

**T.C
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

VERİ MADENCİLİĞİ YÖNTEMLERİYLE DEPREMLERİN ANALİZİ

YÜKSEK LİSANS TEZİ

Özal YILDIRIM

Anabilim Dalı: Bilgisayar Mühendisliği

Programı: Yazılım

Tez Danışmanı: Yrd. Doç. Dr. Galip AYDIN

TEMMUZ-2010

**T.C
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

VERİ MADENCİLİĞİ YÖNTEMLERİYLE DEPREMLERİN ANALİZİ

YÜKSEK LİSANS TEZİ

Özal YILDIRIM

07229102

Anabilim Dalı: Bilgisayar Mühendisliği

Programı: Yazılım

Tez Danışmanı: Yrd. Doç. Dr. Galip AYDIN

Tezin Enstitüye Verildiği Tarih: 13 Temmuz 2010

TEMMUZ-2010

**T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

VERİ MADENCİLİĞİ YÖNTEMLERİYLE DEPREMLERİN ANALİZİ

YÜKSEK LİSANS TEZİ

Özal YILDIRIM

07229102

**Tezin Enstitüye Verildiği Tarih: 13 Temmuz 2010
Tezin Savunulduğu Tarih: 29 Temmuz 2010**

Tez Danışmanı : Yrd. Doç. Dr. Galip AYDIN (F.Ü)
Diğer Jüri Üyeleri : Yrd. Doç. Dr. Mustafa TÜRK (F.Ü)
Doç. Dr. Mehmet KAYA (F.Ü)

TEMMUZ-2010

ÖNSÖZ

Büyük deprem levhalarının çevrelediği ülkemizin geniş bir bölümü deprem kuşağında yer almaktadır. Bu nedenle ülkemiz geçmişte büyük deprem felaketlerine maruz kalmıştır. Büyük ölçekte maddi ve manevi hasarlara yol açan bu yıkımlar, depremin ne derecede büyük bir felaket olduğunu gözler önüne sermektedir. Depremin diğer doğal afetlere oranla büyük bir tehdit unsuru olmasının en önemli nedeni önceden bir uyarısının olmamasıdır. Bu alanda yapılan çalışmalar genellikle deprem tahmini üzerine yoğunlaşmıştır.

Bu tez çalışmasında, deprem alanında yapılan tahmin ve analiz çalışmalarının gelişimine katkıda bulunmak için farklı uygulamaları bünyesinde barındıran bir çalışma yapılmıştır. Depremlerin analizi ve tahmini için bilgisayar bilimlerinin popüler konularından veri madenciliği teknikleri deprem verileri üzerinde uygulanmıştır.

Tez çalışması süresince yardım ve desteğini hiçbir zaman esirgemeyen çok değerli danışmanım Sayın Yrd. Doç. Dr. Galip AYDIN'a, çalışmalarım boyunca bana büyük katkısı olan değerli hocam Sayın Öğr. Gör. Dr. Murat KARABATAK'a, tecrübeleriyle bana yol gösteren değerli meslektaşım Sayın Arş. Gör. Şengül DOĞAN'a, çalışmalarım süresince sürekli olarak bana rehberlik eden ve akademik çalışma adına bana büyük katkı sağlayan mesai arkadaşım Sayın Yrd. Doç. Dr. Hüseyin ERİŞTİ'ye, her zaman desteklerini sağladıkları için Bilgisayar Mühendisliğinin kıymetli öğretim üyelerine, değerli çalışma arkadaşlarıma, çalışmalarımındaki en büyük destekçilerim olan değerli aileme teşekkürü bir borç bilirim.

Özal YILDIRIM

ELAZIĞ–2010

İÇİNDEKİLER

Sayfa No

ÖNSÖZ	II
İÇİNDEKİLER.....	III
ÖZET	V
SUMMARY	VI
ŞEKİLLER LİSTESİ	VII
TABLolar LİSTESİ	XI
KISALTMALAR LİSTESİ.....	XII
1. GİRİŞ	1
1.1. Genel	1
1.2. Tezin Amacı	2
1.3. Tezin Organizasyonu ve Katkıları	4
2. VERİ MADENCİLİĞİ.....	6
2.1. Veri Madenciliği Süreci	7
2.1.1. Problemin Tanımlanması	8
2.1.2. Verilerin Hazırlanması	8
2.1.2.1. Toplama.....	9
2.1.2.2. Birleştirme ve Temizleme	9
2.1.2.3. Dönüştürme	9
2.1.3. Modelin Kurulması ve Değerlendirilmesi	9
2.1.4. Modelin Kullanılması.....	11
2.1.5. Modelin İzlenmesi.....	11
2.2. Veritabanlarından Bilgi Keşfi Süreci	11
2.3. Veri Madenciliğinde Kullanılan Modeller	13
2.4. Veri Madenciliğinde Kullanılan Yöntemler	14
2.4.1. Genetik Algoritmalar	14
2.4.2. K-En Yakın Komşu.....	16
2.4.3. Karar Ağaçları.....	16
2.4.4. Yapay Sinir Ağları	17
2.4.5. Birliktelik Kuralları.....	18
2.4.5.1. Apriori Algoritması.....	21
2.5. Veri Madenciliğinin Diğer Disiplinlerle İlişkisi	23
2.5.1. Örnek Karşılaştırma	24
2.5.2. Veri Madenciliği Yaklaşımları	24
2.6. Veri Madenciliği Türleri	24
2.6.1. Doğrudan Veri Madenciliği.....	25
2.6.2. Varsayım Deneme ve Varsayımı Daha İyi Hale Getirme	25
2.6.3. Dolaylı Veri Madenciliği.....	25
2.7. Bilgi Bulma Süreci.....	25
2.7.1. Bilgi Bulma Sürecinin Önemli Özellikleri	25
2.7.2. Bilgi Bulma Sürecinde Önemli Sorunlar.....	26
2.7.3. Bilgi Bulma Sürecinde Kullanılan Yöntemler.....	27
2.7.3.1. Seçim.....	27
2.7.3.2. İşlem Öncesi	27
2.7.3.2.1. İşlem Öncesi Süreçler(Ön Hazırlık).....	28
2.7.3.3. Veri Madenciliği Aşaması.....	28
2.7.3.4. Yorum.....	28

