T.C.
ERCİYES ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ
ÇEVRE MÜHENDİSLİĞİ ANABİLİM DALI

YAĞIŞ SERİLERİNDE KÜMELEME VE ANORMALLİK TESPİTİ İÇİN YOĞUNLUK TABANLI VERİ MADENCİLİĞİ ALGORİTMALARININ GELİŞTİRİLMESİ

Hazırlayan

Ahmet ÖZEKES

Danışman

Doç. Dr. Filiz DADAŞER ÇELİK

Yüksek Lisans Tezi

Ocak 2021

KAYSERİ

T.C.

ERCİYES ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ

ÇEVRE MÜHENDİSLİĞİ ANABİLİM DALI

YAĞIŞ SERİLERİNDE KÜMELEME VE ANORMALLİK TESPİTİ İÇİN YOĞUNLUK TABANLI VERİ MADENCİLİĞİ ALGORİTMALARININ GELİŞTİRİLMESİ

(Yüksek Lisans Tezi)

Hazırlayan

Ahmet ÖZEKES

Danışman

Doç. Dr. Filiz DADAŞER ÇELİK

Ocak 2021

KAYSERİ

BİLİMSEL ETİĞE UYGUNLUK

Bu çalışmadaki tüm bilgilerin, akademik ve etik kurallara uygun bir şekilde elde edildiğini beyan ederim. Aynı zamanda bu kural ve davranışların gerektirdiği gibi, bu çalışmanın özünde olmayan tüm materyal ve sonuçları tam olarak aktardığımı ve referans gösterdiğimi belirtirim.

Ahmet ÖZEKES

YÖNERGEYE UYGUNLUK

"**Yağış Serilerinde Kümeleme ve Anormallik Tespiti İçin Yoğunluk Tabanlı Veri Madenciliği Algoritmalarının Geliştirilmesi**" adlı yüksek lisans tezi, Erciyes Üniversitesi Lisansüstü Tez Önerisi ve Tez Yazma Yönergesi'ne uygun olarak hazırlanmıştır.

Tezi Hazırlayan	Danışman
Ahmet ÖZEKES	Doç. Dr. Filiz DADAŞER ÇELİK

Çevre Mühendisliği ABD Başkanı

Prof. Dr. Oktay ÖZKAN

Doç. Dr. Filiz DADAŞER ÇELİK danışmanlığında **Ahmet ÖZEKES** tarafından hazırlanan "**Yağış Serilerinde Kümeleme ve Anormallik Tespiti için Yoğunluk Tabanlı Veri Madenciliği Algoritmalarının Geliştirilmesi**" adlı bu çalışma, jürimiz tarafından Erciyes Üniversitesi Fen Bilimleri Enstitüsü **Çevre Mühendisliği** Anabilim Dalında **Yüksek Lisans** tezi olarak kabul edilmiştir.

15/01/2021

JÜRİ:

Danışman : Doç. Dr. Filiz DADAŞER ÇELİK

Üye : Dr. Öğr. Üyesi Ahmet Şakir DOKUZ

Üye : Dr. Öğr. Üyesi Evrim KARAÇETİN BELL

ONAY:
Bu tezin kabulü Enstitü Yönetim Kurulunun ……...… tarih ve ……..…… sayılı kararı ile onaylanmıştır.

……. /……../ ……

Prof. Dr. Mehmet AKKURT

Enstitü Müdürü

TEŞEKKÜR

Öncelikle bu tezin konusunun belirlenmesinden bitimine kadar tüm süreçlerde bana yol gösteren ve araştırmayı sevdiren Doç. Dr. Filiz DADAŞER ÇELİK hocama saygılarımı ve şükranlarımı sunuyorum. Çalışma sürecim boyunca rehberliğini eksik etmeyen ve karşılaştığım problemleri anlamamda ve çözmemde yardımcı olan Doç. Dr. Mete ÇELİK hocama da çok teşekkür ediyorum. Değerlendirmelerdeki katkılarından dolayı ayrıca Doç. Dr. Ali Ümran KÖMÜŞÇÜ ve Arş. Gör. Fatma Özge ÖZKÖK'e teşekkür ediyorum.

Özellikle tez yazım sürecinde manevi desteklerini her zaman üzerimde hissettiğim kıymetli aileme ve sevgili eşim Büşra ÖZEKES'e teşekkür ediyorum.

YAĞIŞ SERİLERİNDE KÜMELEME VE ANORMALLİK TESPİTİ İÇİN YOĞUNLUK TABANLI VERİ MADENCİLİĞİ ALGORİTMALARININ GELİŞTİRİLMESİ

Ahmet ÖZEKES

Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Çevre Mühendisliği Anabilim Dalı
Yüksek Lisans Tezi, Ocak 2021

Danışman: Doç. Dr. Filiz DADAŞER ÇELİK

ÖZET

İklim değişikliğinin en önemli göstergelerinden biri yağışlarda meydana gelen değişikliklerdir. Bu tez çalışmasının amacı, veri madenciliği yöntemlerinden biri olan ve kümeleme ve anormallik keşfi için sıklıkla kullanılan yoğunluk tabanlı VDBSCAN algoritmasının geliştirilmesi ve Türkiye yağış serilerindeki anormal yağışların keşfedilmesidir. Çalışmada geliştirilen autoVDBSCAN algoritması, VDBSCAN algoritmasının ihtiyaç duyduğu girdiler olan minimum nokta sayısı (MinPts) ve komşuluk yarıçap (Eps) değerlerini otomatik olarak belirler. Ayrıca bu algoritma ilk adımda aşırı anormal değerlerin keşfedildiği ve ikinci adımında bu değerlerin ayrı tutularak yeniden kümelemenin yapıldığı iterasyonlu yaklaşımı ile gizlenmiş anormal değerleri keşfeder ve nispeten daha verimli bir kümeleme yapar. Türkiye'deki 195 meteoroloji istasyonunun 1980-2015 dönemine ait yıllık yağış verileri ile yapılan mekânsal analiz çalışmasında algoritma 3 aşırı anormal, 14 gizlenmiş anormal ve 8 küme belirlemiştir. Bununla birlikte, bu istasyonlara ait aylık yağış verileri kullanılarak zamansal analiz yapılmıştır. Dönem boyunca meydana gelen anormal yağışların Ege Bölgesi'nde, Akdeniz Bölgesi'nde, Doğu Anadolu Bölgesi'nde ve İç Anadolu Bölgesi'nin güney kısmında yoğunlaştığı tespit edilmiştir ve Muş Varto, Kahramanmaraş ve Hakkari Merkez istasyonlarının ilk sıralarda oldukları görülmüştür. Bir diğer yandan, 352 anormal yağışın keşfedildiği 2010 yılı en yüksek anormal yağış sayısına sahip yıl olmuş ve ardından 2014, 2009, 1997 ve 2015 yılları gelmiştir. 1993, 1980 ve 2007 ise en az sayıda anormal yağışın görüldüğü yıllardır. Bununla birlikte en çok anormal yağış Eylül aylarında, en az anormal yağış ise Nisan aylarında keşfedilmiştir.

Anahtar Kelimeler: Yağış, Veri Madenciliği, Anormallik Tespiti, Kümeleme, VDBSCAN, autoVDBSCAN

DEVELOPMENT OF DENSITY BASED DATA MINING ALGORITHMS FOR CLUSTERING AND ANOMALY DETECTION IN PRECIPITATION TIME SERIES

Ahmet OZEKES

Erciyes University, Graduate School of Natural and Applied Sciences M.Sc. Thesis, January 2021

Supervisor: Assoc. Prof. Dr. Filiz DADAŞER ÇELİK

ABSTRACT

One of the most important indicators of climatic change is the changes in precipitation. The purpose of this study is to improve density based VDBSCAN algorithm, which is one of the data mining methods and is often used for clustering and anomaly detection. The algorithm is aimed to be used for exploring anomalies in precipitation time-series of Turkey. AutoVDBSCAN algorithm developed in the study automatically determines the minimum points (MinPts) and neighborhood radius values (Eps), which are the input parameters required by the VDBSCAN algorithm. It also contains a level-wise approach. First, extreme anomalies of dataset are determined and excluded from the dataset in order to find optimum input parameters. Secondly, it determines clusters and suppressed anomalies. Spatial analysis with annual precipitation data, collected from 195 meteorological stations of Turkey for the period of 1980-2015, was performed and 3 extreme anomalies, 14 suppressed anomalies, and 8 clusters were discovered. In addition, temporal analysis was performed with monthly precipitation data. It was determined that the anomalous precipitation occurring during the period is concentrated in the Aegean Region, Mediterranean Region, Eastern Anatolia Region and the southern part of the Central Anatolia Region. Mus Varto, Kahramanmaras and Hakkari stations were the top of the list in terms of having the highest number of anomalies. On the other hand, the year 2010 had the highest number of anomalies with 352 anomalies and was followed by 2014, 2009, 1997, and 2015. Also, 1993, 1980 and 2007 were the years with the least number of anomalies. At the same time, the highest number of anomalies was detected in September, and the least number of anomalies was detected in April.

Keywords: Precipitation, Data Mining, Anomaly Detection, Clustering, VDBSCAN, autoVDBSCAN

İÇİNDEKİLER

2. BÖLÜM

YÖNTEM

3. BÖLÜM

BULGULAR

4. BÖLÜM

TARTIŞMA, SONUÇ VE ÖNERİLER

TABLOLAR LİSTESİ

ŞEKİLLER LİSTESİ

GİRİŞ

Dünyanın doğal dengesi 19. yüzyıldan itibaren insan kaynaklı sebeplerden dolayı bozulmaya başlamıştır. Bunun sonucu olarak iklimsel değişiklikler ortaya çıkmakta ve giderek etkileri şiddetlenmektedir. Doğal dengeyi bozan faaliyetlere karşı ciddi önlemler alınmadığı sürece dünya iklim sistemlerindeki değişikliklerin artacağı ve sonuçlarının ağır olacağı açıktır [1].

2019 yılında yayınlanan Hükümetlerarası İklim Değişikliği Paneli (IPCC) raporuna [2] göre sanayi devrimi öncesi dönemden bu yana, kara yüzeyi hava sıcaklığı, küresel ortalama yüzey sıcaklığının neredeyse iki katı yükselmiştir (yüksek güven). Ekstrem olayların sıklığı ve yoğunluğundaki artışlar da dâhil olmak üzere iklim değişikliği, karasal ekosistemleri olumsuz etkilemiş, ayrıca birçok bölgede çölleşme ve arazi bozulmasına neden olmuştur. 1850-1900'den 2006-2015'e kadar ortalama kara yüzeyi hava sıcaklığı 1.53°C (çok büyük olasılıkla 1.38°C ila 1.68°C aralığında) artarken, küresel ortalama yüzey sıcaklığı 0.87°C (muhtemelen 0.75°C ila 0.99°C aralığında) artmıştır. Küresel ısınma, çoğu kara bölgesindeki ısı dalgaları da dâhil olmak üzere ısıyla ilgili olayların sıklığının, yoğunluğunun ve süresinin artmasına neden olmuş ve yağış desenlerinin değişmesine yol açmıştır (yüksek güven). Güney Amerika'da, Batı Asya'da, Akdeniz'de ve Afrika'da görülen kuraklıkların yoğunluğunda ve sıklığında artış belirlenmiştir (orta düzeyde güven). Yoğun yağış sıklığı ise küresel ölçekte artmıştır (orta düzeyde güven).

Türkiye'nin mevsimsel ortalama sıcaklık değerlerine göre, kış aylarında bazı bölgelerde artış bazılarında ise azalma eğilimi görülmüştür. İlkbahar mevsiminde ortalama hava sıcaklıklarında Türkiye genelinde bir artış vardır ve özellikle Akdeniz Bölgesi'nde bu artış daha kuvvetlidir. Soğuma eğilimleri ise genel olarak Karadeniz Bölgesi'nde görülmektedir. Kentleşmenin hızla arttığı İstanbul, Akdeniz ve Ege Bölgeleri'nin kıyı şeritleri ve Güneydoğu Anadolu Bölgesi ısınma eğilimleri açısından ön plandadırlar [3].

İklim değişikliği hem küresel hem de yerel ölçekte önemli değişikliklere sebep olmaktadır. Farklı iklimsel olaylarının bölgesel dağılımında, şiddetinde ve sıklığında farklılıklar meydana gelmektedir. Bu değişimden Türkiye'ye düşen pay, istatistiksel olarak anlamlı ısınma eğilimleri yaşayan Güneydoğu Anadolu, Ege, Marmara ve Akdeniz Bölgeleri'nde açıkça görülmektedir [4].

İklimsel değişikliklere özellikle yarı-nemli ve çok-kurak bölgeler oldukça hassastır. Türkiye'nin de içinde bulunduğu Akdeniz Havzası'nda ve alt-tropikal iklim kuşağında 1970'li yıllardan itibaren yağışlarda azalma eğilimleri ve kuraklıklar ortaya çıkmıştır. Bununla birlikte, Türkiye'de Ege, Marmara, Güneydoğu Anadolu, Akdeniz Bölgeleri'nde ve Doğu ve İç Anadolu'nun güney kısımlarında özellikle ilkbahar ve kış mevsimlerindeki toplam yağışlarda açık bir azalma eğilimi yani kuraklaşma görülmektedir [5].

Türkiye'de ve çevresinde yapılan birçok çalışmaya ve bölgesel ve küresel iklim modellerine göre Türkiye iklim değişikliğinin etkilerini önemli düzeyde yaşamaktadır ve bu etkiler gelecekte artarak devam edecektir [6].

Meteoroloji istasyonları sayesinde hemen hemen tüm dünyada günlük olarak sıcaklık ve yağış gibi iklimsel veriler toplanmaktadır. Uzun yıllar boyunca elde edilen bu veriler iklimsel değişikliklerin takip edilmesine olanak sağlamaktadır. Bu noktada ise klasik yöntemlerin ötesinde olan "veri madenciliği" yöntemleri büyük veri setlerinin incelenmesinde oldukça elverişli bir araçtır.

Veri madenciliği, veri setlerinden önceden bilinmeyen bilgileri ayıklama, ilginç ve anlamlı kalıpları tespit etme sürecidir [7]. Veri madenciliği, büyük veri havuzlarından yararlı bilgiler elde etme bilimi, bilgisayar biliminde genç ve disiplinler arası bir alan olarak ortaya çıkmıştır. Bu alanın teknikleri endüstri, bilim, mühendislik ve sağlık gibi birçok alanda yaygın olarak kullanılmaktadır [8]. Veri madenciliği teknolojisi, büyük miktardaki veriyi otomatik ve etkin bir şekilde dönüştürmek için araçlar sağlamak ve kullanıcılara ilgili bilgiyi vermek amacıyla geliştirilmiştir. Çoğunlukla ilişkilendirme kuralları, karar ağaçları veya kümeler şeklinde ifade edilen yöntemler, karar verme süreçlerini kolaylaştırmak için verilerde derinlemesine gömülü ilginç kalıpların ve düzenlerin bulmasına olanak tanımaktadır [9].

Tezin Önemi ve Amacı

Bu tezin çalışmasında yağış serilerinde kümeleme ve anormallik tespiti için veri madenciliği tekniklerinin geliştirilmesi amaçlanmıştır. Bu amaçla, veri madenciliğinin "kümeleme" algoritmalarından biri olan DBSCAN algoritmasının bir varyantı geliştirilerek, veri analizlerinde daha doğru sonuçlar elde edebilmesi hedeflenmiştir.

Bunun yanında, geliştirilen algoritmanın Türkiye yağış serilerinin analizleri için kullanılması amaçlanmıştır. Türkiye, üç tarafında deniz olması, çeşitlilik içeren topografyası ve orografik özellikleri sebebiyle oldukça karışık bir iklim yapısına sahiptir. Bu farklılıklardan dolayı iklim değişikliğinin Türkiye üzerindeki etkisi her bir bölgede farklı derecede ve şekilde ortaya çıkacaktır. Bu nedenle iklimin en önemli göstergelerinden biri olan yağış serilerinin incelenmesi bu tezin en önemli katkılarından biridir.

Tezin İçeriği

Bu tez çalışması dört bölümden oluşmaktadır. Birinci bölümde veri madenciliği hakkında genel bilgiler verilmekte ve yağış serilerinde veri madenciliği uygulamaları ile ilgili literatür özeti sunulmaktadır. İkinci bölümde bu çalışmada kullanılan veri seti ve geliştirilen autoVDBSCAN algoritması tanıtılacaktır. Üçüncü bölümde bulgulara yer verilmektedir. Bu kapsamda yapılan çalışmalar arasında Türkiye yağış serilerindeki değişimlerin mekânsal ve zamansal olarak incelenmesinde, ayrıca anormal yağışların tespiti ve bunların mekânsal ve zamansal dağılımının belirlenmesi yer almaktadır. Dördüncü ve son bölümde sonuçlar ve öneriler sunulmaktadır.

1. BÖLÜM

GENEL BİLGİLER VE LİTERATÜR ÖZETİ

1.1. Veri Madenciliği Hakkında Genel Bilgiler

Gelişen teknoloji ile günümüzde veri toplanması ve bunların depolanması olanakları oldukça artmıştır. Bununla birlikte, yararlı bilgilerin çıkarılmasının son derece zor olduğu kanıtlanmıştır. Veri kümesinin çok büyük olması çoğunlukla geleneksel yöntemlerin kullanılmasını kısıtlamaktadır. Öyle ki bazen verilerin geleneksel olmayan yönleri sebebiyle veri kümelerinin küçük olması durumunda dahi bu yöntemler kullanılamamaktadır. Yeni yöntemlerin geliştirilmesi gereği bu ve benzeri ihtiyaçlardan doğmaktadır.

Veri madenciliği, büyük hacimli verileri işlemek için geleneksel veri analizi yöntemlerini gelişmiş algoritmalarla harmanlayan bir teknolojidir ve ayrıca eski veri türlerini yeni yöntemlerle analiz etmeye de olanak sağlamaktadır [10].

Aşağıda veri madenciliğinin tanımı, veri madenciliğinin yöntemleri ve bu tez çalışmada kullanılan kümeleme algoritmaları hakkında genel bilgiler sunulmaktadır.

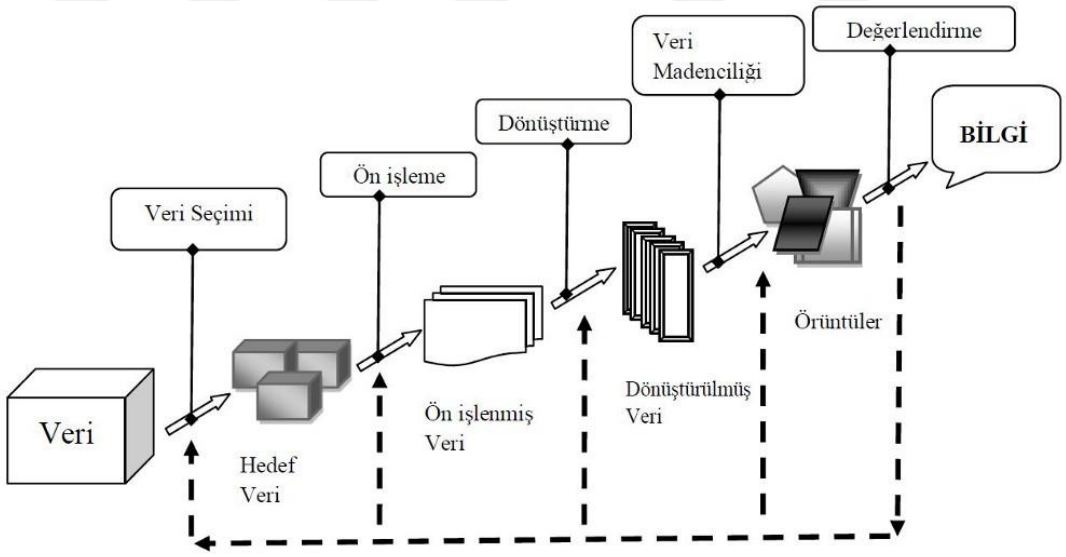
1.1.1.Veri madenciliği nedir?

