul perceived in people's eyes. The aim of the workshop is to find out what the themes of the content of the end users in each shopping mall are, to provide information for the proposal systems to be made in the shopping centers and to guide the shopping centers managers on their subjects and strategies. In addition to this, the text mining study presented the difficulties for Turkish language and suggestions for future studies.



1. GİRİŞ

Günümüz toplumlarında giderek yaygınlaşan tüketme olgusu, yanında birçok oluşumun hayatımıza girmesine sebep olmuştur. Bu oluşumlardan en bilineni şüphesiz alışveriş merkezleridir. Dünyada olduğu gibi Türkiye’de de hızlı gelişen alışveriş merkezleri kentsel hayatın ayrılmaz bir parçası haline gelmekte ve sayıları giderek artmaktadır. Türkiye’nin alışveriş merkezleri ile yaklaşık 30 yıllık bir tanışıklığı olsa da ABD gibi gelişimini tamamlamış ülkelerde bu süre 60 yıldan fazladır.

Alışveriş Merkezlerinin günümüz internet çağında sosyal medya platformlarında yer almaları diğer rakiplerinin gerisinde kalmamaları açısından neredeyse zorunluluk haline gelmiştir. Alışveriş Merkezlerinin insanlara sunduğu olanaklar her geçen gün artmakta ve AVM’ler bu platformlar üzerinden etkinliklerini, haberlerini, yarışmalarını duyurmakla birlikte müşteriyle birebir temas kurabilir hale gelmiştir. Aynı şekilde müşteriler de bu sayede AVM’lere daha kolay ulaşabilmektedir.

Sosyal medya platformlarından olan Foursquare, tüm dünya üzerindeki mekanların kendi sayfalarını oluşturabildiği ve insanların tercihlerini etkileyebildiği bir noktada yer almaktadır. Bu mekan sayfalarına, kullanıcıların puan verebilmesi, yorum yapabilmesi ve Foursquare ile bağlantılı olan bir diğer uygulama olan Swarm üzerinden check-in yapabilmesi mümkündür. Bu çalışmada Foursquare’deki İstanbul AVM’lerine bırakılan yorumlar, veri madenciliği içerisinde yer alan metin madenciliği ile konu modelleme yapılarak incelenmiştir.

Veri madenciliği, büyük boyuttaki, karmaşık ve bilgi bakımından zengin verilerin anlaşılması açısından çok önemlidir. Bu verilerden anlamlı bilgiler çıkarılması; problem çözme, karar verme gibi bir çok alanda fayda sağlar. Çalışmamızda Foursquare uygulaması üzerinden bırakılan yorumlar üzerinden veri madenciliği çalışmaları gerçekleştirilmiştir.

Modern alışveriş merkezlerindeki işletmelerin yoğun rekabet koşulları altında pazar paylarını koruyabilmeleri ve/veya arttırabilmeleri için potansiyel tüketicilerinin profilini bilmesi, söz konusu tüketicilerin ihtiyaç ve isteklerini saptayabilmesi, kısaca

mevcut ve potansiyel tüketicilerinin satın alma davranışı hakkında bilgi elde ederek, pazarlama stratejilerine elde edilen bilgilere göre karar vermesi gerekmektedir [1].

Bu çalışmada en etkili sosyal medya platformlarından biri olan Foursquare üzerinde İstanbul'daki AVM'lere bırakılan yorumlar analiz edilerek mekân popülerlikleri analizi ve AVM'ler için konu modelleme yapılmıştır. Bu çalışmanın çıktıları ile İstanbul'da hayatını sürdüren AVM'lerin insanların gözünden nasıl algılandığı ortaya çıkarılmıştır. Çıktılar bir yandan AVM yöneticilerine hangi alanda konuşuldukları konusunda bilgi vermektedir. Böylelikle gelecekte verecekleri stratejik kararlarda yol gösterici olmak hedeflenmektedir.

Çalışma doğrultusunda Sosyal Medya ve Kullanıcı Yorumlarının Gelişimi başlığı altında Sosyal medyanın günümüze kadar olan yolculuğu ve insan etkileşiminin artmasıyla bırakılan kullanıcı yorumların önemi üzerine bilgi verilmiştir. Literatür bölümünde Alışveriş Merkezleri, Foursquare, Veri Madenciliği başlığı altında Metin Madenciliği-Konu Modelleme ve Duygu Analizi üzerine yapılan çalışmaların bilgileri yer almaktadır. Metodoloji bölümünde Veri Madenciliği süreci tanıtılmış ve Metin Madenciliği Konu Modelleme yöntemlerinden biri olan Gizli Dirichlet Tahsisi (Latent Dirichlet Allocation) tanıtılmış ve ardından uygulamaya geçilmiştir. Uygulamadan elde edilen çıktılar Sonuç ve Öneriler bölümünde değerlendirilmiştir.

2. SOSYAL MEDYA GELİŞİMİ VE KULLANICI YORUMLARI

2.1 Web Teknolojileri ve Sosyal Medya

Teknolojik anlamda yaşanan gelişmeler birçok yüksek teknoloji ürün ve hizmeti gündelik hayatın önemli bir parçası haline getirmiştir. Teknolojiye bağlı olarak geliştirilen hizmetler ayrı birer sektör haline gelmiş ve bu yeni ürün ve hizmetler tüketicinin de değişmesine sebep olmuştur. Özellikle, internet ve internete bağlı gelişen teknolojiler, tüketicilerin alışkanlarını etkilemekle kalmayıp kendi kitlelerini de oluşturmuştur. 1991 yılından bu yana kullanılmakta olan Dünya Çapında Ağ (World Wide Web), bilgi paylaşımının yapılabilmesi, web sayfalarının yayınlanabilmesi ve oluşturulan sayfalara hızlı, ucuz yollarla erişebilmeyi sağlaması ile ivedi bir büyüme yaşamıştır.

Web servis uygulamalarının günden güne artmasından dolayı web artık statik olmaktan ziyade dinamik bir yapıya kavuşmuştur. Web kullanımının yaygınlaşmasıyla ve talebin her geçen gün artması ile birlikte bu alanda kullanılabilecek cihazlar da çeşitlenmeye başlamıştır.

Web teknolojileri denildiğinde, bir süre HTML (Hyper Text Markup Language) yani monolog olarak metinlerin yer aldığı ve tek taraflı kullanıcı hizmetinde olan bir yapı anlaşılırken zaman içerisinde bu dil, XHTML ve HTML 5 gibi sürümlere kadar geliştirilmiştir. Web teknolojilerinde; Web 1,0, Web 2,0 ve Web 3,0 olarak adlandırılan gelişimler yaşandı. Bu gelişim, aşağıda verilen Çizelge 2.1'deki gibidir:

Çizelge 2.1: Web Teknolojilerinin Gelişim Dönemleri.

Web Teknolojisi	Geçerli Olduğu Yıllar	Odak Alanı
Web 1.0 (monolog)	1995-2000	Belge odaklı
Web 2.0 (etkileşimli)	2000-2010	İnsan odaklı
Web 3.0 (yapay zekalı)	2010-2020	Bilgi odaklı

Web 1.0 internette tek yönlü bir iletişime olanak sağlar ve verinin okunabilmesi anlamına gelir. Web 2.0 teknolojisi insan odaklı yapısıyla interaktif verinin yazılabilmesini sağlar. Web 3.0 ile de internetin akıllanıp dinamik uygulamaların,

interaktif servislerin ve makineden makineye etkileşimli uygulamaların çalışabilmesi anlamına geldiğini söylemek mümkündür [2].

2.1.1 Web 1.0

Web 1.0, internetin ilk dönemi için telaffuz edilir. Kullanıcılar bu dönemde ihtiyaç duyduğu bilgiyi bulmak için internet siteleri üzerinde araştırma yapar, istediği bilgiyi alır ve siteden ayrılır. Kullanıcı olarak etkileşime girme yani yorum yapma, katkı sağlama, içerik üretme gibi bir imkânı yoktu. Web 1.0'da kullanıcılar sadece okuyucuydu ve sadece bilgiyi alabilen olarak konumlandırılıyordu. Bu durumun sebebi teknolojinin bu kadarına izin verebiliyor olmasıydı.

Dönem içerisinde internet kullanımının da artmasıyla beraber internette bulunan kişiler ihtiyaçlarını karşılayacak içeriklere ulaşamamaya başladılar, sunulan içeriklerin yetersizliği, bunlara müdahale edebilme isteği, kendi içeriğini oluşturma, paylaşma ve etkileşim hali, kendini internet içinde belirli bir sosyal sınıfa ait görme ihtiyacı arttı ve Web 2.0'ın doğum süreci başladı [3].

2.1.2 Web 2.0

Web 2.0, Web 1.0'ın ihtiyaçlara yeterli gelmemesi ile ortaya çıkmıştır. İnsanların aktif bir şekilde rol alıp web ile etkileşime geçtiği web teknolojileri olarak nitelendirilebilir. Aslında bir nevi internet kullanıcılarının ortak olduğu ve paylaşarak meydana getirdikleri sistemi ifade eder. Kullanıcılar tarafından internette içeriğin oluşturulması ve başka insanların bu içerikleri görebilmesine olanak sağlar. Sosyal ağ siteleri, web tabanlı wikipedia gibi özgür ansiklopediler, bloglar, iletişim araçları gibi çevrimiçi araçlarla etkileşim içinde ve paylaşım ile gerçekleşir. İnsanların sosyalleşme ihtiyaçların karşılanmasına yönelik, blog, wiki, podcast vb. teknolojiler ve uygulamalar ortaya çıkmış ve kullanımı yaygınlaşmıştır [2]. Web 2.0 ile birlikte tasarım alanında da gelişmeler olmuştur.

En basit haliyle Web 2.0, Wikipedia, Ekşi Sözlük, Youtube, Foursquare, Facebook, Instagram, Twitter, Pinterest gibi siteler, kullanıcılar arasında en popüler internet siteleridir.

2.1.3 Web 3.0

Web 3.0, internet kontrolünün insanın elinden çıktığı ve dışarıdan bir müdahaleye gerek duymadan kendini yönetebildiği bir dünyadır. Bu internet üzerindeki tüm bilgi ve hareketlerin makineler tarafından da anlaşılabilir olması anlamına gelmektedir. Makineler, bilgileri hafızasına alacak ve bir çok alanda en uygun olanlarını insanlara sunacaktır.

Web 3.0, oluşturulan içeriklerin kontrolünün artık insan elinde olmadığı yazılımların kontrolüne geçtiği bir dünyadır. Web 3.0, semantik web'in bir diğer söylemi "yapay zekalı web" dir. Birbirleri ile iletişime geçen veritabanları ile, bir insanın neyi sevip neyi sevmediğini çıkaran, buna göre içerik üreten ve müşteri ilişkilerini temel baz alan bir web dünyasıdır. Bu dünya, içinde internet olan tüm cihazları da kapsayacağı için, Web 3.0'ı bizi anlayan robotik teknolojilerin bütünü olarak düşünülebilir [4].

2.1.4 Sosyal medya

Web 2.0 ile birlikte Bloglar, Facebook, Twitter, Foursquare gibi platformlar üzerinde sosyal paylaşım yaşanmaya başladı. İnsanlar artık ne düşündükleri ne hissettiklerini yazabilmek ve bu içerikleri başka insanların görebilmesini istiyordu. Bu düşüncenin artmasıyla ve imkânların el vermesiyle birlikte dünyada milyonlarca yazar ortaya çıktı. Artık, insanlar gittiği yemek yerlerini, kafeleri ve hatta yediklerini sosyal ağlarda paylaşmaya başladı, bu tecrübeyi yaşamak isteyenlere de fikir vermeye başladı. Bir yandan bunlar yaşanırken şuan hayatımızın olmazsa olmazı mobil cihazlar da yaygınlaşınca ve kolaylaşınca bu etkileşim giderek hızlandı. Artık, herkese açık ansiklopedilerde ve vikilerde insanlar bildiklerini yazarak içerik oluşturuyor, diğer insanlar bu içerikleri inceleyip onları doğrulayabilir, yanlışlayabilir veya eklemeler yapabilir hale gelmiştir.

Sosyal Medya, bir çok insanın mobil ve çevrimiçi olduğu dijital çağ da paylaşım ve iletişim sınırı olmadan özgür ve özgün tartışma zemini yaratan platformdur [5]. Sosyal medya kelimesi yerine bazı kaynaklarda Web 2.0' da kullanılmaktadır. Sosyal medyanın temelinde paylaşım ve iletişim vardır. Sosyal medya kelimelerle, resimlerle, seslerle ve videolar ile teknolojiyi bir araya getirir. Tüketicilerin çoğu sosyal medyayı arkadaşlarıyla iletişime geçmek, günceli takip etmek, insanların düşüncelerinden faydalanmak ve ürünler hakkında bilgi edinmek için kullanırlar. Sosyal medyanın 3 boyutu vardır;

1-Medya Boyutu: Medya, haberleşme teknolojileriyle beraber gelişen ve haberleşmenin her döneminde çok geniş iletişim araçları ile yapılan iletişim ve etkileşim faaliyetlerine konu olan bir kavramdır [6]. Eskiden medya gereçleri sadece gazete, dergi, televizyon iken; artık yeni medya araçları hem gazete, dergi ve televizyon görevi gören hem de insanlara bu gereçlerin izlenebildiği ortamlara ulaştıran konumundadır. Yeni medya araçlarının bu özelliği onlara sosyal bir anlam yüklemiştir [7].

2-Kullanıcı Boyutu: Bu boyut tüketiciler yani son kullanıcılar tarafından oluşturulan çeşitli medya türlerinin içeriği olarak ele alınabilir. Kullanıcı tabanlı içerik aşağıdaki üç önemli özelliğe sahiptir:

Yayınlama Zorunluluğu: Kullanıcı tarafından oluşturulan içeriğin temel özelliği, oluşturulan bu içeriğin yayınlandığı herhangi bir ortamın var olmasının zorunluluğudur. Herkese açık olan bir web sitesi veya belirli kullanıcıların görebildiği sosyal ağlar bu ortamlara girer. E-posta ve anlık mesajlaşmalardan farkı da bu şekilde yayınlanabilmesinden gelmektedir.

Yaratıcı Çaba: Bu boyut yeni bir içeriğin oluşturulmasında veya var olan içeriklerden, çalışmalardan yeni bir içerik ortaya çıkarılması sırasında belli bir yaratıcı çabanın ortaya konulması gerektiğini açıklar. Bu bir anlamda içeriğinin kullanıcıya özel olması, kendinden bir şeyler katması anlamına gelir. Başka birinin yaptığı çalışmayı alıp farklı bir ortama yüklemek kullanıcının üretmiş olduğu bir içerik anlamına gelmez.

Profesyonel rutin ve uygulamalardan bağımsız olmak: Kullanıcı tabanlı içerikler, kişi tarafından kurumsal veya ticari içerik oluşturmadıkları ve kar amacı gütmediklerinden belirli bir şablona yani profesyonel bir rutine ve uygulamaya bağlı kalma gereğinin dışında oluşturulmaktadır. Kullanıcı tabanlı içerikleri oluşturan kişilerin güdülleri genelde, kendilerini ifade etmek, tanımadıkları insanlarla iletişime geçmek, tanınmak ve itibar sahibi olmak olarak sıralanabilir.

3-Teknoloji Boyutu: Bu boyut ise teknolojinin çok hızlı gelişmesi ve farklı birçok imkan sunması, bu durumun akabinde insanların internette daha fazla zaman geçirmeye başlaması olarak açıklanabilir [8].

2.1.5 Sosyal medya türleri

2.1.5.1 Bloglar

Blog, internet kullanıcılarını sosyalleşme adına paylaşma ve katılıma yönelten araçlar olarak tanımlanır. İnsanlar blog yazarak; bir diğer ifadeyle hissettiklerini, tecrübelerini çevrimiçi günlüğüne yazarlar. Günümüzde birçok alanda bloglar yazılmakta ve diğer insanlar tarafından takip edilmektedir. Takipçi sayısı fazla olan bir seyahat blogunda yazılan ve paylaşılan yazılar, blogu takip eden ve bir ülke hakkında bilgi almak isteyen kişi tarafından yorumlanır, oylanır ve değerlendirilir. Böylece bu yorumlar sayesinde bilgi almak isteyen ve bilgi vermek isteyen taraflar bir araya gelerek bir fikir alışverişi ve etkileşim ortamı sağlanmış olur [9].

2.1.5.2 Sosyal ağlar

Son zamanlarda sosyal medya olarak tanımlanan internet tabanlı uygulamalarla ortaya çıkan ve birçok insanın ortak bir platform üzerinde enformasyon paylaşabildiği veya pasif olarak da olsa bu mecralar üzerinde iletişimde bulunduğu ortamlar sosyal paylaşım ağları olarak adlandırılmaktadır. Örneğin: LinkedIn, Foursquare, Facebook vb.

Oldukça sık kullanılan akıllı telefon uygulamalarından biri de Foursquare'dir. Foursquare, bulunulan lokasyonu belirleyerek, çevrede hangi mekanların olduğunu gösterebilen bir sosyal ağ olarak uzun yıllar mobil platformlarda devam eden popülerliğini korumaktadır. Foursquare sayesinde bulunan konum, kullanılan tüm sosyal ağlar üzerinden paylaşılabilir. 2009 yılında kurulan Foursquare ciddi bir kullanıcı sayısına sahiptir. Sadece mekan bildirmekle kalmayıp, gidilen bir yeri unutmamak adına işaretleyip, kaydetmek gibi bir özelliği de bulunan bu uygulama hemen herkesin cep telefonunda bulunmaktadır. Mayıs 2014'te, hizmetin sosyal paylaşım ve konum paylaşım özelliklerinden bazılarını ayıran ve Foursquare'e eşlik eden bir mobil uygulama olarak Swarm hayatımıza girdi [10]. Foursquare'in 2015 yılında açıklanan toplam kullanıcı sayısı 55 milyon iken [11] 2016 yılında uygulamada yapılan check-in sayısı 10 milyarı geçmiştir [12]. Türkiye'de ise 7.2 milyon Foursquare kullanıcısı günde 1.8 milyon check-in yapmaktadır [13].

Google+ ve Facebook'un aksine, Foursquare yalnızca konuma dayalı hizmetleri sunar, bu nedenle kullanıcılar diğer iletişimlerinden rahatsız olmaz. Foursquare'deki yerler

herkes tarafından tanımlanabilir ve dokuz yüksek seviye kategorisine yaklaşık 663 alt düzey kategori ile hiyerarşik bir taksonomi altında sınıflandırılabilir. Dolayısıyla, Lokasyon Bazlı Sosyal Ağlar (LBSA)'da, bireysel kullanıcıların özelliklerini ve kaydetmeyi seçtikleri ziyaret yerlerini araştırmak için uygun bir araç sağlar [14].

Yorumlar, Foursquare'da kullanılan ve kullanıcının yazdığı bir terimdir ve diğer kullanıcılara orada neyin iyi veya kötü olduğunu bildiren kısa bir mesajdır [15]. Foursquare yorumları, gelecekteki ziyaretçilerin davranışlarını doğrudan etkileyebilir ve işletme sahiplerine değerli geri bildirimler sağlayabilir [16]. Yorumlar kullanıcılar tarafından gönüllü olarak gönderilir. Lokasyon bazlı sosyal ağlarda gayretli olarak adlandırılan kullanıcılar her gittikleri yerde check-in veya yorum bırakmaktadır [17]. Bırakılan yorumlar, olumlu veya olumsuz olabilir. Bu yorumları gören kullanıcılar yapılacaklar listesine ekleme veya yapıldı olarak işaretleyerek, bir bakıma bu yorumlara geri dönüş yaparlar. Bu özellikler kullanıcıların deneyimlerini paylaşmaları ve diğerlerinden öğrenmeleri bakımından çok değerlidir [18]. Değerli çıktılarından dolayı Foursquare ile ilgili çeşitli araştırmalar yapılmıştır.

2.1.5.3 Forumlar

Forum, blog gibi tek bir kullanıcı tarafından içeriklerin oluşturulmasından ziyade birden fazla insanın bir araya gelerek bilgilerini, tecrübelerini ve aklımıza gelebilecek her konuda konu açtığı veya yorum yazıp paylaşabildiği internet ortamlarıdır. Diğer bir deyişle forumlar sanal dünyanın sosyalleşmiş siteleridir. Günümüzde forumlarda tanışan insanlar gerçek dünyada da bir araya gelerek iletişimlerini devam ettirmektedirler [19]. Örneğin, teknolojik bir alanda desteğe ihtiyacı olan biri Donanım Haber Forumunu mutlaka ziyaret ediyor olacaktır.

2.1.5.4 Vikiler

Viki ismi Hawaii dilinde "hızlı" anlamına gelen wiki kelimesinden türetilmiştir. Ortak bir amaç etrafında toplanan insanların web tarayıcı üzerinden yeni sayfalar oluşturmaya, sayfalarda düzenlemeler yapmasına ve bu sayfaları birbirine bağlamaya olanak sağlayan web sitelerinin genel adıdır. Viki sitelerinde kullanıcılar büyük dokümantasyonları kolayca oluşturabilir. Vikilerin sahip olduğu Diff özelliği sayesinde sürüm yönetimi yapılarak belgeler arasındaki farklılıklar görülür ve önceki sürümler incelenebilir. Sayfalar arasındaki bağlantılar ve sayfa biçimlemeleri sistem tarafından otomatik olarak yapılandırılır. Böylelikle bilgiye erişme ve bilgi belgeleme

wiki ile son derece kolaylaşmaktadır ve ayrıca son yapılan araştırmalar sonucunda en çok tıklananlar arasında yerini almıştır [20].

2.1.5.5 Wordpress

Wordpress, ücretsiz olan ve açık kaynak kodlu yapısıyla dünyanın en çok kullanılan blog altyapısı sistemlerinden biridir. Alexa istatistiklerine göre en iyi 1 Milyon Web sitesinin %22'si Wordpress altyapısı ile oluşturulmuştur. Wordpress altyapısının en önemli özelliklerinden biri web sitesinin kısa sürede aktifleştirilebilmesidir. Her türlü yapıdaki içerik yayınlanabilir. Açık kaynak kodlu olmasının avantajı çok sayıda geliştirici tarafından desteklenip, her geçen gün daha da kolay kullanılabilir, daha gelişmiş özelliklere sahip olmakta olan sağlam bir bloglama sistemi olmasıdır [21].

2.1.5.6 Podcast

Podcasting, ses, görüntü ve yazı içeren yayınların yayın zamanından sonra izlenmek üzere bilgisayar ve taşınabilir cihazlara indirilebilecek şekilde yayınlanmasıdır [22]. Pazarlama aracı olarak postcastler günümüzde giderek daha da önem kazanmaya başlamıştır. Postcastler pazarlama aracı olarak düşünüldüğünde hedeflenebilir, esnektir, masrafı daha azdır, ölçümlenebilir, kontrol edilmesi kolaydır.

2.2 Kullanıcı Yorumları

İnternetin hızlı gelişimiyle insanların buna adapte olması paralel ilerlemektedir. Web 2.0 ile birlikte “user generated content” yani kullanıcı tarafından oluşturulan içerikler artmaya başlamış ve bu durum hem diğer kullanıcılar hem de şirketler için bugünün en önemli konularından biri haline gelmiştir. İnsanlar artık bir ürünü veya hizmeti beğendiklerinde, beğenmediklerinde eskiye göre daha fazla sesini duyurabilmekte ve atılan bir yorum, görüş büyük bir kitleye ulaşabilir hale gelmiştir. Bu duruma bir bakıma dijital ağızdan ağıza pazarlama diyebiliriz. Bundan dolayı artık şirketler insanların görüş bildirdiği bu mecralarda bulunmaya özen göstermekte ve müşterisiyle yakından ilgilenmek durumundadır.

Son yıllarda kullanıcılar, markaların hazırladıkları reklamlar ve pazarlama çalışmalarından daha az etkilenmektedir. Ünlü kişilerin maddi çıkar karşılığında reklamlarda oynaması da bu güveni azaltmıştır. Bu durumların oluşmasından dolayı, kullanıcılar başka kullanıcıların ya da firmaların müşterilerinden gelen yorumları daha

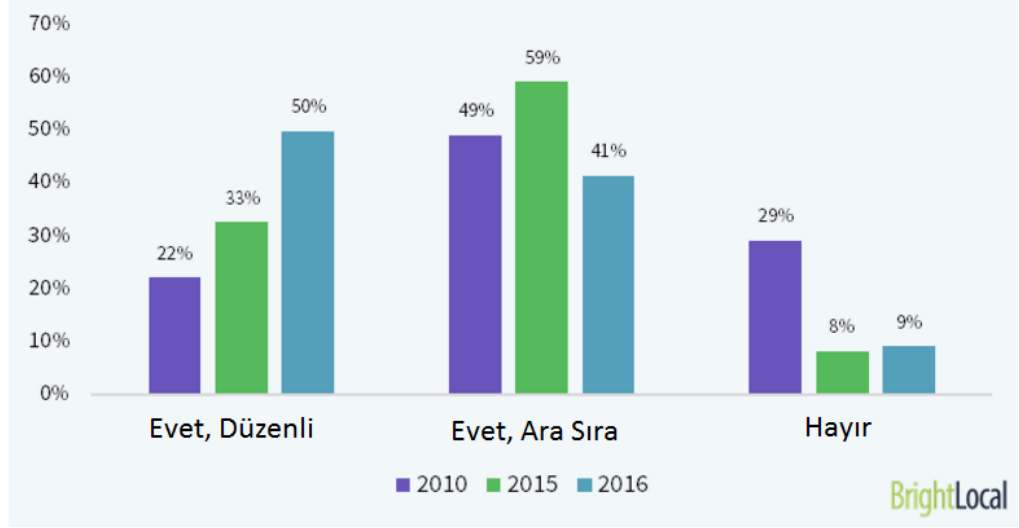
inandırıcı ve doğru bulmaya başlamıştır. Ardından satınalma süreçlerinde ve mekân tercih ederken bu bilgilere göre davranış sergilemeye özen göstermektedir. Bu konuda yapılan çalışmalarda;

- İnsanların yüzde 88’lik bir bölümünün çevrimiçi yorumlara, diğer insanlardan aldıkları tavsiyeler kadar güvendikleri,
- Tüketicilerin yüzde 72’lik bir kısmının, firma ile ilgili olumlu yorumlar okuduktan sonra satın alma işlemini gerçekleştirdikleri,
- Yüzde 72’lik bir bölümün, bir mağaza ile ilgili okudukları olumlu yorumların, mağazaya olan güvenlerini artırdığı,
- İnsanların yüzde 92’lik bölümünün, 5 yıldız üzerinden en az 4 yıldız oy almış işletmeleri tercih ediyor ya da etmeyi düşündüğü,
- Tüketicilerin yüzde 86’lık bir bölümünün ise firmayla ilgili olumsuz yorumların olması durumunda, o firmadan alışveriş yapma konusunda tereddüt ettikleri sonuçları çıkarılıyor [23].

Sosyal Medya bu konuda çok etkilidir. Kullanıcılar çoğunlukla Facebook, Twitter, Foursquare ve diğer sosyal medya kanallarında bir işletme hakkında okudukları yorumlardan etkilenebiliyorlar. Ayrıca forumlar ve sözlükler de büyük ölçüde belirleyici oluyor. Örnek olarak ülkemizde popüler olan Ekşi Sözlük’te “x rezaleti” formatında açılan başlıklar bir işletme ve müşteri arasında yaşanan olumsuz bir olayı anlatmak için sık sık başvurulmuş bir yöntem olarak dikkat çekmektedir.

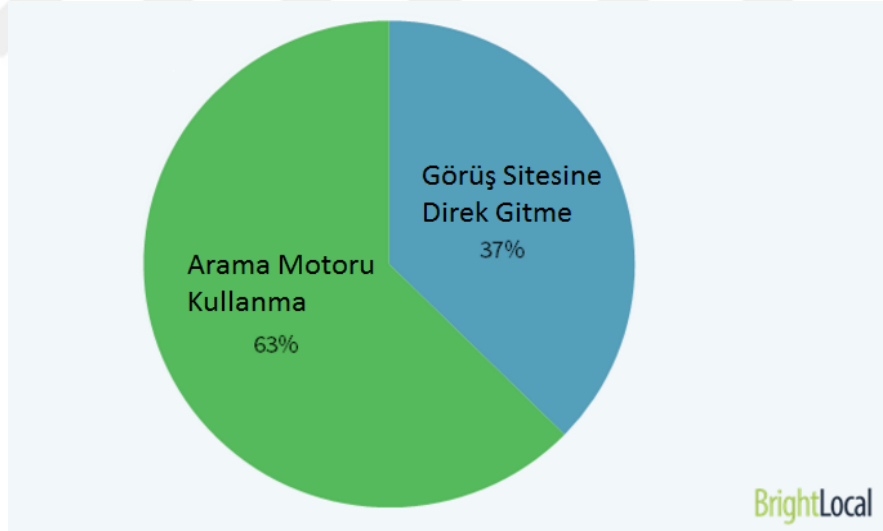
Bir diğer araştırma olan BrightLocal’in 2016 anketinde, kullanıcıların çevrimiçi yorumları nasıl okuduğu ve kullandığı ile ilgili birçok analiz yapılıyor.

Şekil 2.1’de görülen analizde giderek çevrimiçi yorumları düzenli okuyan kullanıcıların arttığını görebiliyoruz. Kullanıcıların %91’i çevrimiçi yorumları okuyor ve sadece %9’luk kesim okumuyor.



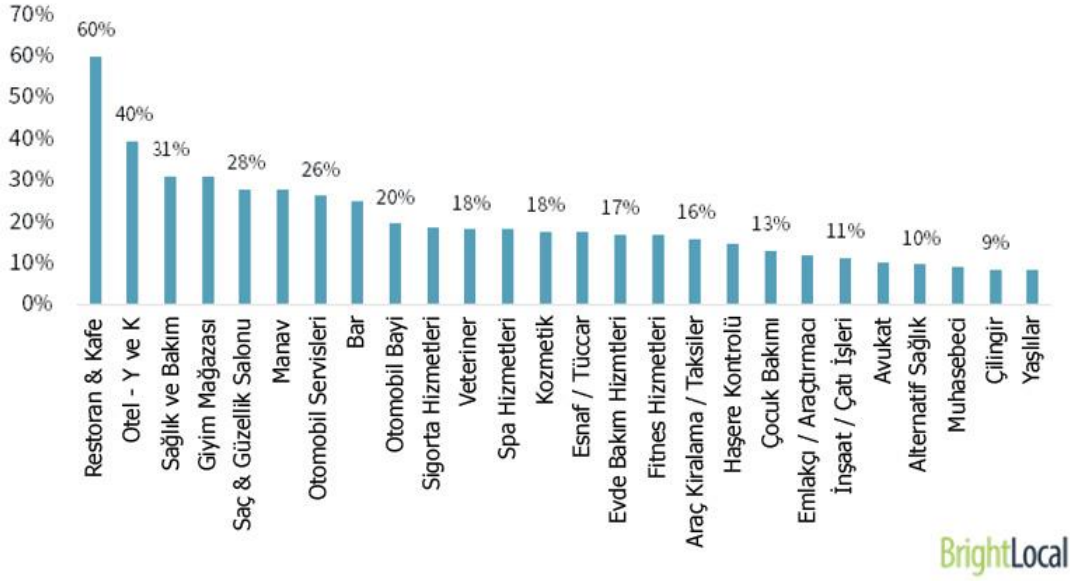
Şekil 2.1: İşletmelerin İyi veya Kötü Olduğunu Öğrenmek İçin Çevrimiçi Yorum Okuyor Musunuz? [24]

Şekil 2.2’de yapılan analizde de kullanıcıların bu yorumlara hangi şekilde ulaştığını anlamaya çalışmaktadır. %63 Google, Bing gibi arama motoru üzerinden ulaşırken, %37’lik kısım Foursquare, TripAdvisor gibi görüş sitelerinden istediği bilgiye ulaşmakta olduğu görülüyor.



Şekil 2.2: Çevrimiçi Yorumlara Nereden Ulaşırsınız? [24]

Çevrimiçi yorumların hangi alanlarda yapıldığını gösteren Şekil 2.3’te ise %60’lık payla en çok restoran ve kafeler, ardından %40’lık payla Oteller/Yatak ve Kahvaltı konusu için yapılan yorumlar incelenmektedir. Bu alanda insanların yaptığı araştırmalarda Foursquare’ın de büyük bir paya sahip olduğu söylenebilir.



Şekil 2.3: Hangi Alanda Çevrimiçi Yorumları Okursunuz? [24]

Kullanıcıların içerik üretmesinin hem avantajları hem de dezavantajları vardır. Kullanıcıların içerik üretmelerinin avantajları;

- Uygulama sitesine içeriklerin kesintisiz eklenmesi hem bu sitenin güncel olmasını hem de içeriğin gelişerek zenginleşmesini sağlar.
- Kullanıcı yorumlarının olması veya sık sorulan sorulara kullanıcılar tarafından cevaplar yazılması ile aradıkları bir konu sebebiyle sayfayı ziyaret eden kullanıcılar, sorularının cevaplarını ve ürünler hakkında ek bilgilere ulaşabilirler.
- Yorumların gerçek kullanıcılar tarafından oluşturulması, insanların ürünler hakkında yazılan bilgileri daha cazip bulmasını sağlar.
- Kullanıcılar, kendi kullanıcı deneyimlerini paylaşarak, bir yandan ilgili ürün veya işletme adına reklam yapmış olurlar. Böylelikle sayfayı ziyaret eden başka kullanıcılar fikir sahibi olabilirler.
- Kullanıcılar tarafından içerik üretilmesi, kullanıcıların site içinde geçirdiği sürenin artmasına yardımcı olur.
- Kullanıcılar içerik üretmek için siteyle daha çok etkileşime girmeye başlayacak, bunun bir getirisi olarak da bunları okuyan kullanıcılar da site ile etkileşime geçmeye başlayabilir.
- Kullanıcı tarafından oluşturulan içerikler (User generated content) sayesinde kullanıcılar hakkında daha fazla bilgi toplanabilir ve bu bilgiler ışığında

kullanıcılar hakkında birçok bilgi elde edilebilir. Bir yandan da kullanıcıların alışveriş alışkanlıkları incelenebilir ve bu bilgiler incelenerek site ve ürünler üzerinde geliştirmeler yapılabilir.

- Kullanıcılar içerikleri oluşturduğu için bu durum markaya ekstra bir maliyet çıkartmaz [25].

Kullanıcıların içerik üretmelerinin dezavantajları ise şöyledir;

- Kullanıcıların oluşturdukları içeriklerin yönetimi ve kontrolü fazladan bir iş gücü gerektirebilir. Üretilen bu içeriklerin yazım dilinin ve içeriğinin düzenli olarak kontrol edilmesi gerektiği için markanın üzerine bir yükümlülük binebilir.
- Kullanıcılar iyi yorumlar ile markayı veya ürünü yükseltebilecekleri gibi markanın ya da ürünün prestijini olumsuz yönde etkileyebilecek içerikler de yazabilirler.
- Yönetimi yapılmayan bu tip kötüleyen içeriklerin yayınlanması başka kullanıcıların da kötü bir şekilde etkilenmesine neden olabilir.
- Kullanıcılar diğer kullanıcıları yanıltabilecek sahte yorumlar bırakabilirler ve bu diğer kullanıcıları olumsuz etkileyebilir.
- İçerik girişinin moderasyon yapılmadan yapılması, spam gibi görünen birçok yorum veya mesajın görünmesine neden olabileceği için sayfayı ziyaret eden kullanıcılar bu tür mesajlardan rahatsız olabilir [25].

Alışveriş sırasında veya öncesinde yorumlara bakarak ürün veya hizmeti satın alan tüketici sayısı günümüzde oldukça artmıştır. Bu durum diğer kullanıcılara tecrübelerden yararlanmak adına büyük fayda sağlamaktadır. Yapılan araştırmalarda insanlar artık bir ürün satın almadan önce o ürünün kullanıcı yorumlarını incelemeye oldukça önem veriyor ve alışverişini bu bilgiler ışığında yapıyor. Dünya üzerinde bir ürün almadan önce yorumlara bakan müşteri sayısı %75 civarında ve bu oran giderek artmaktadır. Bu sebeple insanların alışverişteki bilinçli davranışları üretici veya satıcı firmaları da yakından ilgilendirmeye başlamıştır [26]. EMarketer'e göre, müşteri yorumları, potansiyel müşteriler için üreticiler tarafından sağlanan orijinal ürün açıklamasından 12 kat daha güvenilirdir [27].

Günümüzde problemler daha karmaşıklaşmakta, bir filmi izleyen iki farklı kişinin görüşleri arasındaki farkın ne kadarının film oyuncularından kaynaklandığı, filmin

yönetmeninin veya senaryo yazarının izleyicilerin görüşüne etkisi gibi daha karmaşık çıkarımlar artık sorgulanabilmektedir. Veya bir internet satış mağazasında ürünler hakkında yapılan yorumların ne kadarı satıcının etkisi ile ne kadarı kargo şirketinden ne kadarı ürünün kendisinden kaynaklanmaktadır gibi sorular artık birer çalışma alanı olarak belirmektedir [28].

Çalışmamızda İstanbul'daki Alışveriş Merkezleri kullanıcılarının (tüketicilerinin) Foursquare uygulamasına bıraktıkları yorumlar incelenmiştir. Foursquare denince akla check-in gelse de keşfet araması ile insanların, en iyi Latte'si olan kafe veya rakipsiz burgere sahip olan mekânlarda bıraktıkları yorumlar girişimciler için önemli bir varlıktır. Bu veriler bir mekân hakkında ipuçlarını vermesi veya fikir vermesinden çok, kullanıcı tarafından bir fiil oluşturulduğu için çok değerlidir. İnsanlar özellikle arkadaşları tarafından bırakılan yorumları daha güçlü bulmaktadır [29]. 2016'da açıklanan verilere göre Foursquare üzerinde 100 milyon mekân ve bu mekânlara ait 87 milyon yorum bulunmaktadır [30]. Uygulamaya her ay ortalama bir milyondan fazla yorum eklenmektedir [31].