3.	DEPREM KATALOGLARI.....	30
3.1.	SCSN(Güney Kaliforniya Sismografik Ağ) Kataloğu.....	30
3.2.	SCEDC(Güney Kaliforniya Deprem Veri Merkezi) Kataloğu.....	31
3.3.	Türkiye Deprem Kataloğu.....	33
4.	TÜRKİYE DEPREM PORTALI(TDEP).....	35
4.1.	Giriş.....	35
4.2.	Türkiye Deprem Portalı'nın Çalışma Mimarisi.....	36
4.3.	Deprem Verileri Üzerinde Sorgulama ve İstatistiksel İşlemler.....	37
4.4.	Grafik Ekranı	40
4.5.	Türkiye Deprem Kataloğu Üzerinde Yapılan İstatistiksel Çalışmalar.....	41
4.6.	Bölüm Değerlendirmesi	49
5.	DEPREM VERİLERİ ÜZERİNDE BİRLİKTELİK KURALLARI	50
5.1.	Giriş.....	50
5.2.	Kural Üretimi İçin Parametre Seçimi.....	52
5.3.	Çalışma Bölgesinin Seçimi ve Hücrelere Bölünmesi	53
5.4.	Verilerinin Elde Edilmesi ve Uygun Formata Dönüştürülmesi.....	55
5.5.	Birlikteliklerin Bulunması İçin Geliştirilen Algoritma Adımları	55
5.6.	Deprem Verileri Üzerinde Birliktelik Uygulaması.....	57
5.7.	Bölüm Değerlendirmesi	63
6.	DEPREM TAHMİNİ İÇİN KURAL TABANLI BİR YAKLAŞIM	64
6.1.	Giriş.....	64
6.2.	Kural Veritabanının Oluşturulması.....	65
6.3.	Deprem Tahmin Algoritması.....	67
6.4.	Türkiye Deprem Kataloğu Üzerinde Tahmin Çalışması.....	70
6.5.	Bölüm Değerlendirmesi	73
7.	DEPREM VERİLERİNİN KÜMELENDİRİLMESİ.....	75
7.1.	Giriş.....	75
7.2.	Türkiye Deprem Portalı Üzerinde Kümeleme İşlemleri	75
7.3.	Deprem Konum Tabanlı K-Means Algoritması	78
7.4.	Türkiye Deprem Verileri Üzerinde Kümeleme Çalışması	80
7.5.	Büyük Ölçekli Depremlerin Başlangıç Küme Merkezi Olarak Atanması	83
7.6.	K-Means-KNN Algoritması ile Deprem Verilerinin Kümelendirilmesi	87
7.6.1.	Geliştirilen Yöntem.....	87
7.6.2.	Türkiye Deprem Verileri Üzerinde Uygulaması	89
7.7.	Bölüm Değerlendirmesi	99
8.	SONUÇLAR.....	101
	KAYNAKLAR	103
	ÖZGEÇMİŞ	

ÖZET

Bu tez çalışmasında, deprem hareketliliklerinin veri madenciliği yöntemlerini kullanarak analiz edilmesi ve gelecekte meydana gelebilecek depremlerin tahmininde kullanılabilecek bir yazılım tasarlanması amaçlanmıştır. Bu kapsamda, tez içerisinde deprem verileri ile ilgili dört farklı uygulama gerçekleştirilmiştir.

Birinci uygulama, deprem verileri üzerinde istatistiksel, sorgulama ve veri madenciliği çalışmalarını yürütmek için hazırlanan Türkiye Deprem Portalı(TDEP) yazılımı ve deprem verilerine ait bazı istatistiksel çalışmalardan oluşmaktadır. Deprem verilerine ait istatistiksel bilgilerin yorumlanması ile gelecekte oluşabilecek depremlerin tahmini amaçlanmıştır. Hazırlanan uygulamanın, kullanım ve uygulama bakımından büyük avantajlara sahip olduğu görülmüştür.

İkinci uygulama, deprem alanları üzerinde gizli kalmış birlikteliklerin tespit edilmesine yönelik geliştirilen çalışmayı kapsamaktadır. Bu uygulamada, veri madenciliği tekniklerinden birliktelik kuralları deprem verileri üzerine başarılı bir şekilde uygulanmış ve çeşitli kurallar bulunmuştur. Üretilen bu kurallar değerlendirildiğinde, sistemin başarılı sonuçlar ürettiği gözlemlenmiştir.

Üçüncü uygulamada, deprem verileri üzerine yapılan birçok çalışmanın ortak noktası olan deprem tahmini için kural tabanlı yeni bir yaklaşım sunulmuştur. Bu yaklaşım paralelinde bir tahmin algoritması tasarlanarak deprem verileri üzerine başarılı bir şekilde uygulanmıştır. Elde edilen tahmin sonuçları değerlendirildiğinde, algoritmanın başarılı sonuçlar verdiği gözlemlenmiştir.

Dördüncü uygulama, deprem verilerinin kümelendirilmesine yönelik geliştirilen çalışmayı kapsamaktadır. Bu bölümde, Türkiye genelinde oluşmuş deprem verilerinin kümelendirilmesi işlemleri yapılmıştır. Depremlerin kümelendirilmesinde k-Means ve KNN algoritmaları veriler üzerinde başarılı bir şekilde uygulanmıştır. Klasik k-Means algoritmasının başlangıç küme merkezi seçme işlemine, deprem verileri için alternatif bir yaklaşım bu bölüm altında sunulmuştur. Bu yaklaşımın verileri kümelendirmede klasik k-Means algoritmasından daha hızlı bir performans gösterdiği tespit edilmiştir.