Disiplinler arası bir konu olan veri madenciliği genel tanımı itibariyle veri içerisindeki kolayca anlaşılamayacak ve işe yarar olan bilginin ortaya çıkarılmasıdır [11]. Veri madenciliği veri yığınından ilişkilerin, örüntülerin, değişimlerin, sapmaların ve istatistiksel açıdan önemli olan yapıların keşfedilmesinde kullanılan birtakım tekniklerin bütünüdür. Örneğin, bir mağazada herhangi bir ön hipoteze dayanmadan satış hacmini etkileyen faktörlerin belirlenmesi, hangi grup müşterilere sahip olduğunun anlaşılması

ya da ürünlerin arasındaki birliktelik ilişkilerinin ortaya çıkarılması için veri madenciliği teknikleri kullanılabilir ve belirli varsayımlar altında kolayca tanımlanamayan örüntüler ve gizli ilişkiler keşfedilebilir [11].

1.1.2.Veri Madenciliği Adımları

Bilginin keşfedilme süreci Şekil 1.1'de gösterilmiştir [12]. Veri seçme ve ön işleme adımlarında veriler toplanır, eğer varsa birden fazla veri kaynağından alınan veriler birleştirilir ve homojen olmayan veriler temizlenir. Veri dönüşümü adımında ise istenilirse orijinal verilerin bütünlüğünden ödün vermeyecek şekilde daha küçük bir temsilini alarak veri azaltımı yapılabilir ve veriler veri madenciliğine uygun formlara dönüştürülür. Veri madenciliği adımında anlamlı kalıplar ve desenler keşfedilir. Değerlendirme kısmında ise veri madenciliği adımında elde edilen örüntüler arasında gerçekten anlamlı olanları seçilir. Son olarak elde edilen bilgi görselleştirme ve benzeri teknikler kullanılarak sunulur.



Şekil 1.1. Veri madenciliği sürecindeki adımlar [12]

1.1.3.Veri Madenciliği Modelleri

Veri madenciliği modelleri, genel olarak tanımlayıcı (descriptive) ve tahmin edici (predictive) modeller olmak üzere ikiye ayrılırlar [13]. Tahmin edici modellerde ön verilerle bir model kurulur ve bu modelden yararlanılarak sonucu bilinmeyen verilerin sonucunun tahmini sağlanır. Örneğin, bir banka yeni bir müşterinin kredi başvurusunu değerlendirirken daha önceki müşterilerinin verilerinden elde edilen modeli kullanabilir. Böylelikle yeni müşterinin özelliklerine göre krediyi geri ödeyip ödemeyeceğini tahmin eder.

Tanımlayıcı modellerde ise ön veri kullanılmaz ve model doğrudan verilerdeki örüntüleri tanımlar. Burada da bir örnek verecek olursak, A-B aralığında maaş alan ve iki çocuklu bir çalışanın, C-D aralığında maaş alan ve bir arabası olan fakat çocuğu olmayan bir emekli ile ödeme kapasitelerinin benzerliğini ortaya koymak tanımlayıcı bir modeldir.

Veri madenciliği modelleri işlevleri açısından üç alt başlıkta toplanmıştır [13]:

- Sınıflandırma (classification) ve Regresyon (regression) ağaçları
- Birliktelik Analizi (association analysis)
- Kümeleme (clustering)

Sınıflama ve regresyon modelleri tahmin edici, kümeleme ve birliktelik kuralları modelleri ise tanımlayıcı modellerdir.

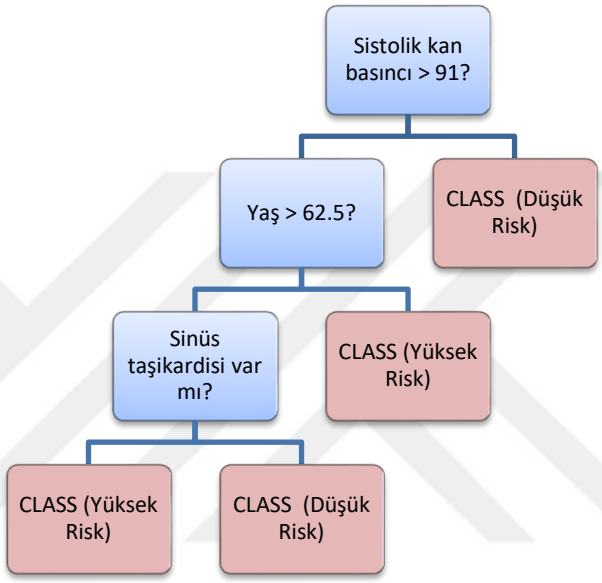
1.1.3.1. Sınıflandırma ve Regresyon

Sınıflandırma ve Regresyon Ağaçları (CART) yöntemi ön veriler ile karar ağaçları oluştur (öğrenme) ve sınıf sayısının da bilinmesini gerektirir. Karar ağaçlarında önceden atanmış sınıflar oluşturur. Böylelikle CART, eldeki verileri ön verilerle oluşturulmuş karar ağaçlarına göre önceden atanmış sınıfları kullanarak sınıflara ayırır.

Karar ağaçları, öğrenme verilerini parçalara ayırmak ve sınıflar oluşturmak için bir takım soruları kullanır. Bu sorular sadece “evet/hayır” cevaplarını içerir. Örneğin “Maaş 5000 TL den fazla mı?” ya da “Kişi üniversite mezunu mu?” gibi sorularla en iyi bölünmeyi oluşturacak değişkenleri ve değerlerini belirlemeye çalışır. Buradaki amaç

veriyi mümkün olduğunca homojen bir şekilde ikiye bölmektir ve bu işlem her bölünmüş kısma tekrar uygulanır.

Şekil 1.2'de San Diego Tıp Merkezi tarafından hastalarını farklı risk seviyelerine göre sınıflandırmak için kullanılan basit bir sınıflandırma ağacı örneği verilmiştir [14].



Şekil 1.2. San Diego Tıp Merkezi hastalarının sınıflandırma ağacı [14]

Her soruda yeni düzeyler oluşturulur ve birçok değişken içeren çok daha karmaşık ağaçlar oluşturulabilir. Verilen örnekten de anlaşılacağı gibi CART hem sayısal hem de kategorik değişkenleri işleyebilir.

1.1.3.2. Birliktelik Analizi

Büyük miktarda toplanan ve depolanan veriler içerisinde mevcut ikili ilişkilerin analiz edildiği yöntemlerdir. Birliktelik kuralları, günümüzde özellikle şirketlerin karar mekanizmalarını etkileyen önemli bir veri madenciliği tekniğidir [15].

Market sepetindeki ürünler üzerine ilişki keşifleri birliktelik kuralının anlaşılmasını sağlayan iyi bir örnektir. Buradaki amaç müşterilerin alışveriş listelerini inceleyerek satın alma biçimlerini keşfetmektir. Örneğin, yumurta alan müşterilerin sıklıkla pirinç

de aldığının tespit edilmesiyle birlikte market yöneticileri raf düzenlerini bu keşfedilmiş bilgiye göre tasarlar ve bunun sonucunda da satış oranlarında bir artış yakalanabilir.

Aşağıdaki birliktelik kuralı örneğinde bir mağazada satılan X ve Y ürünlerinin müşteriler tarafından yapılan tüm alışverişlerde birlikte satın alınması durumu gösterilmiştir [16]:

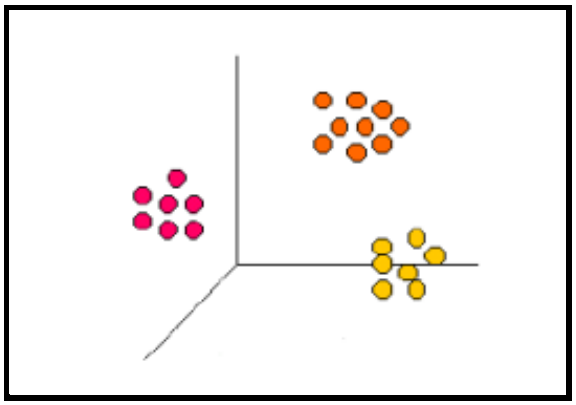
X => Y [destek = %3, güven = %40]

Burada "destek" ile belirtilen, tüm alışverişler içerisinde hangi oranda bu ilişkinin gözlemlendiğini gösterir. Kuraldaki "güven" ise her X ürününü alan müşterinin yüzde kaçının aynı zamanda Y ürününü de aldığını ifade eder. Kullanıcı tarafından belirlenen minimum destek eşik ve minimum güven eşik oranları aşıldığı durumlarda birliktelik kuralları ortaya çıkar.

1.1.3.3. Kümeleme

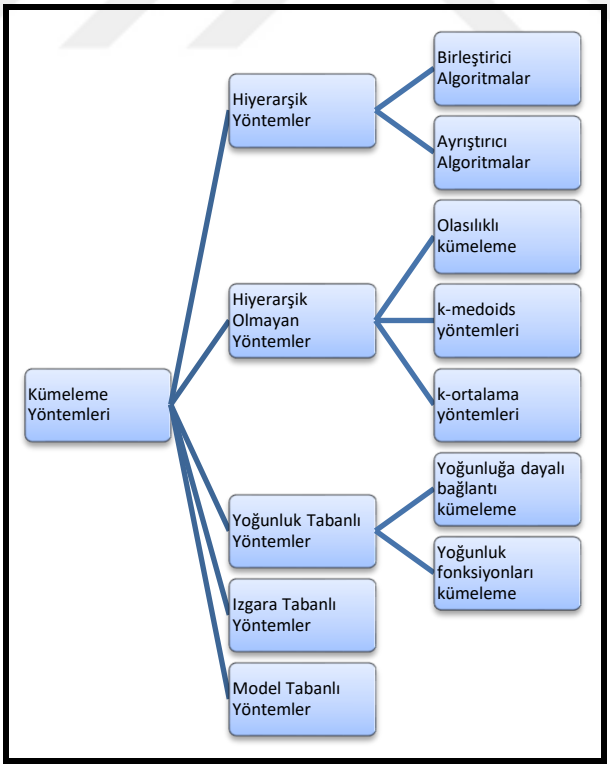
Bir veri setindeki elemanları birbirlerine yakınlıklarına göre gruplara ayırma işlemi kümeleme analizidir. Aynı kümeye ait elemanların en yüksek benzerliği göstermesi ve farklı gruplardaki elemanların ise birbirlerine en az benzemesi sağlanır. Tanımlayıcı modellerden biri olan kümeleme analizleri sınıflandırma yöntemlerinde olduğu gibi ön veri ile model oluşturmaya gerek duymaz, yani denetimsiz/kontrolsüz sınıflandırma yapar [17]. Doğrudan kümelenmek istenen verinin girdisiyle kümeler elde edilir.

Kümeleme işlemlerinde birbirlerine alternatif olarak Manhattan Mahalanobis, öklidyen, Kareli Öklidyen ve Standardize Öklidyen gibi birkaç farklı mesafe ölçü birimleri kullanılmaktadır. Aynı kümeye ait olan elemanlar ortak özellikler gösterirler [18]. Şekil 1.3'te benzer olanların birlikte kümelendiği ve farklı özelliklere sahip kümelerin oluşturulduğu bir örnek gösterilmektedir.



Şekil 1.3. Koordinat düzlemi üzerinde bir kümeleme örneği [18]

Kümeleme teknikleri genel olarak Şekil 1.4'te gösterildiği gibi hiyerarşik, hiyerarşik olmayan, yoğunluk tabanlı, ızgara tabanlı ve model tabanlı yöntemleri içermektedir. Aşağıda bu yöntemler hakkında kısaca bilgi verilmektedir.

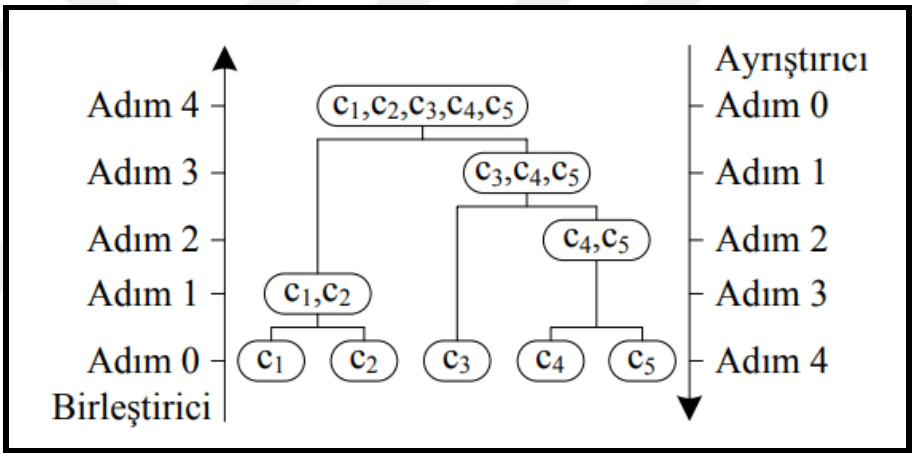


Şekil 1.4. Kümeleme Yöntemleri [19]

1.1.3.3.1. Hiyerarşik Kümeleme

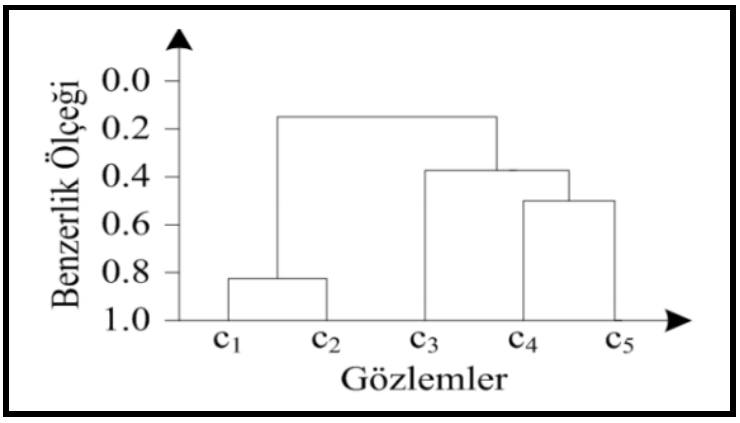
Hiyerarşik kümeleme algoritmaları, temelde birleştirici veya ayrıştırıcı olmak üzere iki ana gruba ayrılırlar [20]. Birleştirici algoritmalar, başlangıçta veri setindeki her bir elemanı küme olarak varsayar ve ardından tekrar tekrar benzer olanları birleştirir. Ayrıştırıcı algoritmalar ise ilk önce tüm elemanların tek bir küme oluşturduğunu varsayar ve daha sonra onları bölerek yeni kümeler oluşturur.

Birleştirici ve ayrıştırıcı hiyerarşik kümeleme yöntemleri Şekil 1.5'te gösterildiği gibi sırasıyla bütüne ulaşan ve bütünü bölen olarak işlev gösterirler.



Şekil 1.5. Birleştirici ve Ayrıştırıcı Hiyerarşik Kümeleme [20]

Bununla birlikte bu yöntemlerden elde edilen çıktılar "dendogram" ile temsili gösterilmektedir. Şekil 1.6'da yukarıdaki örnekten elde edilmiş dendogram gösterilmektedir.



Şekil 1.6. Bütünleştirici Hiyerarşik Kümeleme Örnek Bir Dendogram [20]

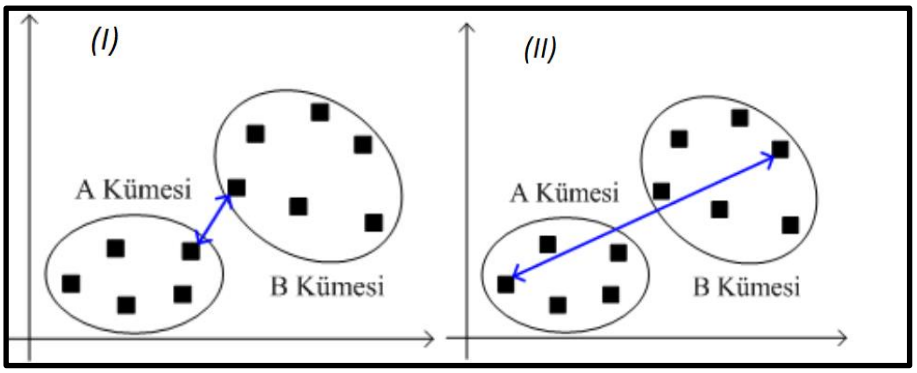
Hiyerarşik kümelemede En yakın komşu, En uzak komşu, Ortalama uzaklık komşu, Ward tekniği, Medyan tekniği ve Centroid tekniği gibi birbirinden farklı birçok yöntem geliştirilmiştir [21]. Bunlardan bazıları aşağıda kısaca açıklanmıştır.

En Yakın Komşu Yöntemi:

Bu yöntem "tek bağlantı kümeleme yöntemi" olarak da bilinmektedir [22]. En yakın komşu algoritması ilk önce veri setindeki elemanlar (i,j) arası uzaklıklar belirlenir ve daha sonra *Min[d(i,j)]* değerini tespit edilir. Bu değerin gözlemlendiği satır birleştirilerek bir küme oluşturulur. Oluşan yeni kümeler üzerinden tekrardan farklı kümelerin birbirlerine en yakın elemanlarının arasındaki uzaklıklar (Şekil 1.7) dikkate alınarak en yakın komşular birleştirilir.

En Uzak Komşu Yöntem:

Bu yöntem ise aynı zamanda "tam bağlantı kümeleme yöntemi" olarak bilinmektedir [22]. En uzak komşu yönteminde ilk kümeleri oluşturmak için aynı şekilde d(i,j) değerleri hesaplanır ve kümeler oluşturmak üzere *Min[d(i,j)]* tespit edilir. En yakın komşu yönteminden farklı olarak burada farklı kümeler arasındaki birbirine en uzak elemanlar arasındaki mesafeler dikkate alınır. Hangi noktalar arası uzaklığın küme birleştirilmesine etki ettiği Şekil 1.7'de gösterilmiştir.