Foursquare yorumları, diğer kullanıcıların ziyaret edilecek yerleri seçmesine yardımcı olmak için geribildirim olarak kullanılabilir. Örneğin, bir otel tavsiye eden veya bir restoranın hizmetini sevmeyen biri, başkalarını kendi tercihleriyle yönlendirir. Yorumlar, kullanıcılarla gerçek işletmeler arasındaki ilişkiyi besler, işletme sahiplerine değerli geribildirim sağlamanın yanı sıra gelecekteki ziyaretçileri çekmek için temel özellikler oluşturur. Yalnızca kullanıcının arkadaşlarıyla görülen check-in'lerin aksine, yorumlar herkes tarafından görülebilir ve bu nedenle çevrimiçi bilgi paylaşımı ve işletme pazarlamasını büyük ölçüde etkileyebilir.

Bu bilgi paylaşımı beraberinde yararsız spam içeriklerin oluşmasına da neden olur. Ancak Foursquare'in yorum beğenerek en popüler yorumlar oluşturması sayesinde bunun önüne geçilmektedir [32]. Foursquare'de başkaları tarafından en çok beğenilen yorumlar yorum listesinin en üst sırasında görünür. Bu sıralamanın arkasında en ilgili ve doğruluk payı yüksek olan yorumu en üste getirmek vardır. Hatta işletmeler diğer işletme mekânlarına yorumlar bırakarak şirketin markasını öne çıkarmaya yardımcı olmanın yanında otoritelerini de genişletebilirler [33].

3. LİTERATÜR ARAŞTIRMASI

3.1 Alışveriş Merkezleri Üzerine Yapılan Çalışmalar

Alışveriş Merkezleri hakkında tüm dünyada birçok alanda inceleme yapılmıştır. Türkiye metropollerinden biri olan Ankara’da, AVM’lerin hayatlara dokunması ile halk yaşamında yakın zamanda yaşanan dönüşümleri ortaya koymak için yapılan bir çalışmada, Bilkent AVM için yapılan gözlemler ve pilot çalışma sonrası, alan ve kullanıcı karakteristiği üzerine 20 soruluk bir anket hazırlanarak 427 veri sağlanmıştır. Verilerin değerlendirilmesi çapraz tablolama ve ki-kare yöntemi ile elde edilmiştir. Öğrencilerin (AVM’nin üniversitenin yanında olmasından) ve çalışan insanların AVM’yi en çok kullanan sınıflar olduğu ortaya konuluyor. Yapılan analizle; ziyaret süresi, meslek ile çok ilişkili çıkıyor. İnsanların alışverişi sevdiği ve bunun cinsiyetle önemli bir ilişkisinin olduğu, sosyalleşme deseninde yaşın çok önemli rol oynadığı, gençlerin kalabalığı daha çok sevdiği, yaş ile AVM’yi diğerlerinden farklı bulmanın ilişkili olduğu, güvenlik konusunda cinsiyetlerin aynı davrandığı ve AVM’yi güvenli bulduğu ortaya konuyor [34].

İncelenen bir çalışmada, akademik literatürdeki alışveriş merkezlerinin sosyal, ticari, mekansal ve yönetsel özellikleri ile ilgili hangi tartışma konularının öne çıktığı analizi yapılıyor. Çalışmanın sonucunda, Türkiye’deki alışveriş merkezleri ile ilgili çalışmalarda üzerinde çok fazla durulmamış olan 4 konu saptanıyor. Bunlar, gençlerin alışveriş merkezlerini tercih etmelerinde bu mekanların sosyal ve ticari özelliklerinin değerlendirilmesi; alışveriş merkezlerinde konfor koşullarının kullanıcı üzerindeki etkilerinin değerlendirilmesi; Türkiye’de ve Batılı ülkelerde alışveriş merkezlerindeki temalı mekan anlayışındaki farklılıkların değerlendirilmesi; değişen sosyo-ekonomik yapının ve tüketim ilişkilerinin alışveriş merkezlerinin mekan kurgusu üzerindeki etkilerinin değerlendirilmesi olarak çıkıyor [35].

Türkiye’de yapılan bir diğer çalışmada, AVM’si olmayan Burdur ile pek çok AVM’ye sahip Antalya ve Isparta tüketicilerinin AVM algıları, karşılaştırmalı olarak incelenmeye çalışılıyor. Araştırmaya, Antalya’dan 168, Burdur’dan 152 ve Isparta’dan 139 olmak üzere, toplam 459 kişi katılıyor. Yüz yüze görüşme yöntemiyle elde edilen

nicel veriler, ANOVA yöntemiyle analiz ediliyor. Ortaya çıkan bulgulara, Burdur tüketicilerinin AVM özlemi içerisinde olduğu ve çoğunlukla AVM'ler hakkında olumlu görüş belirttiği, Antalya tüketicilerinin 'küçük işletmeleri olumsuz etkilediği' ve 'tüketime yönlendirdiği' gibi olumsuz görüşleri daha çok desteklediği ve Isparta tüketicilerinin ise bazı görüşlerde Antalya, bazılarında da Burdur tüketicilerine yakın olduğu görülüyor [36].

AVM'lerin tüketici davranışlarını nasıl etkilediği de araştırılmıştır. Bu alanda, Alışveriş Merkezlerinin sahip olduğu lokasyon belki de en önemli konulardan biridir. İncelenen çalışmada alışveriş merkezi seçim probleminin İstanbul için modellenmesi yapılıyor. İş geliştirme yöneticileri, danışmanlar, akademisyenler ve çalışanlar olmak üzere 15 uzmanla yapılan detaylı görüşmeler sonucu en uygun yer alternatifleri değerlendiriliyor ve sonuç olarak Mecidiyeköy, Levent, Bakırköy, Beylikdüzü, Maltepe ve Kozyatağı olmak üzere 6 uygun yer belirleniyor. Bu yerleri değerlendirmek için de popülasyon karakteristiği, çevresel etmenler, rekabet derecesi, erişilebilirlik, ekonomiklik, toplam maliyet, esneklik ve çekicilik olarak 8 kriter ortaya koyuluyor. Çalışmanın sonucunda ekonomiklik ve toplam maliyet en önemli iki kriter olarak karşımıza çıkıyor. Yapılan AHP ve TOPSIS sonrası Kozyatağı ve Levent seçeneklerinin en iyi alternatifler olduğu çıkarımı yapılıyor [37].

Bir başka AVM lokasyon seçimi çalışmasında CBS (Coğrafi Bilgi Sistemi) uygulaması ile iş dünyasına yönelik yapılıyor. Sistemde dört konum sorunu için sorgular oluşturuluyor: (1) minimum mesafe, (2) maksimum talep kapsamı, (3) maksimum gelir seviyesi ve (4) en iyi merkez. Böylelikle AVM karar vericileri için basit ve doğru bir bilgi kaynağı oluşturma çalışması gerçekleşiyor [38].

İnsanların AVM'leri neden tercih ettiği en çok araştırılan konuların başında gelmektedir. Bu bağlamda Kanada'da yapılan bir başka çalışma, alışveriş merkezi imajını A la Ailawadi ve Keller'in mağaza imajı konsepti ile kavramsallaştırarak, AVM imajının ölçülmesinin geçerliliğini raporlamayı amaçlıyor. AVM trafiğini arttırmak, aynı zamanda AVM içinde yer alan mağaza trafiği ve satışını arttırmak anlamına geldiği için AVM yönetimi müşteri olmaya etki eden faktörleri tanımlamaya çalışıyor ve çeşitli strateji önerilerinin olduğu söyleniyor. AVM'lerin, hedef alışverişçiler için yararlı olan erişilebilirlik, atmosfer, fiyat/promosyon, çapraz kategori ve kategori içi çeşitlilik boyutlarını konumlandırması gerektiği üzerinde duruluyor. AVM imajı mağaza imajını içeren araştırmalara dayalı yapılıyor. Mağaza

imajı, alışverişçilerin mağazayı algılama konseptidir. 861 müşteri ile yapılan anket çalışması sırasında AVM yönetimi özel etkinlikler ve promosyonlar yapmaktan kaçınıyor. AVM imajı ölçümleri için boyutlara yönelik 20 soru, alışverişçi davranışlarına yönelik 6 soru oluşturulup, 7 ölçütlü değerlendirme aracı kullanılıyor. AVM imajındaki boyutlardaki değişimler, tüm AVM imajını etkiler. Çalışmada AVM imajı ölçümü ile üç değişkenin (AVM tutumunun %74 oranında, AVM müşterisi olmanın %33 oranında, Ağızdan ağıza iletişim %65 oranında) önemli ölçüde tahminlenebileceği sonucu çıkarılıyor [39].

Hindistan'da yapılan bir araştırmada ise, faydacı ve hedonik alışveriş değerlerinin kullanıldığı iki alışveriş merkezinden 288 yanıt toplanıyor ve yapısal eşitlik modeli kullanılıyor. Önerilen bir araştırma çerçevesi oluşturularak orta Hindistan Bhopal alışveriş merkezlerinin memnuniyetine, faydacı ve hedonik alışveriş değerlerinin etkisi araştırılıyor. Sonuç alışveriş değerlerinin müşteri memnuniyetini pozitif yönde etkilediğini destekleyen bir şekilde çıkıyor [40].

Yapılan başka bir araştırmada alışveriş merkezlerinin çekiciliğini nelerin sağladığı inceleniyor. Polonya'nın Wrocklaw kentinde bulunan 6 AVM için alışveriş merkezinin cazibesini (duygusal etki, bilişsel etki ve müşterinin ziyaret sıklığı) ölçmek için dokuz öge ve varsayımsal öngörücülerini ölçmek için 49 öge içeren bir anket uygulanıyor. Araştırma sonucunda alışveriş merkezlerinin çekiciliğinin temel nedeni atmosferi ve sosyal konumları olarak çıkıyor. Şaşırtıcı bir şekilde, alışveriş merkezinin daha gürültülü ve kalabalık olmasının çekiciliğini arttırdığı, diğer taraftan ticaretle alakalı özelliklerin, alışveriş merkezinin cazibesine nispeten az katkıda bulunduğu sonucuna ulaşıyor [41].

Dennis, Marshal ve Cockett bu çalışmalarında müşterilerin alışveriş merkezleri hakkındaki görüşlerinin alışveriş davranışlarına olan ilişkilerini araştırmışlardır. Çalışma 287 kişiyle gerçekleştirilmiştir. Katılımcılar temel demografik sorulara cevap verdikten sonra, kendilerine sunulan 38 özellik ile ilgili önem algılarını bildirmiş, seçilen iki adet AVM'yi bu iki kriter üzerinden değerlendirerek karşılaştırmalı sıralama oluşturmuşlardır. Ardından söz konusu iki AVM'nin uzaklığını yaklaşık olarak belirtmiş ve iki AVM için yaklaşık harcamalarını da bildirerek anketi sonlandırmışlardır. Anketten elde edilen verilerle (kriter değerlendirmeleri, uzaklık, harcama oranı) bir ağırlıklı ortalama modeli ile AVM'lerin çekicilik skoru oluşturulmuştur. Katılımcıların harcama verileri ve kriter seti en küçük kareler

regresyon analizi ile incelenmiş ve kriterlerin katılımcıların harcama kararına nasıl etki ettiği araştırılmıştır. Çalışma sonucunda farklı demografik segmentlerde (kadın-erkek, gelir seviyeleri, genç-yaşlı vs.) hangi kriterlerin AVM tercihlerinde önemli olduğu, hangi kriterlerin harcama kararını nasıl etkilediği ortaya çıkarılmıştır [42].

3.2 Foursquare Üzerine Yapılan Çalışmalar

Foursquare günümüzün en yaygın sosyal ağlarından biridir. Bu yaygınlık beraberinde birçok çalışmanın yapılmasını getirmiştir. Çalışmalardan birinde mikro-incelemeler olarak ele alınan yorumların gelecekteki popülerlik seviyesini tahmin sorunu üzerinde çalışılıyor. Foursquare yorumlarının popülerliğini nelerin etkilediğini, izleme süresinin ve hedef tahmin kesinliği ile, şehir ve mekan kategorisine dayalı modelin yararları değerlendiriliyor. Regresyon ve sınıflandırma modelleri kullanılıyor [43].

Başka bir çalışmada bireysel davranışlara ve grup tercihlerine göre etkili mekan önermesi konusu araştırılıyor. Makalede mekan önerme için kişilerin kısmen gözlenen mekan geçişleri ile bir metodoloji geliştiriliyor, tahmin ediliyor. Kullanıcıların demografik özelliklerine ve sıkça ziyaret edilen yerlerine ulaşıyor. Ardından, her bir kullanıcı için övgü alan kategorilere dayalı olarak iki yönlü bir grafik oluşturuluyor. Konumların kişiselleştirilmiş sıralamasını elde etmek için, konum yetkisi skorunun, grubun ziyaret sıklığına ve bazı mesafe kısıtlamalarına göre tekrar tekrar güncellenmesi için, bir mesafe ağırlıklı HITS algoritması önerilmiştir. Kategori öngörüsü ile önerilen iki aşamalı yaklaşım, kurum içi pürüzsüzleştirme yoluyla yer tahmin performansını artırmayı ve aynı zamanda karmaşıklığı azaltmayı amaçlıyor. Gerçek dünyadaki konuma dayalı sosyal paylaşım ağı verilerine dayanılarak elde edilen deneysel sonuçlar, önerilen yaklaşımın mevcut en son teknolojik yöntemleri büyük bir farkla yerine getirdiğini göstermektedir. HITS yöntemi, mesafe kısıtlama faktörünü ve aday konumlarını sıralamak için grup tabanlı popülerliği içermektedir [17].

Uygulamayı kullanan kişilerin özelliklerini anlamaya dair yapılan çalışmalardan birinde, web üzerinden katılan 174 gönüllü Foursquare kullanıcısının kişilik özellikleri ve denetim davranışları, "Büyük Beş" olarak ifade edilen kişilik özellikleri bakımından inceleniyor. Mekânlarda yaptıkları check-inlerden; vicdanlı olma, açıklık ve nevrozizm ile ilgili korelasyonlar bulunurken, diğer özelliklerle (dışadönüklülük, hoşluk) ilgili bir bağlantı bulunamıyor. Çalışma, kişilik özelliklerinin, LBSA

kullanımındaki bireysel farklılıkları ve ziyaret edilen yerlerin türünü açıklamasına yardımcı olduğu sonucuna varıyor [14].

Foursquare kullanıcılarının yorum, yapılacaklar ve yapıldılar ile ilgili özelliklerini nasıl kullandıkları da farklı bir çalışma ile analiz ediliyor. 1.6 milyon mekân ve yarım milyona yakın kullanıcıyı içeren veri seti kullanılarak, kullanıcılar dört gruba ayrılıyor. Bu özellikleri çok etkin kullananların çok ünlü markalar ve büyük işletmeler olduğu görülürken, mekanla ilişkisiz yorumlar yapan spam grubu da çıkarılıyor ve spamları tespit etmek ve kaldırmak için etkili stratejiler tasarlamaya yönelik yapılabilecekler açıklanıyor [18].

Yapılan literatür incelemesinde Türkçe Foursquare yorumlarının incelenmesine dair bir çalışma olmadığı görülmüştür.

3.3 Veri Madenciliği

3.3.1 Duygu analizi üzerine yapılan çalışmalar

Bir metin biriminin duygu içeriğini Doğal Dil İşleme (DDİ), istatistik veya makine öğrenme yöntemlerini kullanarak ayıklamak, tanımlamak veya başka türlü karakterize etmeye “duygu analizi” denir [44]. "Duygu analizi" nden bahseden çok sayıda çalışma, kutupsallıklarına (olumlu ya da olumsuz) göre incelemeleri sınıflandırmanın spesifik uygulaması üzerine yoğunlaşmaktadır; bu durum, bazı yazarların, kısmen bu dar tanımlanmış göreve atıfta bulunduğunu önermelerine neden olmuş gibi görünmektedir. Bununla birlikte, günümüzde birçok kişi terimi daha geniş anlamda metindeki görüş, düşünce ve öznelğin hesaplamalı olarak ele alınması anlamında kullanmaktadır [45].

Bilgi teknolojilerindeki hızlı ilerlemeler ile birlikte, İnternet en önemli bilgi kaynaklarından biri olmuştur. İnternet ortamındaki mevcut veri, hükümetler, iş organizasyonları ve bireysel karar vericiler için oldukça değerli bir bilgi kaynağı olarak işlev görmektedir. Duygu analizi ile politikalar, ürünler ya da hizmetlere ait toplumsal görüşler saptanabilir. Duygu analizi ile yapısal hale getirilen bilgi, karar destek sistemleri ve bireysel karar vericiler için önemli bir veri kaynağı haline gelir. Duygu analizi ile belirli bir ürüne ya da servise yönelik olumlu ve olumsuz yönler belirlenerek, ürün ya da servise ilişkin müşteri memnuniyeti artırılabilir [46].

Duygu analizi, önemli bir güncel araştırma alanıdır. Duygu analizi, belge seviyesi, tümce-seviyesi ve varlık/özellik seviyesi duygu analizi olmak üzere üç temel seviyede incelenmektedir. Belge seviyesi duygu analizinde, tüm metin belgesinin duygu yönünün belirlenmesi amaçlanır. Tümce seviyesi duygu analizinde, belirli bir tümcenin duygu yönü; varlık/özellik seviyesi duygu analizinde ise belirli bir varlığın (ürünün) özelliklerine ilişkin duygu yönünün belirlenmesi amaçlanır.

Duygu analizi çalışmalarında kullanılan teknikler temel olarak makine öğrenmesine dayalı yöntemler ve sözlüğe dayalı yöntemler olmak üzere iki temel grup altında incelenmektedir. Makine öğrenmesine dayalı duygu analizi yöntemlerinde, belirli bir makine öğrenmesi sınıflandırıcısı, görüş kutbu etiketlenmiş veri seti ile eğitilerek, bir sınıflandırma modeli oluşturulmaktadır. Bu sınıflandırma modeli, daha sonraki örneklerin sınıflandırılması amacıyla kullanılmaktadır. Sözlüğe dayalı duygu analizinde ise görüş sözcükleri içeren bir sözlük oluşturularak, duygu analizi bu sözlüğe dayalı olarak gerçekleştirilmektedir.

Foursquare yorumlarının kullanıldığı bir çalışmada, denetimli ve denetimsiz metodlar kullanılarak yorumların polarite (sentiment) analizi yapılıyor. Sözlüğe dayalı (denetimsiz) metodun diğer denetimli modeller kadar iyi sonuçlar verdiği ortaya koyuluyor. Denetlenmeyen sözlük temelli yöntemin, her iki veri kümesindeki en iyi denetlenen yöntemlere (NaiveBayes ve Support Vector Machine) kıyasla daha iyi veya istatistiksel olarak bağlanmış sonuçlar üretir. Karma yöntem, daha da gelişme sağlamaz. Denetlenmeyen yöntem, en iyi denetlenen yöntem üzerinden % 4.8'e kadar geliştirir. Geri çağırma kazanımları 11.9 'a ulaşır. Bu nedenle, odak noktası olumlu ipuçları olduğunda bu yöntem kullanılmalıdır. Odaklanma olumsuz ipuçları üzerindeyse, en iyi yöntem denetlenen olanlar, özellikle NaiveBayes ve Support Vector Machine' dir. Duygusal temsili veri kümesinde, bu denetlenen yöntemler F1'de denetimsiz çözüme göre % 10.8'e kadar kazançlar üretir. Geri çağırma açısından kazançlar daha da etkileyici, bu veri setinde % 36' ya ulaşır. Tüm yöntemler, muhtemelen daha yüksek gürültü seviyesine (örn. Alay) ve ifadelerin üzerinden otomatik etiketlemedeki belirsizliğe bağlı olarak, elle etiketlenmiş veri kümesindeki (yüzde 19.5'e kadar farklılıklar) tüm metriklerde daha iyi performans gösterir [16].

İnternette yer alan platformlar veri açısından çok önemlidir. İnsanların kullandığı bir ürün veya gittiği bir yer için bıraktığı yorumlar veya mikro incelemeler birçok tezin konusu olmuştur. Restoran incelemeleri üzerine yapılan bir tezde, duygu analizi için

mevcut sözlüğün yeterli olmadığını yeni bir sentiks sözlük oluşturulması gerektiği önerilmektedir. İncelenen belgeyi olumlu bir düşünce olarak sınıflandırırken ve denetlenen öğrenme algoritmasını kullanarak olumsuz bir düşünce olarak sınıflandırırken, olumlu sınıflandırma doğruluğunun negatif sınıflandırma doğruluğundan yaklaşık % 10 daha yüksek görünme eğilimi vardır. Bu iki sınıfın doğrulukları ortalama bir değer olarak ifade edildiğinde ortalama doğruluğu azaltma problemi yaratır. Bu problemi azaltmak için geliştirilmiş bir Naïve Bayes algoritması önerilmiştir. Deney sonucunda, bu algoritma kullanıldığında ve bir unigrams + bigram kullanıldığında, pozitif doğruluk ile negatif doğruluk arasındaki boşluk, orijinal Naive Bayes kullanıldığında kıyasla %3,1'e kadar daraltılmış ve SVM kullanımıyla kıyaslandığında % 28.5 boşluk daralabilmiştir [47].

Bir başka çalışmada VIKOR ve duygu analizi birleştirilerek mobil hizmetler için müşteri memnuniyetinin ölçümü için yeni bir çerçeve geliştiriliyor. Çok kriterli karar verme (MKYM) yaklaşımının uzlaşmacı bir sıralama metodu olan VIKOR ile mobil hizmetler için müşteri memnuniyeti, aynı anda maksimum grup yararını ve bireysel üzüntüsü dikkate alan bir duygu analizi şeması ile ölçülebilir. Önerilen çerçeve esas olarak iki aşamadan oluşmaktadır: veri toplamak-ön işleme koymak ve müşteri memnuniyetini ölçmek. Birincisi, veri toplama ve ön işleme aşaması, metin incelemesi, niteliklerin ve duygu sözcüklerinin müşteri inceleme tabanlı sözlüklerini derlemek için kullanılır. Daha sonra, duygu analizi kullanılarak, her bir mobil servis için nitelikler için duygu puanı hesaplanır. İkinci aşamada VIKOR kullanılarak müşteri memnuniyet seviyeleri ölçülür. Örneklem amacıyla, mobil uygulama servislerinin müşteri incelemeleri üzerine deneysel bir vaka çalışması yapılmıştır [48].

Foursquare yorumları, duygu analizi yapmak için sıkça kullanılmaktadır. Bu çalışmalardan birinde, Naive Bayes yöntemi ile konut, alışveriş ve yemek lokasyonu duygularını belirlemek amaçlanmıştır. Araştırma nesnesi olarak, Endonezya'nın Doğu Nusa Tenggara kentinde bulunan Kupang Şhrinde bulunan konaklama, alışveriş ve yemek mekânlarının Foursquare yorumları kullanılıyor. Olasılık sınıflandırıcı modelini üretmek için Naive Bayes kullanılıyor. Bu model, konaklama, alışveriş ve yemek mekânı ile ilgili Foursquare'in ipuçlarını belirlemek için kullanılıyordu. Test setinin doğruluğu % 66,22 olarak bulunuyor. Çıkan sonuçlar, Swiss Belinn Hotel, Aston Hotelin yorumlara göre en iyi seçenek olduğunu, On The Rock Hotel'in ardından T-More Hotel and Lounge'un bu otelleri takip ettiği çıkarılıyor. KFC

Flobamora Alışveriş Merkezi, Palapa Resto ve Kafe, Dapur Nekamese mutfak için en iyi seçenektir. Sonuç, Flobamora Alışveriş Merkezi'nin en iyi alışveriş yeri olduğunu göstermiş. Veriler manuel olarak duygu bakımından etiketleniyor [15].

Duygu analizi, bir başka çalışmada Twitter'dan elde edilen Türkçe mesajlar üzerinde iki kategorili olarak çalışılmıştır. Duygu analizi bir metin sınıflandırma problemi olarak düşünülmüş; duygu analizi tekniklerinin yanı sıra klasik metin sınıflandırma teknikleri de kullanılmıştır. Twitter mesajlarında gözlenen baskın duygunun otomatik olarak tespitinde ise makine öğrenmesi yöntemleri kullanılmıştır. Hem metin sınıflandırma hem de duygu analizi deneylerinin gerçekleştirildiği bu çalışmada, asıl hedef ise duygu önişleme, etiketleme, sınıflandırma ve benzerlik yöntemlerinin etkisini incelemektir. Ayrıca topik bilgisine dayalı etiketleme yöntemi önerilmiş ve en yüksek %92,50 oranında başarı elde edilmiştir. Böylece dil işleme ile ilgili işlemler uygulanmadan duygu analizi başarısı önceki çalışmalara kıyasla daha yüksek elde edilebilmiştir. Bunun yanı sıra, metin sınıflandırma ve duygu analizi süreçlerini otomatikleştirme amacıyla hem Türkçe hem de İngilizce metinsel içerikleri işleyebilen bir yazılım prototipi (OMESIS) geliştirilmiştir [49].

Bir çalışmada, Twitter kullanıcılarının havayolu ulaşımı ile ilgili yorumları derlenerek duygu analizi çalışması yapılmıştır. Kullanıcı yorumları, birçok sosyal medya uygulamasında olduğu gibi Twitter'ın da sunmuş olduğu API (Application Programing Interfaces-Uygulama Programlama Arayüzleri) hizmeti vasıtasıyla java tabanlı program kullanılarak Nisan-Mayıs 2016 tarihleri arasında alınmıştır. Elde edilen 8672 kullanıcı yorumu olumlu, nötr ve olumsuz etiketlerle ayrıştırılmıştır. Elde edilen etiketler etiket bulutunda toplanmış ve sonuçlar Makine Öğrenmesi Yöntemi ve SMO sınıflandırmasında standart ve normalize Kernel Polinomları ile analiz edilmiştir. Analiz sonucu başarı ölçütleri incelendiğinde ise standart Kernel Polinomu'nun SMO algoritması içinde daha iyi bir performans gösterdiği; ancak genel olarak sınıflandırma başarısının düşük çıktığı görülmektedir [50].

3.3.2 Metin madenciliği-konu modelleme üzerine yapılan çalışmalar

İncelenen bir doktora tezinde internette yer alan ürün yorumlarından olumlu ve olumsuz görüşlerin sınıflandırılması sonucu kullanıcı ya da tüketiciye özet bilgi sunmayı hedefleyen bir yazılım geliştirilmiştir. Yazılım, geniş dokümantasyona sahip ve tam nesneye dayalı programlama dili olan Java'da gerçekleştirilmiş. Bu veriler

önceden tespit edilmiş ürün özelliklerine göre sorgulandığında tüketicilerin yaptığı yorumlardan memnuniyet derecesini gösteren sayısal hale dönüştürülmüş oranlar elde edilmiş. Burada sistem performansına en çok etki eden faktörün ağaç yapısındaki bağlanma durumunun olduğu görülmüş [51].

Metin madenciliği toplumsal sorunları çözmek için birçok çalışmada yer almıştır. Bu çalışmalardan birinde çevresel sürdürülebilirlik göstergelerini metin bazlı objektif bilgileri ile adreslenmiş ve bu konu ile ilgili algı yukarıdan aşağıya yaklaşımıyla araştırılmış. Mevcut literatürden 55 gösterge çıkarımı yapılıyor ve şirketlere bu göstergelerin kullanım durumları, yararları hakkında anket çalışması gerçekleştiriliyor. Toplanan göstergelerin metin açıklamaları, metin madenciliği tekniği olan ilişkili konu modelleme ile gizli başlıkları çıkarmak üzere modelleniyor. Göstergeler için 5 kategori çıkarımı yapılıyor. Ayrıca, türetilen kategorilerdeki göstergelerin kullanım durumu ve yarar seviyeleri analiz ediliyor. Gösterge yarar seviyeleri ile şirket özellikleri arasındaki muhtemel ilişkiler de lojistik regresyon ile tanımlanıyor [52].

Bir başka çalışmada hastanelerdeki acil servislerde yaşanan aşırı yoğunluk sorununa metin madenciliği ile tahmin modeli geliştirerek çözüm bulunmaya çalışılmıştır. Acil servise gelen hastaların kısa dönem yatak talebinin erken tahmin edilmesiyle bu aşırı yoğunluğun önüne geçileceği varsayılmıştır. Kapasiteyi yönetmek ve kaynak tahsisi yapabilmek için tahmin yöntemleri kullanılıyor. Bu çözümü bulmaya çalışırken hastaların gözlem altındaki tüm klinik kayıtları ile farklı metin madenciliği yaklaşımları üzerinde denemeler yapılıyor ve en iyi performansın Nu-Support Vector Machine olduğu çıkarılıyor [53].

Bazı çalışmalar da metin madenciliği alanına katkı yapılmıştır. Bu çalışmalardan birinin amacı, farklı veri kaynakları, metin incelemesi yaklaşımları ve öngörü yöntemleri göz önüne alındığında, metin madenciliğinin öngörü için potansiyelini keşfetmektir. Metin incelemesi ve öngörü birleştirmenin potansiyelini açıkladıktan sonra, yol haritalaması ve senaryo geliştirme için örnekler verilmiştir. Sonuçlar metin incelemesinin, öngörünün bilgi tabanını genişleterek ortaya çıkan konuların ve teknolojilerin tespit edilmesini ve incelenmesini kolaylaştırdığını göstermiş [54].

Çalışmamızda yaptığımız gibi, müşteri yorumları üzerinden metin madenciliği çalışmaları da yapılmıştır. Bu çalışmalardan birinde Paris' te akıllı turizm yapmak

adına seyahat edenlerin çevrimiçi yorumları için duygu analizi yapılmış. 19,835 müşteri yorumu ilgili seyahat sitelerinden toplanmış ve 14 grup altında toplanmış. Ardından bu kategorilerle ilişkili olan yorumların duygu analizi yapılarak hangi kategorilerde turistlerin nasıl hissettiği bilgisi çıkarılıyor [55].

Kimlik hırsızlığı üzerine yapılan metin madenciliği çalışmasında, hedef grubun kimler olduğu, hangi alan üzerinden hırsızlığın olduğu, finansal kaybı ve hangi lokasyonlarda olduğu analiz ediliyor. Analiz için 3,500 kimlik çalma haber kaydı kullanılıyor. Bu kayıtlar önışlemden geçiriliyor ve gereksiz bilgilerden arındırılıyor [56].

Sosyal medya ürün ve hizmetleri pazarlamada büyük önem taşımaktadır. Yapılan çalışmada sayfalarda yer alan “beğen”, “oyla” ve benzeri tuşlar sayesinde hem potansiyel müşteriler ürün ve hizmet hakkında bilgi edinmiş olmakta hem de firmalar, markalarının hangi demografik özellikteki kişiler tarafından ne kadar beğenildiğini gözlemleyebilmektedirler. Youtube gibi görsel içerik merkezli sosyal platformlar ise firmalara kendi içeriklerini yayınlama ve müşteri yorumlarını birinci ağızdan alma fırsatı tanımaktadır. Televizyonda yayınlanan bir reklamın müşteri yorumlarını anketlerle toplamaktan çok daha hızlı ve masrafsız olduğu belirtilmiştir. Youtube yorumlarının kutupluluğunu, yorum yapan kullanıcıların ilgi alanlarıyla karşılaştırmaya çalışmışlardır. İlgi alanlarının saptanmasında öncelik “kullanıcının kendi profilinde belirttiği” olarak belirlenmiş, eğer profilde böyle bir öncelik belirtilmemişse de kullanıcının izlemiş olduğu video geçmişi ve yorumlarının analizi kaynak olarak alınmıştır. Yani duygu analizi hem yorumların değerlendirilmesinde hem de kullanıcının analiz edilmesinde kullanılmıştır [57].

Twitter ve Facebook gibi sosyal ağlardan edinilen bilgilerin pazarlama araştırmaları şirketleri, kamuoyu organizasyonları ve diğer metin araştırma kuruluşları için son derece değerli olduğu gösterilmektedir. Bununla birlikte, Web metinleri hem sözlü hem de sözdizimsel düzeyde önemli problemleri temsil ettiği için gürültülü olarak sınıflandırılmıştır. Bir araştırmada tüketicilerin Nokia, T-Mobile, IBM ve DHL gibi tanınmış markalara yönelik duygularını değerlendirmek için rasgele bir 3,516 tweet örneği kullanılıyor. Analiz yapmak için bilinen oryantasyona sahip 6,800 civarında tohum sıfatını içeren önceden tanımlanmış bir sözlük kullanılıyor. Bulgular, birçok ünlü markaya karşı genel olarak olumlu bir tüketici duyarlılığının olduğunu gösteriyor. Markaların tweetlerini analiz etmek için niteliksel ve niceliksel bir metodoloji

kullanan bu çalışma, kozmopolit markalara yönelik tutumlar üzerine tartışmaya genişlik ve derinlik katmaktadır [58].

Çağrı merkezi sektöründe yapılan bir çalışmada ise sesten metne dönüştürülen görüşme kayıtları için Mallet üzerinde metin madenciliği yapılmıştır. Bir yandan görüşmenin içeriğinin duygusu da değerlendirilmiştir. Yapılan metin madenciliği ile metinlerden yeni özellikler çıkarılmış ve bu özelliklerden yararlanılarak sınıflandırma ve regresyon ile içeriği değerlendirmeyi sağlayacak tahmin modelleri oluşturulmuştur [59].

İncelenen bir tezde konu modelleme yaklaşımlarından Gizli Anlamsal Analiz, Olasılıksal Gizli Anlamsal Analiz ve Gizli Dirichlet Tahsisi yöntemleri incelenerek, bu yöntemlerde Türkçe metin üzerinde uygulamalar yapılıyor. Uygulama MATLAB ve R istatistiksel programlama dilleri kullanılarak yapılıyor ve yöntemlerin performansları değerlendiriliyor [60].

Metin madenciliği içerisinde yer alan ve yine bir çok çalışmanın konusu olan konu modelleme ile hem içeriğin geliştirilmesine yönelik hem de farklı alanlarda incelemeler yapıldığı görülmüştür. Yöntemi geliştirmek için yapılan çalışmalardan birinde, veri akımlarındaki konuların zamanla evrimini modellemek için ayrı veriler için çevrimiçi bir konu modeli geliştiriliyor. LDA modeli genişletilerek çevrimiçi LDA modeli geliştiriliyor. Modelin fikri, önceki verileri erişmeye gerek duymadan yeni veri akışından çıkarılan bilgilere göre mevcut modeli aşamalı olarak güncellemektir. Önerilen yaklaşımın dinamikleri aynı zamanda zaman içindeki konuları izlemek ve ortaya çıkan konuları gerçek zamanlı olarak tespit etmek için etkili bir araç sağlamaktadır. Bu modelin küçük belgelerde LDA'dan daha iyi performans gösterdiği söyleniyor [61].

İncelenen bir makalede de konu modelleme kapsamında yer alan iki kategori inceleniyor. İlk kategoride Gizli Anlamsal Analiz, Olasılıksal Gizli Anlamsal Analiz, Gizli Dirichlet Tahsisi ve İlişkili Konu modelleme olarak dört yöntem yer alıyor. İkinci kategoride ise konu değerlendirme modelleri yer alıyor. Konu modelleme yöntemlerinin özellikleri, karakteristikleri ve limitleri konusunda bilgiler veriliyor ve karşılaştırmalı olarak üst düzey bir bakış açısıyla görülebilmesi sağlanıyor [62].

Çalışmada Gizli Dirichlet Tahsisine dayalı konu modelleme yapılmış, makine öğrenmesi sınıflandırıcılarının duygu analizindeki tahmin etme başarımları

değerlendirilmiştir. Türkçe Twitter veri seti, 5300'ü negatif ve 5300'ü pozitif olmak üzere toplam 10,600 Twitter mesajı içermektedir. Çalışmanın sonucunda konu sayısı 50 iken Naive Bayes algoritmasının en yüksek başarıyı sağladığı görülüyor [46].

Konu modellemenin karşılaştığı zorluklarla ilgili çalışmalar da yapılmıştır. Bu zorluklardan biri internet üzerinde her yerde bulunabilen kısa metinlerden konu çıkarmaktır. Bu konunun zorluğunun kısa metinlerde kelime tekrarlarının az olmasından ve LDA gibi geleneksel algoritmaların şiddetli veri seyrekliği çekmesinden kaynaklandığı açıklanıyor. Çalışmada orjinal dokümanlardan daha büyük sahte dokümanlar oluşturularak kısa metinlerin konu modellemesinin yapılabilmesi için bir çerçeve oluşturmuşlardır. Bu çerçevenin iki örneği sunuluyor ve bunların etkililikleri var olan yöntemlerle karşılaştırmalı olarak gösteriliyor [63].

Konu modelleme ayrıca araştırmacıların hangi konularda araştırma yapması gerektiğine de ışık tutar. Buna yönelik yapılan çalışmalardan birinde hidroelektrik araştırmalarının bibliyometri keşfi temelli konu modelleme yapılıyor. 1994-2013 yılları arasında yayınlanan hidroelektrik ile çok ilgili 1726 makale R yazılım paketi ile analiz ediliyor ve 29 konu modeli çıkarılıyor. Hidroelektrik sonrası inşaat konularının, inşaat teknolojisinin kendisinden daha çok bilim insanları için cazip olduğunu ve hidroelektrik araştırmalarının disiplinlerarası bir eğilimi olduğu sonucuna ulaşıyor [64].