Anahtar Kelimeler: Deprem, Veri Madenciliği, Birliktelik Kuralları, Kümeleme, Deprem Tahmin

SUMMARY

Analysis of Earthquakes By Means of Data Mining Methods

The goals of this study can be summarized as analysis of seismic activities through data mining methods and the development of software which can be used to predict earthquakes before they take place. In this regard four different implementations about the seismic data are realized.

The first application is the Turkey Earthquake Portal (TDEP) which is created to carry out statistical processing, querying and data mining studies on seismic data. Prediction of earthquakes is aimed by interpreting the statistical information about the seismic data. Significant advantages of the application in terms of usage and interpretation is observed.

The second application consists of the study which is developed to reveal the hidden associations between earthquake zones. In this application one of the data mining techniques, association rules are successfully applied on the seismic data and various rules are extracted. Evaluation of these rules shows that the system produces successful results.

In the third application a novel approach to earthquake prediction is proposed. In parallel to this approach a prediction algorithm is designed and successfully applied to seismic data. Evaluation of the prediction results shows that the algorithm produces successful results.

The fourth application consists of the clustering of seismic data. In this section clustering results of earthquakes occurred in Turkey is discussed. For clustering k-Means and K-NN algorithms are effectively used. An alternative approach to traditional k-Means initial cluster center selection is proposed in this section for seismic data. It is observed that this approach demonstrates better performance in data clustering than the traditional k-Means algorithm.

Key Words: Earthquake, Data Mining, Association Rules, Clustering, Earthquake Prediction

ŞEKİLLER LİSTESİ

	<u>Sayfa No</u>
Şekil 2.1. Veri madenciliği Süreci	1
Şekil 2.2. Veri tabanından bilgi keşfi aşamaları	1
Şekil 2.3. Genetik algoritma süreçleri	1
Şekil 2.4. Karar ağacının yapısı	1
Şekil 2.5. Basit bir yapay sinir ağı modeli.....	1
Şekil 2.6. İşlemler, satın alınan ürünler, destek değerleri ve güven değerleri	1
Şekil 2.7. Bir marketten satın alınan ürünler listesi	1
Şekil 2.8. Apriori algoritması ile aday öğelerin ve sık kullanılan öğelerin belirlenmesi.....	1
Şekil 2.9. Veri madenciliğinin ilişkili olduğu diğer disiplinler	1
Şekil 3.1. SCSN deprem katalog formatı	1
Şekil 3.2. SCSN deprem katalogunda tutulan veri örneği.....	1
Şekil 3.3. SCEDC katalog formatı	1
Şekil 3.4. SCEDC deprem katalogunda tutulan veri örnekleri	1
Şekil 3.5. Türkiye deprem katalogunda tutulan veri örnekleri	1
Şekil 4.1. Türkiye deprem portalı'nın web ara yüzü.....	35
Şekil 4.2. Türkiye deprem portalı'nın çalışma mimarisi.....	36
Şekil 4.3. Portalın sorgulama ve istatistik için çalışan web ara yüzü.....	38
Şekil 4.4. Portal üzerinde deprem verilerinin yerleşimi	38
Şekil 4.5. Deprem büyüklüklerini temsil eden simgeler	39
Şekil 4.6. Portalda seçilen bölge üzerindeki depremlerin sorgulanması	39
Şekil 4.7. Türkiye Deprem Portalında bulunan istatistiksel grafik ekranı	40
Şekil 4.8. 01.01.1900–01.01.2008 tarihleri arasındaki Türkiye genelinde oluşan $M_c > 0$ büyüklüğündeki depremlerin yıllık sayı dağılımları	41
Şekil 4.9. 01.01.1900–01.01.2008 tarihleri arasında Türkiye genelinde olulan $M_c > 0$ büyüklüğündeki depremlerin yıllık büyüklük ortalamaları	42
Şekil 4.10. 01.01.1960–31.12.2003 tarih aralığında Türkiye genelinde oluşan $M_c > 0$ büyüklüğündeki depremlerin yıllık sayı dağılımları	43
Şekil 4.11. 01.01.1960–31.12.2003 tarihleri arasında Türkiye genelinde oluşan $M_c > 0$ büyüklüğündeki depremlerin büyüklüklere göre sayı dağılımları	43
Şekil 4.12. Marmara bölgesi ve çevresinde yapılan sorgu ve istatistiksel çalışma.....	44

Şekil 4.13. 01.01.1960–01.01.2008 tarihleri arasında Marmara bölgesi ve çevresinde oluşan $M_c > 0$ büyüklüğündeki depremlerin yıllık sayı dağılımları.....	45
Şekil 4.14. 01.01.1960–01.01.2008 tarihleri arasında Marmara ve çevresinde oluşan $M_c \geq 4$ büyüklüğündeki depremlerin yıllık ortalama büyüklük grafiği.....	45
Şekil 4.15. 01.01.1960–01.01.2008 tarihleri arasında Marmara ve çevresinde oluşan $M_c \geq 4$ büyüklüğündeki depremlerin büyüklük aralıklarına göre sayı dağılımları .	46
Şekil 4.16. Marmara ve çevresinde meydana gelen 4-5M büyüklüğündeki depremlerin frekans grafiği	47
Şekil 4.17. Sorgulama işlemlerinin daha küçük ölçekteki bölgeler üzerinde uygulanması	47
Şekil 4.18. 01.01.1965–01.01.2008 tarih aralığında seçimi yapılan bölgeye ait istatistiksel grafikler	48
Şekil 5.1. Birliktelik kural üretimi için hazırlanan sistemin çalışma mimarisi.....	50
Şekil 5.2. Türkiye deprem portalının apriori uygulaması için çalışan web ara yüzü ...	51
Şekil 5.3. Parametre seçimi için portal üzerinde hazırlanan panel görüntüsü	52
Şekil 5.4. Çalışma alanının portal arayüzünde seçimi.....	53
Şekil 5.5. Çalışma bölgesinin farklı derecelerde hücre aralıklarına bölünmesi.....	54
Şekil 5.6. Hücreler arasındaki birlikteliklerin tespit edilmesi için geliştirilen algoritma	56
Şekil 5.7. Kural üretimi için belirlenen örnek çalışma bölgesi.....	57
Şekil 5.8. Üretilen kuralların txt formatında gösterimi	59
Şekil 5.9. Sık geçen deprem bölgelerinin tespit edilmesi ve Türkiye deprem haritasındaki konumu	60
Şekil 5.10. Doğu Anadolu Bölgesi üzerinde belirlenen çalışma alanındaki tekrarlı öğelerin bulunması	61
Şekil 5.11. Çalışma bölgesi üzerinde 01.01.1964 ve 31.12.2007 tarihleri arasında elde edilen kurallar	62
Şekil 6.1. Portal üzerinde deprem tahmin algoritmasının yerleşimi	64
Şekil 6.2. Tahmin bölgesinin seçimi ve hücrelere bölünmesi	65
Şekil 6.3. Deprem tahmin algoritmasının kural üretim bölümü	67
Şekil 6.4. Deprem tahmini için geliştirilen algoritma yapısı	68
Şekil 6.5. Deprem tahmin algoritması için tasarlanan akış şeması	69
Şekil 6.6. Algoritmanın örnek veri için çalışması.....	70