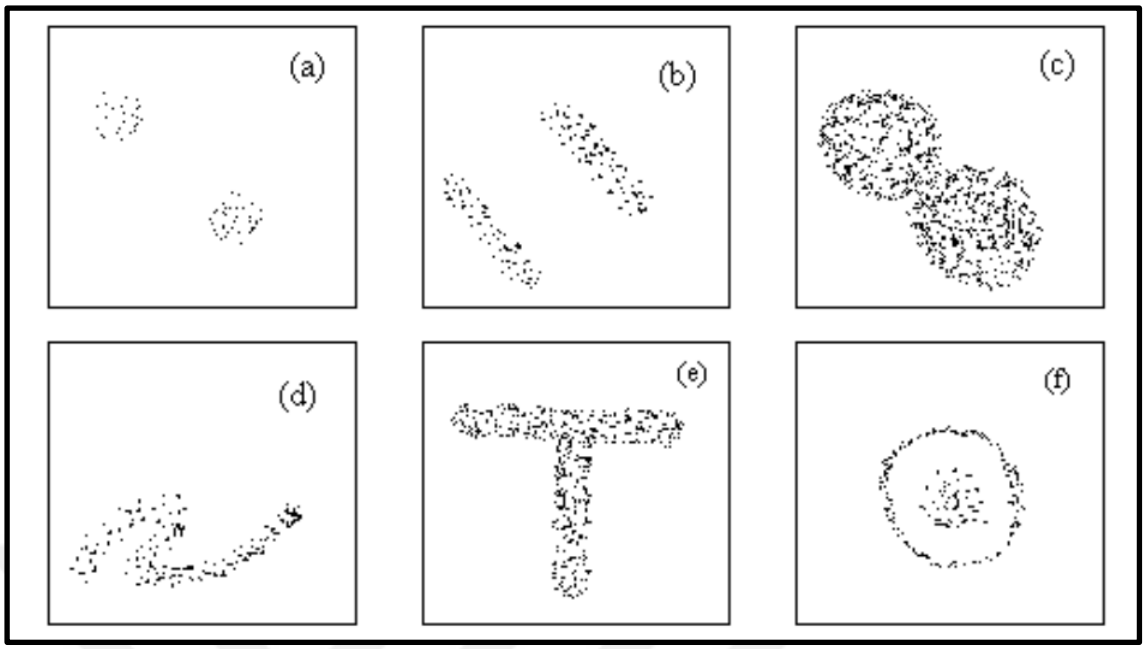
Şekil 1.7. En Yakın ve (II) En Uzak Gözlemler Arası Uzaklıklar [22]

Ward Yöntemi:

Ward, minimum varyans metodu olarak da adlandırılmaktadır. Teknik, kümeler içi kareler toplamı minimum olan (grup içi varyans minimum) iki kümeyi birleştirmeye çalışmaktadır [23]. Ward tekniği, değişkenliği minimum olan kümeleri birleştirir ve yapısı gereği diğer tekniklere kıyasla daha istatistikseldir. $\bar{x}r$ ve $\bar{x}s$ *Br* ve *Bs* için küme ortalama vektörleri ve iki kümenin birleşimi ile elde edilen küme ortalamasını ifade eden $\bar{x}q = (nr\ \bar{x}r + ns\ \bar{x}s) / (nr + ns)$ ise; sözkonusu küme $Bq = Br \cup Bs$ dir.

Yukarıda açıklanan hiyerarşik yöntemlerin her biri farklı durumlar için avantajlar veya dezavantajlar gösterir. Aynı veri setlerini özellikle kümelerin şekillerinden dolayı farklı şekillerde kümelerler.

Örneğin, x_1 ve x_2 değişkenlerinin olduğu ve nesnelerin bu değişkenlere göre değerlendirildikleri düşünüldüğünde bazı olası durumlar Şekil 1.8'de verilmiştir [24]. (a) ve (b) durumları herhangi bir tutarlı algoritma tarafından bulunabilir, (c) durumunda da bazı algoritmalar ara noktalardan dolayı iki kümeyi tespit etmede yanılabilir ve (d),(e),(f) gibi durumlarda birçok algoritma kümelemede zorlanabilir.



Şekil 1.8. İki Küme Oluşturan Noktaların Bazı Durumları [24]

1.1.3.3.2. Hiyerarşik Olmayan Kümeleme

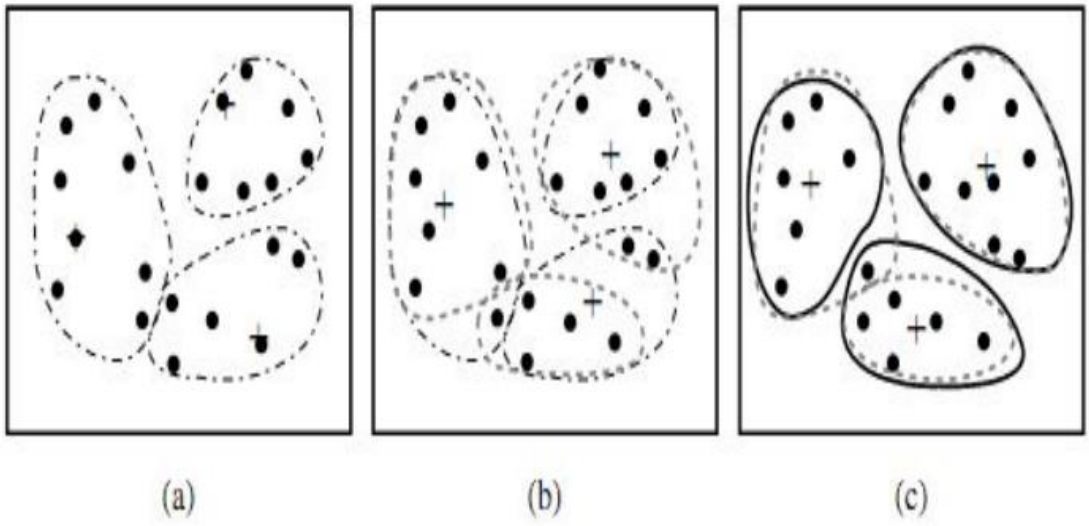
Hiyerarşik olmayan (bölümlemeli) kümeleme yöntemleri her elemanı doğrudan tek seviyeli kümelere atar. Bu yöntemde, hiyerarşik kümeler yoktur ve her kümenin temsili bir merkezi vardır. Bölümlemeli yöntemler verimlidir ve kolay uygulanabilir olduklarında iyi sonuç üretirler [25]. Bu kümeleme yöntemleri kullanıcının küme sayısını girmesini ister. Bu yüzden, tutarlı küme sayısının belirlenmesi için ayrı bir çalışma yapılmalıdır. Bölümlemeli kümeleme yöntemlerinden en yaygın kullanılanları K-ortalamalar, K-medoids ve olasılıklı kümelemedir.

K-ortalamalar Yöntemi:

K-ortalamalar algoritması, verileri kümelemek için yinelemeli bir yol izlemektedir. Başlangıçta, oluşturulacak küme sayısını ifade eden k değeri belirlenir. Ardından, veri içerisinden k adet rasgele küme merkezleri belirlenir ve her bir elaman uzaklık hesaplarına göre kendisine en yakın bulunan merkezin temsil ettiği kümeye atanır. İlk kümeler oluştuktan sonra ise her bir küme için yeni birer merkez tayin edilir ve tekrar bu merkezlere en yakın olan elemanlar kümelenir. Bu işlem artık kümelerin merkezleri

değişmeyene kadar yinelenir. K-ortalamalar yönteminde her bir eleman yalnızca bir kümeye ait olabilir.

Şekil 1.9'da (a) dan (c) ye ulaşana kadar yapılan bir yineleme gösterilmektedir.



Şekil 1.9. K-ortalamalar algoritması yineleme işlemi [26]

K- ortalamalar algoritmasının akış şeması Şekil 1.10'da verilmektedir. Algoritmanın adımlarını aşağıdaki gibi sıralayabiliriz [26]:

Girdiler

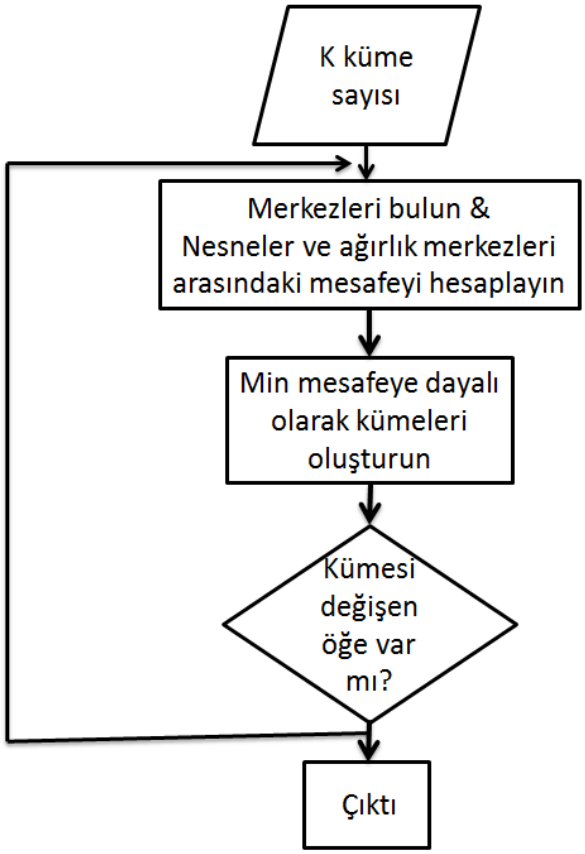
D = { d_1, d_2... d_n} // n adet öğe içeren veri seti

K // istenilen küme sayısı

Output: k adet küme dizisi

Adımlar:

1. D den rasgele k adet küme merkezi seçin;
2. Tekrar;
3. Her bir d (i) öğesini en yakın ağırlık merkezine sahip kümeye atayın;
4. Her küme için yeni ortalamayı hesaplayın;
5. Yakınsama kriterleri karşılanana kadar 3 ve 4 tekrar edin.



Şekil 1.10. K-Ortalamalar Algoritması Akış Şeması [26]

K-ortalamalar kümeleme yönteminin avantajları [27]:

- Basit ve esnektir,
- Anlaşılması ve uygulanması kolaydır.

Dezavantajları [27]:

- Kullanıcının önceden küme sayısını belirtmesi gerekir,
- Performansı başlangıç merkezlerine bağlıdır, bu nedenle algoritmanın optimum çözüm için garantisi yoktur.

1.1.3.3.3.Izgara Tabanlı Kümeleme

Izgara tabanlı hesaplama yönteminde birden fazla bilgisayarın iş birliği sağlanır. Bu işlem çoğunlukla, yüklü bilgisayar işlemlerinde ve büyük verilerin işlenmesinde kullanılır. Yüksek çözünürlüklü sonlu sayıda bir ızgara sisteminden yararlanılır. Izgara sisteminin sağladığı en önemli avantaj işlem sürelerinin kısa olmasıdır. İşlem süresi

verideki nesne sayısına bağlı olmaksızın kullanılan grid (ızgara hücresi) sayısına bağlı olmaktadır. STING, WaveCluster ve CLIQUE, ızgara tabanlı yaklaşıma örnek olarak verilebilirler.

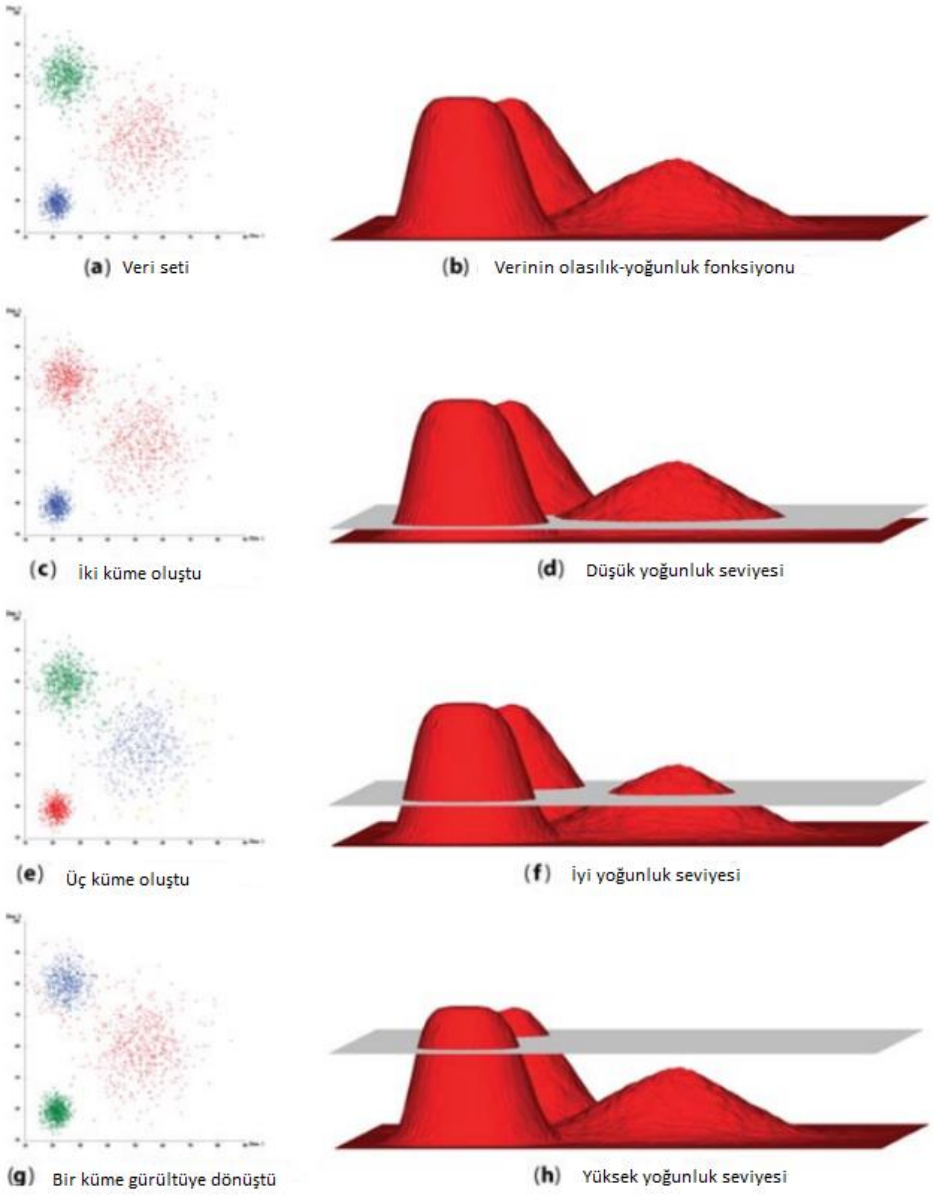
1.1.3.3.4. Model Tabanlı Kümeleme

Model tabanlı kümelemede veri setine en uygun olan model seçilerek kümeleme işlemi yapılır. Veri yapısının bazı matematiksel modellere uyumu için optimizasyon yapılır. Çoğunlukla bütün verilerin dağılımının herhangi bir olasılık dağılımına benzediği kabul edilir. Model tabanlı kümelemede iki farklı yaklaşım vardır: istatistiksel yaklaşım ve yapay sinir ağları.

1.1.3.3.5. Yoğunluk Tabanlı Kümeleme

Yoğunluk tabanlı yöntemler, yoğunluk kavramı esas alınarak geliştirilmiştir. Kümeler, seyrek bölgelerden ayrılmış yoğun alanlar olarak oluşur. Ana fikir, komşuluktaki yoğunluk (nesnelerin veya veri noktalarının sayısı) bir eşiği aştığı sürece, belirli bir kümeyi sürekli olarak büyütmektir. Yoğunluk tabanlı algoritmaların işleyişi düşünüldüğünde, gürültü olan değerleri filtrelemek ve rastgele şekle sahip kümeleri keşfetmek için elverişli oldukları anlaşılır.

Şekil 1.11'de gösterilen veri noktalarının dağılımını göz önünde bulundurulduğunda, belirli bir yoğunluk seviyesindeki veriler “kesim” çizgisinin üzerinde kalır ve yoğunluğa dayalı kümeler olarak düşünülebilir [28]. (c) ve (d) şekillerinde seviyenin düşük seçilmesi ile üç yerine iki küme keşfedildiğini görürüz. (g) ve (h) de ise seviyenin yüksek seçildiğinden dolayı bir kümenin elemanları gürültü olarak algılanır. (e) ve (f) şekillerinde ise en uygun yoğunluk seviyesine karşılık üç ayrı küme keşfedilmiştir.



Şekil 1.11. Farklı Yoğunluk Seviyeleri Eldesi için Kesmeler[28]

Bölümleme yöntemleri sadece dış bükey şekilli kümeleri keşfetmekte başarılıyken yoğunluk tabanlı algoritmalar rasgele şekillerdeki kümeleri de keşfedebilirler. Gerçek hayat koşullarında toplanan veriler genelde rasgele ve çok çeşitli şekillere sahiptirler. Yoğunluk tabanlı yöntemlerin bir diğer olumlu yönü ise küme sayısına girdi olarak ihtiyaç duymazlar. Son olarak, yoğunluk tabanlı yöntemler veri seti içerisindeki anormal değerleri veya sensör arızalanması gibi herhangi bir nedenden dolayı oluşan gürültüleri tespit edebilirler.

Temelde yoğunluk tabanlı algoritmalar şunları içerir: Yoğunluğa dayalı bağlantı analizine göre kümeleri büyüten DBSCAN algoritması, DBSCAN'dan genişletilmiş olan ve parametre ayarlarını içeren OPTICS ve nesneleri bir dizi yoğunluk dağılımı işlevine göre kümeleyen DENCLUE'dir [29]. Bu tez çalışmasında DBSCAN algoritması temel alınarak yeni bir algoritma geliştirilmektedir. Yöntem bölümünde DBSCAN algoritması ile ilgili alt yapı bilgileri ve bu tez çalışmasında geliştirilen AutoVDBSCAN algoritmasının tanıtımı sunulacaktır.

1.2. Literatür Taraması

Yağış verilerinin analizinde veri madenciliği yöntemleri çeşitli çalışmalarda kullanılmıştır. Aşağıda bu çalışmalardan bazıları açıklanmaktadır.

Ruivo ve ark. [30] Brezilya'da gerçekleşen Santa Catarina aşırı yağışları, Amazon kuraklıkları olaylarının sebeplerini araştırmak üzere veri madenciliğinin sınıflandırma yöntemlerini kullanmışlardır. İlk adım olarak, aşırı hava olaylarını tanımlayan ve tahmin edebilen en önemli iklim değişkenleri alt kümesini belirlemişlerdir. İkinci adımda ise bu alt kümeleri "zayıf, orta, güçlü" olarak yağış yoğunluk sınıflarına ayırmak için bir tahmin modeli olarak karar ağaçlarından yararlanmışlardır.

Crane [31] Amerika Birleşik Devletleri'nin kuzeydoğusunda bulunan 104 istasyonun 100 yıllık yağış verileri ile çalışmıştır ve yapay sinir ağlarının alt uygulamalarından biri olan kendi özdüzenleyici haritalar (Self-Organizing Maps (SOMs)) yöntemini kullanmıştır. Böylelikle istasyonların yağış kayıtlarını orantılı bir şekilde bölgesel bir veri kümesinde birleştirmek için kullanmışlardır.

Ahmad ve ark. [32] Malezya Yarımadasında bulunan 59 istasyondan topladıkları 30 yıllık yağış döneminin verisini hiyerarşik kümeleme yöntemleri ile incelemişlerdir. Çalışma, Maleyza Yarımadasında, farklı doğal özelliklere ve yağmur düzenine sahip alanlara karşılık gelen A, B ve C olmak üzere üç homojen yağış bölgesini açıklamaktadır.

Waldow ve ark. [33] İsviçre'deki büyük ölçekli sel olaylarının oluşumunu anlamak için matematiksel bir model olan Ripley'in K fonksiyonunu kullanmışlardır. Bu çalışmada, 420 istasyondan alınan 1 km x 1 km'lik ızgaralı günlük yağış verilerinin zamansal

kümelenmesi incelenmiştir ve ardından kümelemeden sorumlu dinamikler tanımlanmıştır.