Bu çalışma akademik araştırmada kişisel bilgilerin mahremiyeti ile ilgili eğilimleri incelemektedir. Scopus DB kullanarak, 1972 ve Ağustos 2015 yılları arasında yayınlanan gazete makaleleri, incelemeler, kitap bölümleri, konferans kağıtları ve çalışma kağıtlarını kapsayan 2,356 belge inceleniyor ve konuların belirlenmesi için çıkarılan belgelerin özetlerine Latent Dirichlet Allocation (LDA) uygulanıyor. Tüm belgelerde keşfedilen konular içinde ağırlıklı olarak teknoloji üzerine odaklanıldığı, Facebook gizliliğinin ve çevrimiçi sosyal ağların öne çıkan diğer konular olduğu bulunuyor. Bunun aksine, gazetede çıkan makalelerin hem e-ticaret hem de sağlık hizmetleri üzerinde daha fazla durduğu gözlemleniyor. Bu sonuçlar, kişisel bilgi gizliliği alanında bir araştırma boşluğunu belirtir ve gelecek araştırmalar için bir yön verir [65].

Yapılan başka bir çalışma ise ulaşım üzerine yapılıyor. 22 lider ulaşım dergisinden 1990-2015 yılları arasında yayınlanan 17,263 makale analiz ediliyor. Mallet paketinde

50 konu üzerinden LDA modeli uygulanıyor ve sonuç olarak sürdürülebilirlik, seyahat davranışı gibi konuların zamanla popüler hale geldiği, hangi alanlarda daha az araştırma yapıldığı gözler önüne seriliyor. Bu çalışma, akademik topluluklar için gelecek çalışmalarına yönelik yararlı bilgiler içeriyor [66].

Pratisyen hekimler için klinik otomatik özetleme sistemine entegre edilebilen semantik olarak tutarlı sözcük kombinasyonlarını (konular) keşfetmek için olasılıksal konu modelleme kullanılıyor. Klinik raporlar, Mallet üzerinde LDA modeli kurularak 50, 100 ve 150 konu için analiz ediliyor. Ve çıkan konular pratisyen hekimler ve öğrenciler tarafından manuel değerlendiriliyor. Öğrenciler ve pratisyen hekimler arasındaki performans farklılıkları her bir görev için Mann-Whitney U testleri kullanılarak test edilirken, Eşleştirilmiş Örnekler için Wilcoxon İşaretli Sıra Testleri farklı konu modelleri arasındaki farkları değerlendirmek için kullanılmıştır. 150 konu modeli en iyi log olabilirliğini üretirken, katılımcılar 100 konu modelinin öğrendiklerini içermeyen kelimelerin belirlenmesinde en doğru, 100 konunun kullanılan veri seti için keşfedilen semantik temaların daha iyi nispi parçacıklarını sağladığını düşündürmüştür. Modeller, belgelerin içeriğini temsil etme yetenekleri bakımından karşılaştırılabilir niteliktedir. Pratisyen hekimlerin her iki görevde öğrencilerden belirgin biçimde daha iyi performans gösterdiği sonucu çıkarılıyor [67].

Yiyecek İlaç İdaresi (Yİİ) onaylı ilaç etiketleri, olumsuz ilaç reaksiyonlarından, ilaç etkinliğine, risk-fayda düşüncesine ve daha fazlasına kadar geniş bir bilgi dizisi içeriyor ve bu etiketlerin üzerindeki bilgiler özgür metinlerden oluştuğu için karşılaştırmak/ analiz etmek zor bir durum haline gelmiştir. Çalışmada 79 Yİİ onaylı ilaç etiketi kullanılmış ve LDA ile konu modellemesi 100 konu üzerinden uygulanmıştır. Sonuçlar, konulara göre gruplandırılmış ilaçların, aynı güvenlik endişeleri ve / veya terapötik kullanımlarla istatistiksel olarak anlamlı ($p < 0.05$) ilişkili olduğunu göstermektedir. Konu modellenmesinin, ilaç güvenliği, farmakovijilans ve ilaç yeniden konumlandırılması için metinsel bir veritabanından bilgi almak, organize etmek ve entegre ederek biyomedikal dokümanları madenciliği için umut verici bir denetlenmeyen öğrenme tekniği olduğu sonucuna varılıyor [68].



4. METODOLOJİ

4.1 Veri Madenciliği

Pek çok alanda yaygın olarak kullanılan Veri Madenciliği, Türkiye’de ve dünyada özellikle son yıllarda oldukça önem kazanmıştır. Veri madenciliği, büyük miktardaki veriler içinden istenilen bilgiye ulaşma işidir. Diğer bir deyişle, veri madenciliği tek başına bir şey ifade etmeyen veriler içindeki gizli örüntüleri ve ilişkileri ortaya çıkarmak için istatistik, yapay zekâ ve makine öğrenmesi gibi yöntemlerin ileri veri çözümleme araçlarıyla kullanılmasını kapsayan süreçler topluluğudur. Veri madenciliği uygulamaları başta pazarlama, bankacılık, tıp, mühendislik, endüstri, borsa analizleri ve ulusal güvenlik alanlarında kullanılmaktadır [69].

Veri Madenciliği çok büyük veri yığınlarından kritik bilgileri elde etmeyi sağlar. Böylelikle normal şartlar altında uzun zaman süren araştırmalarla doğruluğu kesin olmayacak şekilde elde edilen bilgi veri madenciliği ile kısa sürede ve kesin olarak elde edilir. Elde edilen bu bilgi objektif değerlendirmeler yapılmasında ya da stratejik kararlar almada kullanılır. Bu bilgiler kurumsal veri kaynaklarının iyi analiz edilmesine ve iş dünyasındaki yaklaşımlara ilişkin tahminlerde bulunulmasına yardımcı olur. Kısaca veri madenciliği sayesinde şirketler stratejik adımlar atarken çok büyük veri yığınları arasından kendilerine yol gösterecek kritik verileri ayıklayarak analiz edebilir [70]. Temelde amaç, daha önceden fark edilmemiş veri desenlerini tespit edebilmektir.

VM, temelde iki ana başlıkta incelenmektedir. Birincisi, elde edilen örüntülerden sonuçları bilinmeyen verilerin tahmini için kullanılan tahmin edici, diğeri ise eldeki verinin tanımlanmasını sağlayan tanımlayıcıdır [71]. Tanımlayıcı modeller veri setinin içinden ilişkileri çıkarır. Tanımlayıcı modellerde kullanılan veri madenciliği teknikleri ise, kümeleme, özetleme, birliktelik kuralları, sıralı dizilerdir. Tahmin edici modeller ise, sonuçları önceden bilinen durumlardan bir model geliştirir ve bu model ile sonuçları bilinmeyen veri kümelerinden yeni sonuçlar elde etmektedir. Tahmin edici

modellerde kullanılan veri madenciliği teknikleri sınıflandırma, eğri uydurma, zaman serileridir [69].

4.2 Veri Madenciliği Süreci

Veri madenciliği çalışmalarının hızlı, daha az maliyetli ve daha verimli gerçekleştirilmesi için CRISP-DM (CRoss Industry Process for Data Mining) yaklaşımı geliştirmiştir. Bu yaklaşım altı adımdan oluşmaktadır. Bu adımlar; İş anlamak, veriyi anlamak, veriyi hazırlamak, modelleme, değerlendirme ve yaymadır [72].

4.2.1 İş anlamak

İlk aşama veri madenciliği çalışmasının hangi amaca yönelik yapılacağını net olarak ortaya konmasıdır. Amaç, net bir şekilde ifade edilmeli, cevap aranılan soruya odaklanılmalı ve sonuç değerlendirme kriterleri tanımlanmış olmalıdır [73].

- a. İş hedeflerini anlaşılır hale getirmek: Çalışmanın birincil amacının belirlenmesi ve bu amacın diğer amaçlardan elden geldiğince ayrıştırılarak net olarak tanımlanması aşamasıdır.
- b. Durumu değerlendirmek: Veri madenciliği konusunun temel amacı verimin arttırılmasıdır. Verimin arttırılması amacı ortaya çıkacak sonuçlar kadar sürecin kendisi için de önemlidir. Yapılacak çalışma sonucunda elde edilecek faydayı değerlendiriyor olmak önemli bir zarurettir [74].

4.2.2 Veriyi anlamak

İkinci aşama verilerin derlenmesi, mevcut verilerin uygunluğunun incelenmesi, sahip olunan veri sayısının yeterli olup olmadığına bakılması, modeli oluşturmak için gerek duyulan farklı veri ihtiyaçlarının tespit edilmesi gibi verinin kalitesine ve yeterliliğine yönelik bilgilerin düşünülmesi aşamasıdır [73].

- a. İlk verileri toplamak: Çalışma kaynaklarında tanımlanmış olan başlangıç verilerinin toplanması aşamasıdır.
- b. Veriyi tanımlamak: Derlenen verinin tanımlanması ve ihtiyaçları ne derece karşıladığının değerlendirildiği aşamadır.
- c. Veriyi keşfetmek: Başlangıç aşamasında derlenen veriler ile ilk hipotezlerin oluşturulması, sınırlı şekilde veriden çıkarımlar yapılması aşamasıdır.

- d. Verinin kalitesini belirlemek: Çalışma için kullanılacak verinin doğru mu, hatalı mı, hatalar içeriyorsa ne tür hatalar içerdiği, verinin tam olup olmadığı, eksikler var mı soruları ile verinin kalitesinin tespit edilmesi aşamasıdır [74].

4.2.3 Veriyi hazırlamak

Başlangıçta elde edilen verilerin, çalışmalara temel oluşturacak son veriler haline getirilmesi aşamasıdır [73].

Veri madenciliği çalışmalarında karşılaşılan en önemli sorunlar verilerdeki eksiklikler, araştırmaya uygun verilerin seçilmemesi, seçilen verilerin arasında yüksek korelasyon bulunması, sapan değerler gibi sorunlardır. Bu sorunlar, yanlış bulgular elde edilmesine ve çalışmaların uzamasına neden olmaktadır. Bunun için veri madenciliği adımlarından olan veri ön işleme başka bir deyişle veri hazırlama bütün veri madenciliği sürecinin en önemli adımıdır [75].

- a. Veri Setini Tanımlamak: Araştırılması düşünülen konu için gerekli olan verilerin neler olacağı ve hangi veri kaynaklarından elde edileceği sorularının cevaplarının belirlendiği aşamadır.
- b. Veriyi Seçmek: Yapılacak analizde kullanılacak verilerin belirlenmesi aşamasıdır. Bu verileri değerlendirirken verinin hedefler ile olan ilişkisine, kalitesine dikkat etmek gerekir.
- c. Veriyi Temizlemek: Elde edilen veri derlemindeki eksik verileri düzenlemek ve düzeltmek, hatalı verileri temizlemek için yapılan çalışmaların hepsidir. Ayrıca gürültülü ve tutarsız verileri çıkararak verinin kalitesini artırma aşamasıdır. Burada verinin türüne göre yabancı kelimelerin çıkarılması, spam içeriklerin kaldırılması çalışmaları da yapılır. Aynı şekilde yanlış girişten veya istisnalardan kaynaklanan verilerin, değerlendirmeden çıkarılması tercih edilir [76].
- d. Veriyi Kurmak: Verileri tanımlayan mevcut değişkenlerde değişiklikler yaparak model için daha kullanılır değişken setleri oluşturma aşamasıdır.
- e. Veri Birleştirmek: Veri uyumsuzluklarının olmasının sebeplerinin başında farklı veri kaynaklarından alınan verilerle madencilik çalışmalarının yapılmak istenmesidir. Bu adımda farklı kaynaklardan toplanan verilerin uyumsuzlukları mümkün olduğunca giderilerek, tek bir veri tabanında toplanması amaçlanır.

- f. Veri Formatlamak: Veri seti oluşturulduktan sonra kullanılacak modele göre anlam değişikliği yapmayacak format düzenlemelerinin yapılması aşamasıdır [74].

4.2.4 Modelleme

Veri madenciliği problemleri için birden çok çözüm tekniği olabilmektedir. Tekniklerden bazıları veriler üzerinde özel çalışma yapmayı gerektirir. Bu nedenle veri hazırlama ve model kurma aşamaları, en iyi olduğu düşünülen modele varılıncaya kadar yinelenen bir süreçtir [73].

- a. Model Tekniğini Seçmek: Çalışmada kullanılması düşünülen veri madenciliği fonksiyonunun ve algoritmasının belirlenmesidir.
- b. Model Test Tasarımı Yapmak: Modeli çalıştırıp sonuçları elde etmeye başlamadan önce, modelin kalitesini ve geçerliliğini test etmek gereklidir.
- c. Modelin Kurulması: Model için kullanılacak algoritmanın/yöntemin/teknikğin hazırlanan veri üzerinde çalıştırılması aşamasıdır. Kurulan ve geçerliliği kabul edilen model doğrudan bir uygulama olabileceği gibi, bir başka uygulamanın alt parçası olarak da kullanılabilir.
- d. Modeli Değerlendirmek: Başarı kriterleri, test sonuçları ve daha önceki tecrübelerle göre modelin değerlendirilmesi aşamasıdır. Tüm çalışmanın değerlendirilmesinden ziyade modelin teknik açıdan değerlendirilmesidir [74].

4.2.5 Değerlendirme

Kurulmuş bir model vardır ve bu aşamada modelin son halinin sunulmasından önce modelin yoğun olarak değerlendirilmesi ve iş hedefleri ile uyumlu olup olmadığının kontrol edilmesi amaçlanır [73].

- a. Sonuçları Değerlendirmek: Bu aşamada modelin iş hedeflerini ne ölçüde karşıladığı değerlendirilir.
- b. Süreci Değerlendirmek: Kalite güvence aşamasıdır. Modelin iş hedeflerini karşılamaya yeterli olduğu kararını aldıktan sonra, modelin doğru kurulup kurulmadığı, sadece eldeki verilerden mi yararlandığı, gelecekte kullanılabilecek farklı verilerin neler olabileceği gibi konularda değerlendirmeler yapılmalıdır.
- c. Gelecek Adımları Planlamak: Çalışmanın geldiği noktanın yeterli olup olmadığı, ek çalışma yapılmasının gerekip gerekmediğinin değerlendirilmesi

aşamasıdır. Atılacak başka adımların neler olabileceği, eğer devam etmek mantıklı ise nereden devam etmek gerektiği gibi konularda değerlendirmeler yapılmalıdır [74].

4.2.6 Yayma

Modelin tamamlanmış olması çalışmanın nihai sonucu olduğu anlamına gelmemektedir. Elde edilen veriler kullanılabilecek şekilde sunulmalıdır [73].

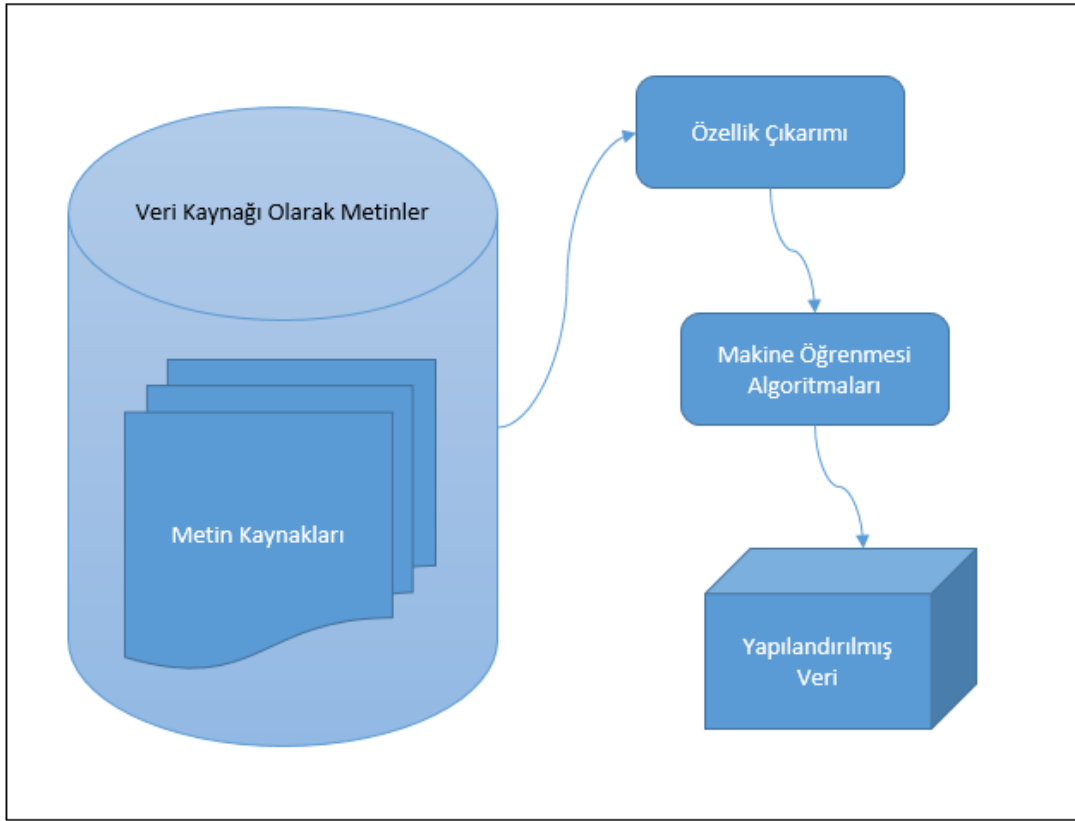
- a. Yayma Planını Oluşturmak: Sonuçları değerlendirerek, yayma stratejisinin oluşturulması aşamasıdır.
- b. Takip ve Bakımı Planlamak: Sistemlerde zamanla üretilen verilerde değişiklikler ve özellik değişimleri yaşanabilmektedir. Bu durum kurulan modellerin sürekli izlenmesini ve gerekiyorsa yeniden düzenlenmesini gerektirecektir. Yanlış verilerin kullanılmasının önüne geçebilmek adına bakım çok önemlidir.
- c. Final Raporu Hazırlamak: Yapılan çalışmanın başkaları tarafından da tekrarlanabilirliğini sağlamak ve sonuçlarını karar vericilere aktarabilmek üzere hazırlanan rapordur. Çalışmanın üçüncü taraflarca denetlenebilmesini ve güvenilirliğini sağlamak açısından önem taşır.
- d. Çalışmayı Değerlendirmek: Yapılan çalışmaya dayandırılan kararların ve sonuçların belli bir zaman sonrasında beklentilerle karşılaştırılması ve gerektiğine çalışmanın yenilenmesi aşamalarını içerir [74].

4.3 Metin Madenciliği

Metin madenciliğinin yaygın olarak kullanılan tanımı, "belirli bir formatta olmayan, yazı tipindeki veriler içerisinde gizli olan nitelikli bilginin çıkarılması, düzensiz haldeki verinin formatlanması süreci" şeklindedir. Metin madenciliği metinlerin sınıflandırılması, metinlerden konu çıkarımının yapılması, kümelendirilmesi, duygu analizi, metin özetleme, varlık ilişki modellemesi gibi çalışmaları hedefler [77]. Bu hedeflere ulaşılması için metin madenciliği çalışmaları kapsamında enformasyon getirme (information retrieval), hece analizi (lexical analysis), örüntü tanıma (pattern recognition), kelime frekans dağılımı (Word frequency distribution), etiketleme (tagging), enformasyon çıkarımı (information extraction), veri madenciliği (data mining) ve hatta görselleştirme (visualization) gibi yöntemleri kullanmaktadır.

Metin madenciliği çalışmaları, diğer bir çalışma alanı olan doğal dil işleme (natural language processing) çalışmaları ile çoğu zaman beraber yürümektedir. Doğal dil işleme çalışmaları daha çok yapay zeka altındaki dil bilim bilgisine dayalı çalışmaları kapsamaktadır. Metin madenciliği çalışmaları ise daha çok istatistiksel olarak metin üzerinden sonuçlara ulaşmayı hedefler. Metin madenciliği çalışmaları sırasında çoğu zaman doğal dil işleme kullanılarak özellik çıkarımı da yapılmaktadır [78].

Şekil 4.1’de klasik bir metin madenciliği çalışması özetlenmiştir:



Şekil 4.1: Klasik Metin Madenciliği

Şekil 4.1’de görüldüğü üzere, bir metin veri tabanından alınan veriler öncelikle bir özellik çıkarımına tabi tutulur. Ardından çıkarılan özellikler üzerinde bir makine öğrenmesi algoritması çalışır (sınıflandırma, kümeleme, tahmin v.b.) ve sonuç olarak yapılandırılmış veri elde edilir.

Metin Madenciliği, veri madenciliği, yapay zekâ, doğal dil işleme (Natural Language Processing – NLP), istatistik, bilgi erişim (Information Retrieval – IR) ve bilgi yönetimi (Knowledge Management) tekniklerini kullanarak “bilgi patlaması / bilgi taşması” (information explosion / information overload) sorununa çözüm bulmayı

amaçlayan bir araştırma alanıdır [79]. Metin madenciliği süreci beş aşamadan oluşur: Belge Erişim, Veri Çıkarımı, Veri Temizleme, Madencilik ve Görselleştirme [80].

4.4 Gizli Dirichlet Tahsisi

Her gün çok miktarda veri toplanmaktadır. Daha fazla bilgi oluştuğunda, aradığımız bilgiye erişmek güçleşir. Bu nedenle, büyük miktarda bilgiyi organize etmek, araştırmak ve anlamak için araçlara ve tekniklerimize ihtiyacımız olur.

Konu modellenmesi, geniş metinsel bilgi koleksiyonlarını organize etme, anlama ve özetleme yöntemleri sunar. Konu modellemesi, koleksiyondaki bilgileri en iyi temsil eden bir dizi belgeden bir grup kelime (yani konu) bulmak için bir yöntem olarak tanımlanabilir. Metinsel materyalde tekrar eden kalıpları elde etmenin bir yolu olarak, bir metin madenciliği biçimi olarak da düşünülebilir.

Son zamanlarda bloglar, haber makaleleri ve literatür gibi çok sayıda elektronik belge koleksiyonu internet üzerinde mevcut hale gelmiştir. Bu durum veri madenciliği topluluğundaki araştırmacılar için bir takım yeni zorluklar ortaya çıkarmıştır. Özellikle, doküman koleksiyonlarını görselleştirmek, analiz etmek ve özetlemek için otomatik tekniklere ihtiyaç artmıştır. Son dönemde de gizli konu modellemesi, büyük belge koleksiyonlarında konu keşfi için tamamen denetlenmeyen bir teknik olarak çok popüler hale gelmiştir [62].

Makine öğrenmede ve doğal dil işlemede bir konu modeli, bir belge koleksiyonunda ortaya çıkan soyut konuları keşfetmek için bir istatistiksel model türüdür. Konu modelleme, metin gövdesinde gizli anlamsal yapıların bulunması için sıkça kullanılan bir metin araştırma aracıdır. Sezgisel olarak, bir belgenin belirli bir konuyla ilgili olması nedeniyle, belgede belirli kelimelerin çok az sıklıkla görünmesini beklersiniz: "köpek" ve "kemik" daha çok köpek başlığı altında, "kedi" ve "miyav" ise kedi başlığı ile ilgili belgelerde daha sık görünecektir [81]. Kısaca özetlersek konu modelleme:

- Koleksiyonda bulunan gizli konu kalıplarını keşfetmek,
- Belgeleri bu konulara göre açıklamak,
- Metinleri organize etmek, aramak ve özetlemek için bu açıklamaları kullanmak konusunda destek olur.

Konu modelleme yöntemleri bir metin veritabanında, hem bir belge içerisinde geçen sözcükleri, hem de farklı belgelerde geçen sözcüklerin birlikte bulunduğu diğer sözcükleri inceleyerek, her belgedeki metni bir veya birden fazla konudan oluşacak şekilde sentezleyen modeli üretir [82].

Konu modelleri elde etmek için kullanılan birçok yöntem vardır. Gizli Anlamsal Analiz, Olasılıksal Gizli Anlamsal Analiz ve Gizli Dirichlet Tahsisi [83] bu yöntemlerdendir.

Çalışmalarda kullanılan en bilinen uygulamalardan birisi literatürde kısaca LDA (Latent Dirichlet Allocation) ve Türkçeye Gizli Dirichlet Tahsisi (GDT) [46][82] olarak çevrilebilecek olan yöntemdir. Bu yaklaşım bir dökümanın hangi konu ya da konularla ilişkili olduğunu tespit etmek için kullanılan olasılık tabanlı bir konu modelleme yöntemidir. Bir başka tanıma göre GDT Bayes teorisini savunan denetlenmeyen bir kümeleme modelidir [84]. Ayrıca GDT modelinin mutlaka metne bağlı olmadığı ve ortak filtreleme, içerik tabanlı görüntü alımı ve biyoinformatik gibi alanlardan alınan veriler de dahil olmak üzere verilerin toplanmasıyla ilgili diğer sorunlara yönelik uygulamaları vardır [83].

Bir dokümanın ilgili olduğu konular, o dokümanda geçen kelimelere bağlıdır. Dokümanlarda geçen kelimeler, gizli olarak belirli konulara işaret eder ve o konulara işaret eden kelimeler, o konularla ilgili kelime grubunu oluşturur. Örneğin; sporla ilgili bir dökümanda futbol, basketbol gibi kelimeler geçerken müzikle ilgili bir dökümanda notalar yer alabilir. Dokümanlarda bir konu grubuna ait kelimelerin sıklıkla geçmesi, dökümanın o konuyla ilgili bilgiler taşıdığını göstermektedir. Ancak, bir dokümanda sadece bir konu grubuna ait kelimeler değil, başka konu grubuna ait kelimeler de sıklıkla geçmektedir. Bu da, bir dokümanın birden çok konuyla ilgili olduğunu göstermektedir. Pek çok dokümanın olduğu bir ortamda hangi kelimelerin hangi konu grubuna dahil olduğunu ve hangi dokümanın hangi konularla ilgili olduğunu bulmak oldukça zor bir işlemdir. Bu atamaları otomatik olarak yapabilmek için geliştirilen yöntem olan GDT'de ilk olarak, verilen doküman kümesindeki kelimeler taranarak kelimelerin konu grupları dağılımı olasılık değerleriyle birlikte bulunur. Daha sonra, bir dokümandaki her kelime incelenir, incelenen kelimenin her konu grubuna ait olma olasılığı bulunur. Dokümandaki her kelime için bulunan olasılık değerleri çarpılarak, ilgili dokümanın konu gruplarına ait olma olasılıkları elde edilir. Olasılık değerleri yüksek olan konu grupları, o dokümanın ilgili olduğu konu gruplarını gösterir.

GDT sürecini temel fikri en basit şekliyle, her doküman konuların karışımı olarak modellenir ve her konu her bir kelimenin konular üzerinde nasıl görüldüğünü tanımlayan kesikli olasılık dağılımıdır. Bu konu olasılıkları dokümanların bir sunumu gibidir ve doküman aslında kelime ve konu istatistiklerinin ötesinde belirli bir yapıya sahip olmayan bir kelime torbasıdır [62].

GDT yönteminde kelimeler, gözlemlenen rasgele değişkenler olarak modellenirken, konular gizli veya saklı rasgele değişkenler olarak işlenir. Diğer üretici olasılıksal modelleme algoritmaları gibi ilk olarak üretim prosedürleri belirlenir. GDT, gözlemlenen ve gizli rasgele değişkenler üzerine bileşik dağılımının ve verilen gözlenen değişkenlerden gizli değişkenlerin çıkarımının yapılabileceğini istatistiksel olarak ilk tanımlayanıdır [85]. GDT modeli belgelerin önceki açıklamalarına veya etiketlemesine ihtiyaç duymaz ve konular orijinal metinlerin analizinden çıkar. GDT modeli, insan açıklaması kullanmanın imkansız olduğu ölçeklerde elektronik arşivleri düzenlememize ve özetlememize olanak tanır [86].

Gizli Dirichlet Tahsisi, metin belgelerinden konuların çıkarılması ve bu konulara göre metinlerin sınıflandırılması olmak üzere iki temel aşamadan oluşmaktadır. Bu iki aşamadan ikincisi olan, metinlerin konulara atanması süreci de iki temel aşama ile gerçekleştirilmektedir. Öncelikle, metin belgesinde yer alan her bir sözcük geçici olarak bir konuya atanır. Ardından, metin belgesindeki sözcüklerin olasılık dağılımlarına göre, metin belgesi, belirli bir konuya atanır. Ancak bazı durumlarda bir kelime birden fazla başlığa ait olabilir, bu yüzden GDT iteratif bir şekilde doğru başlığı bulmak için istatistiksel modelini güncellemektedir. Konuyu daha basit bir örnekle anlatmak gerekirse:

Ben **tavuk** ve sebze yerim.

Tavuklar kümes hayvanlarıdır.

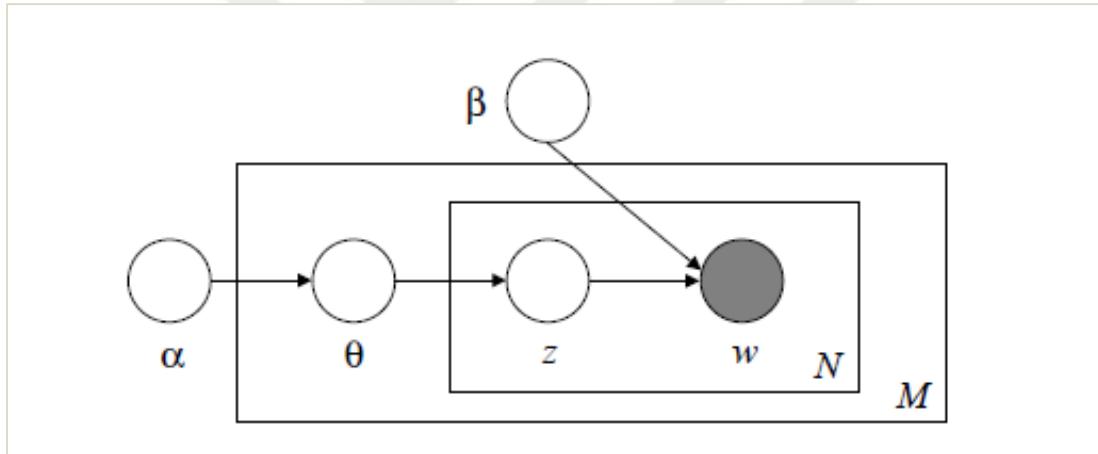
Benim *kedim* **tavuk** yer.

Bu üç cümlede iki farklı fikir çıkarımı yapılabilir. İlk bulduğumuz başlık “yemek” ve ikinci başlık “kümes hayvanı” olarak çıkarılır. Buna göre birinci cümle %100 “yemek” başlığında, ikinci cümle %100 “kümes hayvanı” başlığında, üçüncü cümle %33 “kümes hayvanı” ve %66 “yemek” başlığı altında kabul edilmelidir.

Yukarıdaki üç cümlede geçen “tavuk” kelimesi bir kümes hayvanı veya bir yemek olarak kabul edilebilir. “Benim kedim tavuk yer” cümlesini okuyunca buradaki

“tavuk” kelimesinin bir yemek olduğunu anlayabiliriz ancak bilgisayarın bu analizi yapabilmesi için “yemek” eyleminin tavuğu işaret ettiğini anlaması gerekir. Kısaca elimizdeki bir sözlükle kelimeleri arayarak bu kelimelerin dâhil olacağı sınıfları bulmak çoğu zaman yanlış sonuçlar çıkarmaktadır. Bu sebeple olasılık olarak bir kelimenin birden fazla sınıfa farklı oranlarda dâhil olduğu kabulü yapılır ve her iterasyonda daha doğru sınıfa atama yapılır. Örneğin Türkçede tavuk kelimesi yüksek ihtimalle yiyecek olarak kullanılmakta (%85 varsayarsak) ve daha düşük oranda bir kümes hayvan olarak kullanılmaktaysa GDT öncelikle kelimeyi bir yemek olarak sınıflandıracak ancak daha sonra cümledeki diğer kelimelerin kullanımına göre sınıfını değiştirerek kümes hayvan sınıfına alabilecektir (veya tam tersi) [28].

Şekil 4.2’de Gizli Dirichlet Tahsisi süreci grafiksel olarak özetlenmiştir. Burada, rastgele değişkenler, düğümler ile temsil edilmiştir. Düğümler arasındaki olası bağlantılar ise kenarlar kullanılarak temsil edilmiştir.



Şekil 4.2: Gizli Dirichlet Tahsisi Levha Gösterimi [83]

Notasyonda kullanılan terimleri tanımlarsak,

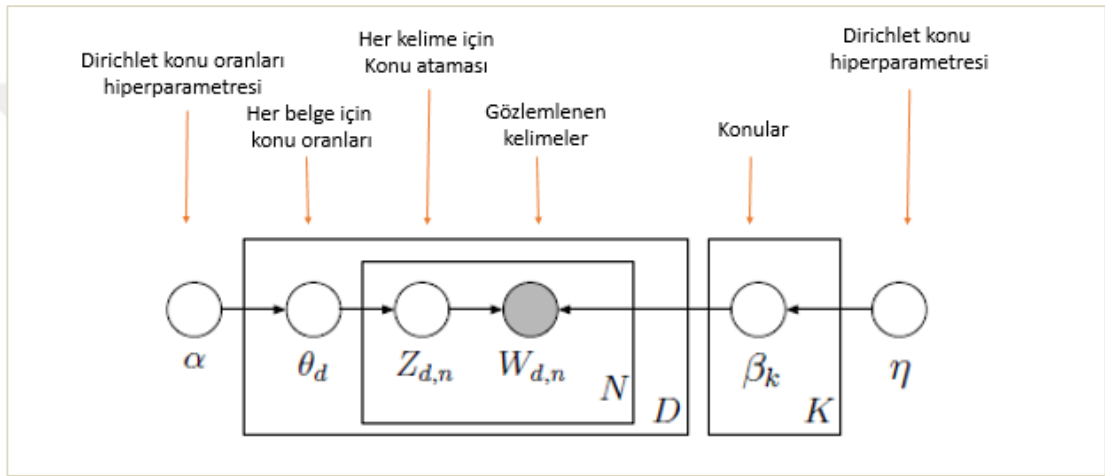
Kelime: $\{1, \dots, V\}$ şeklinde indekslenen bir sözlükten ayrık veriler olarak temsil edilir. Kelimeler, tek bileşeni 1’e eşit olan ve diğer tüm bileşenleri 0’a eşit olan ünite bazlı vektörleri kullanır. Bu nedenle $u \neq v$ iken v . kelime $w^v = 1$ ve $w^u = 0$ olan V -vektör ile gösterilir.

Belge: Bir belgede, w_n , dizideki n . kelime olduğunda, $\mathbf{w} = (w_1; w_2, \dots, w_N)$ ile gösterilen N sözcükten oluşan bir dizidir.

Korpus: $D = \{ \mathbf{w}_1; \mathbf{w}_2, \dots, \mathbf{w}_M \}$ ile gösterilen M belgelerin bütünüdür.

Levha gösterimi ile birçok değişken arasındaki bağımlılıklar gösterilmektedir. Dış plaka belgeleri temsil ederken, iç plaka bir belgedeki konuların ve kelimelerin tekrar tekrar seçilmesini temsil eder. M belgelerin sayısını, N belgedeki kelimelerin sayısını belirtir. Şekil 4.2’de gösterilen levha modelinde, α belge başına konu dağılımındaki Dirichlet parametresini, β konu başına kelime dağılımındaki Dirichlet parametresini, θ doküman seviyesi konu değişkenlerini, w gözlemlenen kelimeleri ve z her bir kelime için atanan konuları göstermektedir. Gölgeledi çember içerisinde yer alan w gözlenen değişkenleri, gölgeledi olmayan z ve θ gizli değişkenleri gösterir.

Bir başka gösterim olan yumuşatılmış levha gösterimi ise Şekil 4.3’teki gibidir.



Şekil 4.3: Yumuşatılmış Levha Gösterimi [87]

Bu gösterim ile konuların nasıl çıktığı daha kolay anlaşılabilir. Burada η Dirichlet konu parametresini temsil eder.

Temel fikir, her konu kelimeler üzerinde bir dağılım ile karakterize edilirken belgelerin gizli konular üzerine rasgele karışımlar olarak gösterilmesidir. Burada, bir "belge", konunun ve kelime istatistiklerinin ötesinde bir yapıya sahip olmayan bir "kelime çantası" dır [62]. GDT, D korpusunda yer alan her w dokümanı için aşağıdaki üretme sürecini varsaymaktadır:

1. β seçimi $\sim \text{Dir}(\eta)$.
2. θ seçimi $\sim \text{Dir}(\alpha)$.
3. w_N içerisindeki her N kelime için (Belgedeki her kelime için);

(a) z_n konusunu seçmek $\sim \text{Çok Terimli}(\theta)$.

(b) $p(w_n | z_n, \beta)$ olasılığından w_n kelimesini seçmek, z_n konusu üzerinde şartlandırılmış çokterimli olasılık. Rastgele ilgili konudan bir kelime seçmek (kelime dağılımı).

Üretici süreç iki aşamadan oluşur. İlk aşamada Dirichlet rasgele değişkenle konu dağılımı vektörü modellenir. Hangi konuların belgelerde daha çok görüldüğü çıkarılır. Sonrasında belgelerde görülen her bir kelime için konu dağılım vektöründen rasgele tek konu seçilir. Kelimeyi üretmek için konuyla ilgili olasılık dağılımı çizilir. Belgedeki her kelime farklı ve rasgele seçilen konularla üretilir ve kelimeler diğer kelimelerden bağımsız üretilir.