Şekil 6.7.	Deprem tahmin algoritmasının test edildiği çalışma bölgesi	71
Şekil 7.1.	Türkiye deprem portalında kullanılan kümeleme ara yüzü.....	76
Şekil 7.2.	Kümeleme için örnek bir gösterim	77
Şekil 7.3.	Kümelerdeki yıllık deprem dağılım grafiği.....	77
Şekil 7.4.	Kümelerin yıllık ortalama deprem büyüklük grafiği	77
Şekil 7.5.	Kümeleme işlemlerinin akış diyagramı	79
Şekil 7.6.	01.01.1900–01.01.2008 tarih aralığında $k=3$, $M_c \geq 4$ için kümeleme sonuçları	80
Şekil 7.7.	01.01.1900–01.01.2008, $k=3$, $M_c \geq 4$ için oluşturulan kümelerdeki yıllık deprem dağılımı	81
Şekil 7.8.	01.01.1900–01.01.2008, $k=3$, $M_c \geq 4$ için oluşturulan kümelerdeki yıllık ortalama deprem büyüklüğü	81
Şekil 7.9.	01.01.1900–01.01.2008 tarih aralığında $k=4$, $M_c \geq 4$ için kümeleme sonuçları	82
Şekil 7.10.	01.01.1900–01.01.2008, $k=4$, $M_c \geq 4$ için oluşturulan kümelerdeki yıllık deprem dağılımı	82
Şekil 7.11.	01.01.1900–01.01.2008, $k=4$, $M_c \geq 4$ için oluşturulan kümelerdeki yıllık ortalama deprem büyüklükleri	83
Şekil 7.12.	Rasgele olmayan başlangıç küme merkezli k-means algoritması	84
Şekil 7.13.	01.01.1900–01.01.2008 tarih aralığında $k=4$ ve $M_c \geq 4$ deprem verilerinin NrK-means ve k-means ile kümelendirilmesi.....	85
Şekil 7.14.	K-means ve NrK-means algoritmalarının 1263 deprem verisini kümelemedeki hız ölçümleri	86
Şekil 7.15.	K-means ve NrK-means algoritmalarının 3119 deprem verisini kümelemedeki hız ölçümleri	86
Şekil 7.16.	K-Means - KNN ile geliştirilen kümeleme yöntemin çalışma yapısı.....	87
Şekil 7.17.	K-Means - KNN kümeleme algoritması	89
Şekil 7.18.	01.01.1900-01.01.2008 tarih aralığında Türkiye deprem kataloğunda yer alan $M_c \geq 7.0$ büyüklüğündeki depremler	90
Şekil 7.19.	01.01.1900-01.01.2008 tarihleri arasında $k=3$ için $M_c \geq 7.0$ büyüklüğündeki depremlerin k-Means ile oluşturduğu kümeler	90
Şekil 7.20.	01.01.1900-01.01.2008 tarih aralığında $M_c \geq 4.0$ büyüklüğündeki depremlerin	

k-means ile önceden oluşan kümelere KNN algoritması ile dahil edilmesi	91
Şekil 7.21. 01.01.1900-01.01.2008 tarih aralığında $k=3$ için $Mc \geq 4.0$ büyüklüğündeki depremlerin K-Means-KNN algoritması ile oluşan kümelere ait grafikler.	92
Şekil 7.22. 01.01.1900-01.01.2008 tarih aralığında $k=4$ için $Mc \geq 7.0$ depremlerin k-Means ile kümelendirilmesi.....	92
Şekil 7.23. 01.01.1900-01.01.2008 tarih aralığında $k=4$ için $Mc \geq 4.0$ büyüklüğündeki depremlerin KNN ile kümelere dahil edilmesi	93
Şekil 7.24. 01.01.1900-01.01.2008 tarih aralığında $k=4$ için $Mc \geq 4.0$ büyüklüğündeki depremlerin K- Means ve KNN algoritmaları ile oluşan kümelere ait grafikler	94
Şekil 7.25. k_0 kümesindeki depremlerin yıllık sayı dağılım grafiği.....	94
Şekil 7.26. 01.01.1900–01.01.2008 tarih aralığında $k=4$ için $Mc \geq 5.0$ büyüklüğündeki depremlerin kümelendirilmesine ait sonuçlar.....	95
Şekil 7.27. 01.01.1900-01.01.2008 tarih aralığında $k=4$ için $Mc \geq 5.0$ büyüklüğündeki depremlerin K- Means ve KNN algoritmaları ile oluşan kümelere ait grafikler	96
Şekil 7.28. 01.01.1900–01.01.2008 tarih aralığında $k=4$ için $Mc \geq 5.0$ büyüklüğündeki depremlerin yıllık sayı dağılımlarının her bir küme için ayrı ayrı gösterimi	97
Şekil 7.29. Kümeleme sonuçlarının karşılaştırılması için küme sonuçlarının gösterimi	99