Kındap ve ark. [34] 113 istasyonun 1951-1998 yılları arasındaki sıcaklık ve yağış verilerini kullanarak hiyerarşik kümeleme analizi yöntemlerinden biri olan Ward tekniği ile Türkiye iklim bölgelerinin yeniden belirlenmesi çalışmasını yapmışlardır.

İyigün ve ark. [35] Hiyerarşik kümeleme yöntemlerinden biri olan Ward yöntemi ile 244 istasyonlarında 1970-2010 yılları arasında toplanan aylık toplam yağış, nem ve sıcaklık serilerini kullanarak Türkiye iklim bölgelerini yeniden sınıflandırmıştır. Analiz sonuçlarına göre Türkiye üzerinde 14 farklı küme oluşturarak bunların iklim bölgelerini yansıtmakta gerçekçi olduğunu tespit etmiştir.

Sönmez ve ark. [36] 1977-2006 dönemine ait Türkiye'nin 148 adet meteorolojik istasyonundan aldıkları aylık yağış verileri üzerinde K-ortalamalar kümeleme yöntemini kullanarak Türkiye'nin yağış bölgelerini tekrar belirlemek istemişlerdir. K-ortalamalar kümeleme analizi sonucunda bu dönem için 6 farklı yağış bölgesini keşfetmişlerdir. Bu çalışmadaki ilginç keşiflerden biriside İç Anadolu Bölgesi ile Doğu Anadolu Bölgesi, Marmara Bölgesi ve Ege Bölgesi'nin bir kısmının aynı yağış kümesinde yer almasıdır.

Şahin ve ark. [37] bulanık C-ortalamalar yöntemini kullanarak Türkiye alt yağış bölgelerini elde etmişlerdir. Bu çalışmada 174 istasyonun 1974-2002 dönemi yağış ve sıcaklık verilerini analiz ederek Türkiye'deki 7 ana iklim bölgesine dağılmış toplam 15 alt yağış bölgesi oluşturmuşlardır.

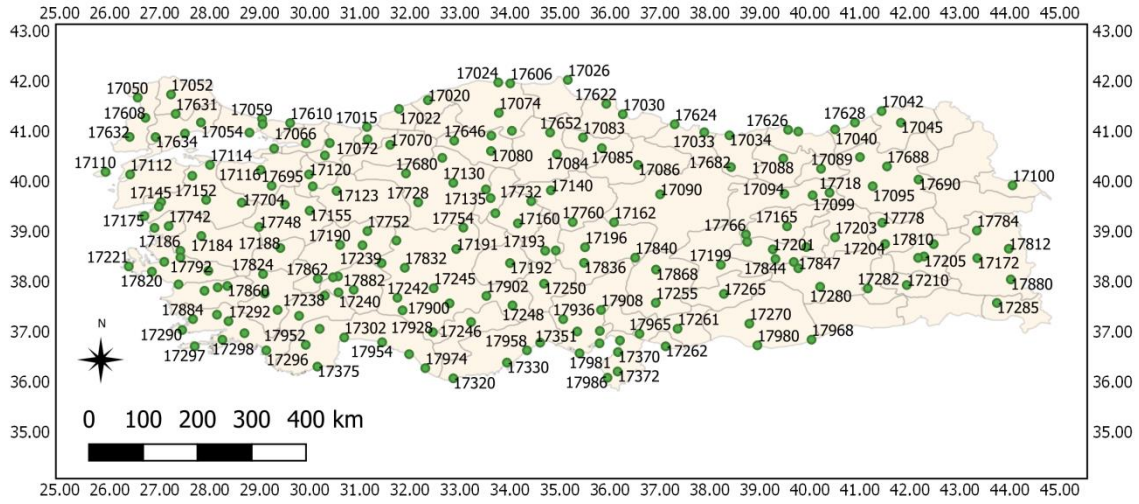
Kumar ve ark. [38] birliktelik kurallarını ortaya çıkaran MOWCATL algoritmasını bazı iklim indekslerini öncül alarak ve Hindistan muson yağışlarını sonuç alarak 1960-2005 dönemi zaman serilerine uygulamışlardır. Kuraklıkların ve sellerin diğer parametrelerle ilişkilerini ortaya çıkarmak için her ikisi için de birliktelik kuralları üretmişlerdir.

2. BÖLÜM

YÖNTEM

2.1. Veri Seti

Bu tez çalışmasında Meteoroloji Genel Müdürlüğü (MGM)'den alınan, Türkiye geneline dağılmış 195 adet gözlem istasyonunun 1980 – 2015 dönemine ait 36 yıllık yağış verisi kullanılmıştır. Yağış verileri her istasyon için aylık toplamlardan oluşmaktadır. Bununla birlikte, her bir istasyonun konum (X,Y) verileri de birer parametre olarak veri setinin özelliklerine eklenmiştir. Şekil 2.1'de istasyonların Türkiye genelinde dağılımı gösterilmiştir.

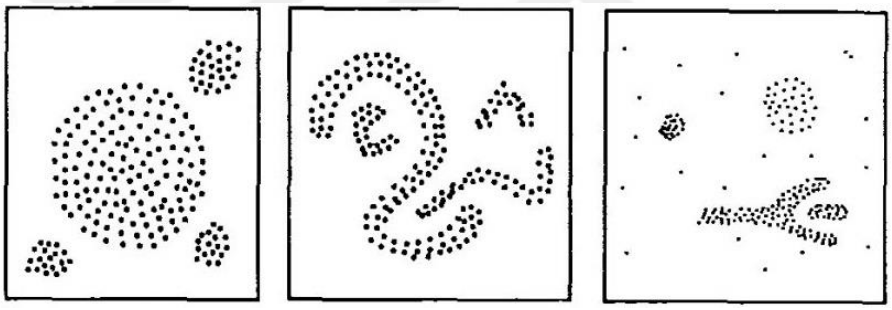


Şekil 2.1. İstasyonların Konumları

2.2. DBSCAN Algoritması

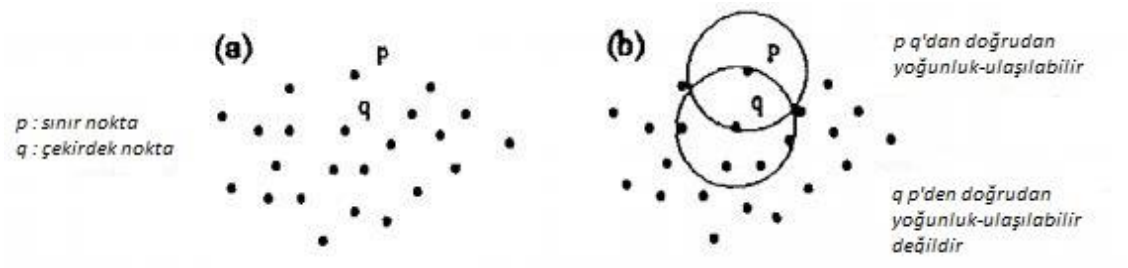
DBSCAN (Density Based Spatial Clustering of Applications with Noise) veri içerisindeki gürültüleri (anormallikleri) keşfetmek ve kümeleme yapmak için geliştirilmiş bir algoritmadır. Yoğunluk tabanlı bu algoritma çok boyutlu uzaysal düzlemde belirli bir yoğunluğa sahip noktaları kümelemekte ve bunların dışındakileri anormallik olarak atamaktadır.

Şekil 2.2'de gösterilen örnek veritabanlarındaki noktaların bazılarının yoğunluk oluşturarak bir kümeye ait olduğunu ve bazılarının da bu yoğunlukların dışında kaldığı görünmektedir. DBSCAN algoritmasının amacı bu yoğunluk durumunu matematiksel olarak ifade etmektir.



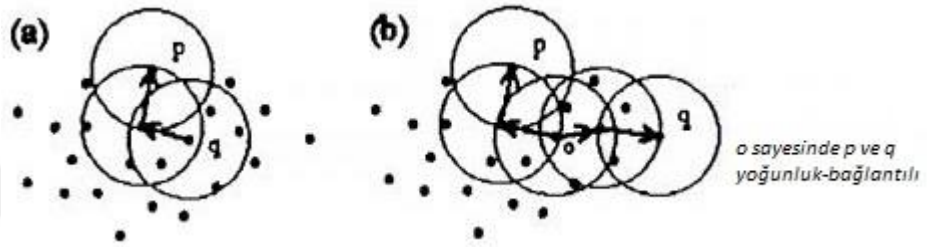
Şekil 2.2. Örnek Veritabanları [39]

DBSCAN algoritması kullanıcının belirleyeceği *MinPts* ve *Eps* değerlerine ihtiyaç duymaktadır. Burada *MinPts* bir küme oluşturabilecek minimum nokta sayısı, *Eps* ise bir noktadan her yöne olan uzaklıktır (radius). Kümeler sınır nokta ve çekirdek nokta olmak üzere iki tip nokta içerirler [39]. Çekirdek nokta, *Eps* uzaklık değeriyle oluşturulan bir çember (*Eps* komşuluğu) içerisinde en az *MinPts* değeri kadar noktaya ulaşabilir, sınır noktası ise *Eps* komşuluğunda *Minpts* eşik değerini aşamaz. Eğer bir nokta çekirdek veya sınır nokta değil ise anormal (noise) olduğu kabul edilir. Şekil 2.3'te gösterilen *p* ve *q* noktaları sırasıyla sınır ve çekirdek noktalardır. Burada *p* noktası *q*'dan doğrudan yoğunluk-ulaşılabilir bir noktadır.



Şekil 2.3. Çekirdek ve Sınır Noktalar [39]

Ayrıca küme içerisinde farklı bağlantılar da oluşur. Şekil 2.4'te *o* noktasından dolayı *p* ve *q* noktaları birbirlerine yoğunluk-bağlantılıdırlar.



Şekil 2.4. Yoğunluk Bağlantılı Noktalar [39]

Veri içerisinde bulunan bir *p* noktasının *Eps* komşuluğunun ifadesi aşağıdaki gibidir [40].

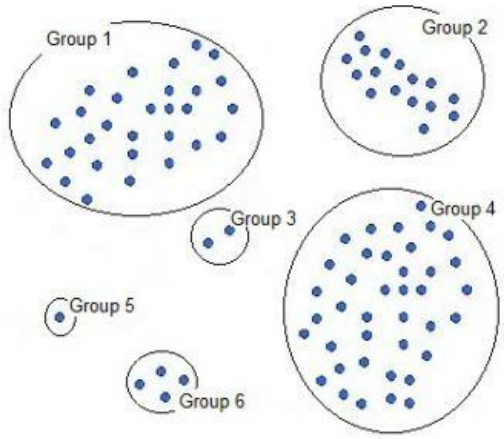
$$N_{Eps} = \{q \in D / dist(p,q) < Eps\}$$

Burada *D* veri setindeki elemanlardır. Bir *p* noktasının *Eps* komşuluğunda en az *MinPts* değeri kadar eleman varsa, *p* noktası bir çekirdek noktadır ve ifadesi aşağıdaki gibidir.

$$N_{Eps} > MinPts$$

DBSCAN algoritmasında *MinPts* ve *Eps* değerlerinin seçimi oldukça önemlidir. Doğru küme ayrımlarının yapılması ve bir elemanın anormal olup olmadığı bu değerlere bağlı olarak değişkenlik gösterebilir. Şekil 2.5'teki örnek veri seti incelendiğinde, *Eps* değerini fazla büyük alınırsa gruplar arası ayrım kaybolabilir ve bazı gruplar tek grupta toplanır. Bu durumun tam tersi olarak *Eps* değeri düşük seçilirse herhangi bir grup

oluşmayabilir. Mesafe ölçüsüne ek olarak, DBSCAN, bir küme olarak etiketlemek için bir grup içinde minimum sayıda nokta gerektirir [40]. Örnek veri seti *MinPts* açısından incelendiğinde, bu değerin 5 olması grup 3, 5 ve 6'daki verileri bir küme oluşturmak için yeterli eleman sayısına ulaşamadığı için anormal olarak atayacaktır. *MinPts* değerinin 4 seçilmesi durumunda ise grup 6 bir küme olmanın şartlarını sağlayabilecektir.



Şekil 2.5. Uzaklığa Dayalı Anormal belirleme Yaklaşımı için Örnek Veri Seti [41]

DBSCAN algoritmasının temsili kodu Algoritma 2.1'de gösterilmiştir. Girdi olarak veri seti, *Eps* ve *MinPts* değerlerini alır ve çıktı olarak kümeleri ve anormal değerleri verir.

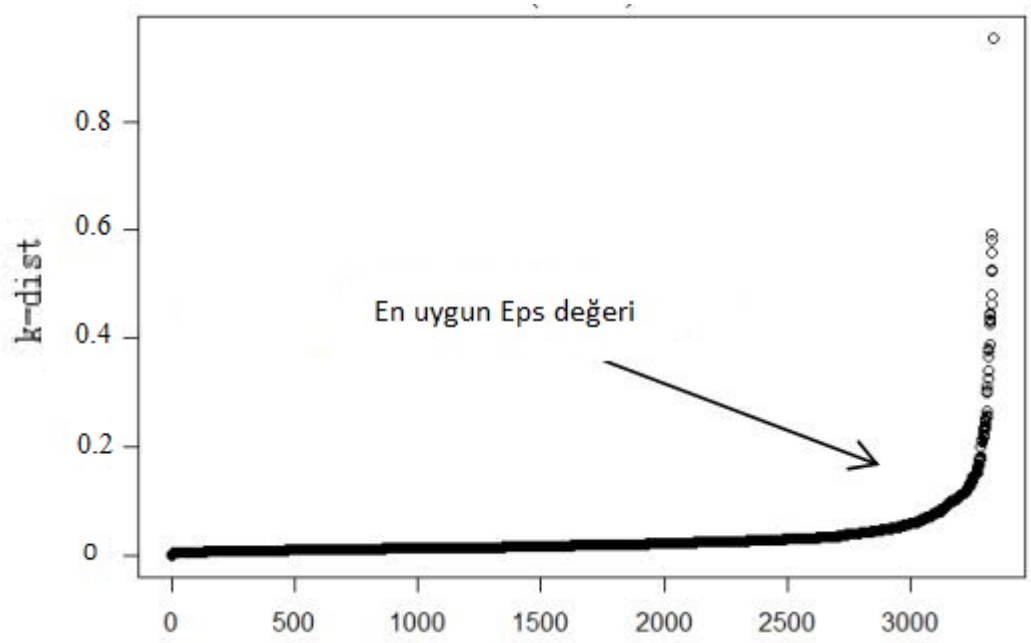
Algoritma 2.1. DBSCAN Temsili Kodu [41]

Girdiler:
D: veri seti
Eps: komşuluk mesafesi
MinPts: minimum nokta sayısı
Çıktılar:
Keşfedilen anormaller ve kümeler
Değişkenler:
m,n: D matrisinin satır ve sütun değerleri
Dist: mesafe vektörü
İndisler: noktaların mesafenin Eps'den daha küçük olduğu değerler
Sınıf_no: kümeleri gösterir
Algorithm:

```
1.   import the data-set into D
2.   for i = 1 to m    //row counter
3.      Dist = distance(i, D)
4.      neighbors= find(Dist =< Eps)
5.      neighbor_count = count(neighbors)
6.      core_neig=check_core_neighbor (neighbors)
7.      if (neighbor_count >=minpts)
8.         class(i) = class_no    //clustered point
9.         while(more points near i)
10.            class(point) = class_no
11.        end while
12.        class_no += 1
13.     else if(neighbor_count<minpts & core_neig==True)
14.        class(i) = 0          //border point
15.     else if (neighbor_count<minpts)
16.        class(i) = -1         //outlier point
17.     end if
18.  end for
19.  return class
```

2.2.1. *Eps* Parametresinin Belirlenmesi

Eps parametresini belirlemek için *k-dist* grafiği yöntemi kullanılır. Bu yöntemde öncelikle uygun bir *MinPts (k)* değeri seçilmelidir ve veri setindeki her bir elemanın *k*(ıncı) en yakın komşuna olan uzaklıkları hesaplanır. Sonuçlar en küçük değerden en büyük değere doğru sıralanır. Şekil 2.6, *k* değerinin 3 alınarak oluşturulan örnek bir *k-dist* grafiğini göstermektedir.



Şekil 2.6. En Yakın 3. Komşuya Olan Uzaklıklardan Oluşan *k-dist* Grafiği Örneği [42]

Oluşturulan *k-dist* grafiği üzerinden ardışık noktalar arası eğimler hesaplanır ve önemli bir sıçramamın gerçekleştiği yer *Eps* değeri olarak kabul edilir.

Eps değerinin otomatik olarak belirlenmesi için bazı yöntemler üretilmiştir. Bunlardan biri de AE-DBSCAN (autoEps) algoritmasıdır ve Algoritma 2.2'de *Eps* değerinin belirlendiği kısmın temsili kodu gösterilmiştir.

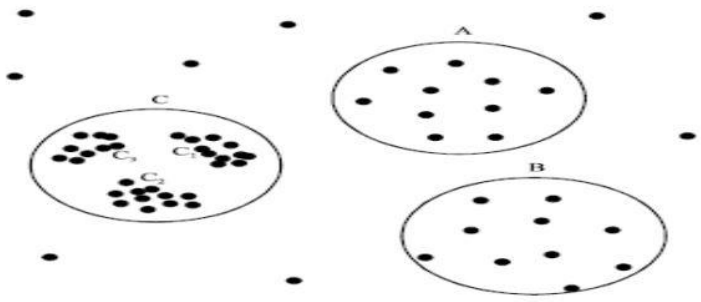
Algoritma 2.2. Otomatik *Eps* Belirleme Temsili Kod [43]

Girdi: Veri seti, *MinPts (k)*
Çıktı: *Eps* Serisi
Eps değerlerinin belirlenmesi
1. *k-dist* değerlerini veri setindeki her bir nokta için hesapla
2. Elde edilen değerleri sırala ve *k-dist* grafiğini çiz
3. *k-dist* grafiğinden noktalar arası her geçişin eğimini hesapla
4. Sıfır olmayan eğimlerin ortalamasını ve standart sapmasını hesapla
5. Ortalama ve standart sapma toplamının üzerinde olan eğimleri bul
6. Bu eğimlere karşılık gelen uzaklıkları *Eps* değeri olarak belirle

2.3. VDBSCAN Algoritması

VDBSCAN (Varied Density Based Spatial Clustering of Applications with Noise) algoritması DBSCAN algoritmasının geliştirilmiş bir versiyonudur ve temel farkı birden fazla *Eps* değeri kullanarak analiz yapmasıdır [44].

DBSCAN algoritmasının girdi olarak ihtiyaç duyduğu parametrelerden biri olan *Eps* değeri, doğru kümelerin elde edilmesinde oldukça kritik bir role sahiptir ve yalnızca bir *Eps* değerinin kullanılması farklı yoğunluğa sahip kümelerin keşfedilmesini engellemektedir. Örneğin, Şekil 2.7'de gösterilen A, B, C kümeleri ve C_1, C_2 ve C_3 kümeleri farklı yoğunluktalardır. Böyle bir veri seti analizinde tek *Eps* değeri ile A, B, C kümeleri oluşturulabilir veya C_1, C_2, C_3 kümeleri tanımlanırken A ve B kümelerindeki elemanlar anormal olarak atanmak zorunda kalır. Aslında olması gereken kümeleme A, B, C_1, C_2, C_3 şeklindedir ve bu kümeleme için birden fazla *Eps* değerine ihtiyaç vardır.