Bu basit modelde bazı basitleştirici varsayımlar yapılmıştır. İlk olarak, Dirichlet dağılımının k boyutu (ve dolayısıyla z konu değişkeninin boyutsallığı) bilinen ve sabit olarak kabul edilir. İkincisi, kelime olasılıkları, tahmin edilecek sabit bir miktar gibi davranan $\beta_{ij} = p(w^j = 1 | z^i = 1)$ olan bir $k \times V$ matrisi β tarafından parametrelendirilir. Ayrıca, N diğer tüm veri üreten değişkenlerden (θ ve z) bağımsızdır. Dolayısıyla, yardımcı bir değişkendir ve genelde sonraki gelişmedeki rasgeleliğini görmezden gelebiliriz.

α parametresi Dirichlet dağılımında k -boyutlu pozitif vektörü tanımlar. Bu parametre bir korpus içindeki tüm dokümanlarda sabit kalır [85]. Dirichlet dağılımı aşağıdaki gibidir:

$$p(\theta | \alpha) = \frac{\Gamma(\sum_i \alpha_i)}{\prod_i \Gamma(\alpha_i)} \prod_i \theta^{\alpha_i - 1} \quad (1)$$

$\Gamma(x)$ bir gama fonksiyonu, α parametresi $\alpha_i > 0$ olan bir k -vektördür.

Dirichlet, simpleks üzerinde uygun bir dağılımdır - üstel ailede bulunur, sınırlı boyutlu yeterli istatistiklere sahiptir ve çokterimli dağılıma eşleniktir. Dağılımların dağılımıdır.

Şekil 4.2’de verilen α ve β parametreleri, konu karışımı θ ’ nın bileşik dağılımı, z konu seti N ve w kelime seti N ile,

$$p(\theta, z, w | \alpha, \beta) = p(\theta | \alpha) \prod_{n=1}^N p(z_n | \theta) p(w_n | z_n, \beta). \quad (2)$$

θ üzerinde bir entegre ve z üzerinde bir toplam ile bir dokümanın marjinal dağılımı elde edilir:

$$p(w | \alpha, \beta) = \int p(\theta | \alpha) \left(\prod_{n=1}^N \sum_{z_n} p(z_n | \theta) p(w_n | z_n, \beta) \right) d\theta. \quad (3)$$

Sonuç olarak, tek bir dokümanın marjinal dağılımının çarpımını alarak, bir korpusun olasılığı elde edilir.

$$P(D | \alpha, \beta) = \prod_{d=1}^M \int p(\theta_d | \alpha) \left(\prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn} | \theta_d) p(w_{dn} | z_{dn}, \beta) \right) d\theta_d. \quad (4)$$

GDT gösteriminde üç seviyenin olduğu görülür. α ve β parametrelerinin olduğu seviye korpus-seviye parametreleridir. Doküman başına bir kez örneklenen değişkenlerden θ_d doküman-seviye değişkenleri ve son olarak her dokümandaki her bir kelime için bir kez örnek alınan z_{dn} ve w_{dn} kelime-seviye değişkenlerdir [83].

Pratikte Beklenti-Maksimizasyon algoritması ve Gibbs örnekleme algoritması parametre tahmin görevini yürütmek için kullanılır. Daha özelleştirilirse, doküman özelinde konu dağılımı olasılığı ve konu özelinde kelime dağılımı olasılığını tahmin etmek için genelde Gibbs örnekleme kullanılarak basit sonsal dağılım yapılır.

α hiperparametresinin optimize değeri Beklenti-Maksimizasyon algoritmasının Maksimizasyon adımı uygulanan Newton-Raphson yöntemi ile bulunur.

GDT modeli, metnin yapısı veya dilin sözdizimsel veya gramersel özellikleri hakkında herhangi bir varsayım yapmaz. Literatürde bulunan diğer metin analiz teknikleri yerine GDT modelinin benimsenmesi şu gerekçelere dayanır. Birincisi, GDT modeli, büyük ölçekli verileri büyük ölçüde ayrıntılı bir biçimde verimli bir şekilde analiz etmede üstünlük sağlar ve böylece farklı müşteri grupları (ör., Erkek vs. kadın) arasındaki boyutların heterojenliğini keşfetmemizi sağlar. Gizli dirichlet tahsisi, "konu modelleri" ailesinin çekirdek modellerinden biridir ve zengin analizler yapmamızı sağlayacak kadar esnektir [88].

Kısaca özetlemek gerekirse;

- Konular, korpusdaki tüm kelimelerdeki dağılımlardır. Her bir konu teknik olarak korpus içerisindeki her kelimenin bir bölümünü içeriyor olsa da, bazı kelimeleri diğerlerinden daha yüksek frekansa sahiptir. Bu sözcüklerin topikal olarak ilişkili olduğu kabul edilir. Her konu en yüksek frekanslı kelimelerin listesini vererek açıklanmaktadır.
- Dokümanlar, korpusun tüm konularında dağılımlar olarak kabul edilir. Her doküman teknik olarak her konunun bir kısmını içerirken diğerlerinden daha

yüksek frekansa sahip bazı konular içermektedir. Bu konular dokümanın ana temalarını temsil etmelidir. Dokümanlar, konunun orantıları vererek açıklanmaktadır.

- Üretken modelde - metinlerin nasıl üretildiğine ilişkin kuramsal model - bir doküman, korpustaki konu oranlarına göre konuların seçilmesi (yani, yüksek oranlı konuların seçilme olasılığı daha yüksektir) ve daha sonra dokümandaki konu oranları ve konudaki kelime oranlarına göre konulardan kelimelerin seçilmesi ile oluşturulur. Bu, belirli bir doküman için seçilmesi en muhtemel kelimelerin yüksek oranlı konular ve yüksek oranlı kelimeler olduğu anlamına gelir.
- Konu modelinde, istatistiksel olarak konuların sayısının ve yığındaki kelimelerin bulunma sayılarının temel alındığı bir tersine mühendislik çalışması yapılır. Kelimelerin tüm dokümanların içinde bulunmalarına göre, tek bir dokümanın içinde bulunmaları gözlenir ve bu gözlem temel alınır. Böylece bu kelimeler belirli konular içinde gruplanır ve dokümanlar da kelimelerdeki oranlara göre konulara atanır [89].

Genel olarak başarılı ve kullanışlı bir yöntem olan GDT'nin bazı dezavantajları bulunmaktadır. Birincisi, konu grupları otomatik olarak belirlendiği için anlamsız başlıklar üretilebilir. Diğerisi ise, doküman seviyesinde inceleme yaptığı için konu grupları üst seviyede kalır, dokümanda geçen özelliklere değinilemez. Örneğin, fotoğraf makinesini anlatan bir dokümanda, fotoğraf makinesi konu grubu olarak bulunabilirken; fotoğraf makinesinin özellikleri olan görüntü kalitesi, şarj ömrü gibi başlıklar konu grubu olarak atanamaz.

Bu çalışmada Andrew McCallum tarafından yazılan MALLET [90] yazılımı kullanılmıştır. MALLET girdi olarak belgelerin sınıflara ait klasörlerin içinde bulunduğu bir ana veri dosyasını alır. Bu dosyayı alırken her dokümanın (yorumun) .txt formatında olması gerekir. Çıktı olarak aşağıdaki dosyaları vermektedir:

- Konu anahtar sözcükleri (keys) ve bunların ağırlıkları,
- Kelimelerin konulara atamaları,
- Her belgenin konular üzerinde dağılımını gösteren olasılıklar.

Bu alanda yapılan bir çalışma kapsamında, farklı konu sayıları için ($k=50, 100, 150, 200$), temsil edilen Türkçe Twitter mesajlarında elde edilen tahmin etme başarımları

değerlendirilmiştir. Farklı konu sayıları ile elde edilen sonuçlar incelendiğinde tüm karşılaştırılan durumlar için en yüksek başarımın ($k=50$) değeri ile elde edildiği gözlenmektedir. Veri setinin temsilinde kullanılan konu sayısı daha az iken tahmin etme başarımının daha yüksek olduğu görülmektedir. Deneysel sonuçlar, gizli Dirichlet tahsisi yönteminin, Türkçe metin belgelerinin sınıflandırılmasında etkin, öz ve uygun bir temsil yöntemi olarak kullanılabileceğini göstermektedir [46].

Twitter ile yapılan bir başka çalışmada yazar kendi twitter hesabını takip eden kullanıcıların hangi konularda tweet attığı inceliyor. Yaptığı denemeler sonucu en iyi sonucu alfa değeri 0,001 iken 40 konu için alıyor. Modelin eğitilmesi zordur ve sonuçlar net bir şekilde yorumlanmalıdır. Denetlenmeyen bir bağlamda, modelin performansını tahmin etmek, elle değerlendirme veya bazı kalite metriğini gerektirir. GDT modeli tarafından bulunan konuların elle değerlendirilmesi kapsamlı değildir. Bununla birlikte izlediğimiz yöntemin ses olduğunu ve yapılması gerekeni yaptığını güçlü bir göstergeyle gösteriyor [91].



5. UYGULAMA

5.1.1 Türkiye’ de AVM sektörü

Alışveriş Merkezleri; merkezi bir birim tarafından tasarlanan, planlanan, yaptırılan ve yönetilen perakende mağazalarının ve çeşitli farklı hizmet kuruluşlarının yer aldığı kompleksler olarak tanımlanır. Bu kompleksler kentsel alanda farklılaşan ihtiyaçların ve toplumun taleplerini görebileceğimiz en önemli alanlardan biridir. Başka bir deyişle alışveriş merkezleri, ziyaretçilerine daha konforlu alışveriş imkânı ve bunun yanında sosyal etkileşim ortamı sağlamak amacıyla tasarlanmış bünyelerinde farklı mağazaları ve kullanımları barındıran, profesyonelce yönetilen perakende kompleksleridir.

Alışveriş merkezi kavramı aslında eskilerden beri var olmuştur. Toplumlar büyüdükçe dükkânlar ve pazarlar gelişip plansız alışveriş merkezleri ortaya çıkmış ve daha sonra planlı alışveriş merkezleri oluşmuştur. Bugünün alışveriş merkezlerinin tarihteki örneklerinden farklılıkları bulunmaktadır. Günümüzdeki alışveriş merkezleri tarihçesi 20. yüzyılın ikinci yarısından itibaren savaşın etkilerini atlatan özellikle Avrupa ve Amerika şehirlerinde yeniden yapılanma sürecine girilmesiyle başlar. Hızla yükselen yaşam standartları, yaygınca kullanılmaya başlanan motorlu taşıtlar, kadınların çalışma hayatına aktif olarak katılmaya başlaması fiziksel ve kültürel olarak değişmesine öncü olmuştur [92].

Günümüz çağdaş tüketim yeri olarak kullanılan alışveriş merkezleri, farklı kültürlerde sosyo-mekansal dinamikleri analiz eden çeşitli araştırmacıların dikkatini çekmektedir. Alışveriş merkezi geliştirme, şu anda birçok ülkede, özellikle ABD ve İngiltere’de iyi bir şekilde kurulmuştur. Alışveriş merkezleri artık Türkiye de dahil olmak üzere birçok ülkede kentsel alanın bir parçası haline gelmiştir.

Türkiye’nin ilk alışveriş merkezi olan Galleria 1988’de açılmıştır. Bu tarihten itibaren geçen süreçte Türkiye’nin yalnızca İstanbul, Ankara gibi metropol kentleri değil, hatta daha küçük yerleşim birimleri bile alışveriş merkezleri ve tüketim kültürü ile tanışmıştır. Türkiye’deki AVM sayılarının çoğalmasıyla beraber artan rekabet ortamında Alışveriş Merkezleri birbirlerinden farklılaşmaya ve böylelikle ayakta kalmaya çalışmaktadır. Çeşitli parametreler bakımından iyi durumda olan AVM’ler

hayatta kalmakta, bazıları ise bu rekabet ortamına daha fazla dayanamayarak kapanmaktadır. Şüphesiz Türkiye içinde yer alan şehirlerden nüfus yoğunluğu ve gelişmişlik bakımından en yüksek dereceye sahip olan İstanbul bu kültürün başını çekmektedir. İstanbul'u, Ankara İzmir gibi metropol şehirler takip etmektedir.

EVA Gayrimenkul tarafından 2016 yılında yapılan bir araştırma sonucunda AVM piyasasındaki rekabetin gittikçe zorlaştığı ve bu nedenle üretimin doğru yerde ve hacimde yapılmasının önemli olduğu çıkarılıyor. Yatırımcıların her sene yeni açılışları ilan ettiği ancak bu ilanların ancak %30'unun gerçekleşiyor. Giderek AVM büyüklüğünün artışa geçtiği, yeni nesil AVM'lerin yaşam alanlarıyla birlikte planlanıyor olduğu ve sosyal medya ile medyanın da aktif şekilde kullanıldığı görülüyor.

AVM sektörünün cirosunun hızla büyüdüğü, 2009 yılında 20,6 milyar TL olan sektör cirosunun 2016 yılında %9.5 büyümeyle 105,9 milyar TL'ye, 2017 yılında %12.4 büyümeyle 119 milyar TL'ye, 2018 yılında ise %10.9 büyümeyle 132 milyar TL'ye ulaşacağı öngörülüyor. Bu gelişimde cadde AVM'leri, AVM bit pazarları, açık hava AVM'leri, tematik AVM'ler ve karma kullanım AVM konseptleri gibi yeni nesil AVM'lerin etkili olduğu belirleniyor.

Türkiye'deki AVM gelişimine bakacak olursak 2011 yılında 264 olan AVM sayısı, 2014 yılında 345'e yükselmiştir. 2016 Mart ayı itibarıyla 369 olan AVM sayısının ise 2016 sonunda 414'e, 2017'de 443'e, 2018'de 448'e yükseleceği öngörülüyor. Son üç yıllık analizlere bakıldığında ise 2016 ve 2017 yıllarında ortalama %10'luk bir büyüme görülürken, 2018 için büyüme oranının azalarak %2 dolayında olması bekleniyor.

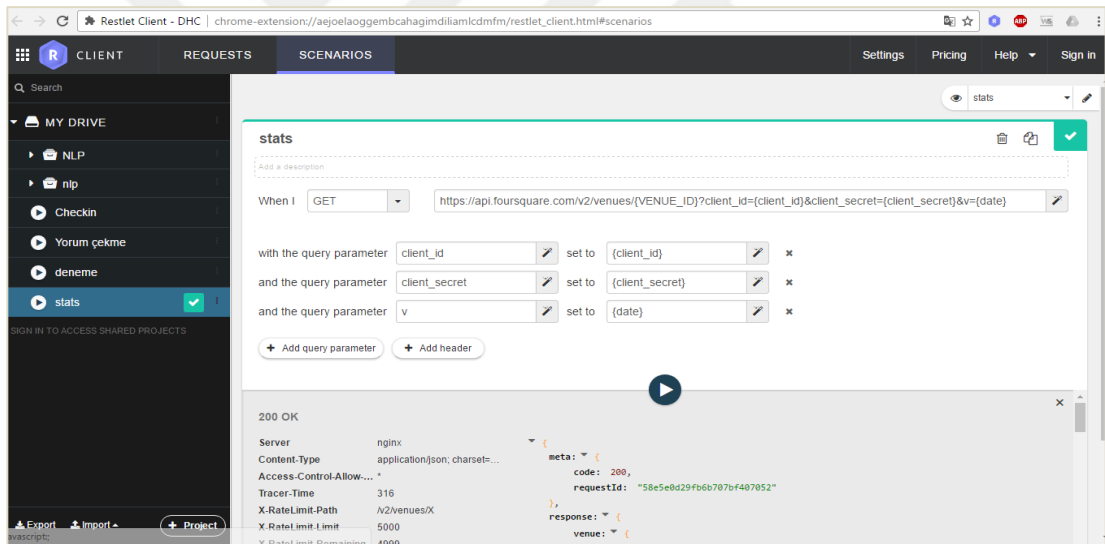
Araştırmaya göre Türkiye'de 2015 yılında 59 şehirde AVM mevcuttur. Deklare edilen açıklamalara istinaden 2018 yılına kadar dört kentin daha AVM ile tanışması öngörülürken, bugünkü planlamalara göre 2018 yılında 18 şehirde AVM olmayacağı ileri sürülüyor [93].

AVM denince akla bir de outlet olup olmadığı gelir. Türkiye'de 35 adet outlet mevcut olduğu ve çok fazla olduğu belirleniyor. Bu sayı içerisinde yer alan outletlerin hakkını vererek işi yapması gerektiği üzerinde duruluyor [94].

5.1.2 Mekân popülerlik analizi

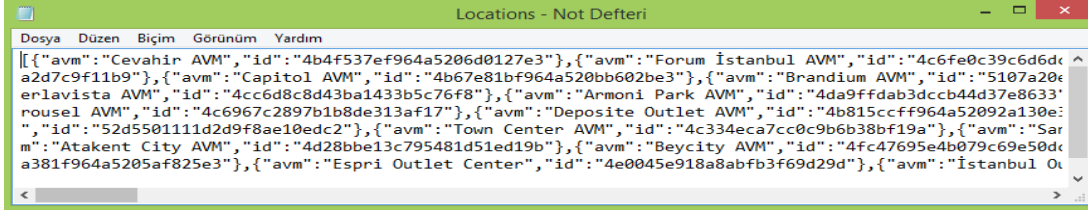
Çeşitli sitelerden İstanbul'da yer alan alışveriş merkezleri araştırılmıştır. Bu araştırmalar sonucu bulunan 118 AVM'ye ait bilgilerin doğruluğu teyit edilerek (kapanan vs.) ve yorum sayısı 10'dan az olan AVM'ler elenerek en son 94 AVM bilgisine ulaşılmıştır [95-98]. Bu AVM'ler için veri çekmede kullanacağımız ID'ler, ilgili mekanın Foursquare sayfasında yer alan adres çubuğundaki veriden elde edilmiştir.

Veriler, Foursquare API üzerinden 10.03.2017 tarihinde bir Google Chrome Eklentisi olan DHC Restlet üzerinden çekilmiştir. Şekil 5.1'de eklenti ekranı görülebilir. Bu eklenti üzerinden elde edilen stats bilgileri ile her bir AVM'nin ID'si girilerek bu AVM'lere ait toplam check-in sayısı, kullanıcı sayısı, yorum sayısı ve ziyaretçi sayısı (AVM'ye giden ama check-in yapmayan) bilgileri alınmıştır. Mekânların Foursquare sayfalarından ayrıca puan ve puanlayan kişi bilgileri alınmıştır.



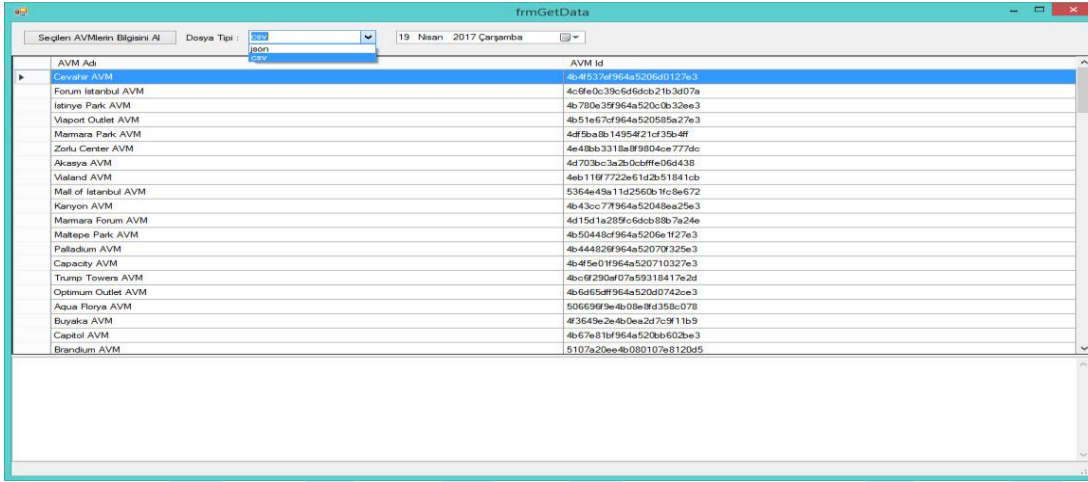
Şekil 5.1: DHC Restlet Eklenti Mekan Bilgileri Çekme Ekranı

Her bir AVM için tüm yorumları manuel olarak çekmek büyük bir operasyon yüküne sebep olacağından ve Foursquare API'nın bir istekte en çok 200 yorum sayısı iletebilmesinden dolayı bu yorumları otomatik çeken bir program yazılmıştır. Program C# ile yazılmıştır. Programda Şekil 5.2'de gösterimi yapılan Locations adlı .json dosyasına girilen Foursquare Mekan ID'si ve mekan adı ile istenilen mekana ait yorumlar csv ya da json formatında çıktıyı anlık olarak çekilebilecek duruma gelinmiştir. Analiz edilen yorum bilgileri, 10.03.2017 tarihine kadar bırakılan yorumlardan oluşmaktadır.



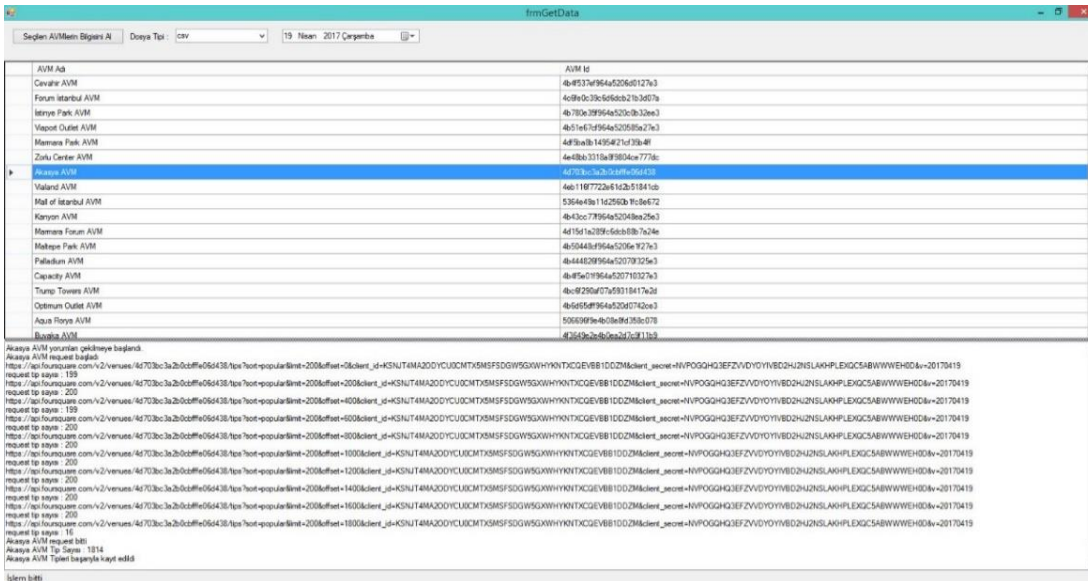
Şekil 5.2: Locations Dosyası İçerik Görünümü

Uygulama ekranı Şekil 5.3'te görüldüğü gibidir. Ekran üzerinde Tüm AVM bilgileri yer almakta ve istenilen AVM veya AVM'ler için csv ya da json formatında istenilen tarihe kadarki yorumlar çekilebilir.



Şekil 5.3: Veri Çekme Program Görünümü

Seçilen bir AVM için yorumların nasıl çekildiği ve sonuçlarını gösteren Şekil 5.4'te verilerin izin verilen sayı olan 200'şerli çekimleri görülmektedir.



Şekil 5.4: Veri Çekme Programı Sonuç Ekranı

Tüm AVM'ler için çekilen yorum bilgileri, AVM'nin Foursquare sayfasındaki bilgileri ve API üzerinden çekilen bilgileri Çizelge 5.1 'deki tabloya yerleştirilmiştir.

Çizelge 5.1: İstanbul'daki AVM Verileri (Verilerin Genişletilmiş Devamı Ek A).

AVM Adı	Yorum Sayısı	Puan	Check-in Sayısı	Tekil Kullanıcı	Ziyaret Sayısı	Frekans (Sıklık)
Marmara Park AVM	1959	8,4	2080674	539114	2945673	3,859
Maltepe Park AVM	1537	8,4	1475814	387029	2030169	3,813
Kule Çarşı Alışveriş Merkezi	26	6,3	12609	3369	13065	3,742
Akbatı AVM	849	8,6	690578	196154	978685	3,520
Palladium AVM	1410	8,9	1215663	366489	1622866	3,317
Capacity AVM	1398	8,4	1491862	460094	1857873	3,242
Perlavista AVM	627	7,3	463817	145621	624464	3,185
Forum İstanbul AVM	2834	8,4	2487754	786500	3348041	3,163
Merkez Kayaşehir AVM	130	5,1	48773	15837	69421	3,079
Akyaka Park Alışveriş Merkezi	61	5,3	62185	20364	119510	3,053
Cevahir AVM	3107	8,4	3048002	1E+06	4056660	3,011

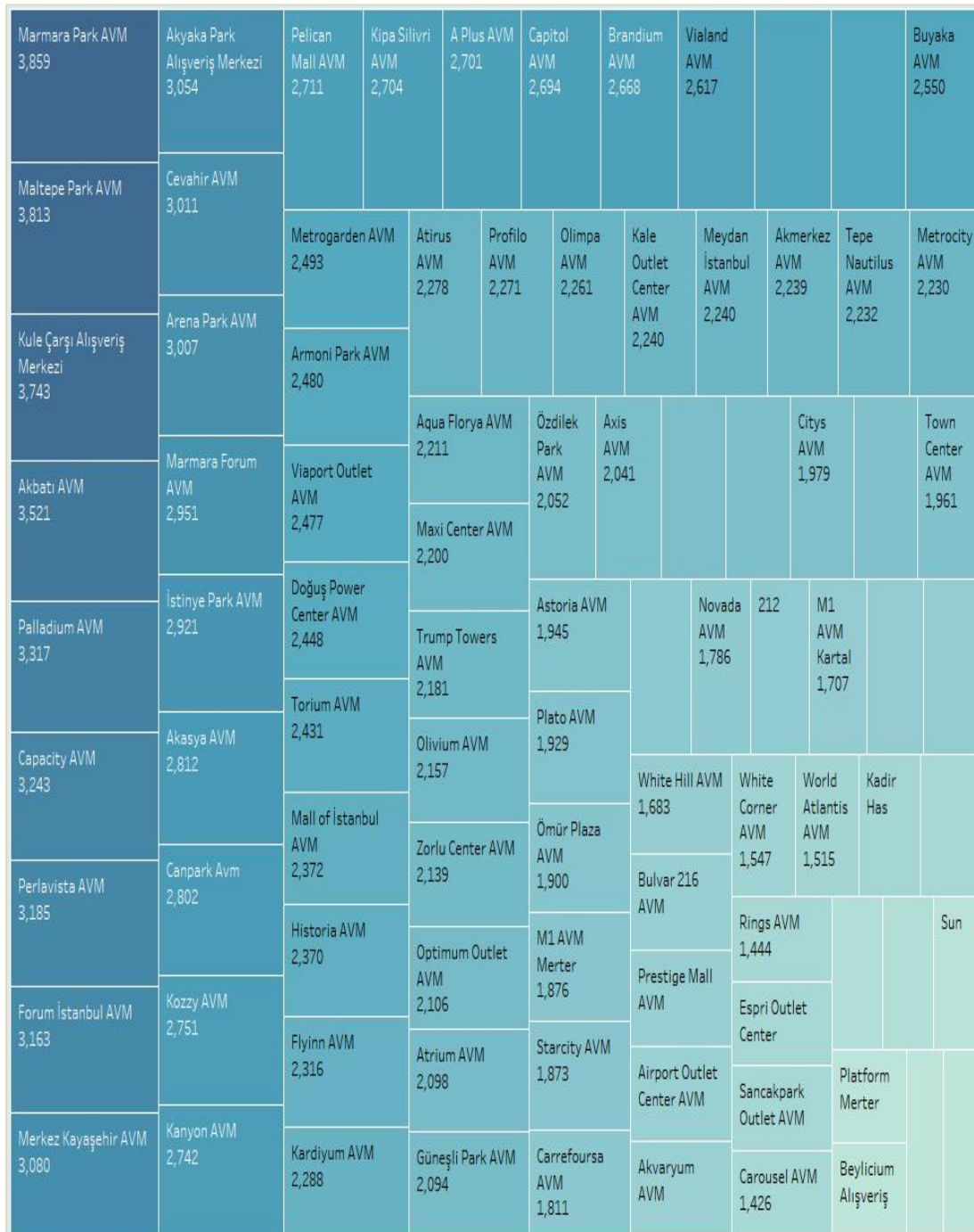
Bu bilgiler ışığında AVM'lerin mekan popülerlik analizi yapılmıştır. Bu analiz yapılırken aşağıdaki ölçüm formülü kullanılmıştır:

- a) x mekanda yapılan kümülatif check-in sayısı (KC_x)
- b) x mekanın kümülatif tekil kullanıcı sayısı (KK_x)
- c) x mekanın kullanıcı check-in sıklığı (KCS_x)

KC_x ve KK_x verileri Foursquare API üzerinden elde edilmiş olup, KCS_x hesaplanmıştır:

$$KCS_x = \frac{KC_x}{KK_x}$$

Şekil 5.5 ‘te görüleceği üzere KCS değeri en yüksek olan AVM, Marmara Park AVM’dir. Ardından Maltepe Park AVM ve Kule Çarşı AVM onu takip etmektedir. Şekilde renk değişimi yoğun olan alanlar yüksek frekansları temsil ederken, giderek açılan renk tonu frekansın azaldığı anlamına gelmektedir. Bu analiz bir bakıma kullanıcıların ne sıklıkta bu AVM’yi tercih ettiğini göstermektedir. Bu açıdan bakıldığında Marmara Park AVM’de bir kişinin en az 3,86 kez burada olduğu görülür.



Şekil 5.5: AVM Popülerliği (Frekansı)

Genellikle daha büyük KCS bir kullanıcıdan ortalama olarak daha fazla check-in yapmayı gösterir. Mekanların %13'ünde bir kullanıcı 3'ten fazla, %47'si 2'den fazla ve geri kalanları 1'den fazla check-in yapmıştır.

Bu frekans analizinin yanında,

- AVM'lerin puan sıralaması,
- AVM'lerin yorum sayısı yüzdelik dağılım grafiği analizleri yapılmıştır.

Analizler ile ilgili görseller Tableau programında hazırlanmıştır.

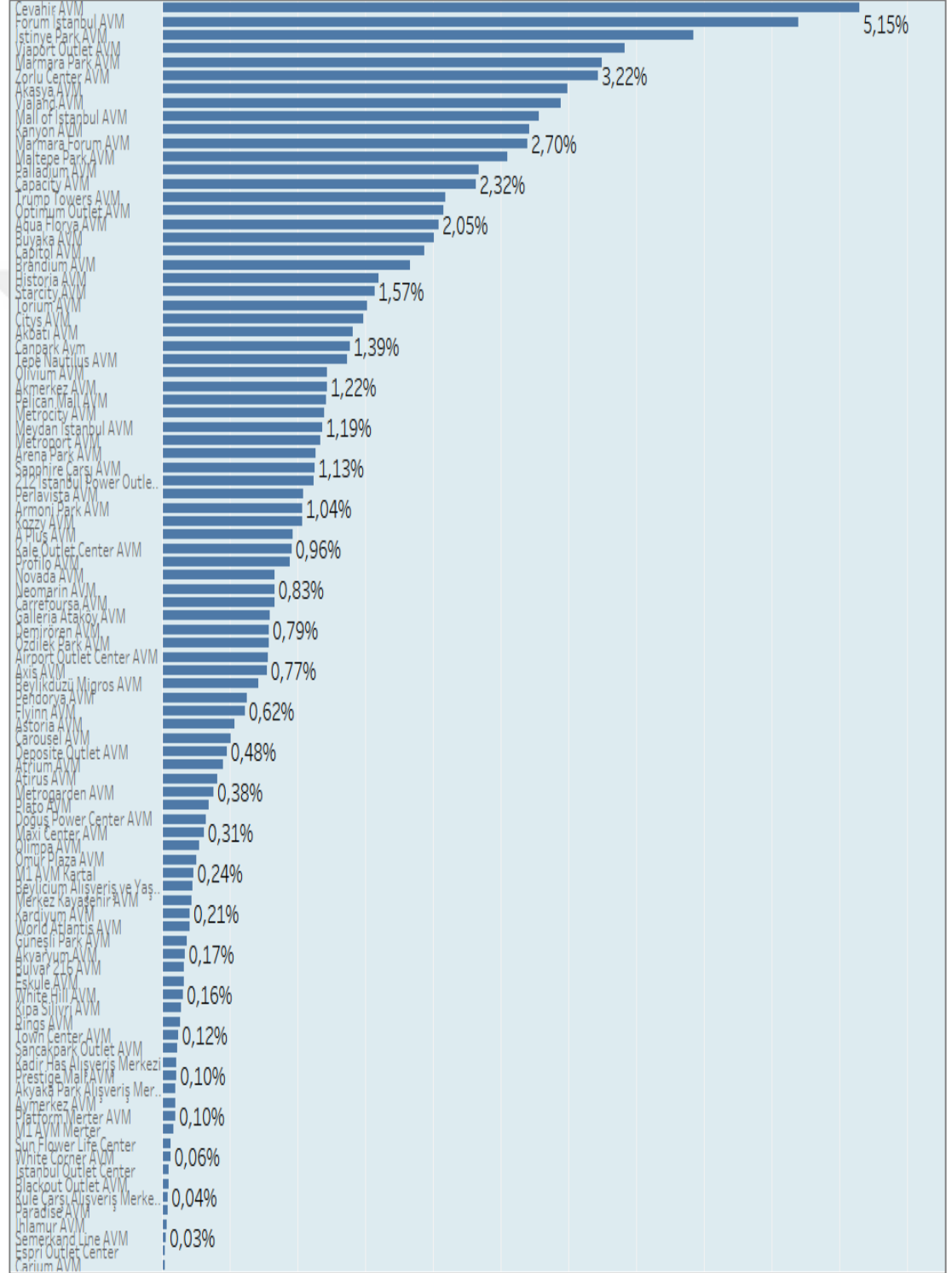
Şekil 5.6'da AVM Foursquare Puan Durumu gösterilmiştir. Her AVM için Foursquare uygulamasında "Puan Ver" özelliği kullanılarak bir puanlama yapılmaktadır. Analiz ettiğimiz 94 AVM için puan durumları çekilmiştir.



Şekil 5.6: Puan Durumu Genel Görünümü

Şekil 5.7'de AVM'lerin aldıklarını puanlara göre içte en yüksek puan olmak üzere dışa doğru azalan puanlara sahip AVM'ler gösterilmiştir.

Şekil 5.8’de ise AVM’lere bırakılan yorumların yüzdeler grafiği çizilmiştir. İlgili AVM’nin tüm AVM’ler içerisinde yorum bazında yüzde kaçlık bir paya sahip olduğunu gösterir. Bu bakımdan incelendiğinde %5,15 yüzdelikle Cevahir AVM açık bir şekilde en çok yoruma sahip olan AVM’dir. Onu Forum İstanbul ve İstinye Park AVM takip etmektedir.



Şekil 5.8: Yorumların AVM Bazlı Yüzdelik Dağılımı

Çizelge 5.2: Örnek Düzenlenmiş Yorum Verisi.

TipId	TipText
519b053f498e27ed86efb347	En üst kat çocuklu aileler için keyifli sinema katında buz pisti olduğunu bilmeyenler asıl eğlenceyi kaçırıyorlar Ailenizle keyif alabileceğiniz bir yer Apaçiler her yerde
517906d3e4b0c03eedb149be	Kırkpınar lokantası 10 numara hele Laz böreği mükemmel ciğer muhteşem çalışanlar temiz ve güler yüzlü tebrikler Kırkpınar
5236d5c6498e2b20adcb74f4	Güzel burası Daha güzel bir sey biliyorum ama Şu popeyes'in dürüm tavuğu Yok böyle lezzet Lavaş ekmek ve tavuğun muhteşem uyumu Deneyin

Bir yandan düzenlenen veriler içerisinde türkçe yazım yanlışları, harf eksiklikleri olan kelimelerin var olduğu tespit edildi. Bu içeriklerin düzenlenebilmesi için Türkçe Doğal Dil İşleme Zincirleri araştırılmıştır [99]. Türkçe için doğal dil işlemeyi gerçekleştiren yazılımlardan ITU NLP [100] ile çalışılmıştır. Ancak yeterli düzeyde doğruluk içermediği gözlemlenerek bu çalışmanın manuel olarak yapılmasına karar verilmiştir. 55,303 veri Microsoft Office programlarından Word yazım denetimi ile düzenlenerek olabilecek en düzgün veri seti elde edilmiştir.

Konu modellemeyi daha anlamlı yapmak için verilerin köklerine ayrıştırılması için araştırmalar yapıldı. Bu incelemede, Türkçe için geliştirilmiş açık kaynak kodlu doğal dil işleme kütüphanesi olan Zemberek [101] ile ayrıştırma çalışması yapıldı. Verilen bir kelimenin kök ve eklerine ayrıştırılarak çözümlenmesini, doğru yazılmayan kelimeler için öneriler sunarak yanlış yazımların düzeltilmesini sağlamayı amaçlayan bu yazılım ile tüm veri köklerine ayrıştırılmıştır. Ancak elde edilen çalışma incelendiğinde kök ayrımlarının doğru yapılmadığı görülmüş olup bu işlemde vazgeçilmiştir.