TABLÖLER LİSTESİ

	<u>Sayfa No</u>
Tablo 3.1. Türkiye Deprem Kataloğundaki referans kısaltmaları.....	3
Tablo 3.2. Türkiye Deprem Kataloğunda yer alan magnitüd bilgileri.....	3
Tablo 5.1. Deprem verilerinin boolean tipte gösterimi.....	55
Tablo 5.2. 01.01.1900–01.01.1984 aralığında çalışma bölgesinde bulunan kuralların bir kısmı.....	58
Tablo 5.3. Üretilen kuralların test sonuçları.....	59
Tablo 6.1. Kural üretimi için deprem verilerinin düzenlenmesi	66
Tablo 6.2. Tahmin bölgesi üzerinde üretilen kuralların bir bölümü.....	66
Tablo 6.3. 01.01.1985–01.01.1995 test aralığında $mc \geq 4$ büyüklüğündeki depremler için tahmin sonuçları.....	71
Tablo 6.4. 01.01.1995–01.01.2000 test aralığında $mc \geq 4$ büyüklüğündeki depremler için tahmin sonuçları.....	72
Tablo 6.5. 01.01.1985–01.01.2000 test aralığında $mc \geq 3$ büyüklüğündeki depremler için tahmin sonuçları.....	73

KISALTMALAR LİSTESİ

Mc, M	: Deprem Büyüklüğü
Conf.	: Güven metriği
Supp.	: Destek metriği
Minconf.	: Minimum güven metriği
Minsup.	: Minimum destek metriği
TDEP	: Türkiye Deprem Portalı

1. GİRİŞ

1.1 Genel

Türkiye, deprem kuşağı üzerinde yer alan bir bölge konumunda olmasından dolayı tarih boyunca büyük depremlere maruz kalmıştır. Türkiye’de, yakın tarihte meydana gelen ve büyük hasarlara yol açan bazı önemli depremler bulunmaktadır. Bunlardan 13.03.1992 tarihinde gerçekleşen ve büyüklüğü 6.8 M olan Erzincan depremi 653 kişinin ölümü ve 3.850 kişinin yaralanmasına neden olmuştur. 17.08.1999 tarihinde meydana gelen ve büyüklüğü 7.4 M olarak belirlenen Düzce depremi 17.127 kişinin ölümü ve 43.953 kişinin yaralanması ile sonuçlanarak Türkiye tarihine büyük bir felaket olarak geçmiştir [1]. Yakın tarihte gerçekleşen ve 42 kişinin ölümüne yol açan 6.0 büyüklüğündeki 08.03.2010 Elazığ/Karakoçan depremi, ülkemizde depremin büyük hasarlara neden olabilecek bir doğal afet olduğunu tekrar ortaya koymuş ve bu alanda yapılacak çalışmaların ne derecede önemli olduğunu göstermiştir [2].

Deprem en fazla korkulan doğal afettir. Çünkü deprem sel, volkanik patlama ve hortum gibi önceden bir uyarı vermez [3]. Depremlerin maddi ve manevi yıkımlarının boyutları düşünüldüğünde bu alandaki tahmin çalışmalarının önemi ortaya çıkmaktadır. Güvenilir tahmin araştırmaları için çok sayıda çalışma yapılmış ve büyük ölçekte bütçeler ayrılmıştır. Bu alandaki birçok tahmin çalışması ağırlıklı olarak deprem hareketlerinin geçmişte depremlerin olduğu bölgelerde veya yakınlarında oluşma olasılığı üzerine yapılmıştır [4].

Deprem tahminleri için çeşitli yaklaşımlar sunulmuştur[5,6]. Genel olarak bu yaklaşımlar iki sınıfta incelenebilir. Birinci sınıf, öncü yeryüzü hareketlerinin gözlemlenmesi ile tecrübeye dayalı yaklaşımlardır. Bu sınıf altındaki çalışmalarda öncü yeryüzü hareketlilikleri gözlemlenerek gelecekte oluşabilecek depremlerin tahmini gerçekleştirilir. İkinci sınıf, depremlerin istatistiksel örüntüleri üzerine kurulan yaklaşımlardan oluşur. Bu yaklaşımda ise deprem verileri üzerine çeşitli istatistiksel metotlar uygulanarak gelecek ile ilgili tahminler gerçekleştirilir. İkinci sınıftaki yaklaşıma örnek olarak Pattern Informatics metodu gösterilebilir [3,7]. Bu yöntemde, geçmişteki deprem hareketlerinin anormal davranışları analiz edilerek gelecekte oluşabilecek depremler hakkında bilgiler verilir. Bu metot, gelecekte oluşacak depremlerin tam olarak büyüklük ve zamanları hakkında kesin bir bilgi vermez. Sadece belirlenen zaman aralığında ve yine belirlenen büyüklükteki depremlerin oluşabileceği sıcak bölgeler

hakkında bilgi verir. Bu metodun Güney Kaliforniya ve Japonya bölgesi üzerindeki uygulamasının başarılı olduğu gözlemlenmiştir [8]. Benzer şekilde istatistiksel alt yapıya dayanan bir diğer metot Hücrel Depremsellik (Cellular Seismology) metodudur. Tamamen istatistiksel olan bu yöntemde, tahmin bölgeleri için iki farklı katalog verisi kullanılır. Bu kataloglardan birincisi, belirli bir zaman aralığındaki geçmiş deprem verilerinden meydana gelir. Diğer katalog ise daha ileri tarihli deprem verilerinden oluşur. Sonraki depremlerin, geçmiş deprem alanları için belirlenen yarıçaptaki bölge üzerinde meydana gelmesinin gözlemlenmesi ile bu metot istatistiksel bir sonuç üretir [4].

Bu yaklaşımlardan hiç biri kısa süreli (ay gün) tutarlı tahminler sağlayamamıştır [7]. Bu yaklaşımların temel aldığı düşünce, genel olarak gelecekte oluşabilecek depremlerin geçmişte depremlerin oluşmuş olduğu bölgelerde veya yakınlarında olacağı yönündedir [9].