Şekil 2.7. Farklı Yoğunluklara Sahip Kümeler [45]

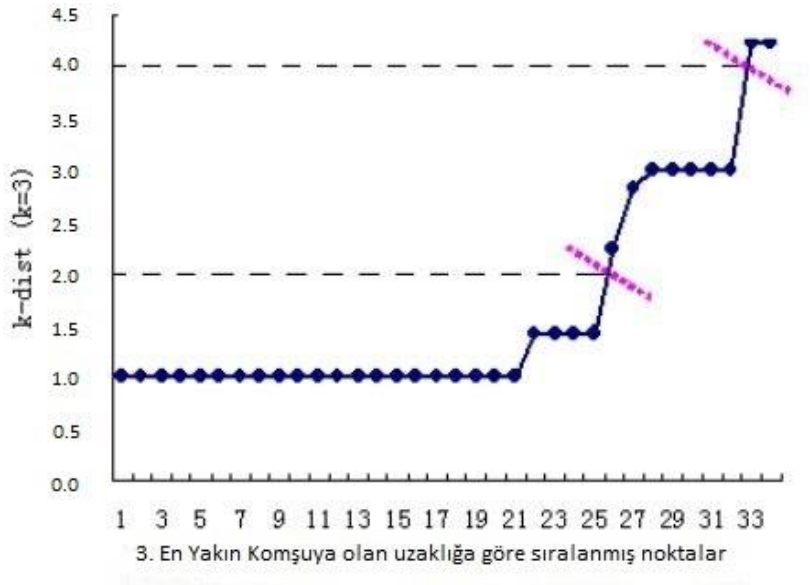
VDBSCAN algoritması iki adımdan oluşmaktadır: *Eps* değerlerinin belirlenmesi ve çeşitli yoğunluklardaki kümelerin keşfedilmesi [44]. Temsili kod Algoritma 2.3'te gösterilmektedir.

Algoritma 2. 3. VDBSCAN Temsili Kodu [44]

Adım 1:
1. *k-dist* grafiğini çiz
2. Sıçramaları $Eps_{(i)}$ olarak belirle (i=1, 2,..., n)

Adım 2:
1. Her bir $Eps_{(i)}$ değeri için (i=1, 2,..., n)
2. Eps= $Eps_{(i)}$;
3. henüz bir kümeye dahil edilmemiş noktalar için DBSCAN algoritmasını uygula
4. Noktaları C_{i-t} olarak işaretle;
5. Tüm işaretlenmiş noktaları karşılık kümeye dahil edin

Birden fazla *Eps* değerinin bulunması için yine *k-dist* yöntemi kullanılır. Şekil 2.8'de örnek bir *k-dist* grafiği gösterilmiştir. Bu örnek, bir veri setindeki elemanların 3(üncü) en yakın komşuya olan mesafelerinin sıralanması ile elde edilmiştir.



Şekil 2.8. *Eps* Değerlerinin Elde Edilmesi İçin Örnek *K-Dist* Grafiği (k = 3) [44]

Şekil 2.8'deki eğimler incelendiğinde iki büyük sıçramanın gerçekleştiği görülür. O halde iki adet *Eps* değeri kullanılarak farklı yoğunluklara sahip iki küme keşfedilebilir. Burada Eps_1 2.0 ve Eps_2 4.0 olarak belirlenmiştir. *Eps* sayısı iki adet olabileceği gibi veri setinin içeriğine göre çok daha fazla sayılarda da olabilir. VDBSCAN algoritması en küçük *Eps* değeri ile ilk kümeleri belirler ve ardından ikinci en küçük *Eps* değerini kullanarak kalan elemanları kümeler. Bu işlem en büyük *Eps* değeri kullanılıp en az yoğunluğa sahip kümeler belirlenerek son bulur ve hala bir kümeye eklenmemiş elemanlar anormal değer alarak atanır.

2.4. AutoVDBSCAN Algoritması

Geliştirilen autoVDBSCAN (Automatic and Level-Wise Varied-Density Based Anomaly Detection Algorithm) algoritması kümenin minimum eleman sayısını gösteren *MinPts (k)* değerini ve *Eps* serilerini otomatik olarak belirlemektedir. Algoritma bu değerlerin belirlenmesi ve kümelemenin yapılmasında iterasyonlu bir yaklaşımı içermektedir.

2.4.1. *MinPts (k)* Değerinin Otomatik Olarak Belirlenmesi

AutoVDBSCAN algoritmasının ilk adımı veri setinin içeriğine dayalı olarak *MinPts* değerinin belirlenmesidir. Otomatik olarak *MinPts* değerin belirlenmesi için veriseti içerisindeki elemanların ortalama uzaklık değerleri kullanılır [46]. Yöntemin adımları aşağıdaki gibidir:

i) n elemana sahip bir veriseti içerisindeki her bir p*(i)* noktasının diğer tüm elemanlara olan uzaklıklarının ortalaması hesaplanır:

$$\text{eleman\_d}_{ort}(\text{pi}) = \frac{\sum_{j=0}^{n-1} \text{mesafe(pi,xj)}}{n-1}$$

ii) Her *p(i)* için hesaplanmış olan ortalama uzaklık değerleri elde edildikten sonra tekrar bu değerlerin ortalaması alınarak bu sefer de veri setine ait ortalama uzaklık elde edilir:

$$\text{veriseti\_d}_{ort} = \frac{\sum_{i=0}^{n} \text{eleman\_d}_{ort}(\text{pi})}{n-1}$$

iii) Bu adımda ise her bir *p(i)* elemanı için *veriseti_*d_{ort} mesafesine kaçıncı en yakın komşusunun en yakın olduğu belirlenir:

K(i) = min │mesafe(veriseti_d_{ort} - xj) │

iv) Son adım olarak, elde edilen *K(i)* değerleri arasında en çok tekrar eden değer(Mod) *Minpts (k)* olarak atanır.

2.4.2. *Eps* Değerlerinin Otomatik Olarak Belirlenmesi

Önceki aşamada belirlenen *MinPts (k)* değeri *k-dist (bknz)* grafiğinin çizimi için kullanılır. Grafikteki sıçramalar birer potansiyel *Eps* değeridir. Bu aşamada sıçramaların büyüklükleri hesaplanır ve ortalaması alınır. Sonuç olarak, bu ortalamadan yüksek olan değerler autoVDBSCAN algoritması tarafından *Eps* serisi olarak belirlenir.

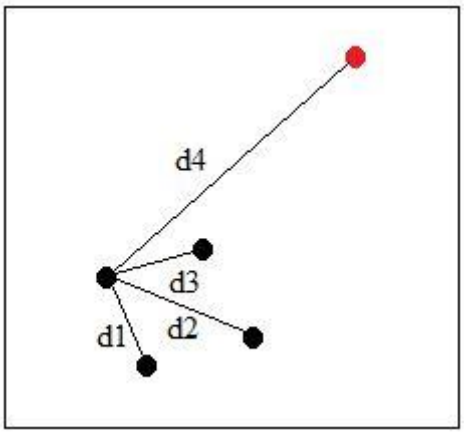
Kullanılan *Eps* değerlerini otomatik olarak belirleme yöntemi göreceli olarak fazla sayıda *Eps* değeri üretir. Bu değerlerden bazıları kümeleme işleminde etkili olmayacaktır. Örneğin, *Eps* serisi arasında *Eps(i)* ve *Eps(i+1)* değerleri birbirlerine oldukça yakın ise VDBSCAN algoritmasının çalışma prensibine göre önce küçük olan *Eps(i)* değeri ile kümeleme yapılacaktır ve ardından *Eps(i+1)* değeri ile henüz kümelenmemiş olan elemanlar analiz edilecektir. Bu noktada *Eps(i+1)* değeri, eğer veri seti birbirine yakın yoğunluklara sahip kümeler içeriyorsa yeni bir küme oluşturacaktır ve tersi durumda işlevsiz kalacaktır. Bu durumun analize tek etkisi kullanıcıyı doğru kümeleme açısından güvenli tarafta tutmasıdır.

2.4.3. Kümeleme ve Anormal Değerlerin Belirlenmesi

Eps ve *MinPts (k)* girdi parametreleri elde edildikten sonra klasik VDBSCAN algoritması uygulanır. Bu aşamada, üretilen *Eps* serisinin tamamı kullanılarak kümeler elde edilir ve hiç bir kümeye ait olmayan elemanlar anormal olarak atanır.

2.4.4. İterasyon ile Doğruluğun Artırılması

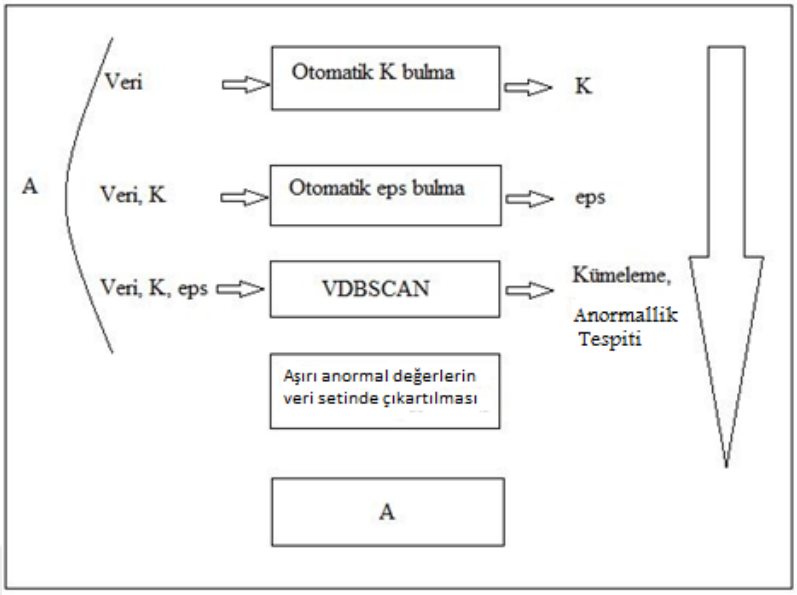
Otomatik olarak *MinPts* değerinin belirlenmesinde matematiksel açıdan veri setinin dağılımın kuvvetli bir şekilde etkili olduğu anlaşılmaktadır. Bununla anlatılmak istenen, küçük veya büyük anormal değerlerin optimum *MinPts* değerini bulmayı açıkça zorlaştırıyor olmasıdır. Örneğin, Şekil 2.9'da gösterilen örnek veri seti incelendiğinde, kırmızı noktanın mesafesi(d4) *Minpts* girdisinin hesaplanmasında kullanılan *eleman_*d_{ort}*(pi)* değerinin olması gerekenden daha yüksek çıkmasına sebep olacaktır.



Şekil 2.9. Aşırı Anormal Değeri İçeren Örnek Veriseti

Bu durumdan *Eps* serisi de etkilenecektir ve aynı şekilde sadece büyük olan *Eps* değerleri belirlenebilecektir. Sonuç olarak bu aşamada sadece aşırı anormal değerler keşfedilecektir. Ayrıca, doğru sayıda küme oluşturulamayacak ve gizlenmiş anormal değerler belirlenemeyecektir.

AutoVDBSCAN algoritması analizin bu kısmında, optimum *MinPts* değerinin belirlenmesi için keşfedilen aşırı anormal değerleri veri setinden çıkarır ve yeni bir *MinPts* değeri ve *Eps* serisi elde eder. Ardından tekrar VDBSCAN algoritmasının uygulanması ile daha doğru sayıda kümeler ve gizlenmiş anormal değerleri keşfeder. Şekil 2.10'da autoVDBSCAN algoritmasının adımları şematize edilmiştir.



Şekil 2.10. AutoVDBSCAN Adımları

Sonuç olarak, autoVDBSCAN algoritması girdi olarak sadece veri setine ihtiyaç duyar ve çıktı olarak kümeleri ve bununla birlikte aşırı anormal ve anormal değerleri keşfeder. Kullanıcı girişli *MinPts* ve *Eps* değerlerinin otomatikleştirilmesi farklı kullanıcıların aynı sonuçları elde edebilmesi noktasında önemli bir yaklaşımdır.

2.4.5. Analizlerde Belirlenen Anormallik Türleri

Bu çalışmada yapılan analizler sonucunda dört farklı anormallik kavramı sunulmuştur. Bunlardan ilk ikisi "aşırı anormal" ve "gizlenmiş anormal" kavramlarıdır ve bulgular bölümünde 3.2'de aktarılan yıllık ortalama yağış analizleri sonuçlarının aktarımında kullanılmıştır. AutoVDBSCAN algoritmasının ilk adımında keşfedilen anormal değerler aşırı anormal değerler olarak adlandırılır. Gizlenmiş anormal değerler ise bu aşırı anormal değerler sebebiyle fark edilemeyen, gizlenen veya baskılanan anormal değerlerdir. Diğer iki tür anormallik tanımı ise "yüksek anormal" ve "düşük anormal" terimleri ile ifade edilmiştir. Bu terimler Bulgular 3.3'te verilen aylık ortalama yağış serilerinde belirlenen anormal değerler için kullanılmıştır. Böylelikle yağış analizlerinde keşfedilen farklılığın (anormalliğin) aşırı miktardaki bir yağış mı yoksa düşük miktardaki bir yağış mı olduğun göstermektedir.

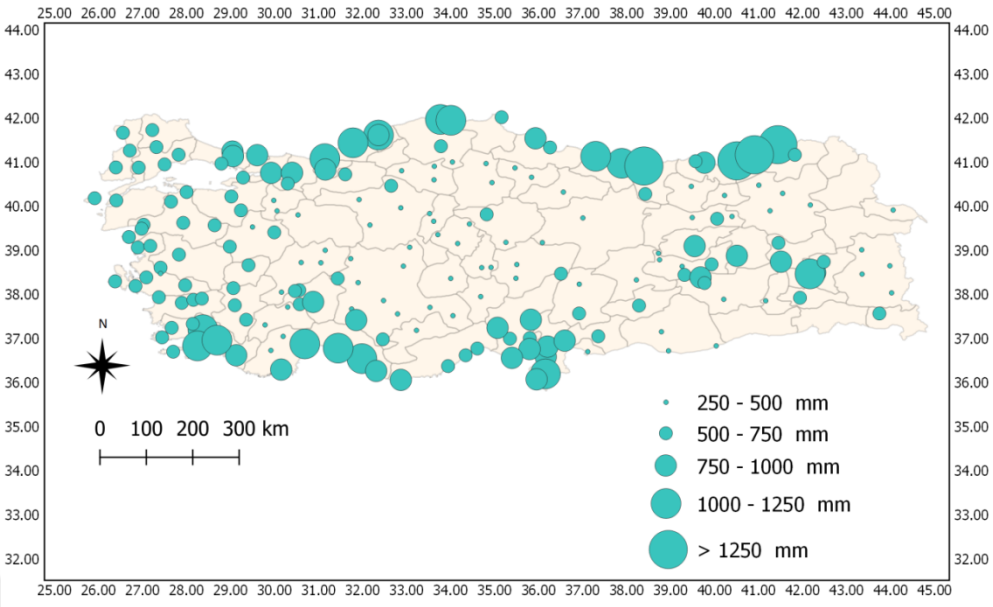
3. BÖLÜM

BULGULAR

Bu tez çalışması kapsamında AutoVDBSCAN algoritması hazırlanmış ve bu algoritma kullanılarak Türkiye yağış serileri analiz edilmiştir. Aşağıda sunulan bulgular üç kısımdan oluşmaktadır. Birinci kısımda kullanılan yağış veri setinin istatistiksel özellikleri açıklanmaktadır. Ayrıca yıllık ve aylık yağışların mekânsal dağılımları incelenmektedir. İkinci kısımda yıllık yağış verileri AutoVDBSCAN algoritması kullanılarak analiz edilmekte ve elde edilen sonuçlar sunulmaktadır. Üçüncü kısımda ise aylık veri serileri kullanılarak anormal değerler tespit edilmiştir.

3.1. Veri Seti Özellikleri

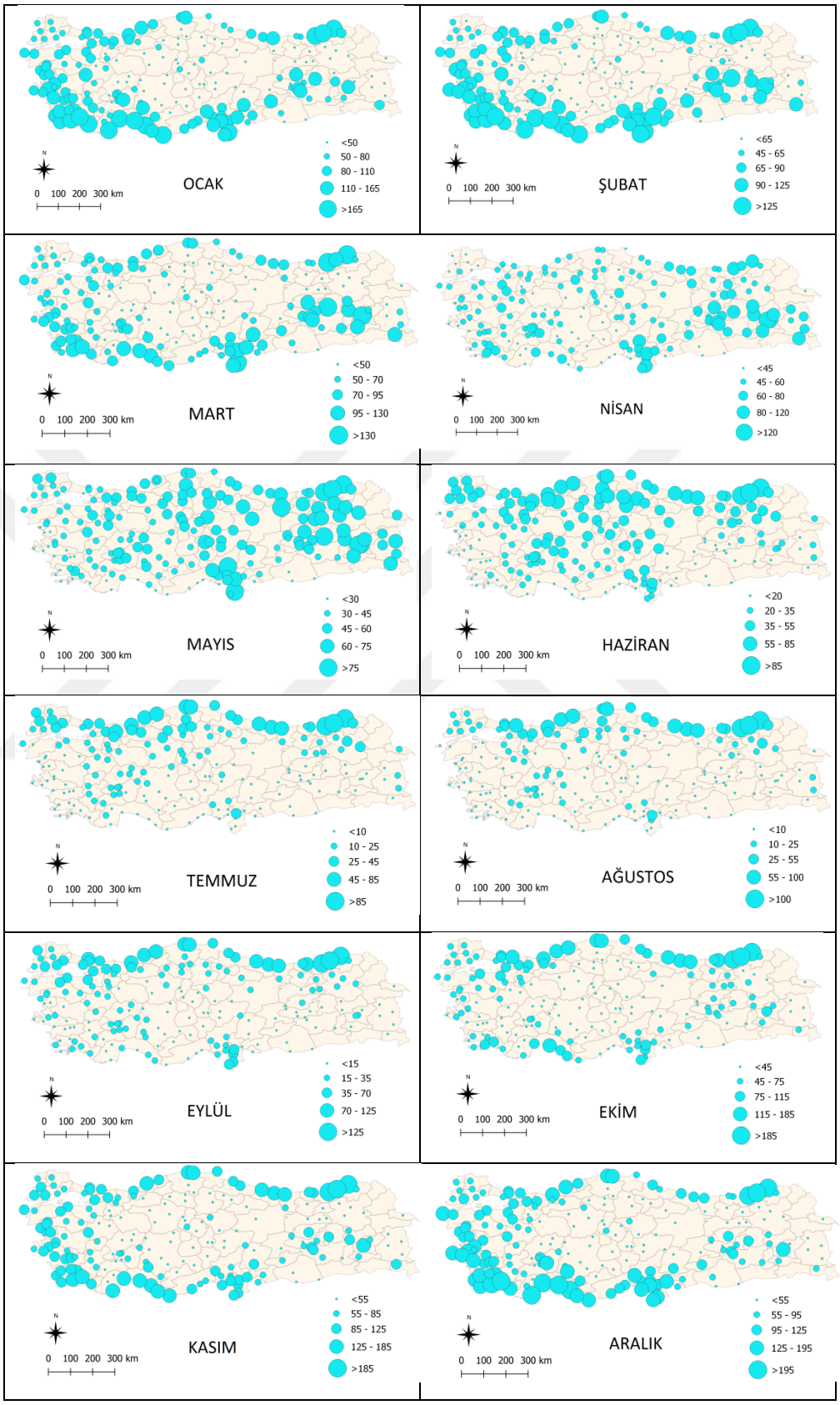
Bu tez çalışmasında MGM tarafından işletilen 195 adet meteoroloji istasyonunun 1980-2015 dönemine ait aylık yağış verileri kullanılmıştır. Veri setinin homojenliği %1 (0.01) anlamlılık düzeyinde Pettitt testi, Standart normal homojenlik testi (SNHT), Buishand aralığı testi ve Von Neumann oran testi ile daha önce yapılan bir çalışmada araştırılmıştır [47]. 195 homojen meteoroloji istasyonunda 1980-2015 dönemi ortalama yıllık yağış değerleri Şekil 3.1'de gösterilmiştir.