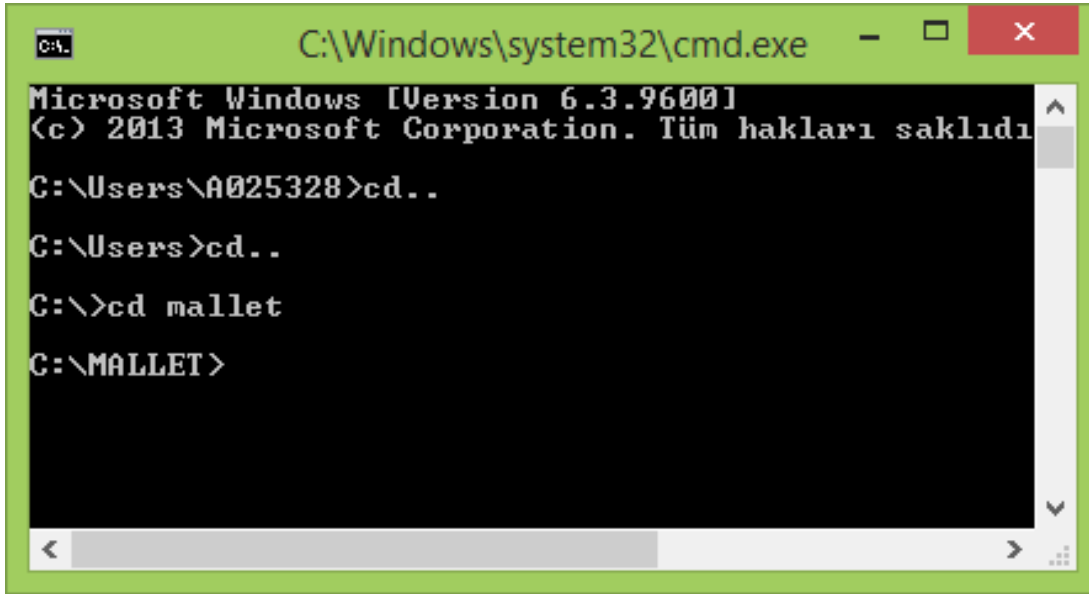
5.1.3.2 MALLET ile analiz

Veri ön işlemesi tamamlandıktan sonra veri Mallet yazılımı aracılığıyla analiz edildi.

MALLET, çeşitli doğal dil işleme görevleri için sağlam ve hızlı uygulamalar sunmayı amaçlayan, Java tabanlı makine öğrenme araç takımıdır. MALLET, konu modelinin özelliklerini çıkarmak için Gibbs Örneklemesini kullanır: Bu, Markov Zinciri Monte Carlo metodu olup, rasgele sayı üretici kullanarak modeldeki bazı parametreleri diğerlerinin şimdiki değerine dayanarak tekrar tekrar örneklemektedir. Her seferinde farklı sonuçlar çıkarması, olasılıksal, istatistiksel bir model olduğu için normaldir.

MALLET komut satırında çalışmaktadır. MALLET aşamaları, verinin .txt formatında içeriye atılması, verinin eğitilmesi ve hiperparametrelerin optimize edilmesi şeklindedir. Mallet her .txt. belgesini tek belge olarak algılar.

Exceldeki 55,303 yorum verisinin her birinin .txt formatına çevrilmesi için Python dilinde yazılan bir komut kullanılmıştır. EK D’de bu kodlara yer verilmiştir. Bu komut ile her bir yorum ID’si .txt dosyanın ismi olarak alınarak tüm yorumlar bu formata çevrilmiştir. Bu işlem tamamlandıktan sonra C sürücüsünde yer alan MALLET uygulaması komut satırından çağırılır



```
C:\Windows\system32\cmd.exe
Microsoft Windows [Version 6.3.9600]
(c) 2013 Microsoft Corporation. Tüm hakları saklıdır.
C:\Users\A025328>cd..
C:\Users>cd..
C:\>cd mallet
C:\MALLET>
```

Şekil 5.10: Komut Ekranından Mallet Çağırımı

İlk adım olarak 55.303 adet .txt dosyayı uygulamanın içine import-dir komutuyla aktarılır. Bu işlem için Şekil 5.10’da yer alan komut kullanılır.

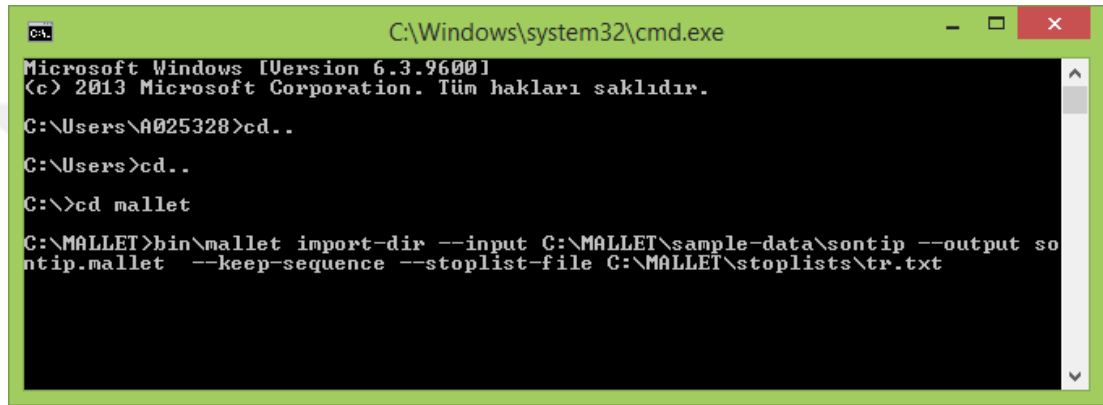
Burada yer alan komutların açıklaması aşağıdadır:

--input: .txt formatındaki verilerin olduğu klasör yolu gösterilir ve veriyi buradan alması gerektiği belirtilir.

--output: Çıktı olarak vereceği .mallet formatındaki dosyanın ismini buraya dosyaismi.mallet şeklinde eklenir.

--keep sequence: Bu komut belgeyi kelime özellik sayılarından oluşan bir vektör yerine bir dizi kelime özellikleri olarak korur. MALLET konu modelleme araç takımı özellik dizileri gerektirir [102].

--stoplist-file: Bu duraklama kelimeleri listesi metin verisi içerisinde yer alan ve sonuç içerisinde gösterilmesi istenmeyen sayı,”ve”, “şey”,”eğer” gibi bulunma frekansları yüksek olan kelimeleri içerir. Duraklama sözcüklerini kaldırma, anlambilim analizinin önemli süreçlerinden biridir. Duraklama listesinde yer alan kelimeler, sistem tarafından filtrelenen kelimelerdir. MALLET içerisinde Türkçe için bir liste bulunmadığından bu liste manuel oluşturulmuş ve eklenmiştir. Bu listeye AVM isimleri de eklenmiştir. Bu liste içerisinde yer alan türkçe karakterleri olan kelimeleri, uygulamanın görmediği tespit edilerek veri seti içerisinde bu karakterlerin düzeltimi yapılarak tekrar içeri atılmıştır.



```
C:\Windows\system32\cmd.exe
Microsoft Windows [Version 6.3.9600]
(c) 2013 Microsoft Corporation. Tüm hakları saklıdır.
C:\Users\A025328>cd..
C:\Users>cd..
C:\>cd mallet
C:\MALLET>bin\mallet import-dir --input C:\MALLET\sample-data\sontip --output sontip.mallet --keep-sequence --stoplist-file C:\MALLET\stoplists\tr.txt
```

Şekil 5.11: Verilerin Mallet Uygulamasına Aktarımı

Veriler içeri aktarıldıktan sonra bu verilerin eğitilmesi aşamasına geçilir. Bunun için Şekil 5.11’de görülen train-topics komutu kullanılır.

Burada yer alan komutların açıklaması aşağıdadır:

--input: bir önceki komutla oluşturulan .mallet dosyası burada girdi olur.

--num-topics: Analiz edilen veriden kaç konu çıkarılması isteniyorsa bu sayı girilir. Manuel girilen tek parametre konu sayısıdır. Konu sayısı tamamen istenilen çıktıya ve veri büyüklüğüne bağlıdır. Genel bir bakış istenirse 10 konu kadar bakılabilir. Ancak daha detaylı incelemek istenirse sayıyı 100-200’e çıkarılabilir [103]. Yapılan denemeler sonucu çalışmamıza en uygun sonucu üreten konu sayısı 6 olarak belirlenmiştir. Ayrıca bazı çalışmalarda konu sayısının Log-Likelihood (LL) değeri baz alınarak yapıldığı görülmüştür. Bu çıkarım yüksek konu sayılarında daha çok işe yaramaktadır. Konu sayısı belirlenirken bu değer de incelenmiştir.

--output-state: Bu komut, konu atamaları ile birlikte korpus içerisindeki kelimeleri içeren sıkıştırılmış bir metin dosyası çıktısı verir. Bu dosya biçimi, Java tabanlı

olmayan yazılımlar tarafından kolayca ayrıştırılabilir ve kullanılabilir. gz ile biten bir dosya adı verilir. Zipli bir dosya çıkartır.

--output-topic-keys: Bu dosya, her konunun üst k kelimelerinden oluşan bir "anahtar" içerir. Bu çıktı, modelin çalışma halinin yanı sıra modelin sonuçlarını kontrol etmede de yararlı olabilir. Buna ek olarak, bu dosya her konunun Dirichlet parametresini bildirir.

--output-doc-topics: Bu çıktı, belgelerin konularla olan ilişkisini ortaya çıkarır.

```
C:\Windows\system32\cmd.exe - bin\mallet train-topics --input sontip.mallet ...
C:\MALLT>bin\mallet train-topics --input sontip.mallet --num-topics 6 --output
-state topic-state.gz --output-topic-keys sontip_keys.txt --output-doc-topics so
ntip_compostion.txt
Mallet LDA: 6 topics, 3 topic bits, 111 topic mask
Data loaded.
max tokens: 27
total tokens: 316185
<10> LL/token: -9,84649
<20> LL/token: -9,6063
<30> LL/token: -9,45104
<40> LL/token: -9,36494

0      0,83333 otopark de-şil kâatı hiğ berbat bîyle otopark- yırlıyen sa
ğma nas-ıl saat şey gıın buras- laz-m ayn- iğın gıardı-şım gııvenlik para
1      0,83333 şey tavsiye buras- harika batşka iğın hiğ hım gereken bıtı
n iyi art-k gızel de-şil arap bugıın tım mıkemmel kayn-yor yerde
2      0,83333 iğın gızel sakın ğok al-şveriş mekton yemek kalabal-k kah
ve yak-n ideal boş pazar sessiz oldu-şu iği sıper keyifli raşmen gıını
3      0,83333 ğok gızel iyi sinema buray- seviyorum mekton de-şil harika ol
muş sıper ferah batşar-ıl- sakın numara sinemas- hoş yemek kıığık mıkemmel
4      0,83333 al-şveriş gızel merkezi ğok iyi bıyık ferah rahat anadolu
kaliteli şeyi yakas-n-n sevdi-şim maşazalar avmsi kıığık maşaza uygun bi
rğok istanbul00un
5      0,83333 ğok gızel kalabal-k yemek ağ-k olmas- iyi olmuş soşuk de
şil hava kâatı buras- bıyık sıper hiğ gereksiz s-k-c- dolu iği

<50> LL/token: -9,31378
<60> LL/token: -9,27735
<70> LL/token: -9,2516
<80> LL/token: -9,22699
<90> LL/token: -9,21231

0      0,83333 otopark berbat kâatı otopark- nas-ıl yırlıyen buras- saat sa
ğma de-şil bîyle iğın gııvenlik kat para laz-m rezalet hiğ önce işreniğ
1      0,83333 şey buras- tavsiye batşka işte yeni harika arap hiğ gereken
art-k hım şimdi hiğbir apaği iğinde bîyle iyidir mıthiş kayn-yor
2      0,83333 iğın al-şveriş yemek gızel sakın kahve yak-n mekton ideal
boş sinema pazar oldu-şu iyi gıını tercih raşmen iği keyifi huzur
3      0,83333 ğok gızel harika iyi mekton olmuş sıper buray- seviyorum si
nema sinemas- mıkemmel hoş gerğekten numara gayet ferah ortam batşar-ıl- mu
hteşem
4      0,83333 al-şveriş merkezi gızel iyi ğok bıyık ferah rahat sakın a
nadolu kaliteli şeyi maşaza yakas-n-n sevdi-şim kıığık olmas- maşazalar
avmsi şey
5      0,83333 ğok kalabal-k gızel de-şil yemek ağ-k iyi olmas- hiğ gere
ksiz soşuk fena kâatı iği hava s-k-c- bıyık buras- maşazalar buz

<100> LL/token: -9,20049
<110> LL/token: -9,18242
<120> LL/token: -9,16967
```

Şekil 5.12: Örnek Veri Eğitme Ekran Görüntüsü

MALLET 1000 iterasyon ile modeli geliştirir. Bu iterasyon sayısını değiştirmek mümkündür. Ancak yeterli iterasyon olarak default sayı olan 1000 kullanılmıştır. Ayrıca modelde α ve β değerlerinin default değerleri olan $\alpha=5/T$, $\beta=0,01$ [104] kullanılmıştır.

Çizelge 5.3'te görüldüğü gibi parametre değeri $\alpha =5/T$ ile bulunmaktadır. Konu sayımız 6 olduğu için her bir konu için parametre değerimiz $5/6= 0,83333$ olarak bulunmuştur. MALLET uygulaması da bu formülü kullanmaktadır.

MALLET Gibbs Örnekleme algoritmasını kullandığından dolayı ve olasılıksal, istatistiksel bir model olduğu için her seferinde farklı sonuçlar vermesi normaldir.

Çizelge 5.3: Output-Topic-Keys Çıktı Verisi.

Konu numarası	Dirichlet parametresi	Anahtarlar
1	0,83333	yemek tavsiye kahve süper harika iyi uygun mutlaka katı mükemmel iyidir keyifi özellikle bol starbucks üst indirim alanı sigara muhteşem
2	0,83333	burayı seviyorum gün pazar apaçi gereken boş hala dolu arkadaş nasıl gez olacak mekanı para günleri gelmiş bide önce mescit
3	0,83333	otopark kötü berbat otoparkı yürüyen saat saçma zor güvenlik gelmeyin iğrenç kat rezalet gordugum giriş sakın otoparka trafik otoparkta gormedim
4	0,83333	kalabalık güzel açık boş gereksiz soğuk arap hava içi sıkıcı dışında dolu pek değer fena buz kışın eski büyük havası
5	0,83333	güzel sakın mekân harika sinema iyi küçük rahat süper yakın sineması ideal kalabalık hoş mükemmel keyifli sessiz nezih merkezi numara
6	0,83333	iyi güzel merkezi büyük ferah anadolu mağazalar mağaza kaliteli sevdiğim yakasının yeni tüm avmsi geniş başarılı istanbul'un mekân markalar marka

Ancak bu modeli daha da geliştirmek ve konuların birbirinden daha çok ayrışmasını sağlamak adına –optimize interval [SAYI] komutu ile hiperparametre optimizasyonu gerçekleştirilmiştir. Bu seçenek, modelin bazı konuların diğerlerinden daha belirgin olmasını sağlayarak verilere daha iyi uymasını sağlayan, hiperparametre optimizasyonunu etkinleştirir. Burada belirlenen sayı her iterasyonda kaçarlık konu paketlerini karşılaştırıp optimize etmesi gerektiğini söyler [105]. Optimizasyonun her 10 iterasyon için yapılması genel kabul görmüş bir durumdur [106].

Özetle, alfa ve beta hiperparametreleri sırasıyla her bir konudaki ve her belgedeki konuların dağılım profilini etkiler. Ve hiperparametreleri belirli bir seviyeye

ayarlamak yerine, hiperparametre optimizasyonu, belirli bir modelleme sürecinde düzeltilmiş hiperparametrelerin etkisinin ne kadar kuvvetli olabileceğini etkiler. Bu komut olmadığında her konunun aynı ağırlıkta olduğunu söyler.

Çizelge 5.4’te görüleceği gibi –optimize interval komutu ile veri eğitimi yapıldığından parametre değerleri üzerinde belirgin farklılıklar ortaya çıkarılmıştır.

Çizelge 5.4: Optimize Edilen Konular.

Konu numarası	Optimize Dirichlet parametresi	Anahtarlar
1	0,08715	kalabalık arap dolu apaçi pazar gez kaynıyor günleri boş gün gelmeyin gel araplar gelmiş arkadaş mekân mekânı kızlar amele git
2	0,08492	anadolu iyi yakasının güzel yeni merkezi avmsi eski avm'si istanbul'un iyisi göz istanbul'un hala olacak yakasında avmlerinden kaliteli büyük metro
3	0,06587	buz sinema film nasıl mescit para bowlingi eskiden geldik saat ediyorum beş yağmur güzeldi pateni gün uzun önce vardı bowling
4	0,37682	güzel iyi kalabalık merkezi mekân sakın büyük ferah yemek harika rahat süper sinema küçük seviyorum mağaza burayı mağazalar gayet kaliteli
5	0,09655	kahve yemek tavsiye harika mutlaka güzel süper burger iyi mükemmel indirim ugramadan starbucks sigara cafe gitmeyin keyifi çay kahvesi uğrayın
6	0,10239	otopark kötü berbat yürüyen saçma otoparkı çıkış saat giriş zor otoparka rezalet otoparkta nasıl gereksiz iğrenç bebek görmedim kat gelmeyin

Konuların çıkarılması işleminden sonra bu konuların belirli etiketler altında toplanması çalışmasına geçilmiştir.

Kelimelerin hangi konulara ait olduğunu gösteren çıktı ile ilgili bir işlem yapılmamıştır.

5.1.3.3 Konuların etiketlenmesi

Çıkarılan konu anahtarları incelenerek en yakın olan etiket belirlenip etiketleme işlemi gerçekleştirilmiştir. Bu etiketleme aşağıdaki şekilde yapılmıştır.

Çizelge 5.5: Konuların Etiketlenmesi.

Konu Numarası	Anahtarlar	Etiket
1	kalabalık arap dolu apaçi pazar gez kaynıyor günleri boş gün gelmeyin gel araplar gelmiş arkadaş mekân mekânı kızlar ameile git	Kitle-Yoğunluk
2	anadolu iyi yakasının güzel yeni merkezi avmsi eski avm'si istanbul'un iyisi göz istanbul'un hala olacak yakasında avmlerinden kaliteli büyük metro	AVM Konumu
3	buz sinema film nasıl mescit para bowlingi eskiden geldik saat ediyorum beş yağmur güzeldi pateni gün uzun önce vardı bowling	Eğlence-Ek Hizmetler
4	güzel iyi kalabalık merkezi mekân sakın büyük ferah yemek harika rahat süper sinema küçük seviyorum mağaza burayı mağazalar gayet kaliteli	Genel AVM Atmosferi
5	kahve yemek tavsiye harika mutlaka güzel süper burger iyi mükemmel indirim uğramadan starbucks sigara cafe gitmeyin keyifi çay kahvesi uğrayın	Yiyecek-İçecek Hizmetleri
6	otopark kötü berbat yürüyen saçma otoparkı çıkış saat giriş zor otoparka rezalet otoparkta nasıl gereksiz iğrenç bebek görmedim kat gelmeyin	Otopark

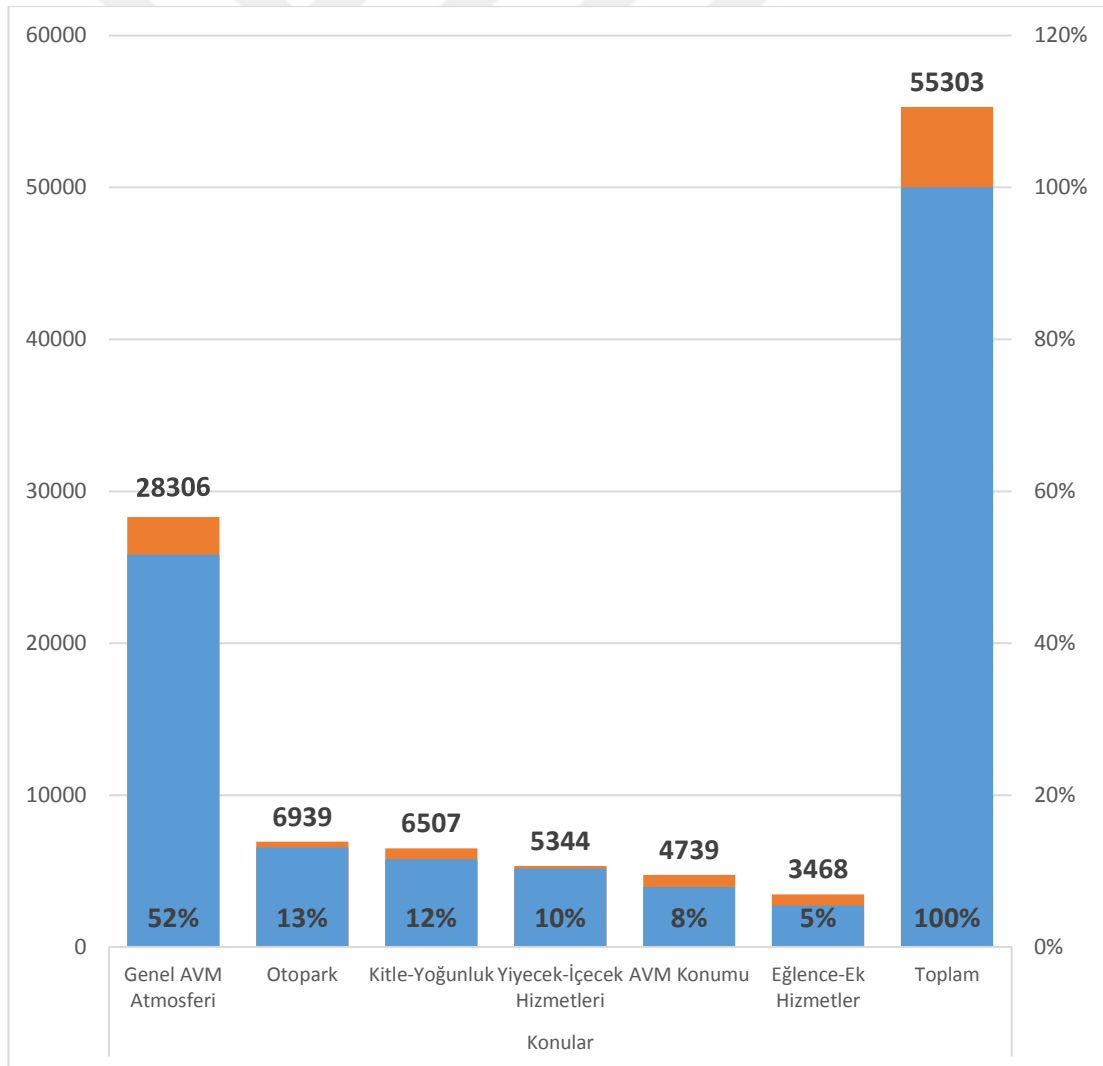
Bu etiketleme işleminin ardından MALLEET çıktılarında biri olan belgelerin konulara göre olasılıkları değerlendirilerek tüm AVM'ler için konu dağılımları çıkarılmıştır.



6. SONUÇ VE ÖNERİLER

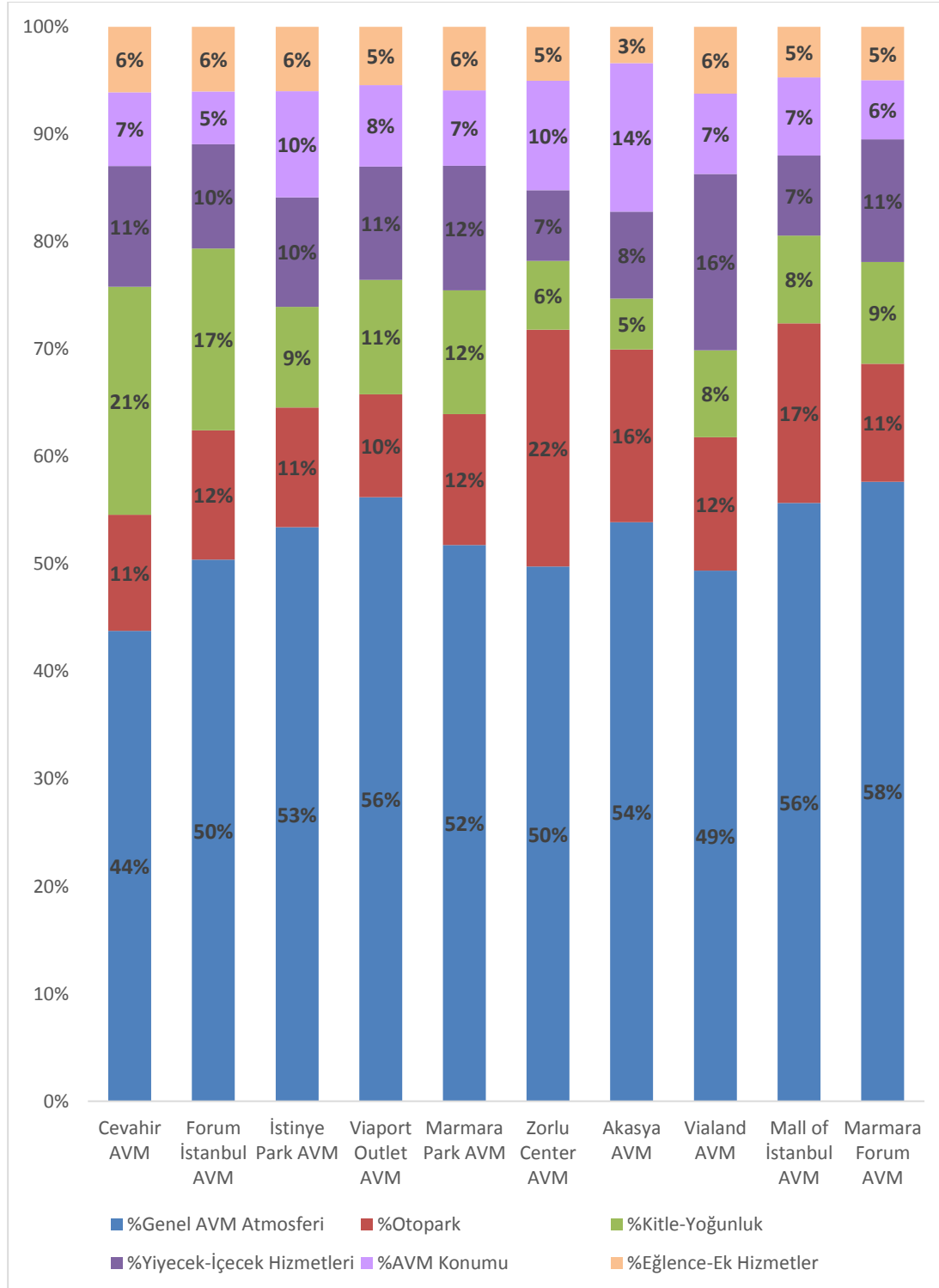
6.1 Sonuçlar ve Yorumlar

Yapılan analiz ve etiketleme işlemi sonucunda tüm yorumlara göre konuların dağılım grafiği Şekil 6.1’de gösterildiği gibidir. Analiz edilen 55,303 yorumun %52’lik ve en yüksek yüzdeye sahip kısmında Genel AVM Atmosferi hakkında görüş bildirilmiş, daha sonrasında sırasıyla Otopark, Kitle-Yoğunluk, Yiyecek-İçecek Hizmetleri, AVM Konumu ve Eğlence-Ek Hizmetler konularında yorum bırakıldığı sonucuna ulaşılmıştır. Şekil 6.1’de her bir konuya ait yorum sayıları turuncu bloklar olarak, bu yorumların tüm yorumlara oranı mavi bloklar içerisinde gösterilmiştir.



Şekil 6.1: Tüm Yorumlar Bazında Konu Dağılım Grafiği

Yorum sayısı en yüksek olan ilk 10 alışveriş merkezi için gösterim Şekil 6.2’de gösterilirken, bu 10 alışveriş merkezinin her biri için detaylı analizler verilmiştir. Kalan 84 alışveriş merkezi için çıkan sonuçlar EK B’de verilmiştir. Ayrıca incelenmesi adına yorum sayıları bazında konu dağılımları EK C’de verilmiştir.



Şekil 6.2: En Yüksek Yoruma Sahip İlk 10 AVM Konu Dağılımları Grafiği

Şekil 6.2’de görüldüğü üzere 10 AVM’de de en çok konuşulan konu Genel AVM Atmosferi üzerine olmuştur. Bu şekilde gösterilen her AVM için detaylı analizler yapılmıştır. Oranları verilen yorumların sayısal değerleri Çizelge 6.1’de belirtilmiştir.

Çizelge 6.1: AVM Yorum Sayılarının Konular Bazındaki Değerleri.

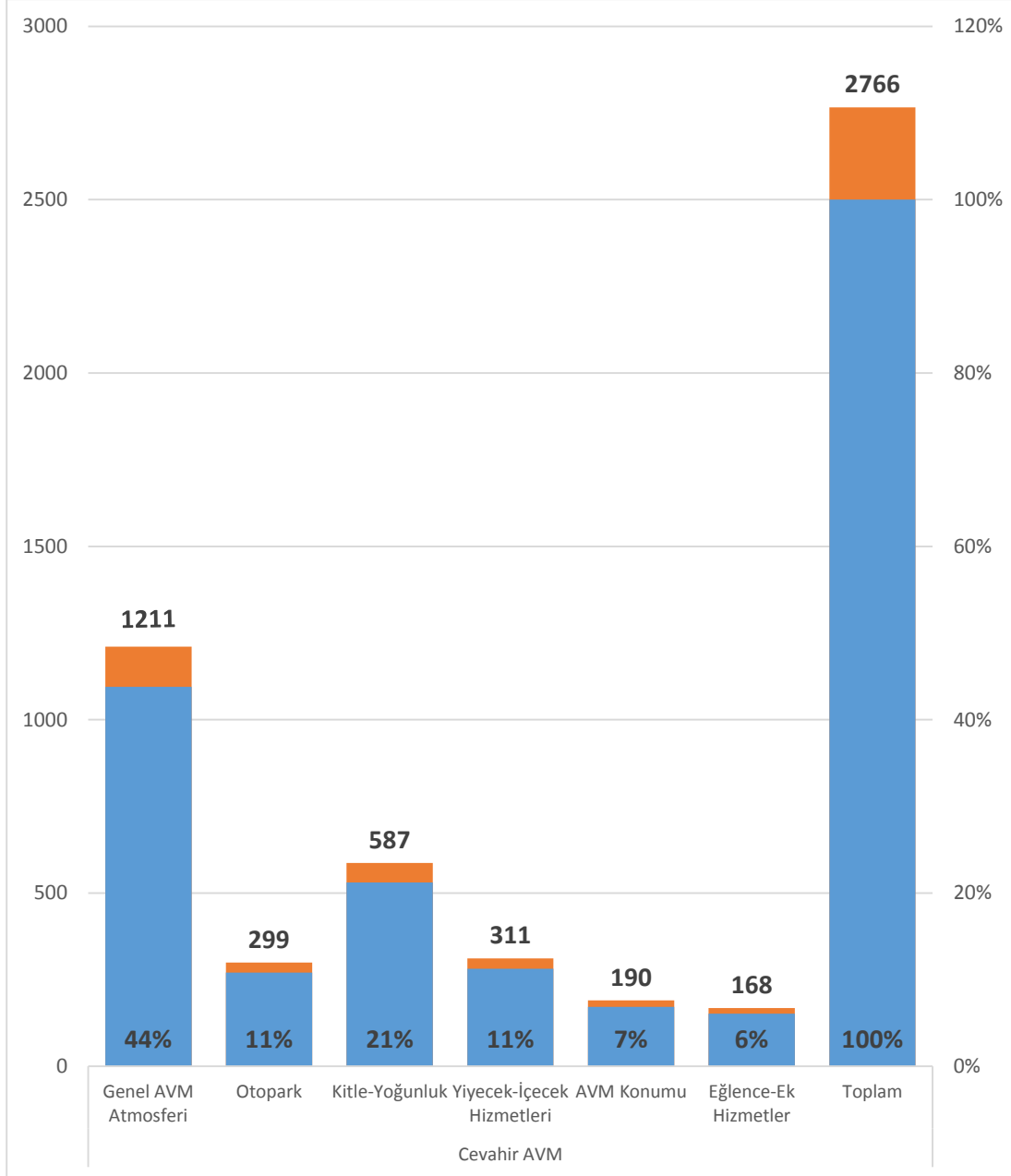
AVM'ler	Genel AVM Atmosferi	Otopark	Kitle-Yoğunluk	Yiyecek-İçecek Hizmetleri	AVM Konumu	Eğlence-Ek Hizmetler
Cevahir AVM	1211	299	587	311	190	168
Forum İstanbul AVM	1329	317	446	256	130	158
İstinye Park AVM	1038	216	182	198	192	116
Viaport Outlet AVM	1082	184	205	203	146	104
Marmara Park AVM	921	217	205	206	125	105
Zorlu Center AVM	864	383	111	114	177	87
Akasya AVM	925	276	81	139	237	58
Vialand AVM	819	206	134	272	124	103
Mall of İstanbul AVM	866	260	127	116	113	73
Marmara Forum AVM	859	164	141	170	82	74
Toplam	9914	2522	2219	1985	1516	1046

Şekil 6.3’te Cevahir AVM için yorum sayıları ve konulara dağılım oranları verilmiştir.

Cevahir AVM’ye bırakılan yorumların %44’ü Genel AVM Atmosferi hakkında olmuşken, ikinci sırada %21’lik oranla Kitle-Yoğunluk konusu hakkındadır. Şekil 6.2’deki grafikte yer alan diğer AVM’ler içerisinde en yüksek orana sahiptir. Bu AVM yorumları üzerinde veri keşfi yaparken yapılan yorumlarda Arap turistlerden çokça bahsedildiği görülmüştü, analizden çıkan bu sonuçla da örtüştüğü görülmektedir. Üçüncü ve dördüncü sırada %11’lik oranla Otopark ve Yiyecek İçecek Hizmetleri,

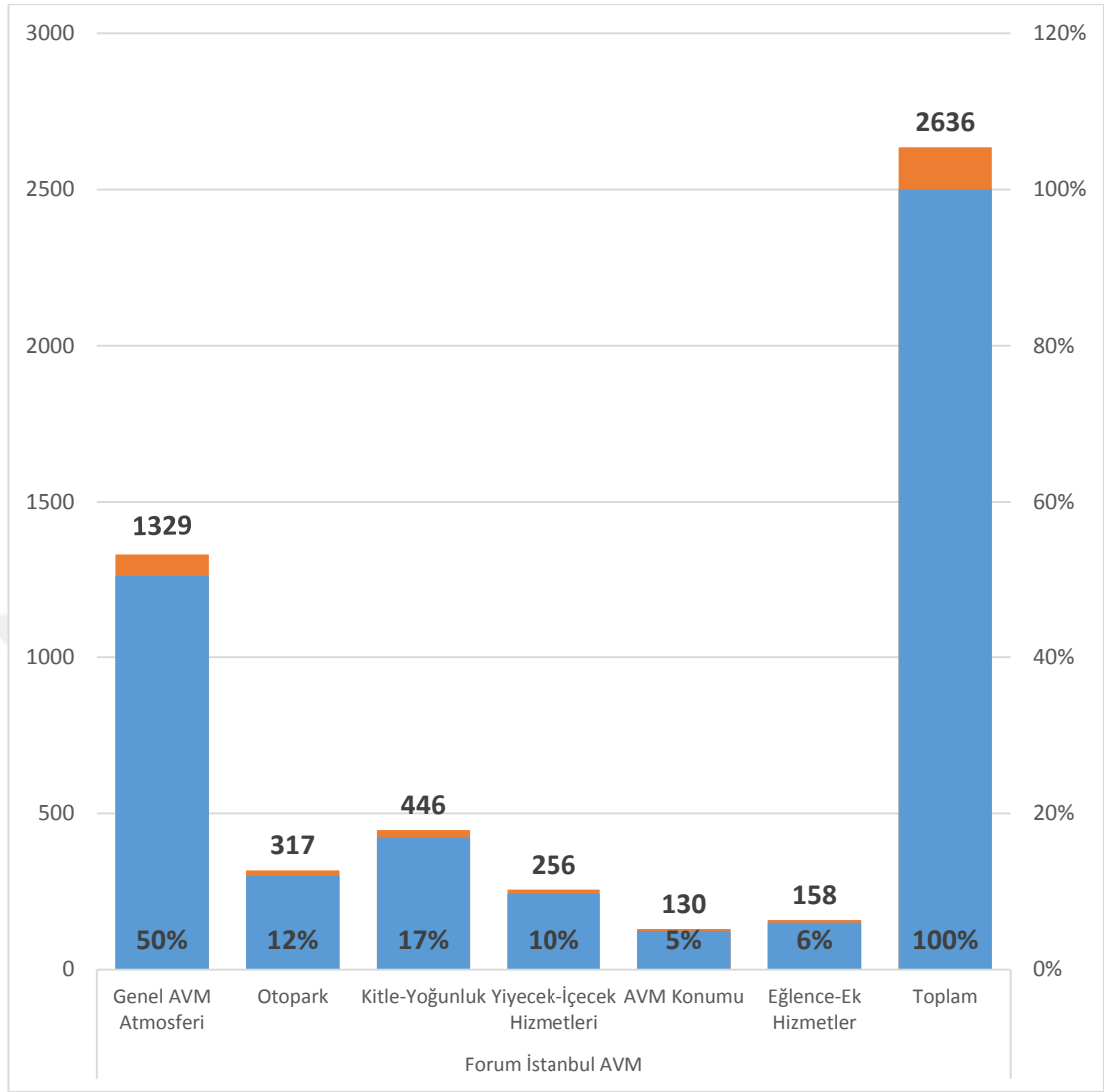
beşinci sırada %7 oranla AVM Konumu ve %6 oranla Eğlence-Ek Hizmetler hakkında konuşulduğu ortaya çıkarılmıştır.