Deprem verileri üzerinde istatistiksel alt yapıya dayanan çalışmaların yanında, veri madenciliği yöntemleri de bu alanda kullanılan etkili uygulamalar arasında yer almaktadır. Veri madenciliği yöntemleri ile coğrafi veri tabanlarında tutulan bilgiler üzerinde geleceğe yönelik tahmin çalışmaları özellikle deprem verileri üzerinde sıklıkla kullanılmıştır [10,11,12]. Veri madenciliği yöntemlerinden sınıflama ve kümeleme [13,14], yapay sinir ağları [12,15,16], bulanık mantık [17] ve istatistiksel alt yapıya dayanan temel yöntemler [18, 19, 20] deprem verilerine etkili bir şekilde uygulanmıştır.

Deprem verilerinin görsel olarak sorgulanması ve bu veriler üzerinde bilimsel uygulamaların gerçekleştirilmesi için tasarlanan bazı sistemler mevcuttur. Bu sistemlere örnek olarak NASA tarafından desteklenen QuakeSim projesi gösterilebilir. Bu proje, depremler ve yeryüzü hareketlerinin anlaşılması için gereken bilimsel uygulamalar, dağıtık hesaplama altyapısı, veri tabanları ve veri yönetimi sistemlerini içeren bir çalışmadır [21].

1.2. Tezin Amacı

Bu tez çalışmasının genel amacı, deprem verileri üzerinde bazı veri madenciliği tekniklerinin uygulanması ve bu verilere ait birtakım istatistiksel çalışmaların gerçekleştirilmesidir. Deprem verileri üzerine uygulanan bu yöntemler yardımı ile bölgelerin deprem karakteristiklerinin tespit edilmesi ve gelecek deprem hareketlilikleri hakkında çeşitli tahminlerin yapılması amaçlanmaktadır.

Bu çalışmada, deprem verileri üzerinde veri madenciliği tekniklerinin uygulanması ve istatistiksel çalışmaların yapılması için, web tabanlı bir sistem olan Türkiye Deprem Portalı(TDEP) geliştirilmesi hedeflenmiştir. Hazırlanan TDEP sistemi içerisinde, deprem verileri üzerinde çeşitli işlemler gerçekleştiren bir takım alt algoritmaların yer alması ve bu algoritmalar aracılığı ile hem veri madenciliği tekniklerinin hem de istatistiksel çalışmaların başarılı bir şekilde deprem verileri üzerinde uygulanması beklenmektedir.

Tez çalışmasının amaçlarından biri, deprem alanları arasındaki hareketlilik ilişkilerinin belirlenmesidir. Bu amaç kapsamında, veri madenciliği yöntemlerinden birliktelik kuralları deprem verileri üzerine başarılı bir şekilde uygulanmıştır. Üretilen kurallar, hareketlilik bakımından birbirleri ile ilişkili bölgelerin tespit edilmesinde ve deprem bakımından riskli alanların belirlenmesinde etkili bir şekilde kullanılmıştır. Kural çıkarım algoritmalarından Apriori algoritması, deprem verileri üzerinde çalıştırılarak TDEP yazılımına entegre edilmiştir. TDEP arayüzü sayesinde tüm kullanıcıların, karmaşık olan bu algoritmayı kolay bir şekilde kullanması ve deprem verileri üzerinde kendi belirledikleri parametreler ile uygulayabilmeleri amaçlanmıştır.

Tez çalışmasının bir diğer amacı, geleceğe yönelik bölgeler bazında deprem tahminleri gerçekleştirmektir. Bu amaç doğrultusunda, kural tabanlı bir tahmin algoritmasının tasarlanması amaçlanmıştır. Tasarlanan bu algoritmanın, geçmiş deprem verileri ile eğitilmesi ve bölgeler arasındaki birlikteliklere göre gizli örüntüleri tespit etmesi ve daha sonra, bu örüntüler aracılığı ile gerçek zamanlı oluşan deprem hareketlerine göre gelecekte deprem oluşabilecek diğer bölgelerin uyarısını yapması beklenmektedir. Hazırlanan bu sistemin, deprem tahmini gerçekleştirebilecek yeni bir yaklaşım olarak tezde sunulması hedeflenmiştir.

Son olarak tezin amaçlarından bir diğeri, Türkiye genelindeki deprem verilerinin oluştukları konumlar baz alınarak belirli kriterlere göre kümelendirilmesidir. Elde edilen kümeler yardımı ile deprem hareketliliği bakımından benzer davranışlar gösteren olayların aynı kümelerde yer alması ve bu veriler üzerine yapılacak çalışmaların daha etkili olması beklenmektedir. Kümeleme işlemlerinde sıklıkla kullanılan k-Means algoritmasının gerçek konum verisine göre çalışacak biçimde kodlanarak TDEP yazılımına eklenmesi ve kullanıma uygun hale getirilmesi amaçlanmıştır. Ayrıca k-Means algoritmasının performans problemlerinden biri olan, başlangıç küme merkezleri atamasına alternatif yeni bir yaklaşımın tez kapsamında sunulması hedeflenmiştir. Bu yaklaşımda, başlangıç küme merkezleri rastgele değil geçmişteki büyük ölçekli deprem verileri baz alınarak atanmıştır.

Performans olarak yeni yaklaşımın klasik k-Means algoritmasına oranla büyük ölçekte başarı sağlandığı tespit edilmiştir.

1.3 Tezin Organizasyonu ve Katkılar

Tezin giriş bölümünde, teze genel bir bakış açısı kazandırmaya yönelik temel bilgiler verilmiştir. Tezde yer alan diğer bölümlerin organizasyonu aşağıda sunulmuştur.

Bölüm 2'de, veri madenciliği özetlenerek, veri madenciliğinde uygulanan teknikler ve uygulama alanları hakkında genel bilgiler verilmiştir. Veri madenciliği kapsamında bulunan bazı algoritmalar ve çeşitli veri madenciliği modellerinden bahsedilmiştir.