Şekil 3.1. Türkiye 1980-2015 dönemi yıllık ortalama yağış haritası

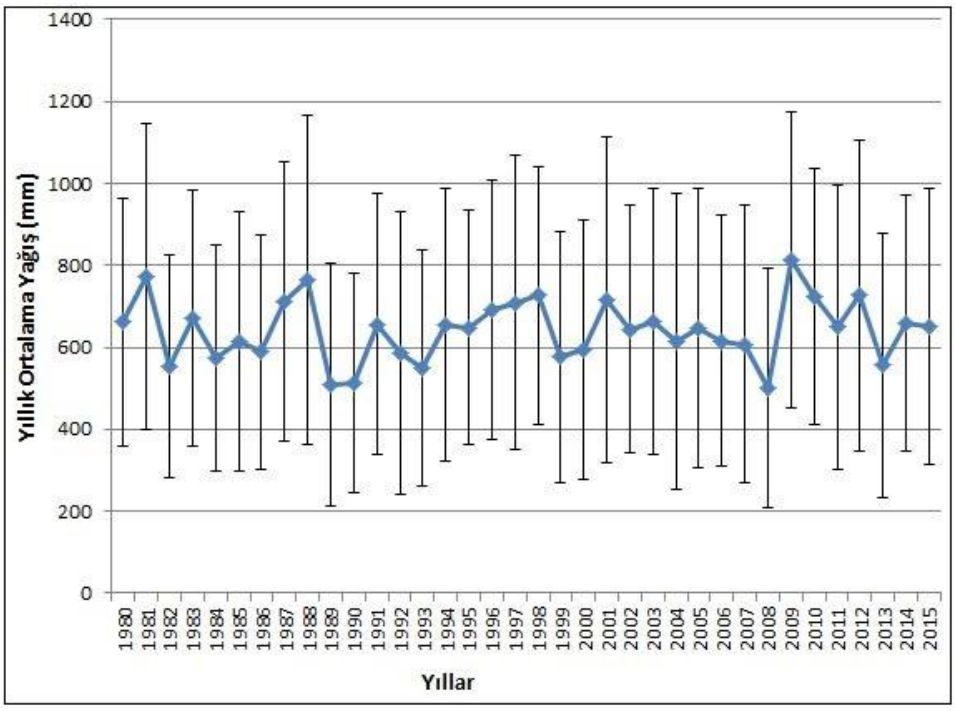
İstasyonların verilerine göre, 1980-2015 döneminde ortalama yıllık yağışlar 263-2264 mm aralığında değişmektedir. 195 istasyonun ortalama yıllık yağışlarının ortalaması 642 mm ve standart sapması 300 mm olarak tespit edilmiştir. Ortalama yıllık yağış miktarları incelendiğinde, 4 istasyonun 1250 mm'nin üzerinde, 43 istasyonun 750-1250 mm aralığında, 148 istasyonun 500-750 mm aralığında ve 73 istasyonunda 500 mm'nin altında yağışa sahip oldukları görülmüştür.

Bununla birlikte 1980-2015 dönemindeki en yüksek aylık ortalama yağış 90 mm ile aralık ayında görülürken, en düşük aylık ortalama yağış 17 mm ile temmuz ayında görülmüştür. Şekil 3.2'de aylara göre aylık ortalama yağış dağılımı gösterilmiştir.



Şekil 3.2. Türkiye 1980-2015 Dönemi Aylık Ortalama Yağışlar (mm)

Şekil 3.3'te yıllara göre 195 istasyonun ortalama yıllık yağışları ve her bir yıl için standart sapma değerleri gösterilmiştir. 195 istasyonun yıllık yağış verilerinin ortalaması dikkate alındığında, en yüksek yağış 2009 yılında 813 mm ve en düşük yağış 2008 yılında 500 mm olmuştur. Yıllık bazdaki standart sapma değerlerinin ise 267-401 mm aralığında olduğu görülmüştür.



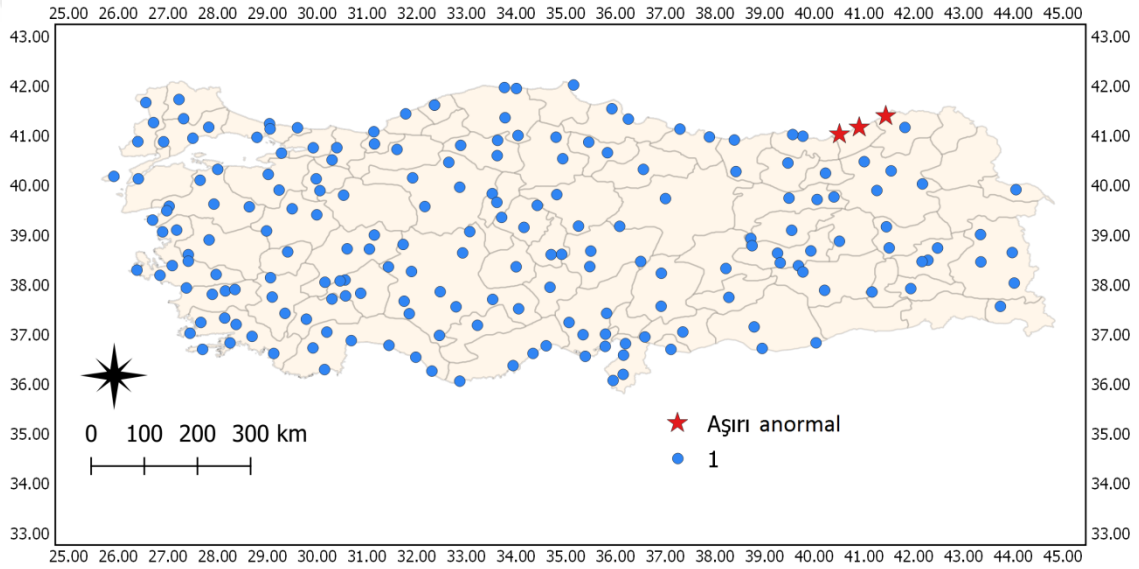
Şekil 3.3. Yıllık Ortalama Yağış ve Standart Sapma Değerleri

3.2. Yıllık Yağış Verilerinin AutoVDBSCAN Algoritması ile Analizi

Türkiye'de 195 meteoroloji istasyonundan toplanmış 1980-2015 dönemine ait yıllık yağış verilerin AutoVDBSCAN algoritması ile analizi yapılmıştır. Veri seti içerisinde bulunan her bir istasyonun 36 yıllık süreçteki ortalama yağış değerleri ve konum özellikleri (X,Y) kullanılmıştır. Böylelikle her istasyondan alınan ortalama yağış değeri ve konum bilgileri ile toplamda 195 x 3'lük veri seti oluşturulmuştur.

Algoritmanın ilk adımında *MinPts* değeri 10 olarak belirlenmiştir. Bu işlemin ardından veri setindeki her bir eleman için 10'uncu en yakın komşuya olan uzaklıklar belirlenerek *k-dist* grafiği elde edilmiştir. Bu grafik kullanılarak, *Eps* dizisi

oluşturulmuştur ve klasik VDBSCAN algoritması uygulanmıştır. Sonuç olarak algoritmanın ilk aşamasında Şekil 3.4'te kırmızı yıldızlarla işaretli üç aşırı anormal değer keşfedilmiştir ve diğer tüm istasyonlar mavi daire ile temsil edilen tek bir kümede toplanmıştır. Aşırı anormal olarak belirlenen değerler Artvin Hopa (istasyon no.17042), Rize merkez (istasyon no.17040) ve Pazar (istasyon no.17628) ilçesindeki istasyonlara aittir ve bu istasyonlarda ortalama yıllık yağış değerleri sırasıyla 2294 mm, 2273 mm ve 2062 mm'dir. Türkiye'nin geneliyle kıyaslandığında oldukça yüksek miktarda yağış alan bu bölgeler anormal olarak tespit edilmesi beklenen bir durumdur.



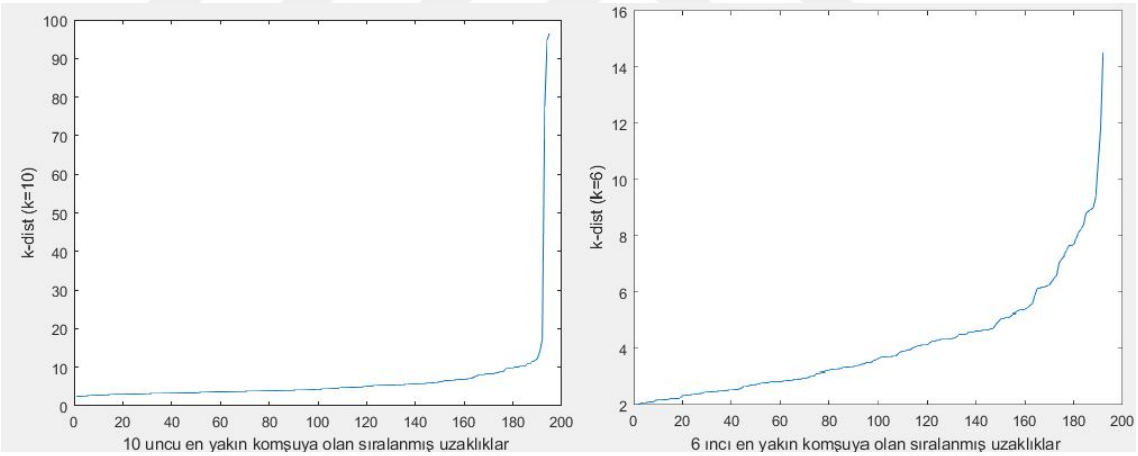
Şekil 3.4. İlk Basamakta Belirlenen Aşırı Anormal İstasyonlar

Bu aşamada belirlenen aşırı anormal değerler veri seti içerisinden çıkarılmıştır ve tekrar *MinPts* değeri ve *Eps* serisi elde edilmiştir. Tablo 3.1'de iterasyon öncesi ve sonrası belirlenen parametreler gösterilmiştir.

Tablo 3.1. İterasyon Öncesi ve Sonrası *MinPts*, *Eps* ve Küme Sayıları

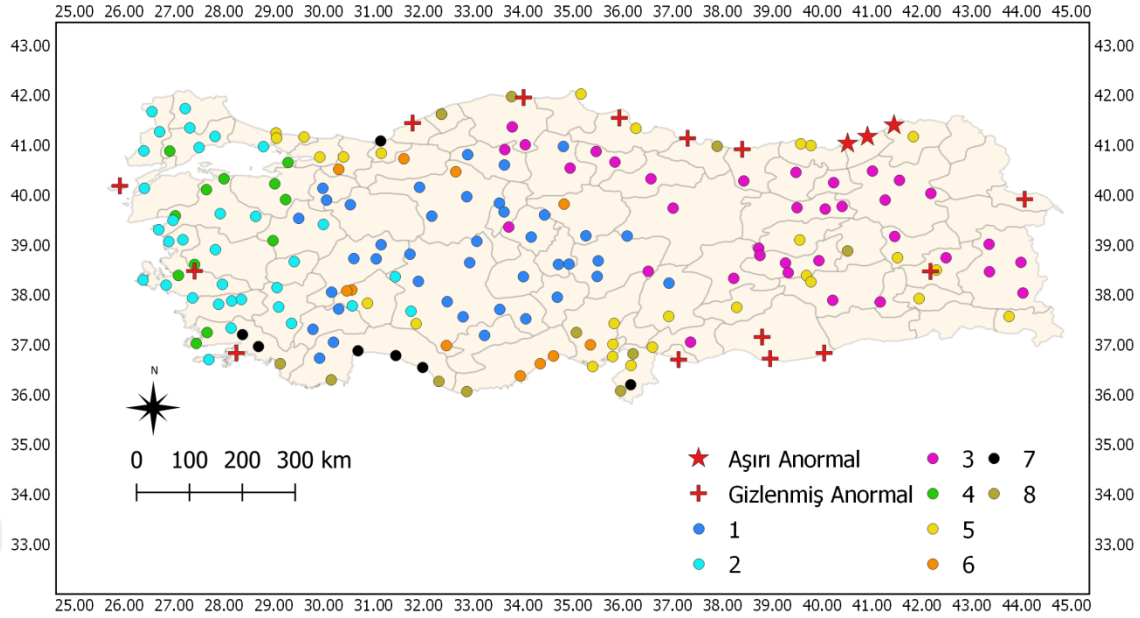
Parametre	**İterasyondan önce**	**İterasyondan Sonra**
MinPts (k)	10	6
***Eps* Serisi**	(8-96) aralığında 10 adet	(2-15) aralığında 30 adet
Küme Sayısı	1 adet	8 adet

Bununla birlikte farklı *MinPts (k)* değerleri için oluşturulan *k-dist* grafikleri Şekil 3.5'te gösterilmiştir. Sonuç olarak aşırı anormal değerlerin etkisi ile k değeri yüksek çıkmakta ve buna bağlı olarak Tablo 3.1'de gösterildiği gibi *Eps* değerleri de yüksek çıkmaktadır. İterasyon öncesi elde edilen *Eps* serisi 8-96 aralığında değişmektedir. Bu değerlerin büyüklüğü sebebi ile yeterince benzer olmayan noktalar bile aynı kümeye dâhil edilmekte ve yalnızca 1 adet küme oluşmaktadır.



Şekil 3.5. k=10 ve k=6 Değerleri için Çizilen *k-dist* Grafikleri

AutoVDBSCAN algoritması iterasyon sonrasında *MinPts (k)* değeri 6 olarak tespit edilmiştir ve 2-15 arasında değerler alan *Eps* serisi ile farklı yoğunluklara sahip 8 adet küme oluşturulmuştur. Şekil 3.6'da iterasyon sonrası oluşturulan kümeleri ve keşfedilen gizlenmiş anormal değerleri göstermektedir.



Şekil 3.6. İterasyon Sonrası 1980-2015 Dönemi Ortalama Yağış Verilerinde Elde Edilen Kümeler

Analiz sonuçları hakkında özet bilgiler Tablo 3.2'de aktarılmıştır.

Tablo 3.2. Yıllık Yağış Verilerinin Kümeleme Sonuç Özeti

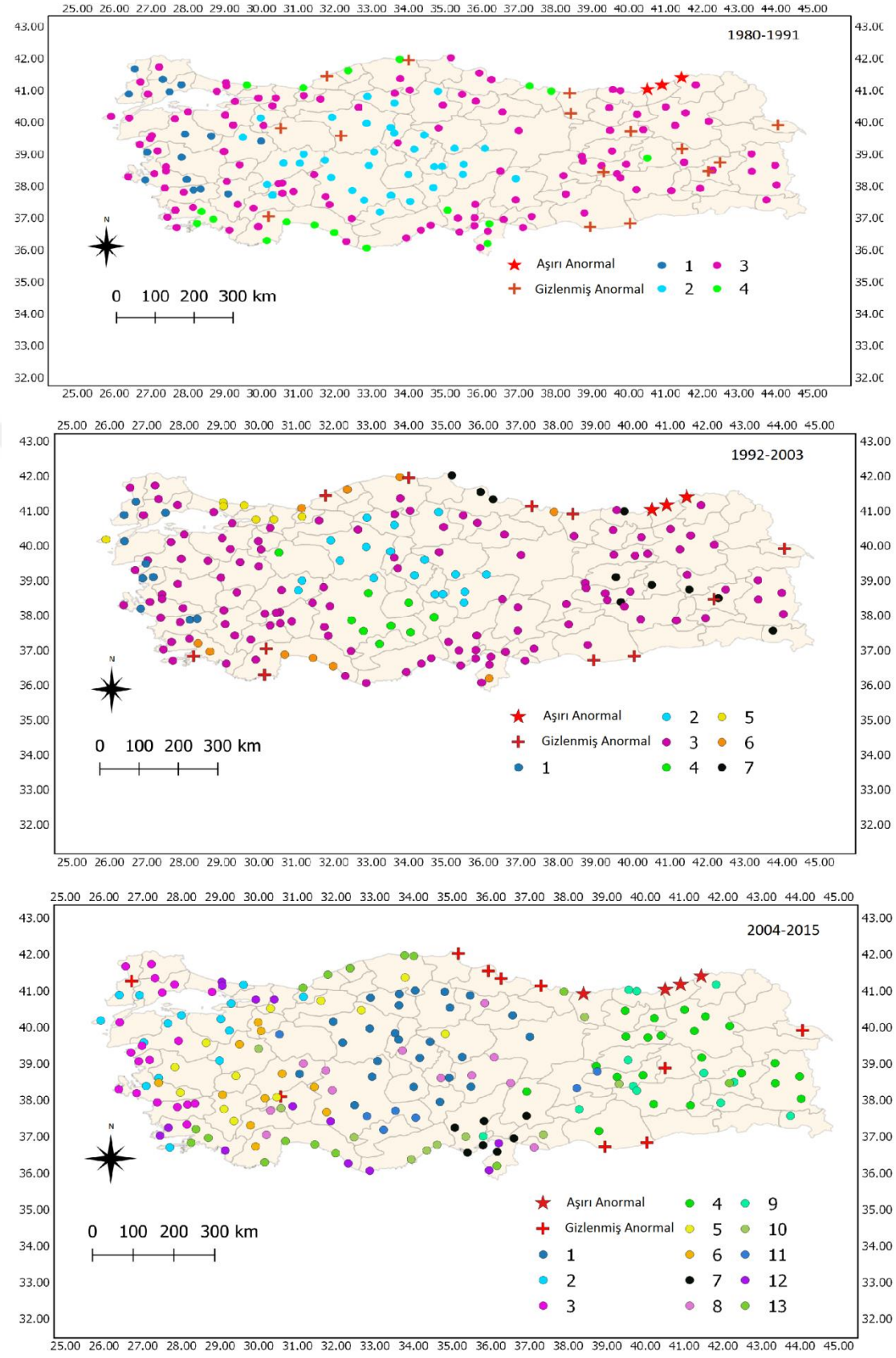
Küme No.	Eleman Sayısı	Küme için Ortalama Yıllık Yağış
1	40 adet	396 mm
2	33 adet	588 mm
3	35 adet	456 mm
4	12 adet	708 mm
5	28 adet	780 mm
6	11 adet	600 mm
7	7 adet	1116 mm
8	12 adet	948 mm
Aşırı Anormal	3 adet	
Gizlenmiş Anormal	14 adet	

Burada aşırı anormal değerlerin baskıladığı gizlenmiş anormal olarak 14 adet istasyon işaretlenmiştir. Gizlenmiş anormal istasyonlar çok küçük miktarda ve çok yüksek miktarda yağış alan istasyonlar olduğu gibi, ayrıca bulunduğu bölgeden farklı miktarlarda yağış alan istasyonlarda olmuştur. Şanlıurfa Akçakale (istasyon no.17980)

istasyonu ortalama yağış miktarı 263 mm'dir ve genel olarak benzer yağışlar küme oluşturamayacak kadar az oldukları için gizlenmiş anormal olarak belirlenmiştir. Bir diğer yandan, Giresun Merkez (istasyon no.17034) istasyonu genele göre yüksek bir değer olan 1296 mm yağış ile gizlenmiş anormaldir.