En az orana sahip olan Eğlence-Ek Hizmetler konusunda AVM yönetimi daha çok konuşulmasını sağlayacak çalışmalar yapabilir. Bu hizmetlerini geliştirebilir. Cevahir AVM'nin toplam 2766 adet yorumunun olduğu görülür.



Şekil 6.3: Cevahir AVM Konu Dağılım Grafiği

Şekil 6.4'te Forum İstanbul AVM için yapılan analiz gösterilmiştir.



Şekil 6.4: Forum İstanbul AVM Konu Dağılım Grafiği

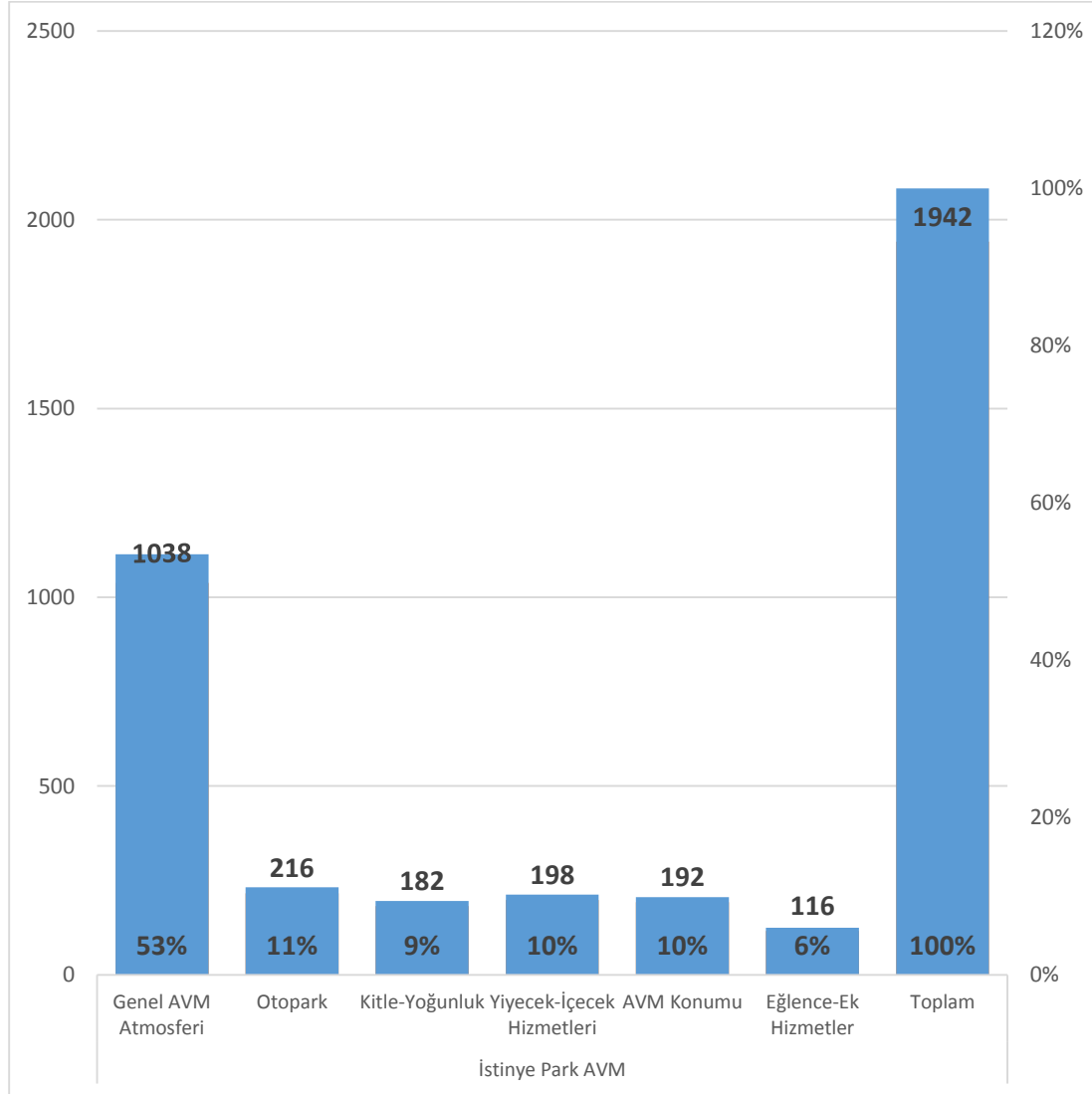
Forum İstanbul AVM için bırakılan yorumların yarısının Genel AVM Atmosferi hakkında olduğu görülüyor. %17'lik oranla Kitle-Yoğunluk konusu ikinci sırada, %12'lik oranla Otopark konusu üçüncü sıradadır. Otopark konusunu %10'luk oranla Yiyecek-İçecek Hizmetleri takip ediyor, %6'lık oranla Eğlence-Ek Hizmetler konusu gelmektedir.

En düşük oran olan %5 ile AVM konumu hakkında görüş bildirilmiş. AVM Konumu konusunda Şekil 6.2'deki grafikte yer alan diğer AVM'ler içerisinde en düşük orana sahip olduğu görülmektedir.

Şekil 6.5'te İstinye Park AVM için yapılan analiz gösterilmiştir.

İstinye Park AVM için bırakılan yorumların yarıdan fazlası yani %53'lük kısmı Genel AVM Atmosferi hakkında olmuştur. %11'lik oranla Otopark konusu ve %10'luk

oranlarla AVM Konumu ve Yiyecek-İçecek Hizmetleri konuları ardından gelmiştir. %9'luk oranda Kitle-Yoğunluk konusu ve en düşük oran olan %6 ile de Eğlence-Ek Hizmetler hakkında görüş bildirildiği sonucuna ulaşılmıştır.



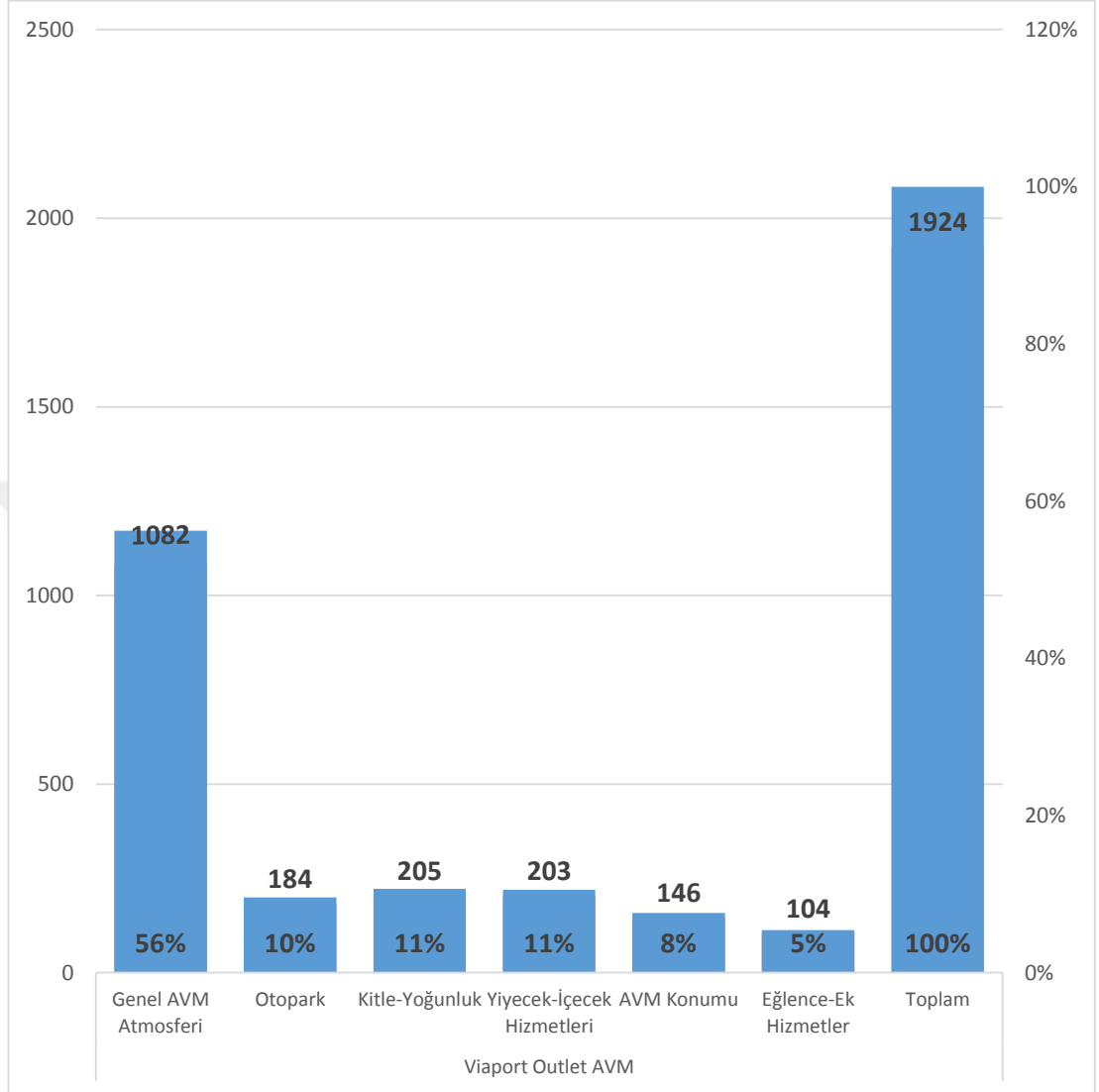
Şekil 6.5: İstinye Park AVM Konu Dağılım Grafiği

Şekil 6.6'te Viaport Outlet AVM için yapılan analiz gösterilmiştir.

Viaport Outlet AVM için bırakılan yorumların %56'lık kısmı Genel AVM Atmosferi hakkında olmuştur. Ardından %11'erlik oranla Kitle-Yoğunluk ve Yiyecek-İçecek Hizmetleri konuları gelmiştir.

%10'luk oranda Otopark konusu, %8'lik oranla AVM Konumu ve en düşük oran olan %5 ile de Eğlence-Ek Hizmetler hakkında görüş bildirildiği sonucuna ulaşılmıştır. Otopark konusunda Şekil 6.2'deki grafikte yer alan diğer AVM'ler içerisinde en düşük orana sahip olduğu görülmektedir. Bu belki de otopark konusunda yeterli hizmeti

sağladığı anlamına gelmektedir, çünkü incelenen yorumlarda otopark konusunda genelde olumsuz bir durum varsa yorum bırakıldığı görülmüştür.



Şekil 6.6: Viaport Outlet AVM Konu Dağılım Grafiği

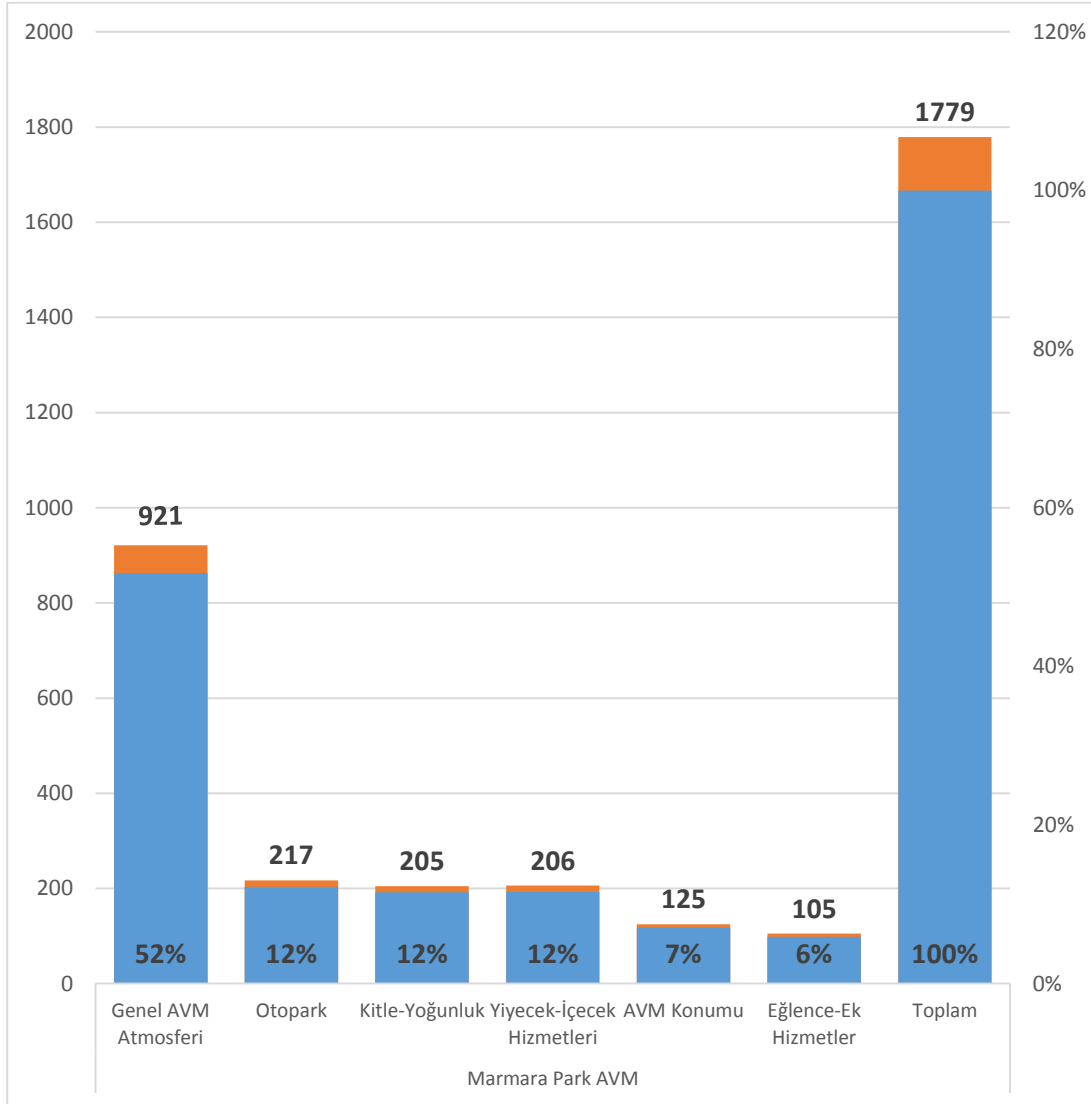
Şekil 6.7’te Marmara Park AVM için yapılan analiz gösterilmiştir.

Marmara Park AVM için bırakılan yorumların %52’lik kısmı Genel AVM Atmosferi hakkında olmuştur. Ardından %12’şerlik oranla Otopark, Kitle-Yoğunluk ve Yiyecek-İçecek Hizmetleri konuları gelmiştir.

%7’lik oranda AVM Konumu konusu, %6’lık en düşük oranla Eğlence-Ek Hizmetler hakkında görüş bildirildiği sonucuna ulaşılmıştır.

Yorum sayılarına baktığımızda en çok yorum sayısı 921 ile Genel AVM Atmosferi hakkında, 217 adet yorumla Otopark konusu, 206 ile Yiyecek-İçecek konusu, 205 ile

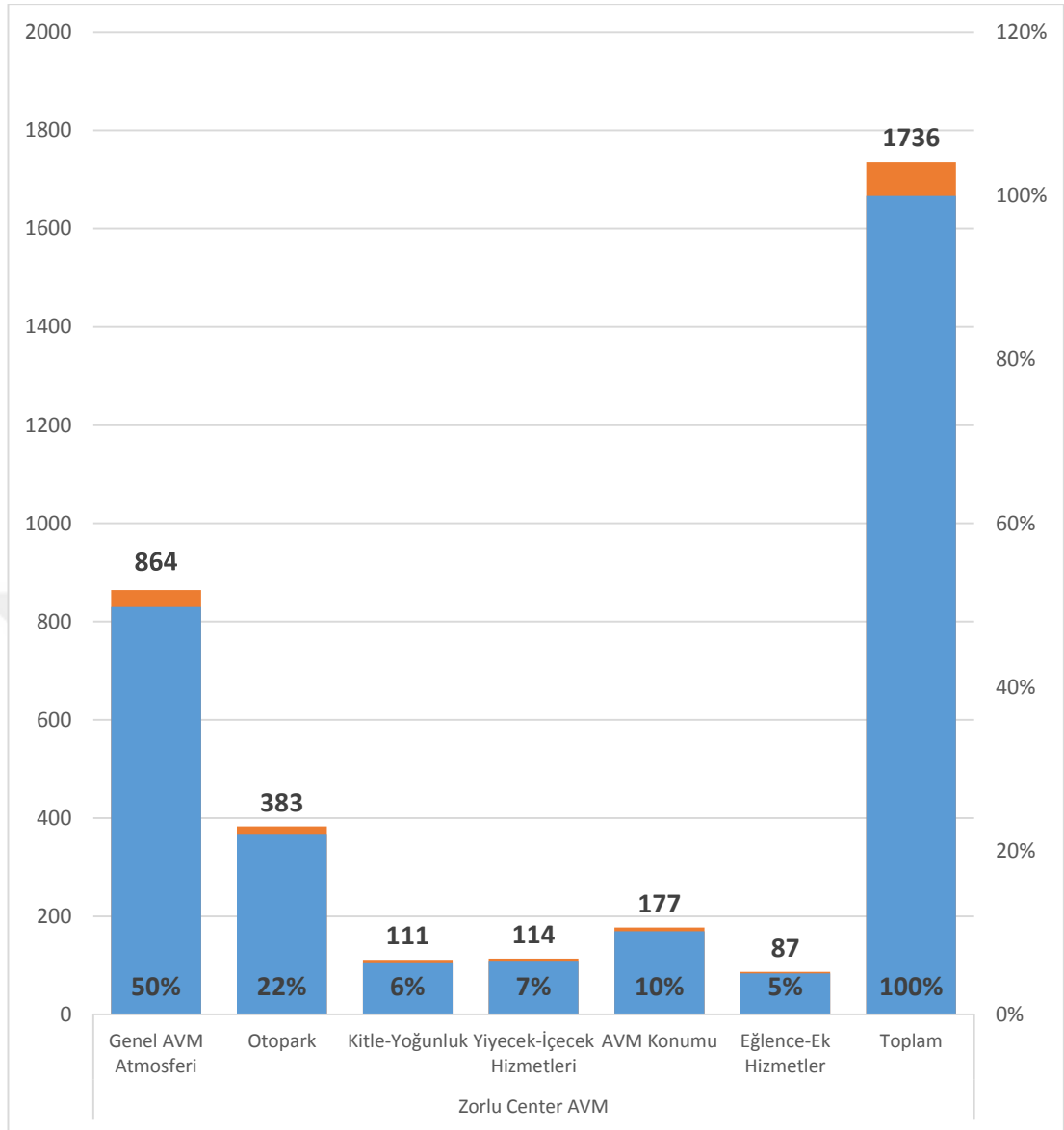
Kitle-Yoğunluk, 125 ile AVM Konumu ve 105 adet ile Eğlence-Ek Hizmetler gelmektedir.



Şekil 6.7: Marmara Park AVM Konu Dağılım Grafiği

Şekil 6.8’de Zorlu Center AVM için yapılan analiz gösterilmiştir.

Zorlu Center AVM için bırakılan yorumların yarısının Genel AVM Atmosferi hakkında olduğu görülüyor. %22’lik oranla Otopark konusu ikinci sırada, %10’luk oranla AVM Konumu konusu üçüncü sıradadır. AVM Konumu konusunu %7’lik oranla Yiyecek-İçecek Hizmetleri, %6’lık oranla Kitle-Yoğunluk konusu takip etmektedir. En düşük olan %5’lik kısımda ise Eğlence-Ek Hizmetler konusunda yorum bırakılmıştır. Otopark konusunda Şekil 6.2’de gösterilen diğer AVM’lere göre en fazla orana sahip olan AVM’dir. %7’lik Yiyecek-İçecek Hizmetleri konusunda Mall of İstanbul ile birlikte en düşük orana sahiptir.

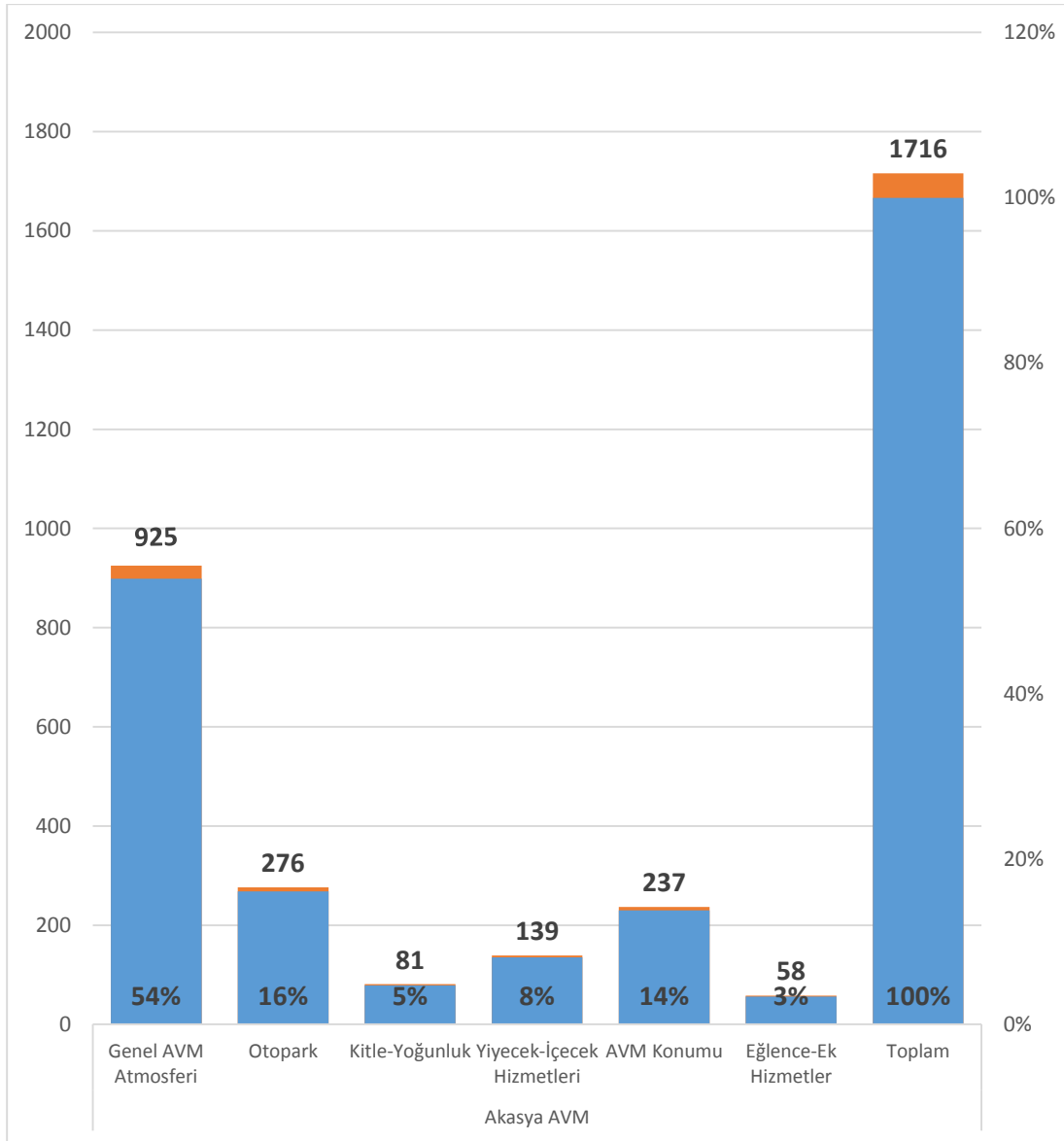


Şekil 6.8: Zorlu Center AVM Konu Dağılım Grafiği

Şekil 6.9’da Akasya AVM için yapılan analiz gösterilmiştir.

Akasya AVM için bırakılan yorumların %54’ünün Genel AVM Atmosferi hakkında olduğu görülüyor. %16’lık oranla Otopark konusu ikinci sırada, %14’lük oranla AVM Konumu konusu üçüncü sıradadır. AVM Konumu konusunu %8’lik oranla Yiyecek-İçecek Hizmetleri, %5’lik oranla Kitle-Yoğunluk konusu takip etmektedir.

En düşük olan %3’lük kısımda ise Eğlence-Ek Hizmetler konusunda yorum bırakılmıştır. Eğlence-Ek Hizmetler konusunda Şekil 6.2’de gösterilen diğer AVM’lere göre en düşük orana sahip olan AVM’dir. Bunun yanında Kitle-Yoğunluk oranı da diğer on AVM içinde en düşük orana sahiptir. AVM Konumu konusunda ise en yüksek orana sahip olmaktadır.



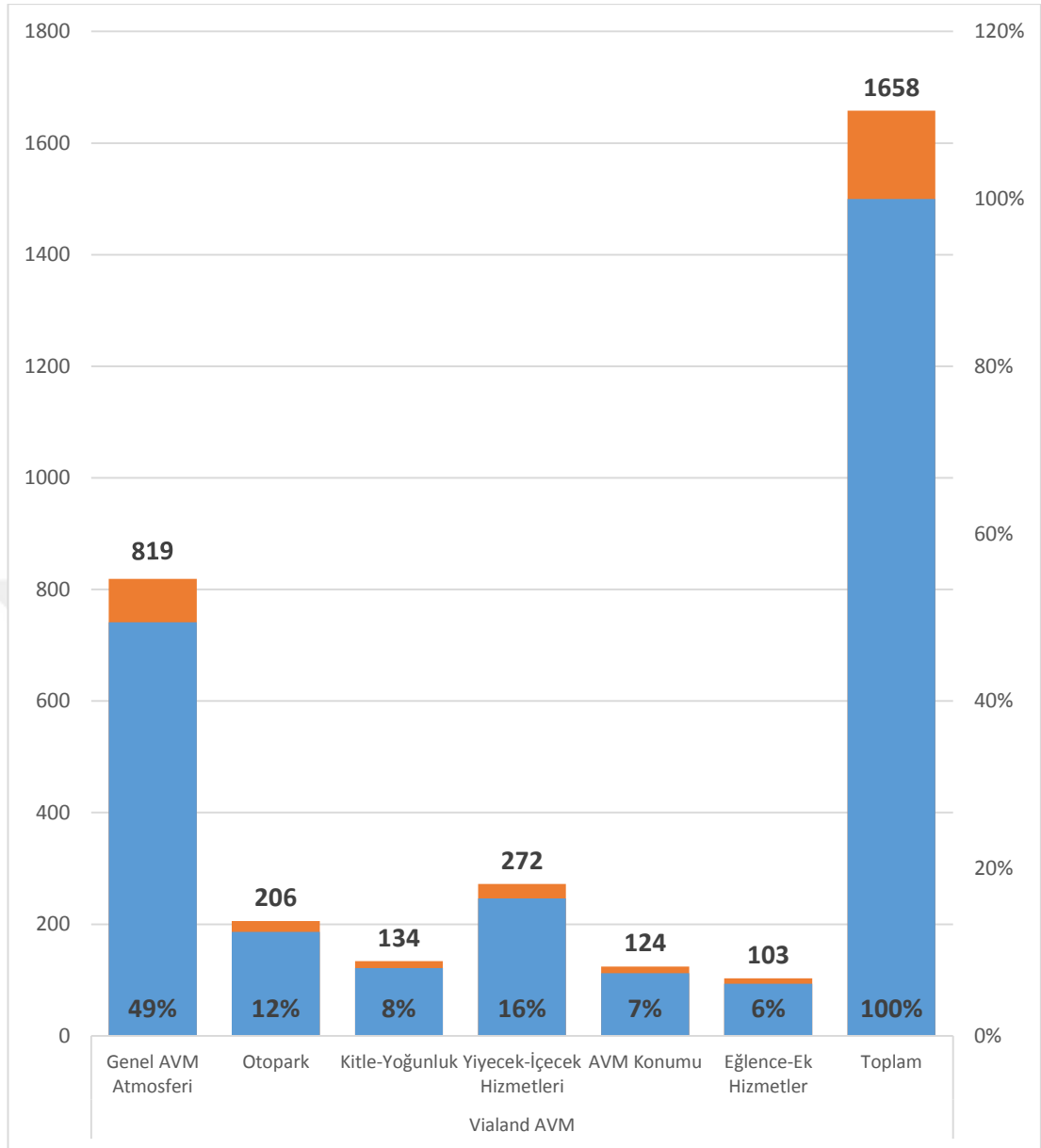
Şekil 6.9: Akasya AVM Konu Dağılım Grafiği

Şekil 6.10’da Vialand AVM için yapılan analiz gösterilmiştir.

Vialand AVM için bırakılan yorumların %49’unun Genel AVM Atmosferi hakkında olduğu görülüyor. %16’lık oranla Yiyecek-İçecek Hizmetleri konusu ikinci sırada, %12’lik oranla Otopark konusu üçüncü sıradadır. Otopark konusunu %8’lik oranla Kitle-Yoğunluk, %7’lik oranla AVM Konumu konusu takip etmektedir.

En düşük olan %6’lık kısımda ise Eğlence-Ek Hizmetler konusunda yorum bırakılmıştır. Yiyecek-İçecek Hizmetleri konusunda Şekil 6.2’de yer alan diğer AVM’lere göre en yüksek orana sahiptir.

Baktığımızda toplam yorum sayısı 1658 olup, bu yorumların 819 adeti Genel AVM Atmosferi hakkında bırakılmıştır.



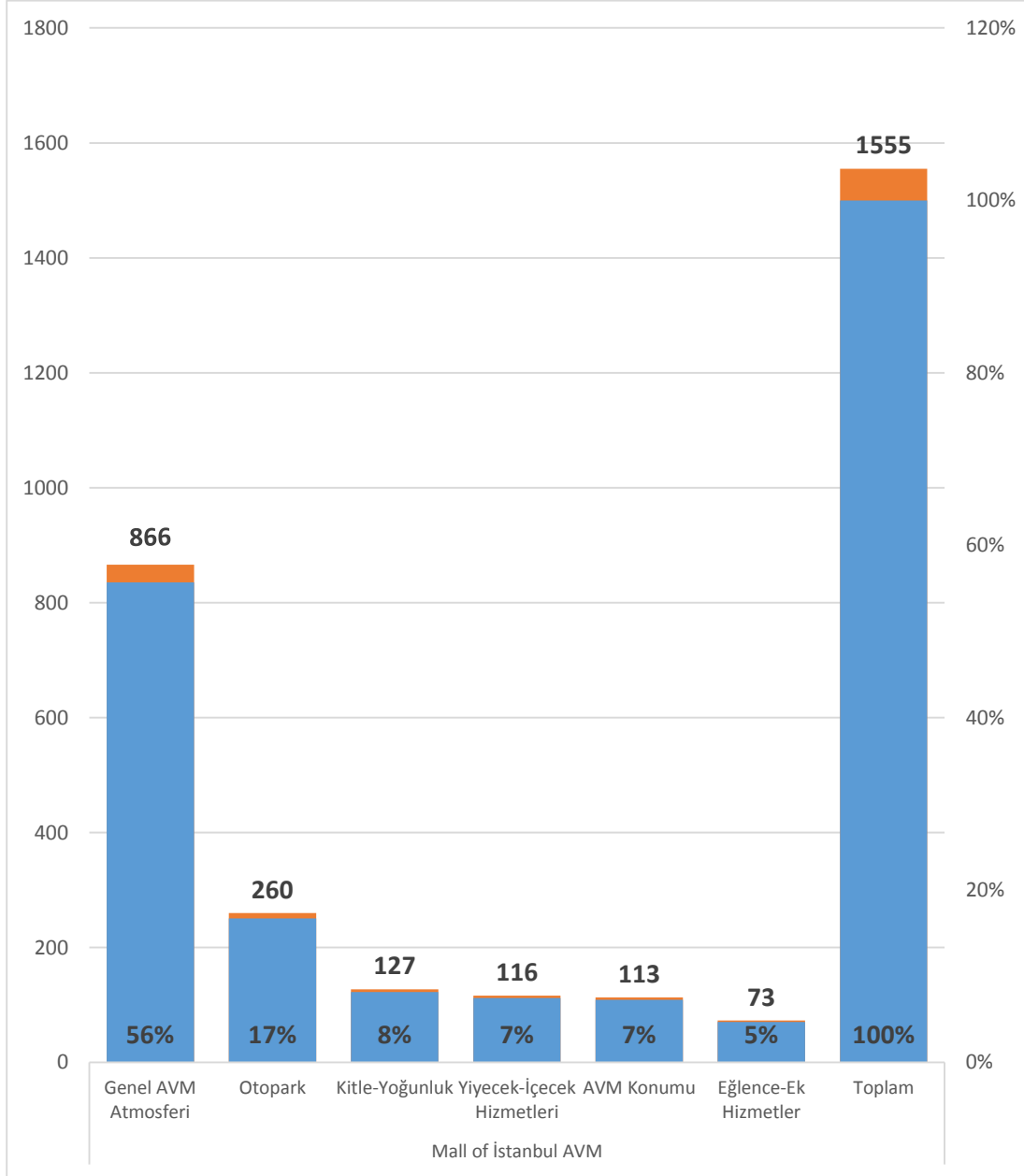
Şekil 6.10: Vialand AVM Konu Dağılım Grafiği

Şekil 6.11’de Mall of İstanbul AVM için yapılan analiz gösterilmiştir.

Mall of İstanbul AVM için bırakılan yorumların %58’inin Genel AVM Atmosferi hakkında olduğu görülüyor. %17’lik oranla Otopark konusu ikinci sırada, %8’lik oranla Kitle-Yoğunluk konusu üçüncü sıradadır. AVM Konumu ve Yiyecek-İçecek Hizmetleri %7’lik oranla, Eğlence-Ek Hizmetler %5’lik oranla takip etmektedir. %7’lik Yiyecek-İçecek Hizmetleri konusunda Zorlu Center AVM ile birlikte en düşük orana sahiptir.

Yorum sayılarına baktığımızda en çok yorum sayısı 866 ile Genel AVM Atmosferi hakkında, 260 adet yorumla Otopark konusu, 127 ile Kitle-Yoğunluk, 116 ile Yiyecek-

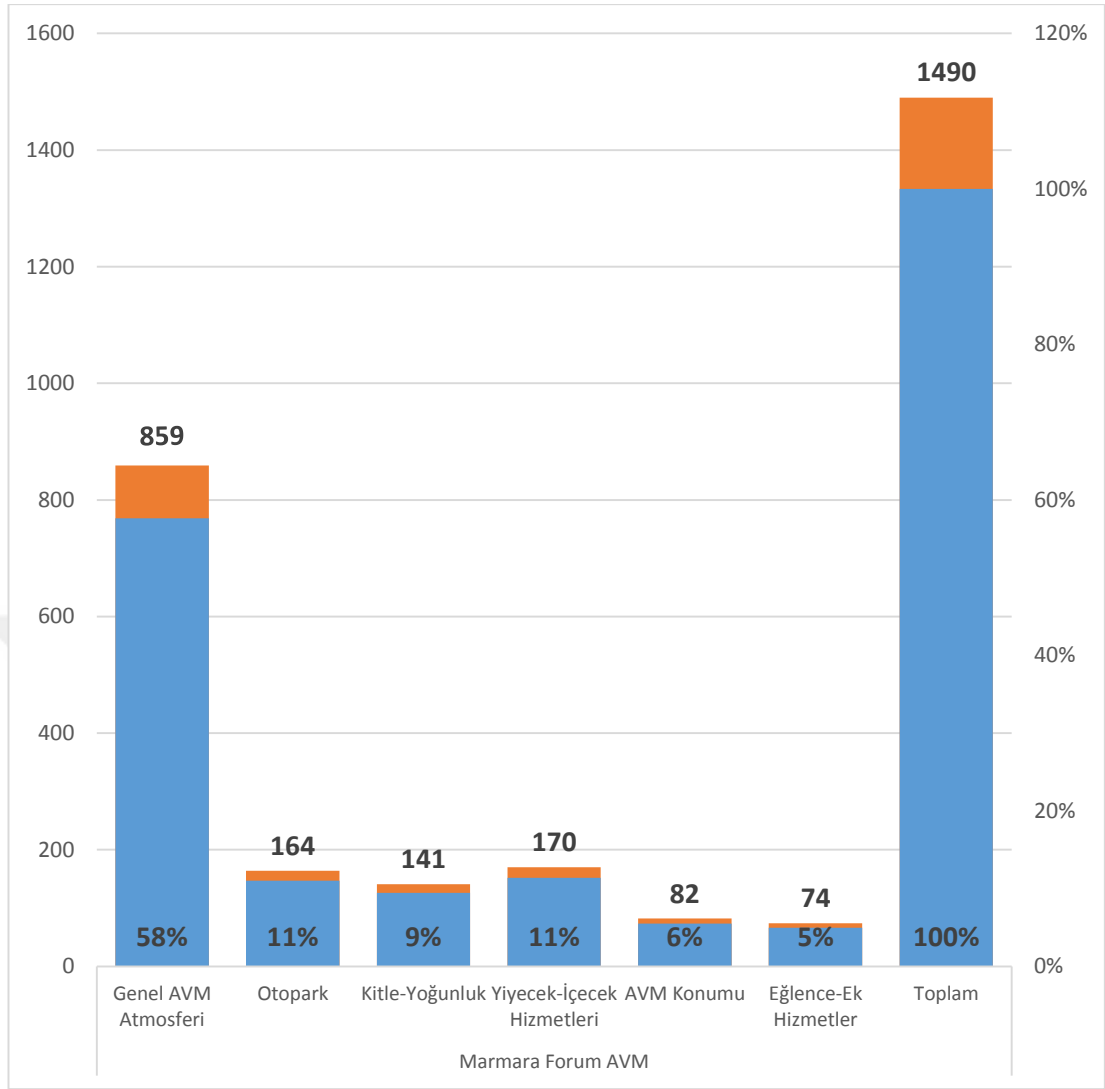
İçecek konusu, 113 ile AVM Konumu ve 73 adet ile Eğlence-Ek Hizmetler gelmektedir.