Bölüm 3'te, deprem verilerinin dünya genelinde tutulduğu çeşitli kataloglar ve bu katalogların veriyi hangi formatlarda sakladıkları hakkında bilgiler sunulmuştur. Ayrıca tez çalışmasında kullanılan Türkiye Deprem Kataloğu ile ilgili detaylı bilgilere bu bölümde yer verilmiştir.

Bölüm 4'te, deprem verileri üzerinde gerçekleştirilecek veri madenciliği ve istatistiksel çalışmalar için tasarlanan Türkiye Deprem Portalı (TDEP) uygulamasından bahsedilmiştir. TDEP bünyesinde yer alan teknolojiler ve bu sistemin çalışma mimarisi hakkında detaylı bilgi sunulmuştur. Ayrıca bu bölümde, deprem verileri üzerine gerçekleştirilebilecek bazı istatistiksel işlemler ve sorgulama işlemleri için hazırlanan menüler hakkında bilgi verilmiştir. Türkiye genelinde ve seçimi yapılan bazı bölgeler üzerinde gerçekleştirilen istatistiksel işlemler ve sorgulama çalışmalarına bu bölümde yer verilmiştir. Elde edilen sonuçlar sunularak tartışılmıştır.

Bölüm 5'te, veri madenciliği yöntemlerinden birliktelik kurallarının deprem verileri üzerinde uygulanma aşamaları ve bu uygulamanın sonuçları tartışılmıştır. Deprem bölgeleri arasındaki hareketlilik ilişkileri ve deprem bakımından en fazla risk oluşturan alanların tespiti bu bölüm altında yapılmıştır. Kural çıkarım algoritmalarından apriori algoritması, bu bölümde deprem verileri üzerinde çalışacak şekilde geliştirilmiştir. TDEP yazılımı içerisindeki web arayüzü aracılığı ile algoritmanın çalışacağı bölge ve çeşitli parametrelerin seçim işlemleri kolaylaştırılmıştır. Algoritmanın çalıştırılması sonucunda

elde edilen kurallar, bölgelerin deprem karakteristiği hakkında önemli bilgiler sunmaktadır. Bunun yanında, bölgeler arasındaki saklı örüntülerin keşfedilmesi, bu bölgelerin deprem karakteristiği hakkında daha geniş bilgilere ulaşılmasına imkan tanımıştır. Ayrıca bu çalışmada, kural çıkarımı on-line olarak gerçekleştirilmekte ve üretilen sonuçlar farklı formatlarda kullanıcılara sunulmaktadır. Bu özellikler, geliştirilen sistemin avantajları içerisinde yer almaktadır.

Bölüm 6'da, deprem tahminine yönelik kural tabanlı bir algoritma sunulmuştur. Geliştirilen bu algoritma, deprem bölgeleri arasındaki birliktelikleri kullanarak gelecekte deprem oluşabilecek bölgelerle ilgili tahminler üretmektedir. Algoritma öncelikle tahmin bölgesi üzerindeki gizli örüntüleri keşfetmektedir. Keşfedilen bu örüntüler, kural veri tabanında saklanmaktadır. Daha sonra, gerçek zamanlı olarak gelen deprem verilerinin oluştuğu bölgeleri tespit ederek, bu bölgelerle ilişkili ve deprem meydana gelebilecek bölgelerin önceden uyarısını yapmaktadır. Algoritmanın, elde edilen sonuçlar değerlendirildiğinde uygun parametreler paralelinde başarılı tahmin sonuçları ürettiği gözlemlenmiştir. Literatürde çeşitli istatistiksel tahmin çalışmaları bulunmaktadır. Bu yaklaşım veri madenciliği yöntemlerini kullanarak tahmin gerçekleştiren farklı bir model olarak sunulmuştur.

Bölüm 7, Türkiye genelinde oluşan geçmiş deprem verilerinin kümelendirilmesi işlemlerini kapsamaktadır. Depremlerin kümelere ayrılmasındaki amaç, benzer davranış gösteren hareketliliklerin aynı küme içerisinde toplanması ve değerlendirilmesidir. Kümeleme işlemleri için k-Means ve KNN algoritmaları kullanılmıştır. Bu algoritmalarındaki mesafe hesaplanmasında, depremlerin meydana geldiği enlem ve boylam bilgisine dayanan gerçek konum verileri kullanılmıştır. K-Means algoritması ile elde edilen kümeleme sonuçlarının iyileştirilmesine yönelik olarak, k-Means ve KNN algoritmalarının ortak kullanıldığı hibrit bir algoritma bu bölüm altında geliştirilmiştir. Her iki algoritma ile elde edilen küme sonuçları görsel bir şekilde sunulmuştur.

Bölüm 8'de, tezin sonuçları irdelenmiş ve tez kapsamında yapılan katkılar vurgulanmıştır.

2. VERİ MADENCİLİĞİ

Teknolojideki hızlı gelişmeler paralelinde, veriyi elde etmek ve saklamak için kullanılan donanımların yaygınlaşması, bu teknolojilerin günlük hayatımızda hemen hemen her yerde kullanımını mümkün kılmıştır. Toplanan veri miktarı büyüdükçe ve toplanan verilerdeki karmaşıklık arttıkça, daha iyi çözümleme tekniklerine olan gereksinim de artmaktadır [22]. Ancak üretilen verilerin artması tek başına bir anlam ifade etmez çünkü bu veriler ancak belli bir amaç doğrultusunda işlendiği zaman anlamlı hale gelmektedir. Veri madenciliği, bu büyük miktardaki veriler arasından gelecek ile ilgili tahminde bulunabilmemizi sağlayabilecek bağıntıların bilgisayar programlarıyla aranması ve bilgiye ulaşma işidir. Yani genel olarak veri madenciliği veriler içerisinden tutarlı tahminler yapmak için çeşitli veri analiz araçlarını kullanarak örüntü ve ilişki çıkarma işlemidir [23]. Veri madenciliği, istatistiği de kapsayan daha geniş bir model olarak görülebilir. VTBK(Veri Tabanlarından Bilgi Keşfi)'nin amacı, istatistikçilerin hipotez seçme yeteneğini de içerecek şekilde bütün veri analizi işlemlerini mümkün olduğunca otomatikleştiren araçlar geliştirmektir [24].