Algoritmanın oluşturduğu 1 no'lu küme 396 mm ortalama yağışa sahip 40 (%20) istasyonu içermektedir ve İç Anadolu Bölgesi'nin orta ve batı kısımlarını ve Akdeniz Bölgesi'nin batı kısmını kapsamıştır. 2 ve 4 no'lu kümeler sırasıyla 588 mm ve 708 mm ortalama yağışa sahip 33 (%17) istasyon ve 12 (%6) istasyondan oluşmaktadırlar ve Ege Bölgesi ile Marmara Bölgesi'nin her ikisine de dağılmış durumdadırlar. 3 no'lu kümede bulunan 35 (%18) istasyon ise 456 mm ortalama yağış ile Karadeniz Bölgesi'nin güney kısmına ve Doğu Anadolu Bölgesi'ne yayılmıştır. 5, 6, 7 ve 8 nolu kümeler ise belirli bir bölgede yoğunlaşmayacak şekilde Akdeniz Bölgesine, Karadeniz ile Marmara Bölgeleri'nin kesiştiği ve Doğu Anadolu ile Güney Doğu Anadolu Bölgeleri'nin kesiştiği alanlara dağılmıştır. Bu kümeler arasında en yüksek yıllık ortalama yağışa sahip olan küme ise 1116 mm ile 7 no'lu küme olmuştur ve ayrıca 7 (%3.5) istasyon ile en istasyon içeren küme olmuştur. Bunula birlikte, Fırat ve ark. [48] hiyerarşik olmayan kümeleme yöntemlerinden biri olan k-ortalamalar ile benzer şekilde 1 ve 3 no'lu kümeleri keşfetmişlerdir. Sariş ve ark. [49] da bütünleştirici hiyerarşik kümeleme yöntemi kullanarak onlar da 1 ve 3 nolu kümeleri oluştumuşlardır ve ayrıca 2 ve 4 nolu kümelerde olduğu gibi Ege Bölgesi'nde iki farklı kümenin hâkim olduğunu belirlemişlerdir.

Türkiye yağış serilerinin 1980-2015 dönemine ait 36 yıllık yağış verisi üç alt dönem için analiz edilmiştir. Böylelikle 1980-1991, 1992-2003 ve 2004-2015 olmak üzere 12'şer yıllık yağış serilerinin kümeleme ve anormallik keşfi sonuçları Şekil 3.7'de gösterilmiştir.



Şekil 3.7. 1980-1991, 1992-2003 ve 2004-2015 Yağış Dönemleri için Kümeleme Sonuçları

Yapılan kümeleme işleminin özet sonuçları Tablo 3.3'te gösterilmiştir.

Tablo 3.3. 12'şer Yıllık Alt Dönemlere Ait Özet Sonuçlar

Dönem	Küme No.	Eleman Sayısı	Küme için Ortalama Yıllık Yağış
1980-1991	1	15 adet	540 mm
	2	24 adet	384 mm
	3	109 adet	612 mm
	4	19 adet	1020 mm
	Aşırı Anormal	3 adet	
	Gizlenmiş Anormal	15 adet	
1992-2003	1	10 adet	612 mm
	2	18 adet	384 mm
	3	116 adet	588 mm
	4	9 adet	324 mm
	5	7 adet	840 mm
	6	11 adet	1104 mm
	7	10 adet	804 mm
	Aşırı Anormal	3 adet	
	Gizlenmiş Anormal	11 adet	
2004-2015	1	25 adet	396 mm
	2	15 adet	744 mm
	3	19 adet	636 mm
	4	22 adet	432 mm
	5	12 adet	576 mm
	6	11 adet	492 mm
	7	7 adet	756 mm
	8	12 adet	444 mm
	9	12 adet	720 mm
	10	10 adet	552 mm
	11	7 adet	312 mm
	12	13 adet	852 mm
	13	16 adet	1104 mm
	Aşırı Anormal	4 adet	
	Gizlenmiş Anormal	10 adet	

Üç dönemin yıllık ortalama yağış verilerinin bazı istatistiksel bilgileri Tablo 3.4'te verilmiştir.

Tablo 3.4. 1980-1991, 1992-2003 ve 2004-2015 Dönemleri Yıllık Ortalama Yağışların İstatistiksel Özeti

Parametre	**1. Dönem**	**2. Dönem**	**3. Dönem**
Ortalama	633 mm	647 mm	647 mm
Maksimum	2237 mm	2243 mm	2420 mm
Minimum	257 mm	250 mm	232 mm
Standart Sapma	293 mm	302 mm	310 mm

1. dönemin yıllık ortalama yağış miktarı 633 mm iken 2. ve 3. dönemlerin 647 mm olmuştur. Bununla birlikte maksimum yağış açısından ilk iki dönem benzerlik gösterirken son dönemde yaklaşık %10'luk bir artış olmuştur. Ayrıca minimum yağışta da son dönem 232 mm ile en düşük yağışı almıştır. Sonuç olarak 3. dönemde maksimum ve miinimum yağış aralığı genişlemiştir. Buna karşılık algoritma, alt dönemlerin yağış analizleri sonucunda 1980-1991 dönemi için 4 küme, 1992-2003 dönemi için 7 küme ve 2004-2015 dönemi için 13 küme oluşturmuştur. Son dönemde oluşan fazla kümeler çoğunlukla Ege Bölgesi, İç Anadolu Bölgesi ve Akdeniz Bölgesinin kesiştiği alanlarda ortaya çıkmıştır. Dadaşer-Çelik ve ark [47] 1980-2015 dönemine ait yağış verileri ile yaptıkları eğilim analizi çalışmasında özellikle eylül ve ekim aylarında Türkiye'nin batısında bulunan bir çok istasyonda yukarı yönlü eğilim belirlemişlerdir. Küme sayısının belirtilen bölgelerdeki artışının bu çalışmanın sonuçları ile uyumlu olduğu görülmüştür.

2004-2015 döneminde ilk iki dönemde de keşfedilen 3 aşırı anormal değere ek olarak Giresun (istasyon no.17034) istasyonu da 1360 mm yıllık ortalama yağış ile 3. dönemde aşırı anormal olarak belirlenmiştir. Bir diğer yandan, gizlenmiş anormal sayısı küme sayılarındaki artışa karşılık sırasıyla 15, 11 ve 10 olacak şekilde azalmıştır. İlk iki dönem boyunca görülen 3 no'lu küme 1980-1991 döneminde 109 (%56) istasyon ve 1992-2003 döneminde 116 (%59) istasyonu temsil etmiştir. 2004-2015 döneminde ise bu küme bölünerek birçok farklı yağış kümesini oluşturmuştur. Yeni oluşan kümeler genelde Ege ve Akdeniz Bölgeleri'nde etkin hale gelirken Karadeniz, Güney Doğu Anadolu ve Doğu Anadolu Bölgeleri ve İç Anadolu Bölgesi'nin çoğunluğunda önemli değişimler gözlemlenmemiştir.

3.3. 1980-2015 Dönemi Aylık Yağış Serilerinin Anormallik Analizi

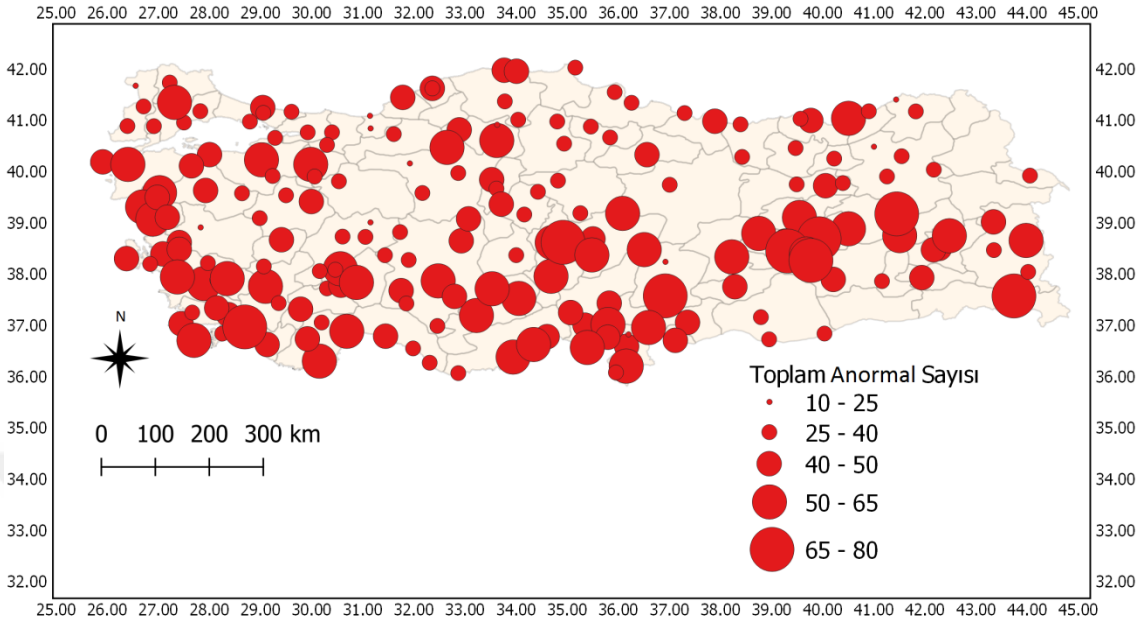
Bu tez çalışmasının ikinci kısmı olarak 1980–2015 dönemine ait Türkiye aylık yağış serilerinin analizi yapılmış ve kümeler ve anormal değerler keşfedilmiştir. Verilerin kullanım şeklini açıklayacak olursak, her bir istasyonun aynı aylarını içeren 36 yıllık serisi öncelikle kendi içerisinde analiz edilmiştir. Böylelikle her bir istasyon için her ayı temsil eden 12 ayrı seri ile toplamda 195 istasyon bulunmasından dolayı 2340 adet serinin analizi yapılmıştır. Analiz sonucunda ise her değer normal, yüksek anormal ve düşük anormal durumlarından biri olarak atanmıştır. Tablo 3.5'te veri seti, analiz ve sonuçların özeti gösterilmiştir.

Tablo 3.5. Türkiye Yağış Serileri Analizinin Özet Sonuçları

Parametre	**Değer**
İstasyon Sayısı	195
Toplam Veri Sayısı	84240 (195x12x36)
Analiz Edilen Seri Sayısı	2340
Keşfedilen Anormal Sayısı	8338 (toplam veride % 9.89)
Keşfedilen Yüksek Anormal Sayısı	7188 (toplam veride % 8.53)
Keşfedilen Düşük Anormal Sayısı	1150 (toplam veride % 1.36)

Analizin sonuçlarına göre 1980 – 2015 dönemine ait Türkiye yağış serilerinde görülen anormal miktarı %9.89'dur ve bunların çoğunluğu yüksek anormal yağışlar olmuştur. Şekil 3.8'de istasyonlarda görülen toplam anormal sayılarına göre derecelendirilmiş haritaya yer verilmiştir.

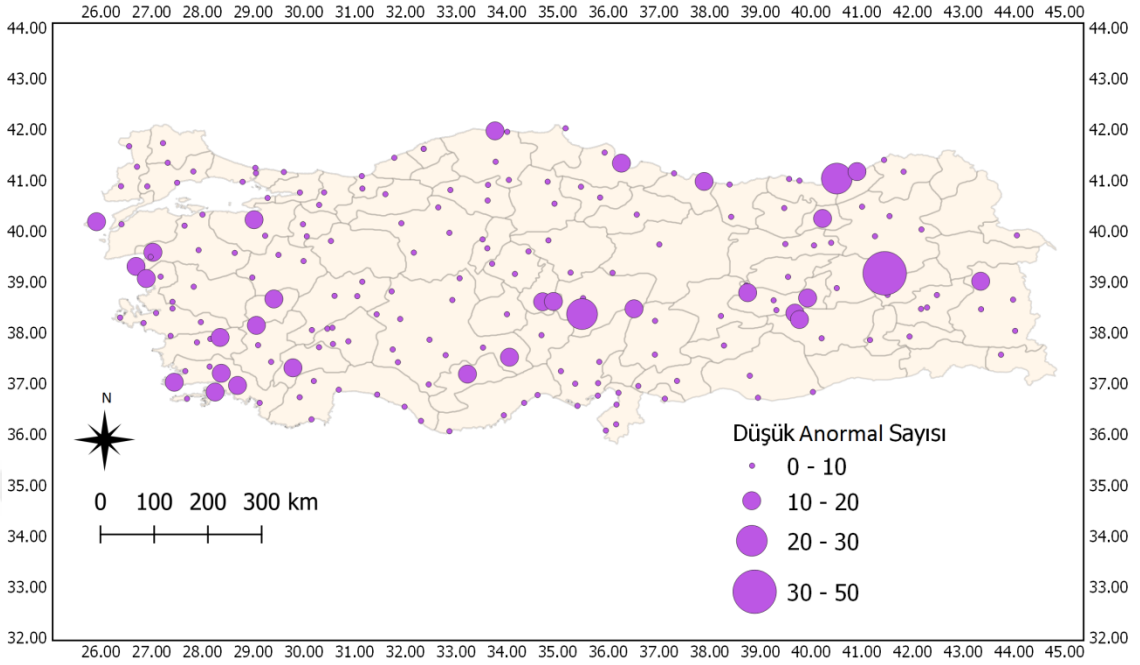
Türkiye üzerinde anormal yağış sayısının en fazla olduğu istasyon 80 ile sahip Muş Varto (istasyon nu. 17778) olmuştur. Bu durumu takiben Kahramanmaraş (istasyon nu. 17255) ve Hakkari merkez (istasyon nu. 17285) istasyonları da keşfedilen 72 anormal değer ile ilk sıralarda yer almıştır. Bununla birlikte en az anormal yağış sayısının 14 olarak Edirne Merkez (istasyon nu. 17050) ve 16 olarak Düzce Akçakoca (istasyon nu. 17015) istasyonlarında olduğu belirlenmiştir. Türkiye genelinde ise toplam anormal yağış sayısının sıklıkla yüksek görüldüğü bölgelerin Doğu Anadolu, Akdeniz, Ege ve İç Anadolu'nun güneyi olduğu görülmüştür.



Şekil 3.8. Toplam Anormal Sayısına Göre Derecelendirilmiş Harita

Yüksek anormal açısından Hakkari Merkez (istasyon no. 17285) 66, Elazığ Sivrice (istasyon no. 17844) 63 ve Mersin Silifke (istasyon no. 17330) 62 sayılarında yüksek anormal yağışlarla en öne çıkan istasyonlar olmuştur. Yüksek anormal değerlerin sayısı toplam anormal sayılarının %85'inden fazlası oluştuğu için aynı şekilde Doğu Anadolu, Akdeniz ve Ege Bölgeleri'yle birlikte İç Anadolu Bölgesi'nin güney kısmında yoğunlaşmıştır.

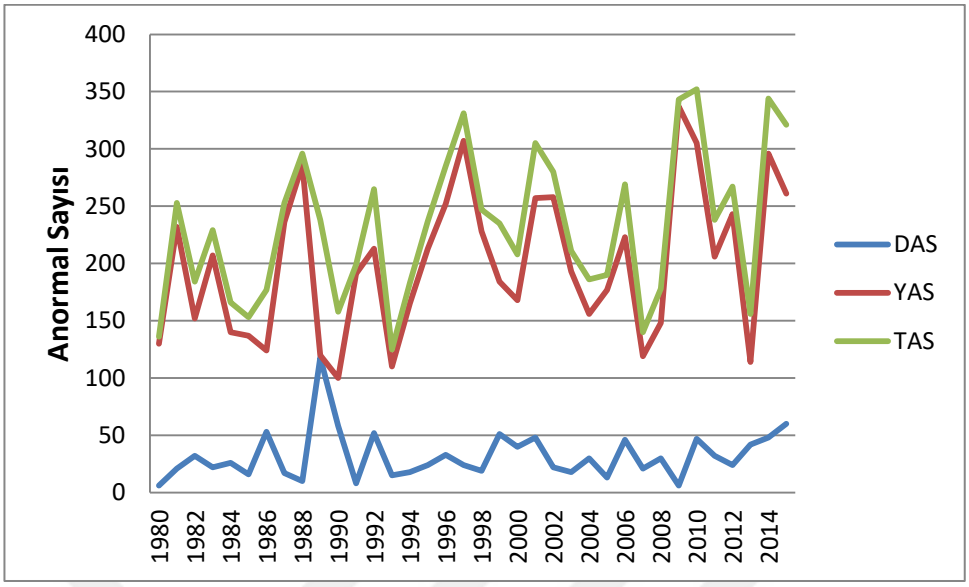
Düşük anormal sayıları incelendiğinde ise bu değerin 31 istasyonda 0 olduğu görülmüştür. En çok sayıda düşük anormale sahip olan istasyon 48 ile Muş Varto (istasyon no. 17778) olmuştur. Tüm istasyonlar arasında Muş Varto, düşük anormal sayısının yüksek anormal sayısından daha fazla olduğu tek istasyon olarak bulunmuştur. İkinci en fazla düşük anormal sayısına sahip olan Rize Merkez (istasyon no. 17040) istasyonu 22 ve üçüncü olan Kayseri Develi (istasyon no. 17836) istasyonu ise 21 düşük anormale sahip oldukları görülmüştür. Şekil 3.9'da düşük anormal sayılarının derecelendirmiş haritası gösterilmiştir.



Şekil 3.9. Düşük Anormal Sayısına Göre Derecelendirilmiş Harita

Sonuç olarak düşük anormal yağışların çoğunlukla bölgesel olarak dağılmamasına karşılık Ege Bölgesi istasyonlarında değerleri 10–20 aralığında bulunan bir yoğunluk oluştuğu fark edilmiştir.

Anormal sayılarının yıllara göre dağılımı incelendiğinde toplam anormal sayısının en çok görüldüğü yıl 352 ile 2010 yılı olmuştur. Bu yıldan sonra ise 341–321 aralığında toplam anormal yağış sayısına sahip yıllar sırasıyla 2014, 2019, 1997 ve 2015 yılları olmuştur. Ayrıca 1993, 1980 ve 2007 yılları en düşük anormal sayılarının görüldüğü yıllar olmuştur ve değerleri 125–140 aralığında değişmiştir. Şekil 3.10'da 1980 – 2015 yılları arasında gerçekleşen toplam anormal yağış sayısı (TAS), düşük anormal yağış sayısı (DAS) ve yüksek anormal yağış sayısı (YAS) gösterilmiştir.