Şekil 6.11: Mall of İstanbul AVM Konu Dağılım Grafiği

Şekil 6.12’de Marmara Forum AVM için yapılan analiz gösterilmiştir.

Marmara Forum AVM için bırakılan yorumların %58’inin Genel AVM Atmosferi hakkında olduğu görülüyor. Bu oran Şekil 6.2’de yer alan AVM’ler içerisindeki en yüksek değerdir. %11’erlik oranla Otopark ve Yiyecek-İçecek Hizmetleri konuları ikinci sırada, %9’luk oranla Kitle-Yoğunluk konusu üçüncü sıradadır. AVM Konumu konusu %6’lık oranla, Eğlence-Ek Hizmetler konusu %6’lık oranla takip etmektedir.



Şekil 6.12: Marmara Forum AVM Konu Dağılım Grafiği

EK B’de yer alan tüm AVM analizlerine bakıldığında ise;

- Genel AVM Atmosferi konusunun en çok konuşulduğu (en çok yoruma sahip) AVM Forum İstanbul AVM,
- Otopark konusunun en çok konuşulduğu AVM Zorlu Center AVM,
- Kitle-Yoğunluk konusunun en çok konuşulduğu AVM Cevahir AVM,
- Yiyecek-İçecek Hizmetleri konusunun en çok konuşulduğu AVM Cevahir AVM,
- AVM Konumu konusunun en çok konuşulduğu AVM Palladium AVM,
- Eğlence-Ek Hizmetler konusunun en çok konuşulduğu AVM Cevahir AVM olarak çıkmıştır.

6.2 Çalışmanın Eksikleri

Türkçe, morfolojik açıdan karmaşık dillerden biridir. Sözcüklerin sonuna ekler eklenir ve bu, gövdedeki bazı harfleri kaybetmesine neden olabilir. Sözcük yapısal olarak incelendiğinde, bir kök ve bir takım soneklerden oluştuğu görülür. Sözcüğün çıplak biçimi kök ve bu çıplak formdan sonraki birimler sonektir. Kelimeler için ekleme sonekleri, anlambilim analizinde yararlıdır. Ortak bir kök içeren terimlerin genellikle benzer anlamları olacaktır. Bu semantik benzerliği belirlemeye yardımcı olur. Bu sürecin bir diğer avantajı da, toplam terim ve boyut sayısını azaltmasıdır. Bu avantajı sağlamak adına Zemberek kütüphanesi ile kök ayrıştırma işlemi yapılmıştır ancak elde edilen verinin sonuçlar üzerinde olumsuz etkisi olacağı belirlenerek kullanılmamıştır. Türkçe için kök ayrıştırma alanında daha başarılı çalışmaların yapılması ile modelin üreteceği çıktı kalitesi artacaktır. Çalışmamızda köklerine ayrılmayan veriler kullanılmak durumunda kalınmıştır. Aynı köke sahip kelimelerin anahtarlar içerisinde görülmesi bu durumdan dolayı yaşanmıştır. Bu etkiyi azaltmak adına anahtar sayısı yazılımın getirdiği default sayı olan 20 olarak kullanılmış, bir değişiklik yapılmamıştır.

Bir diğer konu olan doğal dil işleme alanında da üzerinde denemeler yapılan ITU NLP yazılımı yeterli seviyede olmadığı görülmüştür. Bu sebeple yanlış yazılan Türkçe kelimelerin düzeltilmesi, eksik harflerin tamamlanması işlemleri manuel yapılmış ve zaman kaybı yaşanmıştır. Bu işlemler için doğal dil işleme yazılımlarının geliştirilmesi, manuellikten kurtulup otomatize edilmesinden dolayı sonuçların kalitesine olumlu etki edecektir.

6.3 Gelecek Çalışmalara Öneriler

Yapılan makale ve tez incelemelerinde İngilizce için bir çok anlamsal analiz çalışması bulunmaktadır ancak Türkçe için anlamsal analiz çalışmalarının oldukça az olduğu tespit edilmiştir. Bunun sebebi Türkçenin yapısal zorluklarıdır. Bu konuda dilbilimcilerin katkılarıyla akademik çalışmaların sürdürülmesi gerekmektedir. Bu alandaki çalışmaların artırılması ile birlikte Türkçe metin verilerinin analizi için gereken yazılımlar geliştirilerek daha kaliteli ve efektif sonuçlar üretilebilecektir.

MALLET uygulamasına verilen ve kelime anahtarları içerisinde yer almaması gereken duraklama kelimelerinde Türkçe karakterleri algılamama sorunuyla karşılaşmıştır.

Duraklama listesinde yer alan kelimelerin Türkçe karakterleri İngilizce karakterlere dönüştürülmüş, bu şekilde bir çözüm bulunmuştur. Ancak bu konuda yapılacak iyileştirme ile bu manuel işleme gerek kalmayacaktır.





KAYNAKLAR

- [1] **Akgün, V. Ö.** (2010). Modern Alışveriş Merkezlerinin Tüketici Davranışları Üzerindeki Etkisi ve Konya İlinde Bir Uygulama, *KMÜ Sos. ve Ekon. Araştırmalar Dergisi*, 12 (19), 153–163.
- [2] **Url-1** <<https://dralabay.wordpress.com/2014/08/29/web-teknolojilerinin-gelisimi-ve-hayatimiza-etkileri/>>, erişim tarihi 08.04.2017.
- [3] **Url-2** <<http://www.ayberkburc.com/internet/web-1-0-2-0-3-0-nedir.html>>, erişim tarihi 08.04.2017.
- [4] **Url-3** <<http://www.volkaninanc.com/web-3-0-nedir/>>, erişim tarihi 08.04.2017.
- [5] **Url-4** <<http://sosyalmedyakulubu.com.tr>>, erişim tarihi 08.04.2017.
- [6] **Url-5** <<http://sosyalsen.blogspot.com.tr/2014/05/sosyal-medyanin-medya-boyutu.html#.WOoeqfmLTIU>>, erişim tarihi 08.04.2017.
- [7] **Url-6** <https://prezi.com/w8utzzy4-967_/sosyal-medyanin-pazarlama-uzerindeki-etkileri/>, erişim tarihi 08.04.2017.
- [8] **Keskin, S. ve Baş, M.** (2015). Sosyal Medyanın Tüketici Davranışları Üzerine Etkisinin Belirlenmesi, *Gazi Üniversitesi İktisadi ve İdari Bilim. Fakültesi Derg.*, 3, 51–69.
- [9] **İşlek, M. S.** (2012). *Sosyal Medyanın Tüketici Davranışlarına Etkileri Türkiye'deki Sosyal Medya Kullanıcıları Üzerine Bir Araştırma* (Yüksek lisans tezi). Karamanoğlu Mehmetbey Üniversitesi, Sosyal Bilimler Enstitüsü, Karaman.
- [10] **Url-7** <https://business.yell.com/knowledge/why-your-business-needs-to-use-four-square/>, erişim tarihi 14.04.2017.
- [11] **Url-8** <<http://expandedramblings.com/index.php/by-the-numbers-interesting-foursquare-user-stats/2016>>, erişim tarihi 20.03.2017.
- [12] **Url-9** <<http://venturebeat.com/2016/09/13/foursquare-users-have-checked-in-over-10-billion-times/>>, erişim tarihi 20.03.2017.
- [13] **Url-10** <<http://webrazzi.com/2015/04/16/foursquare-turkiye-kullanici-infografik/>>, erişim tarihi 20.03.2017.

- [14] **Chorley, M. J., Whitaker, R. M. ve Allen, S. M.** (2015). Personality and location-based social networks, *Comput. Human Behav.*, 46, 45–56.
- [15] **Aliandu, P.** (2015). Sentiment Analysis to Determine Accommodation, Shopping and Culinary Location on Foursquare in Kupang City, *Procedia Comput. Sci.*, 72, 300–305.
- [16] **Moraes, F., Vasconcelos, M., Prado, P., Dalip, D., Almeida, J. M. ve Gonçalves, M.** (2013). Polarity detection of Foursquare tips, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 8238, 153–162.
- [17] **Chen, J., Li, X., Cheung, W. K. ve Li, K.** (2016). Effective successive POI recommendation inferred with individual behavior and group preference, *Neurocomputing*, 210, 174–184.
- [18] **Vasconcelos, M. A., Ricci, S., Almeida, J., Benevenuto, F. ve Almeida, V.** (2012). Tips, done's and todos: uncovering user profiles in FourSquare,” *Wsdm*, 653,.
- [19] **Url-11** <<http://seonun.blogspot.com.tr/2014/01/forum-siteleri-nedir-ne-ise-yarar.html>>, erişim tarihi 09.04.2017.
- [20] **Url-12** <<https://tr.wikipedia.org/wiki/Viki>>, erişim tarihi 09.04.2017.
- [21] **Url-13** <<https://aykutekinci.com/wordpress-nedir-ne-ise-yarar/>>, erişim tarihi 09.04.2017.
- [22] **Url-14** <<https://www.turkcebilgi.com/podcasting?q=podcasting>>, erişim tarihi 09.04.2017.
- [23] **Url-15** <<https://www.ideasoft.com.tr/musteri-yorumlarinin-onemi/>>, erişim tarihi 08.04.2017.
- [24] **Url-16** <<https://www.brightlocal.com/learn/local-consumer-review-survey/>>, erişim tarihi 14.04.2017.
- [25] **Url-17** <<http://eticaretmag.com/user-generated-content-avantaj-ve-dezavantajlari/>>, erişim tarihi 14.04.2017.
- [26] **Url-18** <<http://www.yemrekoc.com/kullanici-yorumlarinin-onemi/>>, erişim tarihi 08.04.2017.
- [27] **Url-19** <<http://www.bazaarvoice.com/social-commerce-statistics>>, erişim tarihi 14.04.2017.
- [28] **Seker, S. E.** (2016). Duygu Analizi (Sentimental Analysis), *YBS Ansiklopedi*, 3 (3), 1-16 .
- [29] **Url-20** <<http://avc.com/2016/01/foursquare-trip-tips/>>, erişim tarihi 14.04.2017.

- [30] **Url-21** <<http://expandedramblings.com/index.php/by-the-numbers-interesting-foursquare-user-stats/>>, erişim tarihi 14.04.2017.
- [31] **Url-22** <<https://www.fastcompany.com/3015168/foursquares-tips-growing-faster-than-yelps-reviews>>, erişim tarihi 14.04.2017.
- [32] **Vasconcelos, M., Almeida, J. ve Gonçalves, M.** (2014). What Makes your Opinion Popular? Predicting the Popularity of Micro-Reviews in Foursquare, *Proc. 29th Annu. ACM Symp. Appl. Comput.*, 598–603.
- [33] **Url-23** <<http://www.socialmediaexaminer.com/foursquare-marketing-tips/>>, erişim tarihi 14.04.2017.
- [34] **Erkip, F.** (2005). The rise of the shopping mall in Turkey: The use and appeal of a mall in Ankara, *Cities*, 22 (2), 89–108.
- [35] **Vural Arslan, T.** (2009). Türkiye’deki Alışveriş Merkezleri İncelemelerine Eleştirel Bir Bakış: Yorumlar, Eleştiriler, Tartışmalar, *Uludağ Üniversitesi Mühendislik-Mimarlık Fakültesi Derg.*, 14 (1), 147–159.
- [36] **Köksal, Y. ve Aydın, E. E.** (2015). Tüketicilerin Alışveriş Merkezleri Algısı Üzerine Karşılaştırmalı Bir İnceleme: Göller Bölgesi Örneği, *Uluslararası Yönetim İktisat ve İşletme Derg.*, 11 (24), 231–248.
- [37] **Önüt, S., Efendigil, T. ve Soner Kara, S.** (2010). A combined fuzzy MCDM approach for selecting shopping center site: An example from Istanbul, Turkey, *Expert Syst. Appl.*, 37 (3), 1973–1980.
- [38] **Cheng, E. W. L., Li, H. ve Yu, L.** (2007). A GIS approach to shopping mall location selection, *Build. Environ.*, 42 (2), 884–892.
- [39] **Chebat, J. C., Sirgy, M. J. ve Grzeskowiak, S.** (2010). How can shopping mall management best capture mall image?, *J. Bus. Res.*, 63 (7), 735–740.
- [40] **Kesari, B. ve Atulkar, S.** (2016). Satisfaction of mall shoppers: A study on perceived utilitarian and hedonic shopping values, *J. Retail. Consum. Serv.*, 31, 22–31.
- [41] **Debek, M.** (2015). What Drives Shopping Mall Attractiveness?, *Polish J. Appl. Psychol.*, 13 (1), 67–118.
- [42] **Dennis, C., Marsland, D. ve Cockett, T.** (2001). Data mining for shopping centres-customer knowledge-management framework, *J. Knowl. Manag.*, 5 (4), 368–374.
- [43] **Vasconcelos, M., Almeida, J. M. ve Gonçalves, M. A.** (2015). Predicting the popularity of micro-reviews: A Foursquare case study, *Inf. Sci. (Ny)*, 325, 355–374.
- [44] **Url-24** <<https://lct-master.org/files/MullenSentimentCourseSlides.pdf>>, erişim tarihi 02.04.2017
- [45] **Pang, B. ve Lee, L.** (2006). Opinion Mining and Sentiment Analysis, *Foundations and Trends in Information Retrieval*, 1 (2), 91–231.

- [46] **Onan, A.** (2017). *Türkçe Twitter Mesajlarında Gizli Dirichlet Tahsisine Dayalı Duygu Analizi*. Erişim Mart 15, 2017, ab.org.tr/ab17/bildiri/77.pdf
- [47] **Kang, H., Yoo, S. J. ve Han, D.** (2012). Senti-lexicon and improved Naive Bayes algorithms for sentiment analysis of restaurant reviews, *Expert Syst. Appl.*, 39 (5), 6000–6010.
- [48] **Kang, D. ve Park, Y.** (2014). Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach, *Expert Syst. Appl.*, 41 (4), 1041–1050.
- [49] **Çoban, Ö.** (2016). *Turkish Twitter Sentiment Analysis using Text Classification Techniques* (Yüksek lisans tezi). Atatürk Üniversitesi, Fen Bilimleri Enstitüsü, Erzurum.
- [50] **Koçak, B. B., Polat, İ. ve Koçak, C. B.** (2016). Twitter Kullanıcılarının Havayolu Pazarına Yönelik Duygu Kutuplarının Belirlenmesi Bir Fikir Madenciliği Örneği, *Global İşletme Araştırmaları Kongresi*, İstanbul, Türkiye, Mayıs 26-27.
- [51] **Ergün, K.** (2012). *Metin Madenciliği Yöntemleri ile Ürün Yorumlarının Otomatik Değerlendirilmesi* (Doktora tezi). Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Sakarya.
- [52] **Park, K. ve Kremer, G. E. O.** (2017). Text mining-based categorization and user perspective analysis of environmental sustainability indicators for manufacturing and service systems, *Ecol. Indic.*, 72, 803–820.
- [53] **Lucini, F. R., Fogliatto, F. S., da Silveira, G. J. C., ..., Schaan, B. D.** (2017). Text mining approach to predict hospital admissions using early medical records from the emergency department, *Int. J. Med. Inform.*, 100, 1–8.
- [54] **Kayser, V. ve Blind, K.** (2016). Extending the knowledge base of foresight: The contribution of text mining, *Technol. Forecast. Soc. Change*,
- [55] **Kim, K., Park, O., Yun, S. ve Yun, H.** (2017). What makes tourists feel negatively about tourism destinations? Application of hybrid text mining methodology to smart destination management, *Technol. Forecast. Soc. Change*.
- [56] **Zaeem, R. N., Manoharan, M., Yang, Y. ve Barber, K. S.** (2017). Modeling and analysis of identity threat behaviors through text mining of identity theft stories, *Comput. Secur.*, 65, 50–63.
- [57] **Jang, H., Sim, J., Lee, Y., ve Kwon, O.** (2013). Deep sentiment analysis : Mining the causality between personality-value-attitude for analyzing business ads in social media, *Expert Syst. Appl.*, 40 (18), 7492–7503.
- [58] **Mostafa, M. M.** (2013). More than words : Social networks ' text mining for consumer brand sentiments, *Expert Syst. Appl.*, 40 (10), 4241–4251.
- [59] **Yiğit, İ. O., Güvercin, M., Ateş, A. F. ve Ferhatosmaoğlu, F.** (2017). Çağrı Merkezi Metin Madenciliği Yaklaşımı. *25th Signal Processing and Communications Applications Conference*, Antalya, Türkiye, Mayıs.
- [60] **Uzun, V.** (2014) *Semantic Text Mining and An Application in Turkish Documents* (Yüksek lisans tezi). Dokuz Eylül Üniversitesi, Fen Bilimleri Enstitüsü,

İzmir.

- [61] **Alsumait, L., Daniel, B. ve Domeniconi, C.** (2008). On-Line LDA: Adaptive Topic Models for Mining Text Streams with Applications to Topic Detection and Tracking, *Acdm*, 3-12.
- [62] **Alghamdi, R. ve Alfalqi, K.** (2015). A Survey of Topic Modeling in Text Mining, *Int. J. Adv. Comput. Sci. Appl.*, 6 (1), 147–153.
- [63] **Bicalho, P., Pita, M., Pedrosa, G., Lacerda, A. ve Pappa, G. L.** (2017). A general framework to expand short text for topic modeling, *Inf. Sci. (Ny)*, 393, 66–81.
- [64] **Jiang, H., Qiang, M. ve Lin, P.** (2016). A topic modeling based bibliometric exploration of hydropower research, *Renew. Sustain. Energy Rev.*, 57, 226–237.
- [65] **Choi, H. S., Lee, W. S., ve Sohn, S. Y.** (2017). Analyzing research trends in personal information privacy using topic modeling, *Computers & . Security*, 67, 244-253.
- [66] **Sun, L. ve Yin, Y.** (2016). Discovering themes and trends in transportation research using topic modeling, *Transp. Res. Part C*, 77, 49–66.
- [67] **Arnold, C. W., Oh, A., Chen, S. ve Speier, W.** (2016). Evaluating topic model interpretability from a primary care physician perspective, *Comput. Methods Programs Biomed.*, 124, 67–75.
- [68] **Bisgin, H., Liu, Z., Fang, H., Xu, X. ve Tong, W.** (2011). Mining FDA drug labels using an unsupervised learning technique - topic modeling. *Eighth Annual MCBIOS Conference Computational Biology and Bioinformatics for a New Decade College Station, TX, USA*. Nisan 1-2.
- [69] **Sarıman, G.** (2011). Veri Madenciliğinde Kümeleme Teknikleri Üzerine Bir Çalışma: K-Means ve K-Medoids Kümeleme Algoritmalarının Karşılaştırılması, *Süleyman Demirel Üniversitesi, Fen Bilim. Enstitüsü Derg.*, 3, 192–202.
- [70] **Apaydın, E.** (2000). *Zeki Veri Madenciliği: Ham Veriden Altın Bilgiye Ulaşma Yöntemleri*. Erişim Nisan 20, 2017, https://www.cmpe.boun.edu.tr/~ethem/index_tr.html
- [71] **Erpolat, S.** (2012). Otomobil Yetkili Servislerinde Birliktelik Kurallarının Belirlenmesinde Apriori ve FP-Growth Algoritmalarının Karşılaştırılması, *Anadolu Üniversitesi Sos. Bilim. Derg.*, 12 (2), 137–146.
- [72] **Url-25** <<http://mustafaacungil.blogspot.com.tr/2007/08/veri-madenciliinde-crisp-dm-yaklam.html>>, erişim tarihi 10.04.2017.
- [73] **Argüden, Y. ve Erşahin, B.** (2002). *Veri Madenciliği*. İstanbul, Türkiye, ARGE Danışmanlık.
- [74] **Kumdereli, Ü. C.** (2012). *Tıp Bilişimi Ve Veri Madenciliği Uygulamaları: Eeg Sinyallerindeki Epileptiform Aktiviteye Veri Madenciliği Yöntemlerinin Uygulanması* (Yüksek lisans tezi). Trakya Üniversitesi, Fen Bilimleri Enstitüsü, Edirne.

- [75] **Gemici, B.** (2012). *Veri Madenciliği ve Bir Uygulaması* (Yüksek lisans tezi). Dokuz Eylül Üniversitesi, Sosyal Bilimler Enstitüsü, İzmir.
- [76] **Kaplan, Ö. ve Gözen, G.** (2013). 'Oracle Data Miner' ile Mantarların Zehirliliği Üzerine Bir Veri Madenciliği Uygulaması. Erişim Nisan 19, 2017, akademi.itu.edu.tr/kiris/DosyaGetir/92643/12%20GozenKaplan%202013.pdf
- [77] **Url-26** <<http://biyo-informatik.blogspot.com.tr/2015/10/metin-madenciligi-text-mining-nedir.html>>, erişim tarihi 06.04.2017.
- [78] **Url-27** <<http://bilgisayarkavramlari.sadievrenseker.com/2014/06/15/metin-madenciligi-text-mining/>>, erişim tarihi 06.04.2017.
- [79] **Tunalı, V.** (2011). *Metin Madenciliği İçin İyileştirilmiş Bir Kümeleme Yapısının Tasarımı Ve Uygulaması* (Doktora tezi). Marmara Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- [80] **Demirdümen, B.** (2008). *Data And Text Mining Techniques On Real Medical Data* (Yüksek lisans tezi). Çankaya Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- [81] **Url-28** <https://en.wikipedia.org/wiki/Topic_model>, erişim tarihi 17.04.2017.
- [82] **Öztürk, S., Sankur, B., Güngör, T., ..., Ahat, M.** (2014). Türkçe Etiketli Metin Derlemi, 2014 IEEE 22nd Signal Process. Commun. Appl. Conf. SIU 2014 - Proc., Trabzon, Türkiye, Nisan 23-24.
- [83] **Blei, D. M., Ng, A. Y. ve Jordan, M. I.** (2003). Latent Dirichlet Allocation, *J. Mach. Learn. Res.*, 3, 993–1022.
- [84] **Url-29** <<https://www.slideshare.net/WayneLee9/lda-oct3-2013>>, erişim tarihi 16.04.2017
- [85] **Ma, B., Zhang, N., Liu, G., Li, L. ve Yuan, H.** (2016). Semantic search for public opinions on urban affairs: A probabilistic topic modeling-based approach, *Inf. Process. Manag.*, 52 (3), 430–445.
- [86] **Blei, D., Carin, L., ve Dunson, D.** (2010). Probabilistic topic models, *IEEE Signal Process. Mag.*, 27 (6), 55–65.
- [87] **Blei, D.M.** (2009). *Topic Models*, http://videolectures.net/mlss09uk_blei_tm/
- [88] **Guo, Y., Barnes, S. J. ve Jia, Q.** (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation, *Tour. Manag.*, 59, 467–483.
- [89] **Miller, I. M.** (2013). Rebellion, crime ve violence in Qing China, 1722-1911: A topic modeling approach, *Poetics*, 41 (6), 626–649.
- [90] **Url-30** <<http://mallet.cs.umass.edu/>>, erişim tarihi 10.03.2017.
- [91] **Url-31** <<http://alexperrier.github.io/jekyll/update/2015/09/04/topic-modeling-of-twitter-followers.html>>, erişim tarihi 16.07.2017.

- [92] **Vural, T. ve Yücel, A.** (2007). Çağımızın Yeni Kamusal Mekanları Olan Alışveriş Merkezlerine Eleştirel Bir Bakış, *İtü Dergisi/a Mimar. Planlama, Tasarım*, 5 (2), 97–106.
- [93] **Url-32** <<http://www.emlaktagundem.com/turkiyede-2016-2018-avm-arastirmasi-yayinlandi/>>, erişim tarihi 15.04.2017.
- [94] **Url-33** < <http://www.emlaktasondakika.com/haber/avm/turkiyede-avm-sektoru-ne-durumda/106802>>, erişim tarihi 15.04.2017.
- [95] **Url-34** <<https://menuburada.com/araclar/avm-listesi-ve-restoranlari/istanbuldaki-avmler>>, erişim tarihi 15.01.2017.
- [96] **Url-35** <<http://www.avmlife.com.tr/haber/4008/istanbul-avrupa-yakasi-avm-leri-nin-listesi.html>>. erişim tarihi 15.01.2017.
- [97] **Url-36** <<http://www.liste.gen.tr/istanbuldaki-avmler-listesi/>>. erişim tarihi 15.01.2017.
- [98] **Url-37** <https://www.google.com/maps/d/viewer?mid=1897Hzhnm2miVOqEP7YN4gmdgZ_M&hl=tr&ll=41.030938231895654%2C28.287337790039032&z=10>, erişim tarihi 15.01.2017
- [99] **Güngör, T.** (2016). An Overview of Resources Available for Turkish Natural Language Processing Applications, 75–78.
- [100] **Url-38** < <http://tools.nlp.itu.edu.tr/>>, erişim tarihi 16.03.2017.
- [101] **Url-39** < <http://zembereknlp.blogspot.com.tr/>>, erişim tarihi 20.03.2017.
- [102] **Url-40** < <http://stackoverflow.com/questions/29023108/why-mallet-lda-need-to-keep-sequence>>, erişim tarihi 18.04.2017.
- [103] **Url-41**<<http://stackoverflow.com/questions/3636067/topic-modeling-using-mallet?rq=1>>, erişim tarihi 18.04.2017.
- [104] **Url-42**<<https://lists.cs.princeton.edu/pipermail/topic-models/2012-February/001744.html>>, erişim tarihi 18.04.2017.
- [105] **Url-43** < <https://dragonfly.hypotheses.org/1051>>, erişim tarihi 18.04.2017.
- [106] **Url-44** <<http://mallet.cs.umass.edu/topics.php>>, erişim tarihi 18.04.2017.



EKLER

EK A: İstanbul'daki Alışveriş Merkezleri Verileri

EK B: Alışveriş Merkezlerinin Konu Dağılımları

EK C: Alışveriş Merkezlerinin Konu Dağılımına Etki Eden Yorum Sayıları
Çizelgesi

EK D: Exceldeki Her Bir Yorum Satırını .txt Dosya Formatına Çeviren Kod



EK A**Çizelge A.1: İstanbul'daki Alışveriş Merkezleri Verileri.**

AVM Adı	İlçe	Yorum Sayısı	Puan	Puanlayan Kişi Sayısı	Check-in Sayısı	Tekil Kullanıcı	Ziyaret Sayısı	Frekans (Sıklık)
Marmara Park AVM	Esenyurt	1959	8,4	23.556	2080674	539114	2945673	3,859
Maltepe Park AVM	Cevizli	1537	8,4	19.325	1475814	387029	2030169	3,813
Kule Çarşı Alışveriş Merkezi	4. Levent	26	6,3	191	12609	3369	13065	3,743
Akbatı AVM	Bahçeşehir	849	8,6	9.957	690578	196154	978685	3,521
Palladium AVM	Ataşehir	1410	8,9	22.091	1215663	366489	1622866	3,317
Capacity AVM	Bakırköy	1398	8,4	20.040	1491862	460094	1857873	3,243
Perlavista AVM	Beylikdüzü	627	7,3	6.249	463817	145621	624464	3,185
Forum İstanbul AVM	Bayrampaşa	2834	8,4	34.434	2487754	786500	3348041	3,163
Merkez Kayaşehir AVM	Başakşehir	130	5,1	628	48773	15837	69421	3,080
Akyaka Park Alışveriş Merkezi	Ümraniye	61	5,3	282	62185	20364	119510	3,054
Cevahir AVM	Şişli	3107	8,4	46.629	3048002	1012244	4056660	3,011
Arena Park AVM	Küçükçekmece	682	7,2	7.365	513810	170895	802411	3,007
Marmara Forum AVM	Bakırköy	1626	8,6	26.476	1741693	590214	2344356	2,951
İstinye Park AVM	Sarıyer	2366	9,4	36.381	1600634	548007	2161576	2,921
Akasya AVM	Acıbadem	1805	8,9	29.123	1664126	591751	2342565	2,812
Canpark AVM	Ümraniye	837	6,6	7.128	594790	212281	804368	2,802
Kozzy AVM	Kozyatağı	625	7,8	5.835	278758	101345	405004	2,751
Kanyon AVM	Levent	1634	9,1	22.726	1210096	441327	1544298	2,742
Pelican Mall AVM	Avcılar	732	7,5	8.142	644974	237885	891413	2,711
Kipa Silivri AVM	Silivri	83	5,6	1.148	112948	41774	17855	2,704
A Plus AVM	Ataköy	581	8,6	6.687	368584	136486	449155	2,701

Çizelge A.1 (Devam): İstanbul'daki Alışveriş Merkezleri Verileri.

AVM Adı	İlçe	Yorum Sayısı	Puan	Puanlayan Kişi Sayısı	Check-in	Tekil Kullanıcı	Ziyaret Sayısı	Frekans (Sıklık)
Capitol AVM	Altunizade	1169	8	14.680	859604	319089	1214842	2,694
Brandium AVM	Ataşehir	1103	7,7	15.049	787216	295049	1177375	2,668
Vialand AVM	Eyüp	1776	8,7	17.443	891792	340785	1320320	2,617
Beylikdüzü Migros AVM	Beylikdüzü	427	7,2	4.664	300000	115269	412988	2,603
Semer kand Line AVM	Beylikdüzü	18	5,6	33	4384	1714	4891	2,558
Buyaka AVM	Ümraniye	1209	8,5	17.386	893167	350287	1228814	2,550
Metrogarden AVM	Çekmeköy	228	6,8	3.251	280731	112622	424680	2,493
Armoni Park AVM	Sefaköy	626	6,3	5.388	415259	167476	649852	2,480
Viaport Outlet AVM	Pendik	2062	8,6	27.470	1542412	622790	2048239	2,477
Doğuş Power Center AVM	Sarıyer	197	6,8	1.739	98268	40148	123225	2,448
Torium AVM	Esenyurt	915	7,3	12.101	742087	305284	1033617	2,431
Mall of İstanbul AVM	Başakşehir	1679	8,8	18.264	1367036	576213	2021057	2,372
Historia AVM	Fatih	965	6,7	8.978	546712	230717	725455	2,370
Flyinn AVM	Florya	371	7,5	3.706	185068	79921	244024	2,316
Kardiyum AVM	Taşdelen	125	5,9	987	60818	26586	87667	2,288
Atirus AVM	Büyükçekmece	247	5,9	2.465	176070	77299	267567	2,278
Profilo AVM	Mecidiyeköy	569	6,4	6.397	360376	158687	469399	2,271
Olimpa AVM	Başakşehir	165	5,6	1.331	69284	30638	107606	2,261
Kale Outlet Center AVM	Güneşli	578	6,4	5.336	368506	164485	574289	2,240
Meydan İstanbul AVM	Ümraniye	715	7,9	8.260	410878	183416	530660	2,240
Akmerkez AVM	Etiler	733	8,4	11.028	513861	229457	725308	2,239
Tepe Nautilus AVM	Kadıköy	823	7,6	17.520	904802	405303	1373332	2,232

Çizelge A.1 (Devam): İstanbul'daki Alışveriş Merkezleri Verileri.

AVM Adı	İlçe	Yorum Sayısı	Puan	Puanlayan Kişi Sayısı	Check-in	Tekil Kullanıcı	Ziyaret Sayısı	Frekans (Sıklık)
Metrocity AVM	Levent	721	7,4	11.252	598702	268445	814361	2,230
Aqua Florya AVM	Bakırköy	1233	9,2	16.827	845365	382420	1050923	2,211
Maxi Center AVM	Silivri	188	5,8	2.068	135820	61746	193392	2,200
Trump Towers AVM	Mecidiyeköy	1261	7,6	13.857	780131	357634	1045426	2,181
Olivium AVM	Zeytinburnu	736	6,7	6.950	431642	200118	636363	2,157
Zorlu Center AVM	Zincirlikuyu	1941	8,6	29.934	1256721	587559	1804970	2,139
Optimum Outlet AVM	Yenisahra	1253	7,1	19.453	968535	459954	1455708	2,106
Atrium AVM	Ataköy	273	6,6	2.993	150207	71589	234605	2,098
Güneşli Park AVM	Güneşli	110	6,2	1.048	76636	36606	10304	2,094
Özdilek Park AVM	Levent	475	8,2	8.019	529788	258184	844373	2,052
Axis AVM	Kağıthane	467	7,8	650	243659	119383	410005	2,041
Metroport AVM	Bahçelievler	706	6,4	8.441	550058	271738	737414	2,024
Neomarin AVM	Pendik	501	6,2	5.031	278815	139351	415572	2,001
Citys AVM	Nişantaşı	897	8,9	17.194	587666	296980	768961	1,979
Blackout Outlet AVM	Şişli	30	5,1	135	5857	2984	6432	1,963
Town Center AVM	Bakırköy	74	4,9	231	13764	7020	13768	1,961
Astoria AVM	Esentepe	321	6,6	2.759	151170	77739	186461	1,945
Plato AVM	Sultanbeyli	209	5,4	1.457	93833	48641	134656	1,929
Ömür Plaza AVM	Bahçelievler	154	5,9	1.907	127556	67142	147627	1,900
M1 Merter AVM	Merter	53	4,8	512	16766	8938	17793	1,876
Starcity AVM	Yenibosna	946	6,6	9.281	426951	227964	693928	1,873
Carrefoursa AVM	İçerenköy	500	6,8	9.658	418502	231130	758485	1,811

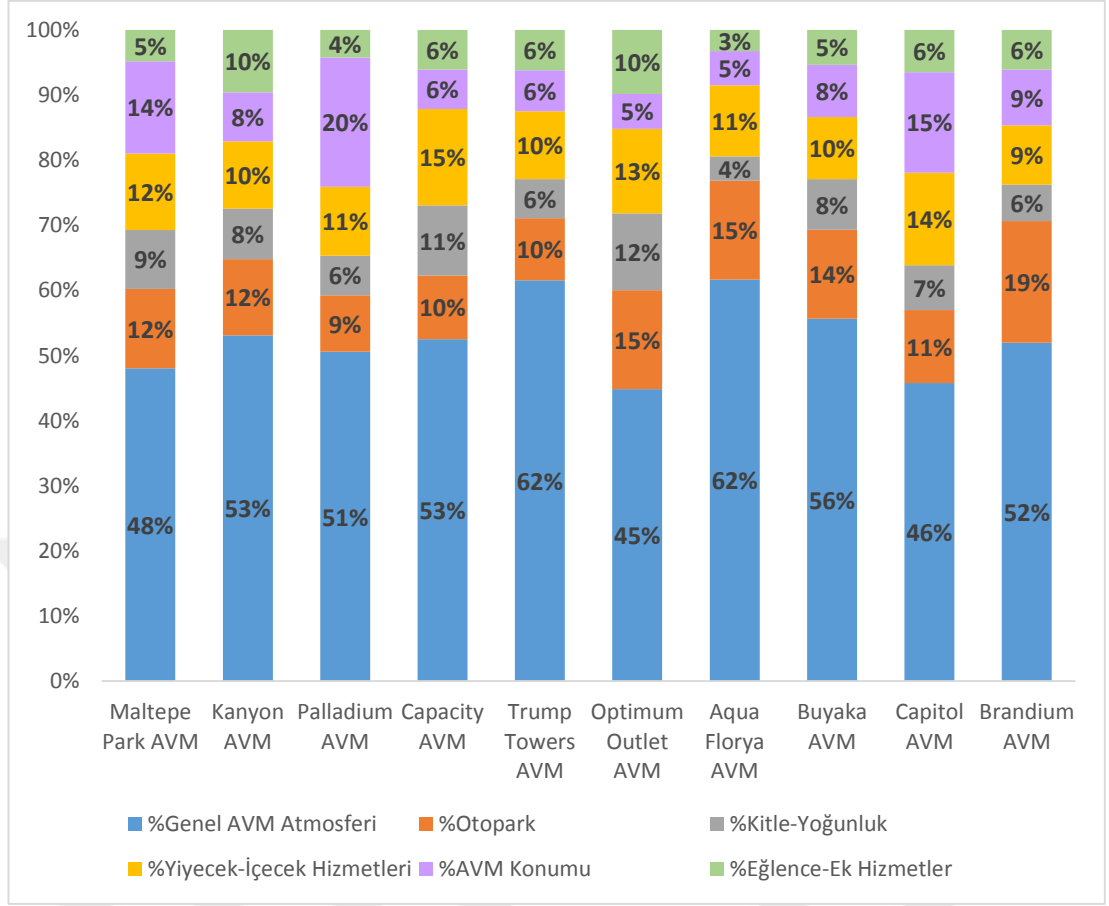
Çizelge A.1 (Devam): İstanbul'daki Alışveriş Merkezleri Verileri.