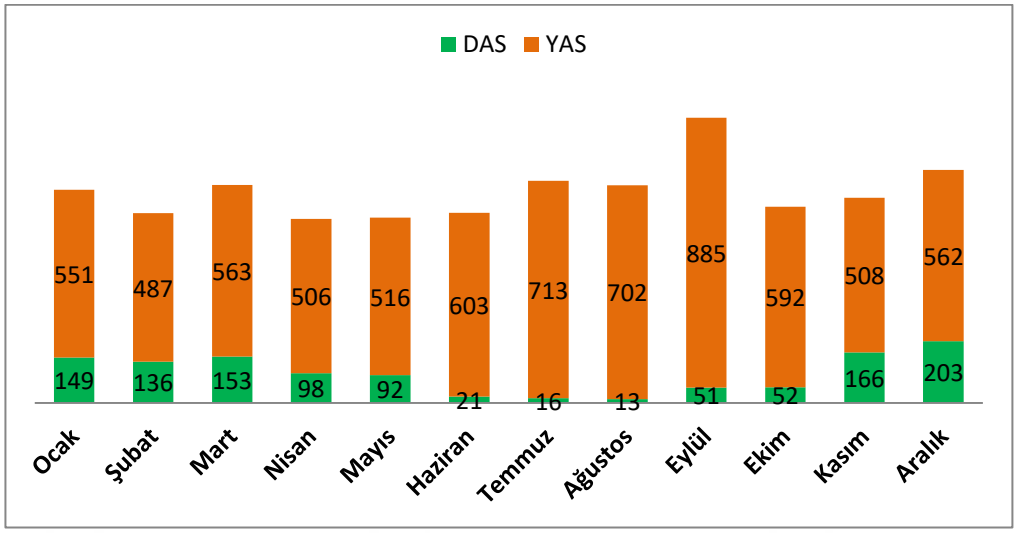


Şekil 3.10. Düşük, Yüksek ve Toplam Anormal Sayılarının Yıllara Göre Dağılımı (DAS: Düşük Anormal Sayısı, YAS: Yüksek Anormal Sayısı, TAS: Toplam Anormal Sayısı)

YAS'a göre 2009, 1997 ve 2010 yılları ön plana çıkmıştır ve değerleri 337, 307 ve 305 olarak kaydedilmiştir. YAS'ın en düşük olduğu yıllar ise 1990, 1993 ve 2013 yılları olmuştur ve değerlerin 100 civarında kaldığı görülmüştür.

DAS açısından 1989 yılı diğer yıllara göre önemli bir farkla öne çıkmıştır. Öyle ki ikinci en yüksek DAS'a sahip 2015 yılında 60 kez bu durum görülmüşken 1989 yılında bu değerin 118 olduğu görülmüştür. Bununla birlikte en az DAS'a sahip yıllar 2009, 1980 ve 1991 yılları olmuştur ve tek haneli rakamlarda kaldıkları görülmüştür.

Bir diğer söz konusu zamansal dağılım ise anormal sayılarının aylar bazında incelenmesi olmuştur. En çok sayıda anormalin gözlemlendiği ay Eylül ayı olmuştur. Eylül ayında 36 yıllık süre boyunca 936 kez anormal yağış gözlemlenmiştir ve bu yağışların 885 tanesi yüksek anormal yağışlar olmuştur. En az sayıda anormal yağış ise 604 kez olmak üzere Nisan ayında ve 608 ile Mayıs ayında olduğu görülmüştür. Şekil 3.11'de Anormal sayılarının aylara göre dağılımı gösterilmiştir.



Şekil 3.11. Anormal Sayılarının Aylara Göre Dağılımı (DAS: Düşük Anormal Sayısı, YAS: Yüksek Anormal Sayısı)

Yaz aylarının ortak noktası az sayıda düşük anormal değerlerin gözlemlenmesi olmuştur. Düşük anormal yağışların özellikle kış aylarında meydana geldiği tespit edilmiştir ve Aralık aylarının 203 ile maksimum sayıda düşük anormale sahip olduğu belirlenmiştir.

4. BÖLÜM

TARTIŞMA, SONUÇ VE ÖNERİLER

Bu tez çalışmasında yoğunluk tabanlı kümeleme algoritmalarından biri olan ve anormallik tespiti için de kullanılabilen VDBSCAN algoritmasının, girdi parametrelerinin otomatik olarak belirlenmesi ve iterasyonlu bir yaklaşımla daha verimli kümeler elde edilmesi için geliştirilmesi yapılmıştır. Sonuç olarak autoVDBSCAN algoritması hazırlanmıştır. Yapılan deneysel çalışmalar sonucunda önerilen algoritmanın veri seti içerisindeki aşırı anormal ve gizlenmiş anormal değerleri belirlediği ve mantıklı kümeler oluşturduğu görülmüştür. autoVDBSCAN algoritması kullanılarak Türkiye yağış serilerinin analizleri yapılmıştır.

1980-2015 dönemine ait ortalama yağış verileri ile yapılan analizlerde 8 farklı yağış bölgesi, 3 aşırı anormal ve 14 gizlenmiş anormal keşfedilmiştir. Aşırı anormallerin hepsi ve gizlenmiş anormallerin bir kısmı Doğu Karadeniz Bölgesi'nde keşfedilmiştir. Gizlenmiş anormal yağışların düşük olanları ise çoğunlukla Güney Doğu Anadolu Bölgesi'nde gözlemlenmiştir.

1980-1991, 1992-2003 ve 2004-2015 alt dönemlerinde sırasıyla 4, 7 ve 13 adet küme oluşturulmuştur. Son dönemde oluşan kümelerin, ilk iki dönemdeki en büyük kümeyi oluşturan 3 nolu kümenin (sırasıyla 109 ve 116 adet istasyon içeren) çoğunlukla Ege Bölgesi, Akdeniz Bölgesi ve İç Anadolu Bölgesi'nin bu iki bölge ile kesiştiği alanlarda bölünmesi ile oluştuğu görülmüştür. Algoritmanın verdiği sonuçlara göre 1980-2015 döneminde yağışların çeşitliliği artmıştır. Ayrıca ilk iki dönemde belirlenen 3 aşırı anormal değer: Artvin Hopa (istasyon no.17042), Rize merkez(istasyon no.17040) ve Rize Pazar (istasyon no.17628) istasyonlarındadır. Üçüncü dönemde ise Giresun Merkez istasyonu da aşırı anormal yağış sınıfına girmiştir.

Bu tez çalışmasında yapılan diğer bir analiz ise 1980-2015 dönemi aylık yağış verilerinin sadece anormallik yönünden incelenmiş olmasıdır. Sonuç olarak keşfedilen anormal değerlerin neredeyse % 85'inin yüksek değere sahip anormaller olduğu görülmüştür. En çok sayıda anormal yağış Muş Varto (istasyon no. 17778), Kahramanmaraş (istasyon no. 17255) ve Hakkari merkez (istasyon no. 17285) istasyonlarında gözlemlenmiştir. Bununla birlikte en çok sayıda anormal yağışın 2010, 2014, 2019, 1997 ve 2015 yıllarında olduğu tespit edilmiştir. Aylar açısından ise Nisan ayı en az sayıda anormal yağış görülen ay olmuştur ve Eylül ayının en çok sayıda anormal yağışa sahip olduğu belirlenmiştir.

Çalışmada iterasyon yöntemi ile yapılan analizlerde bu işlem yalnızca bir kez uygulanmıştır. Gelecekte yapılacak olan çalışmalarda veri setine bağlı olarak kaç iterasyonun yapılacağı üzerine teknikler geliştirilebilir. Bununla birlikte anormallik analizleri sonucunda çoğu anormalliğin yüksek değerdeki anormaller olduğu ve çok az sayıda düşük değere sahip anormalliğin keşfedildiği görülmüştür. Buradaki problem düşük yağışların yıllar içerisinde tekrarlaması ve artık birer küme oluşturmalarıdır. Bu durumun farklı bir çalışmada incelenmesi faydalı olacaktır.

KAYNAKÇA

1- Öztürk, K., 2002. Küresel iklim değişikliği ve Türkiye'ye olası etkileri. **Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi**, **22** (1): 47-65

2- Intergovernmental Panel on Climate Change (IPCC), 2019. **IPCC Special Report on Climate Change**. 2 - 6 August in Geneva, Switzerland.

3- Türkeş, M., Sümer, U. M., Demir, D., 2002. Re-evaluation of trends and changes in mean, maximum and minimum temperatures of Turkey for the period 1929-1999. **International Journal of Climatology**, **22**(8): 947-977.

4- Türkeş, M., 2012. Türkiye'de gözlemlenen ve öngörülen iklim değişikliği, kuraklık ve çölleşme. **Ankara Üniversitesi Çevre Bilimleri Dergisi, 4**(2): 1-32.

5- Türkeş, M., Tatlı, H. 2009. Use of the standardized precipitation index (SPI) and modified SPI for shaping the drought probabilities over Turkey. **International Journal of Climatology**, **29**(15): 2270–2282.

6- Altınsoy, H., Öztürk, T., Türkeş, M., Kurnaz, M. L., 2011. Projections of climate change in the Mediterranean Basin by using downscaled global climate model outputs. **International Journal of Climatology**, **35**(14): 4276-4292.

7- Vijayakumar, V., Nedunchezhian, R. A, 2012. A study on video data mining. **International Journal of Multimedia Info Retrieval**, 1:153–172

8- Cavoukian, A., 1998. Data Mining: Staking a Claim on Your Privacy. Information and Privacy Commissioner's Report, Ontario, Canada, 350pp.

9- Bertıno, E., Fovino, I.N., Provenza, L.P., 2005. A framework for evaluating privacy preserving data mining algorithm, **Data Mining and Knowledge Discovery, 11**(2): 121–154,

10- Tan, P. N., Steinbach, M., Kumar, V., 2005. Introduction to Data Mining. Pearson Addison-Wesley, Boston, 165pp.

11- Gorunescu, F., 2011. Data Mining Concepts, Models and Techniques. Springer-Verlag Berlin Heidelberg, 360pp.

12- Han, J., Kamber, M., Pei, J., 2011. Data Mining: Concepts and Techniques, 3rd edition, Morgan Kaufmann, 744pp.

13- Akpınar, H., 2000. Veri tabanlarında bilgi keşfi ve veri madenciliği. **İ.Ü İşletme Fakültesi Dergisi, 29**(1): 1-22.

14- Timofeev, R. V., 2004. Classification And Regression Trees Theory And Applications. Humboldt Üniversitesi, A Master Thesis, Berlin, 39pp.

15- Han, J., Fu, Y., 1999. Mining multiple-level association rules in large databases, **IEEE Transactions on Knowledge and Data Engineering, 11**(5): 798-805.

16- Zaki, M. J., 1999. Parallel and distributed association mining: a survey. **IEEE Concurrency, special issue on Parallel Mechanisms for Data Mining, 7**(5): 14-25

17- Dinçer, E., 2006. Veri Madenciliğinde K-Means Algoritması ve Tıp Alanında Uygulanması. Kocaeli Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Kocaeli, 101s.

18- Arslan, H., 2008. Sakarya Üniversitesi Web Sitesi Erişim Kayıtlarının Web Madenciliği İle Analizi. Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Sakarya, 80s.

19- Berkhin, P., 2006. A Survey of Clustering Data Mining Techniques, pp. 27-29. In: Grouping Multidimensional Data (Eds. Kogan J., Nicholas C., Teboulle M.). Springer, Berlin, Heidelberg.

20- Yeşilbudak, M., Kahraman, H. T., Karacan, H., 2011. Veri madenciliğinde nesne yönelimli birleştirici hiyerarşik kümeleme model. **Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi, 26**(1): 27-39.

21- Birant, D., 2019. Farklı bağlantı yöntemleri ile hiyerarşik kümeleme topluluğu. **Selçuk Üniversitesi Mühendislik Bilim ve Teknik Dergisi, 7**(1): 154-164.

22- Özkan, Y., 2008. Veri Madenciliği Yöntemleri. Papatya Yayıncılık, İstanbul, 233s.

23- Latin, J. M., Carroll, D. J., Green, P. E., 2003. Analyzing Multivariate Data. Thomson Brooks-Cole, USA, pp283.

24- Akın, Y. K., 2008. Veri Madenciliğinde Kümeleme Algoritmaları ve Kümeleme Analizi. Marmara Üniversitesi, Sosyal Bilimler Enstitüsü, Doktora tezi, İstanbul, 189s.

25- Jain, A. K., Murty, M. N., Flynn, P. J., 1999. Data clustering: A review. **ACM Computing Surveys, 31**(3): 264-323.

26- Bhoomi M., 2014. Enhanced k-means clustering algorithm to reduce time complexity for numeric values. **International Journal of Computer Science and Information Technologies, 5**(1): 876-879

27- Derban, G., Moldovan, G. S., 2006. A comparison of clustering techniques in aspect mining. **Studia University, L1**(1): 69-78.

28- Kriegel, H. P., Kröger, P., Sander, J., Zimek, A., 2011. Density-based clustring. **WIREs data mining and knowledge discovery, 1**(3): 231–240

29- Amini, A., Wah T. Y., Saboohi, H., 2014. On density-based data streams clustering algorithms: a survey. **Journal Of Computer Scıence And Technology 29**(1): 116–141.

30- Ruivo, H. M., Velho, H. F., Sampaio, G., Ramos, F. M., 2015. Analysis of extreme precipitation events using a novel data mining approach. **American Journal of Environmental Engineering , 5**(1A): 96-105

31- Crane, R. G., 2003. Clustering and upscaling of station precipitation records to regional patterns using self-organizing maps (SOMs). **Climate Research, 25**(2): 95–107.

32- Ahmad, N. H., Othman, I. R., Deni, S. M., 2013. Hierarchical cluster approach for regionalization of peninsular malaysia based on the precipitation amount. **Journal of Physics: Conference Series, 423**(1) 012018.

33- Barton, Y., Giannakaki, P., Waldow, H., Chevalier, C., Pfahl, S., Martius, O., 2016. Clustering of regional-scale extreme precipitation events in southern switzerland. **Monthly weather Review, American Meteorological Society, 144(1)**: 347–369.

34- Kındap, T., Ünal, Y., Karaca, M., 2003. Redefining the climate zones of Turkey using cluster analysis. **International Journal of Climatology, 23**(9): 1045–1055.

35- İyigün, C., Türkeş, M., Batmaz, İ., 2013. Clustering current climate regions of Turkey by using a multivariate statistical method. **Theoretical and Applied Climatology 114**(1-2): 95–106.

36- Sönmez, İ., Kömüşçü, A.Ü., 2008. K-ortalamaları kümeleme yöntemi ile Türkiye yağış bölgelerinin yeniden tanımlanması ve altperiyodlardaki değişimler. **İklim Değişikliği ve Çevre, 1**(1): 39-49.

37- Şahin, S., Cigizoğlu, H., K., 2012. The sub-climate regions and the sub-precipitation regime regions in Turkey. **Journal of Hydrology, 450–451**: 180-189.

38- Kumar, N., D., Dhanya, C., T., 2009. Data Mining and It's Applications For Modelling Rainfall Extremes. **ISH Journal Of Hydraulic Engineering, 15**(1): 25-51

39- Ester, M., Kriegel, H., P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. in Proc. **2nd International Conference on Knowledge Discovery and Data Mining, Portland**, pp. 226-231.

40- Khan, K., Rehman, S., U., Aziz, K., Fong, S., Sarasvady, S., 2014. DBSCAN: Past, present and future. **The Fifth International Conference on the Applications of Digital Information and Web Technologies, Bangalore**, pp. 232-238.

41- Celik, M., Dadaşer-Celik, F., Dokuz, A., S., 2011. Anomaly detection in temperature data using DBSCAN algorithm. **International Symposium on Innovations in Intelligent Systems and Applications, Istanbul**, pp. 91-95,

42- Rahmah, N., Sitanggang, I., S., 2015. Determination of optimal epsilon (Eps) value on dbscan algorithm to clustering data on peatland hotspots in Sumatra, **IOP Conference Series: Earth and Environmental Science, Workshop and International Seminar on Science of Complex Natural Systems, 31**(1), Bogor, Indonesia.

43- Ozkok, F., O., Celik, M., 2017. A new approach to determine eps parameter of DBSCAN algorithm. **International Journal of Intelligent Systems and Applications in Engineering, 5**(4): 247-251.

44- Liu, P, Zhou D., Wu, N., 2007. VDBSCAN: Varied density based spatial clustering of applications with noise. **International Conference on Service Systems and Service Management, Chengdu**, pp. 1-4.

45- Gaonkar, M., N., Sawant, K., 2013. AutoEpsDBSCAN : DBSCAN with eps automatic for large dataset. **International Journal on Advanced Computer Theory and Engineering, 4**(2): 11-16.

46- Parvez, A., W., M., 2012. Data set property based 'K' in VDBSCAN clustering algorithm, in Proc. **World of Computer Science and Information Technology Journal, 2**(3): 115-119.

47- Dadaşer-Çelik, F., Kömüscü, A. U., Uçar, R. 2017. Recent trends in precipitation in Turkey: 1980-2015. **Conference: International 8 th. Atmospheric Sciences Symposium, İstanbul, Turkey,** pp: 645-650.

48- Fırat, M., Dikbaş, F., Koç, A.C., Güngör, M., 2012. K-ortalamalar yöntemi ile yıllık yağışların sınıflandırılması ve homojen bölgelerin belirlenmesi. **İnşaat Mühendisleri Odası Teknik Dergi, 23**(113): 6037-6050.

49- Sariş, F., Hannah, D., M., Eastwood, W., J., 2010. Spatial variability of precipitation regimes over turkey**. Hydrological Sciences Journal, 55**(2): 234-249.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı, Soyadı: Ahmet ÖZEKES
Uyruğu: Türkiye
Doğum Tarihi, Yeri: 02/01/1989 Kayseri
e-posta: ahmetozekes@hotmail.com

EĞİTİM

Derece	Kurum	Başlangıç - Mezuniyet
Yüksek Lisans	Wageningen University&research Environmental Sciences	2021 – Devam ediyor
Yüksek Lisans (3,81/4)	Erciyes Üniversitesi Fen bilimleri Enstitüsü Çevre Mühendisliği Anabilim dalı	2018 - 2021
Lisans (3,68/4)	Erciyes Üniversitesi Mühendislik Fakültesi Çevre Mühendisliği Bölümü	2015 - 2018
Lise	Kocasinan Lisesi	2002 - 2005

YABANCI DİL

İngilizce (B2)

PROJELER

Çalışma-Etkinlik Adı	Açıklama	Tarih
Sultan Sazlığına Yayılı Kaynaklardan Ulaşan Kirletici Yüklerinin SWAT Modeli ile Modellenmesi ve Su Kirliliğini Önleyici Stratejilerin Geliştirilmesi	TÜBİTAK 1001 Araştırma Projesi Proje Yürütücüsü : Doç. Dr. Filiz DADAŞER ÇELİK Proje No : 114y595 (Bursiyer)	10.2017/06.2018
Kayseri Kent Merkezinde NOx Düzeyinin Mekânsal Dağılımı Ve İncelenmesi	2209/A TÜBİTAK Üniversite Öğrencileri Araştırma Projeleri Destekleme Programı 2017/1 (Araştırmacı)	2017-2018
Kayseri Kent Merkezinde Ozon Seviyelerinin Mekânsal Dağılımının İncelenmesi	Erciyes Üniversitesi Mühendislik Fak. Çevre Müh. Bölümü Bitirme Ödevi	2017-2018

YAYINLAR

Özekes, A., Celik, M., Özkök, F.Ö. , Komuscu, A.U., Dadaser-Celik, F., "AutoVDBSCAN: An Automatic and Level-Wise Varied-Density Based Anomaly Detection Algorithm", 7th International Conference on Advanced Technologies, 28 Nisan - 1 Mayıs 2018, Antalya.