AVM Adı	İlçe	Yorum Sayısı	Puan	Puanlayan Kişi Sayısı	Check-in	Tekil Kullanıcı	Ziyaret Sayısı	Frekans (Sıklık)
Pendorya AVM	Pendik	377	6,1	3.563	178993	99275	238493	1,803
Novada AVM	Ataşehir	502	6,4	2.185	121094	67818	165021	1,786
212 İstanbul Power Outlet AVM	Bağcılar	676	6,6	8.818	436561	248407	699417	1,757
M1 Kartal AVM	Kartal	142	5,3	1.721	64059	37517	10376	1,707
Sapphire Çarşı AVM	Levent	680	4,8	3.688	326294	191262	506403	1,706
Galleria Ataköy AVM	Bakırköy	479	8,1	7.416	295338	173895	425121	1,698
White Hill AVM	Alibeyköy	95	5	501	22890	13597	34503	1,683
Bulvar 216 AVM	Ataşehir	98	8,4	1.900	87398	52822	117246	1,655
Prestige Mall AVM	Bahçeşehir	63	6,3	564	24659	14917	33476	1,653
Airport Outlet Center AVM	Bakırköy	473	6,4	6.460	322041	199687	447259	1,613
Akvaryum AVM	Bayrampaşa	104	5,4	416	12192	7631	12522	1,598
White Corner AVM	Beylikdüzü	37	5,3	363	16153	10443	22172	1,547
World Atlantis AVM	Pendik	124	5,1	748	29250	19311	38120	1,515
Kadir Has Alışveriş Merkezi	Bahçelievler	64	5,4	669	29101	19627	35302	1,483
Deposit Outlet AVM	İkitelli	287	5,6	2.511	99884	68947	158826	1,449
Rings AVM	Sancaktepe	79	6,3	732	40661	28162	66441	1,444
Esprit Outlet Center	Esenler	12	4,6	57	1068	741	1088	1,441
Sancakpark Outlet AVM	Sancaktepe	67	5	518	23872	16591	32864	1,439
Carousel AVM	Bakırköy	307	6,8	5.168	224062	157131	306809	1,426
Aymerkez AVM	Büyükdere	61	5,9	692	32013	24048	54987	1,331
Eskule AVM	Esenyurt	97	5	411	19676	15060	30593	1,307
Sun Flower Life Center	Bahçeşehir	37	5,9	374	13541	11195	36017	1,210

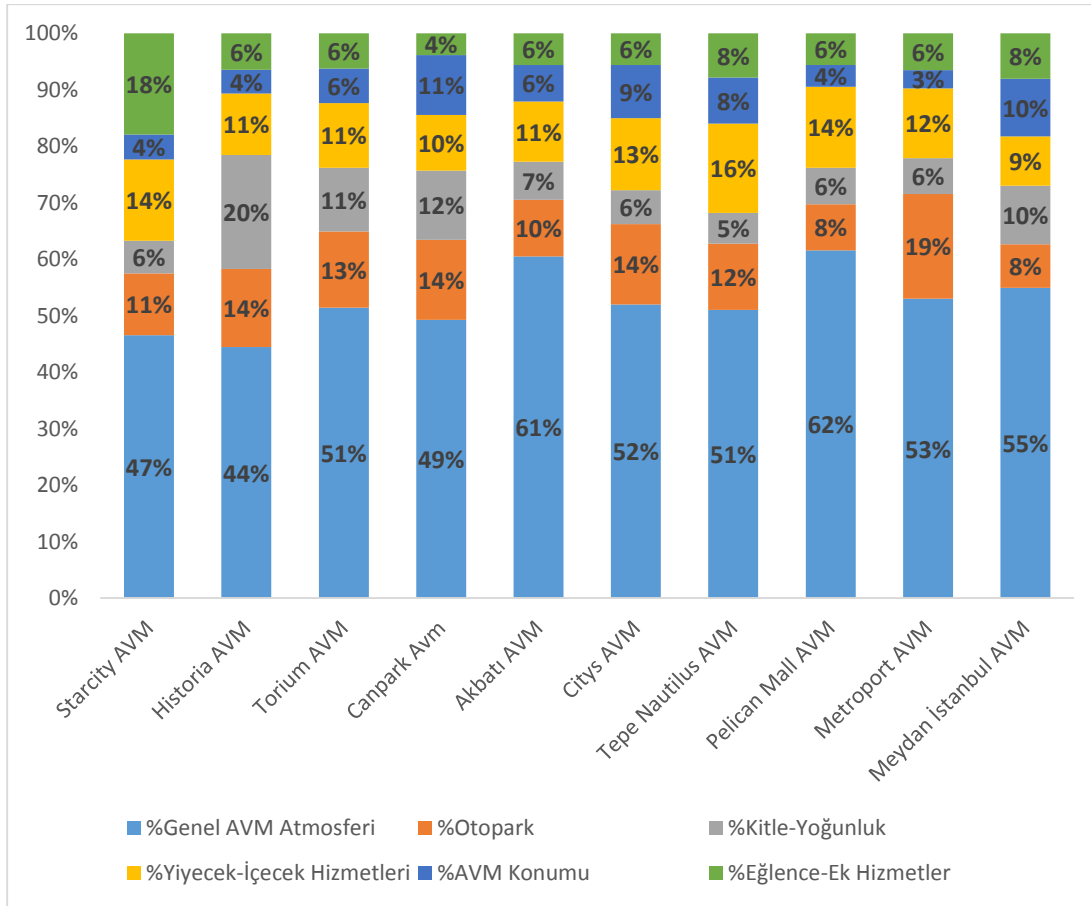
Çizelge A.1 (Devam): İstanbul'daki Alışveriş Merkezleri Verileri.

AVM Adı	İlçe	Yorum Sayısı	Puan	Puanlayan Kişi Sayısı	Check-in	Tekil Kullanıcı	Ziyaret Sayısı	Frekans (Sıklık)
Platform Merter AVM	Güngören	60	5,6	489	19217	16149	29424	1,190
Beylicium Alışveriş ve Yaşam Merkezi	Beylikdüzü	137	5,3	1.221	51784	44445	10370	1,165
İhlamur AVM	Halkalı	23	5,6	163	5788	4997	13299	1,158
Demirören AVM	Taksim	476	6,6	5.145	197841	171576	264184	1,153
İstanbul Outlet Center	Beylikdüzü	31	5,4	367	18007	17959	35447	1,003
Paradise AVM	Büyükcçekmece	24	5,5	266	11700	11549	26681	1,013
Carium AVM	Beylikdüzü	11	5,8	58	2793	2788	5895	1,002

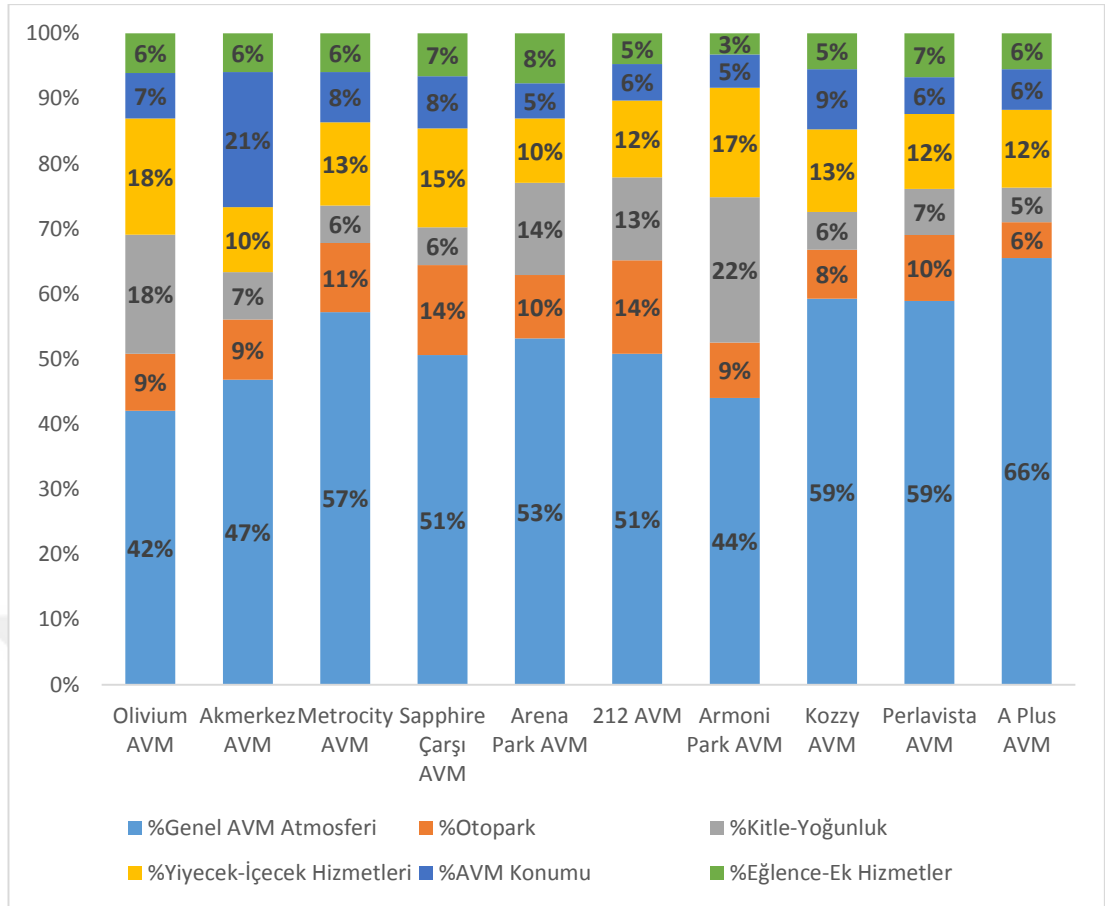
EK B



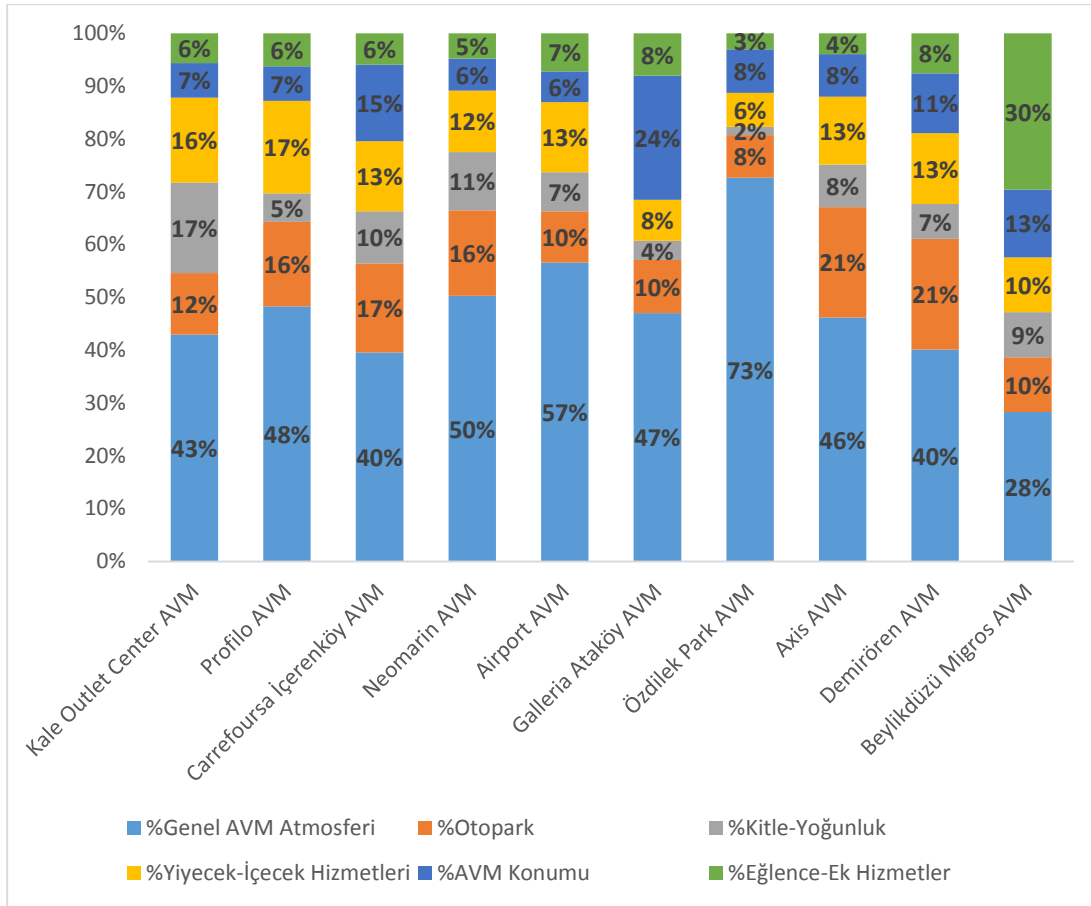
Şekil B.1: 11.-20. Sıradaki AVM'ler İçin Konu Dağılım Grafiği



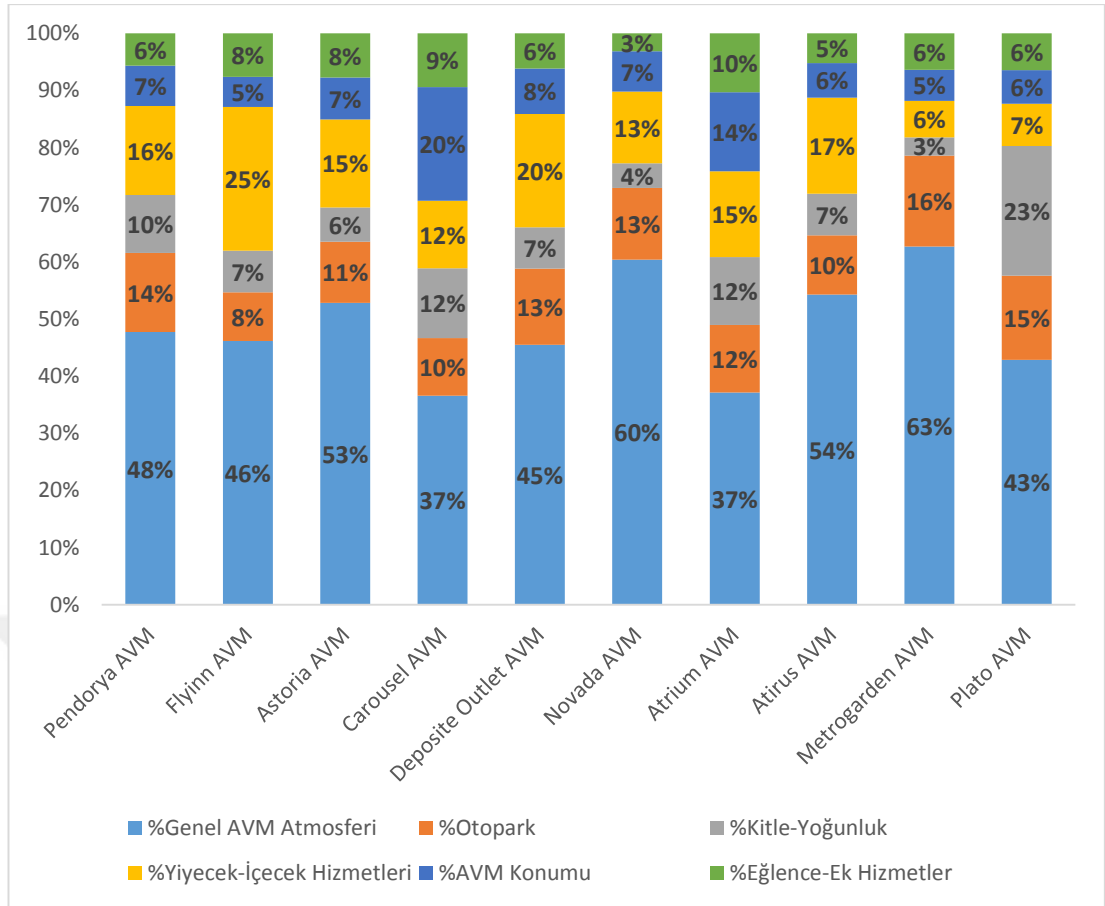
Şekil B.2: 21.-30. Sıradaki AVM'ler İçin Konu Dağılım Grafiği



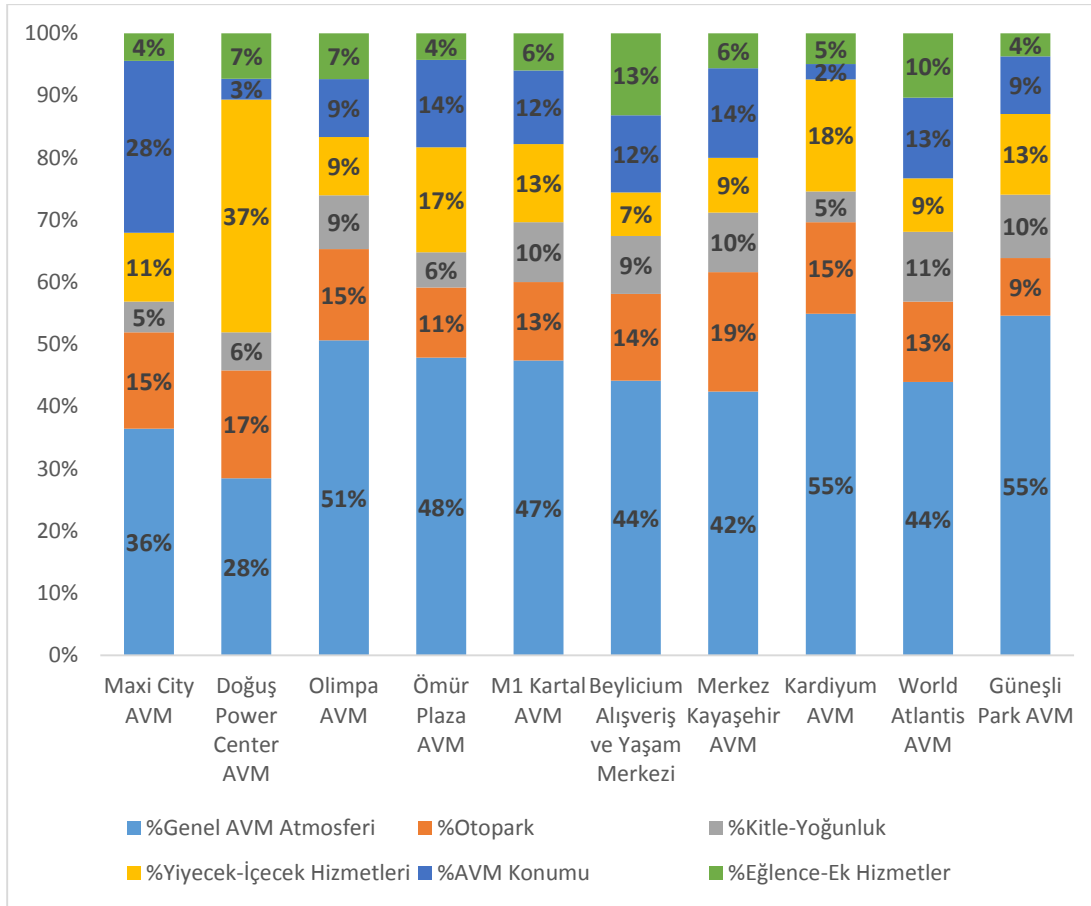
Şekil B.3: 31.-40. Sıradaki AVM'ler İçin Konu Dağılım Grafiği



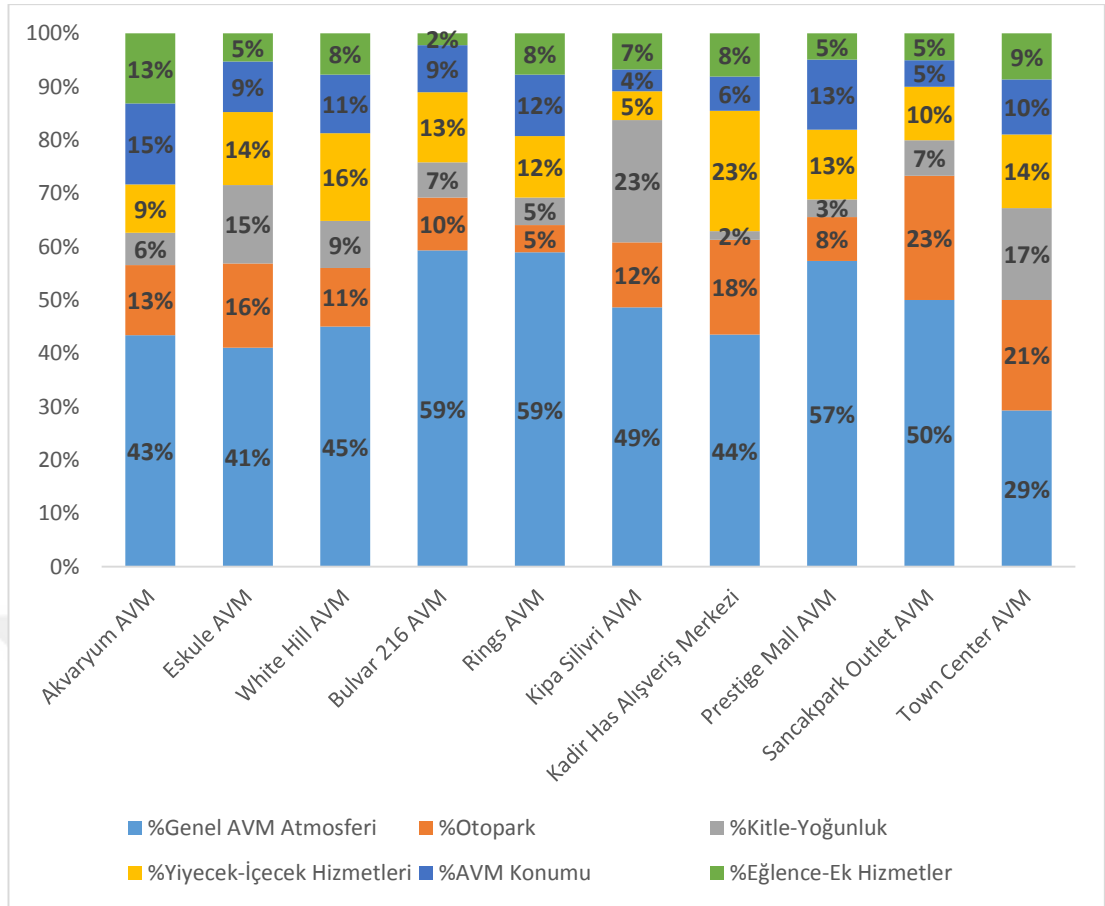
Şekil B.4: 41.-50. Sıradaki AVM'ler İçin Konu Dağılım Grafiği



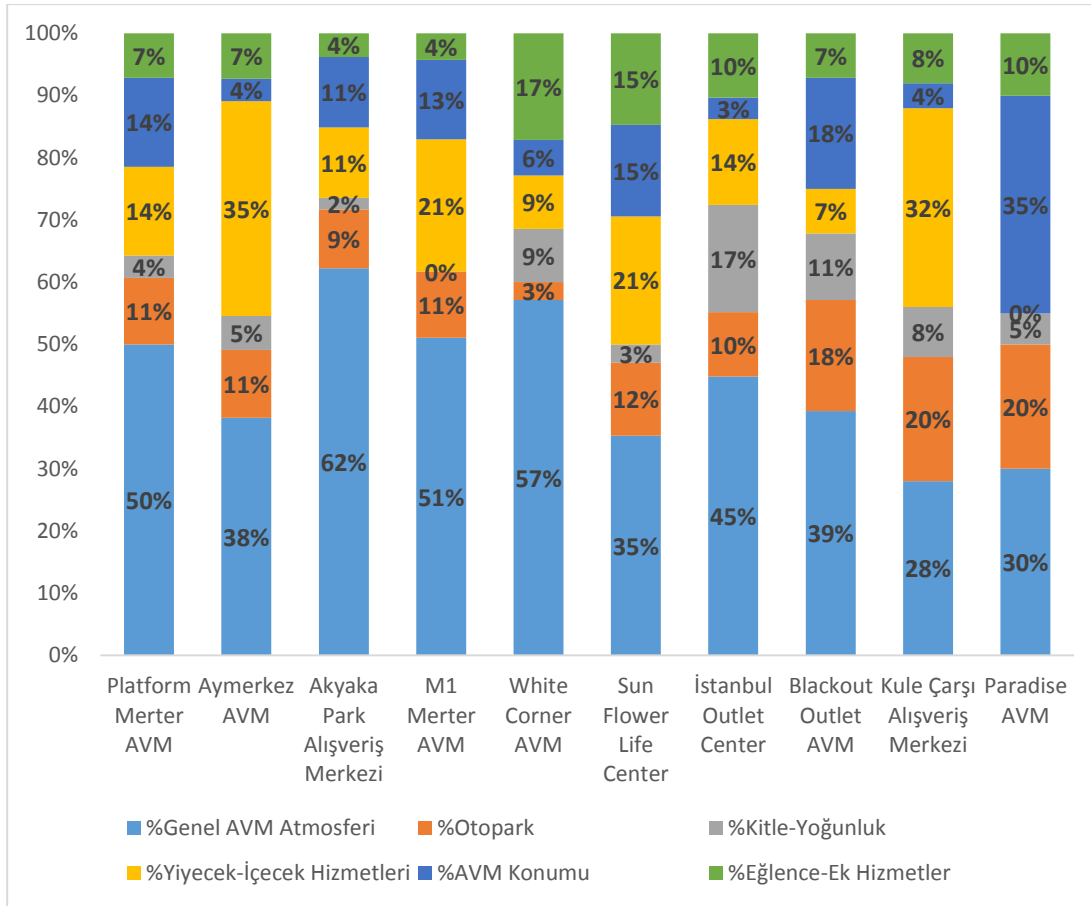
Şekil B.5: 51.-60. Sıradaki AVM'ler İçin Konu Dağılım Grafiği



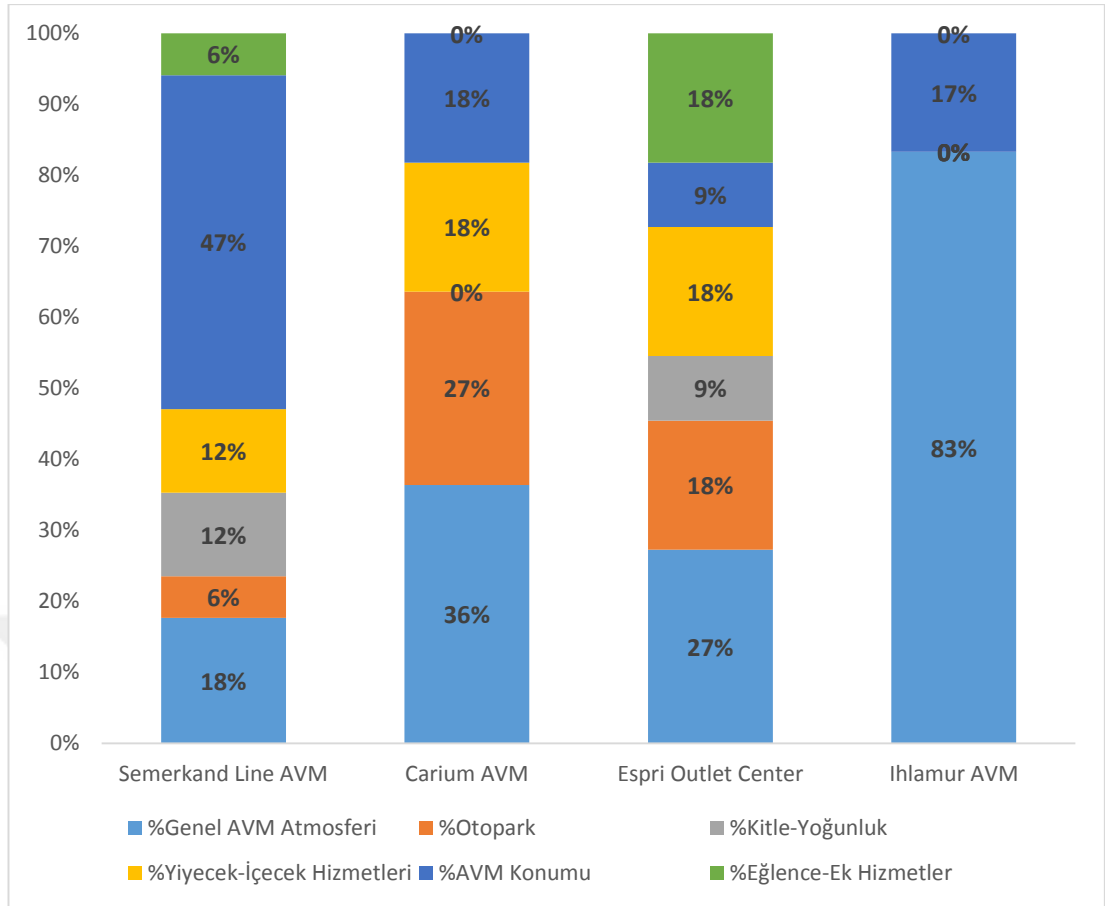
Şekil B.6: 61.-70. Sıradaki AVM'ler İçin Konu Dağılım Grafiği



Şekil B.7: 71.-80. Sıradaki AVM'ler İçin Konu Dağılım Grafiği



Şekil B.8: 81.-90. Sıradaki AVM'ler İçin Konu Dağılım Grafiği



Şekil B.9: 91.-94. Sıradaki AVM'ler İçin Konu Dağılım Grafiği

EK C**Çizelge C.1:** Alışveriş Merkezlerinin Konu Dağılımına Etki Eden Yorum Sayıları
Çizelgesi.

Konular AVM Adı	Genel AVM Atmosferi	Otopark	Kitle- Yoğunluk	Yiyecek- İçecek Hizmetleri	AVM Konumu	Eğlence- Ek Hizmetler
Cevahir AVM	1211	299	587	311	190	168
Forum İstanbul AVM	1329	317	446	256	130	158
İstinye Park AVM	1038	216	182	198	192	116
Viaport Outlet AVM	1082	184	205	203	146	104
Marmara Park AVM	921	217	205	206	125	105
Zorlu Center AVM	864	383	111	114	177	87
Akasya AVM	925	276	81	139	237	58
Vialand AVM	819	206	134	272	124	103
Mall of İstanbul AVM	866	260	127	116	113	73
Marmara Forum AVM	859	164	141	170	82	74
Maltepe Park AVM	700	178	132	171	206	70
Kanyon AVM	739	163	108	143	105	133
Palladium AVM	671	115	81	140	263	56
Capacity AVM	675	126	138	190	78	78
Trump Towers AVM	721	112	70	122	74	72
Optimum Outlet AVM	521	176	137	151	62	114
Aqua Florya AVM	712	175	43	126	61	37
Buyaka AVM	639	157	89	110	92	61
Capitol AVM	504	123	76	156	170	71
Brandium AVM	548	197	59	96	90	64
Starcity AVM	411	97	51	127	39	158
Historia AVM	392	122	178	96	37	57
Torium AVM	439	115	96	98	52	53
Canpark Avm	399	115	99	80	86	31
Akbati AVM	477	79	53	84	51	44
Citys AVM	399	109	46	98	72	43

Çizelge C.1 (Devam): Alışveriş Merkezlerinin Konu Dağılımına Etki Eden Yorum Sayıları Çizelgesi.

Konular AVM Adı	Genel AVM Atmosferi	Otopark	Kitle-Yoğunluk	Yiyecek-İçecek Hizmetleri	AVM Konumu	Eğlence-Ek Hizmetler
Tepe Nautilus AVM	387	89	41	120	62	59
Pelican Mall AVM	430	57	45	100	27	39
Metroport AVM	360	126	43	84	22	44
Meydan İstanbul AVM	371	52	70	59	69	54
Olivium AVM	283	59	123	120	47	41
Akmerkez AVM	309	61	48	66	137	39
Metrocity AVM	377	70	38	84	51	39
Sapphire Çarşı AVM	323	88	37	97	51	42
Arena Park AVM	334	61	89	62	34	48
212 AVM	315	89	79	73	35	29
Armoni Park AVM	259	50	131	99	30	19
Kozzy AVM	346	44	34	74	54	32
Perlavista AVM	343	59	41	67	33	39
A Plus AVM	357	30	29	65	34	30
Kale Outlet Center AVM	230	62	92	86	35	30
Profilo AVM	257	85	29	93	35	33
Carrefoursa İçerenköy AVM	188	80	47	63	69	28
Neomarin AVM	233	75	51	54	28	22
Airport AVM	252	43	33	59	26	32
Galleria Ataköy AVM	206	44	16	34	103	35
Özdilek Park AVM	317	35	7	28	36	13
Axis AVM	201	91	35	56	35	17
Demirören AVM	170	89	28	57	48	32

Çizelge C.1 (Devam): Alışveriş Merkezlerinin Konu Dağılımına Etki Eden Yorum Sayıları Çizelgesi.

Konular AVM Adı	Genel AVM Atmosferi	Otopark	Kitle-Yoğunluk	Yiyecek-İçecek Hizmetleri	AVM Konumu	Eğlence-Ek Hizmetler
Beylikdüzü Migros AVM	112	41	34	41	51	117
Pendorya AVM	169	49	36	55	25	20
Flyinn AVM	158	29	25	86	18	26
Astoria AVM	158	32	18	46	22	23
Carousel AVM	105	29	35	34	57	27
Deposite Outlet AVM	126	37	20	55	22	17
Novada AVM	154	32	11	32	18	8
Atrium AVM	94	30	30	38	35	26
Atirus AVM	126	24	17	39	14	12
Metrogarden AVM	138	35	7	14	12	14
Plato AVM	87	30	46	15	12	13
Maxi City AVM	66	28	9	20	50	8
Doğuş Power Center AVM	51	31	11	67	6	13
Olimpa AVM	76	22	13	14	14	11
Ömür Plaza AVM	68	16	8	24	20	6
M1 Kartal AVM	64	17	13	17	16	8
Beylicium Alışveriş ve Yaşam Merkezi	57	18	12	9	16	17
Merkez Kayaşehir AVM	53	24	12	11	18	7
Kardiyum AVM	67	18	6	22	3	6
World Atlantis AVM	51	15	13	10	15	12
Güneşli Park AVM	59	10	11	14	10	4
Akvaryum AVM	43	13	6	9	15	13
Eskule AVM	39	15	14	13	9	5

Çizelge C.1 (Devam): Alışveriş Merkezlerinin Konu Dağılımına Etki Eden Yorum Sayıları Çizelgesi.

Konular AVM Adı	Genel AVM Atmosferi	Otopark	Kitle-Yoğunluk	Yiyecek-İçecek Hizmetleri	AVM Konumu	Eğlence-Ek Hizmetler
White Hill AVM	41	10	8	15	10	7
Bulvar 216 AVM	54	9	6	12	8	2
Rings AVM	46	4	4	9	9	6
Kipa Silivri AVM	36	9	17	4	3	5
Kadir Has Alışveriş Merkezi	27	11	1	14	4	5
Prestige Mall AVM	35	5	2	8	8	3
Sancakpark Outlet AVM	30	14	4	6	3	3
Town Center AVM	17	12	10	8	6	5
Platform Merter AVM	28	6	2	8	8	4
Aymerkez AVM	21	6	3	19	2	4
Akyaka Park Alışveriş Merkezi	33	5	1	6	6	2
M1 Merter AVM	24	5		10	6	2
White Corner AVM	20	1	3	3	2	6
Sun Flower Life Center	12	4	1	7	5	5
İstanbul Outlet Center	13	3	5	4	1	3
Blackout Outlet AVM	11	5	3	2	5	2
Kule Çarşı Alışveriş Merkezi	7	5	2	8	1	2
Paradise AVM	6	4	1		7	2
Semerkan Line AVM	3	1	2	2	8	1
Carium AVM	4	3		2	2	
Esprı Outlet Center	3	2	1	2	1	2

Çizelge C.1 (Devam): Aışveriş Merkezlerinin Konu Dağılımına Etki Eden Yorum Sayıları Çizelgesi.

Konular AVM Adı	Genel AVM Atmosferi	Otopark	Kitle- Yoğunluk	Yiyecek- İçecek Hizmetleri	AVM Konumu	Eğlence- Ek Hizmetler
Ihlamur AVM	5				1	
Toplam	28306	6939	5344	6507	4739	3468

EK D

```
import pandas as pd

def read_data(input_file, sheet, id_column, text_column):
    data_file = open(input_file, 'rb')
    data = pd.read_excel(data_file, sheetname=sheet)

    # Metin sutunlarini al.
    text_data = pd.DataFrame(data, columns=[id_column, text_column])

    return text_data

def write_data(output_directory, text_data, id_column, text_column):
    for index, row in text_data.iterrows():
        # Cikti dosyasinin adini belirle.
        output_file_name = output_directory + str(row[id_column]) + '.txt'

        # Ciktiyi dosyaya yaz.
        with open(output_file_name, "w") as text_file:
            text_file.write(unicode(row[text_column]).encode('utf8'))

def run():
    print "Foursquare Data Preprocessor"

    input_file = 'input/File Name.xlsx'
    sheet = 'Sheetname'
    id_column = 'TipId'
    text_column = 'TipText'

    output_directory = 'output/'

    print "Reading data ..."
    text_data = read_data(input_file, sheet, id_column, text_column)

    print "Writing data ..."
    write_data(output_directory, text_data, id_column, text_column)

    print "Done."

if __name__ == '__main__':
    ru
```



ÖZGEÇMİŞ

VEŞİKALIK
FOTO

Ad-Soyad : Nergiz Sönmez
Doğum Tarihi ve Yeri : 02.06.1991 Zeytinburnu
E-posta : nergizsonmezz@gmail.com

ÖĞRENİM DURUMU:

- **Lisans** : 2013, Yıldız Teknik Üniversitesi, Makine Fakültesi, Endüstri Mühendisliği Bölümü

MESLEKİ DENEYİM VE ÖDÜLLER:

- Kurumsal Mimar – Türk Telekom Grubu AssisTT A.Ş. (05.2016-Devam)
- Proje Yöneticisi – Türk Telekom Grubu AssisTT A.Ş. (08.2015-04.2016)
- İş ve Sistem Analisti – Türk Telekom Grubu AssisTT A.Ş. (11.2014-07.2015)
- BT Uzman Yardımcısı- Ziraat Hayat ve Emeklilik A.Ş. (08.2013-10.2014)