Veri madenciliği Biyotıp, Gen fonksiyonları ve DNA sıralama desenlerinin veri analizlerinde, hastalık tanısında, telekomünikasyon endüstrisinde, finans analizi, astronomi ve birçok alanda uygulanmaktadır [25].

Bununla birlikte günümüzde veri madenciliği teknikleri özellikle işletmelerde çeşitli alanlarda başarı ile kullanılmaktadır. Bu uygulamalardan başlıcaları ilgi alanlarına göre aşağıda özetlenmiştir [26].

Pazarlama

- Müşterilerin satın alma örüntülerinin belirlenmesi,
- Müşterilerin demografik özellikleri arasındaki bağlantıların bulunması,
- Posta kampanyalarında cevap verme oranının artırılması,
- Mevcut müşterilerin elde tutulması, yeni müşterilerin kazanılması,
- Pazar sepeti analizi,
- Müşteri ilişkileri yönetimi,
- Müşteri değerlendirme,
- Satış öngörüsü.

Bankacılık

- Farklı finansal göstergeler arasında gizli korelasyonların bulunması,

- Kredi kartı dolandırıcılıklarının tespiti,
- Kredi kartı harcamalarına göre müşteri gruplarının belirlenmesi,
- Kredi taleplerinin değerlendirilmesi.

Sigortacılık

- Yeni poliçe talep edecek müşteriler ile ilgili tahmin yapılması,
- Sigorta dolandırıcılıklarının tespiti,
- Riskli müşteri örüntülerinin belirlenmesi.

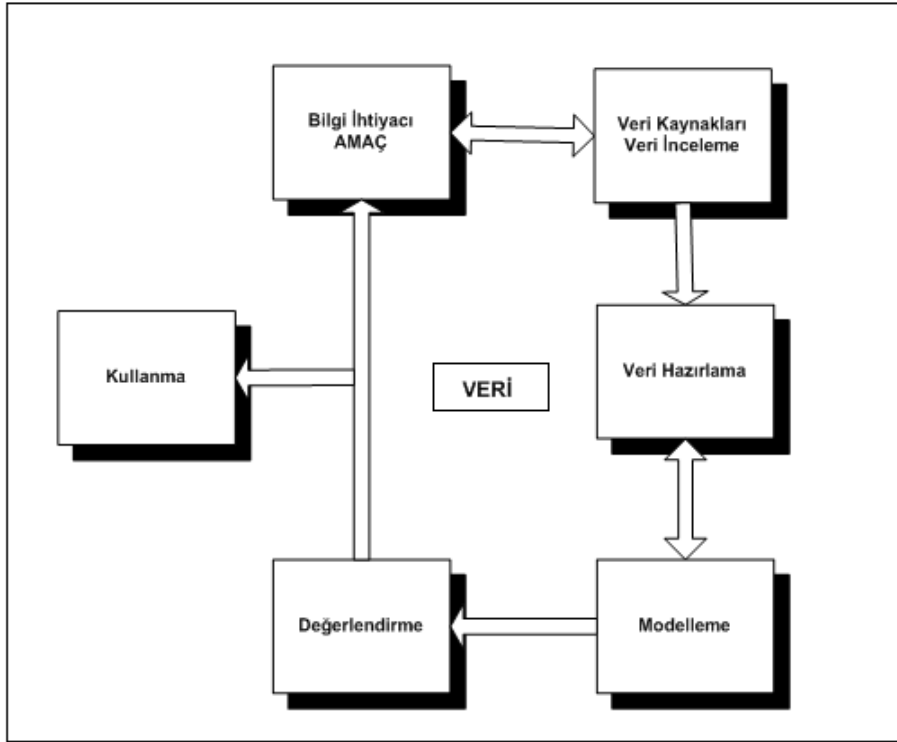
Veri tabanlarındaki veriden anlamlı bilgiler çıkarılırken öncelikle veri yığınları arasından gereken veriler alınır, sınıflandırılır ve daha sonrada işlenir [27]. Veri madenciliğinde sınıflama metotları, büyük verilerde ilginç nesnelerin otomatik tanımlanmasında ve mali piyasalardaki eğilimin sınıflamalarını içeren uygulamalarda bilgi keşfi için kullanılır [28]. Bu verileri sınıflandırırken kullanılan karar ağaçları, birliktelik kuralları, genetik algoritmalar gibi çeşitli yöntemler mevcuttur [29].

2.1. Veri Madenciliği Süreci

Ne kadar etkin olursa olsun hiç bir veri madenciliği algoritmasının, üzerinde inceleme yapılan işin ve verilerin özelliklerinin bilinmemesi durumunda fayda sağlaması mümkün değildir. Bu nedenle yukarıda tanımlanan tüm aşamalardan önce, iş ve veri özelliklerinin öğrenilmesi/anlaşılması başarının ilk şartı olacaktır. Başarılı bir Veri madenciliği projesinde izlenmesi gereken yol aşağıdaki adımlardan oluşur:

- Problemin Tanımlanması,
- Verilerin Hazırlanması,
- Modelin Kurulması ve değerlendirilmesi,
- Modelin Kullanılması,
- Modelin İzlenmesi

Veri madenciliği tıp ve diğer alanlarda bilgi keşfi sürecinin bir parçası olarak değerlendirilmektedir [30]. Veri madenciliği süreci Şekil 2.1’de gösterilmiştir.



Şekil 2.1. Veri Madenciliği Süreci

2.1.1. Problemin Tanımlanması

Veri madenciliği çalışmalarında başarılı olmanın en önemli şartı, projenin hangi işletme amacı için yapılacağına açık bir şekilde tanımlanmasıdır. İlgili işletme amacı işletme problemi üzerine odaklanmış ve açık bir dille ifade edilmiş olmalı, elde edilecek sonuçların başarı düzeylerinin nasıl ölçüleceği tanımlanmalıdır. Ayrıca yanlış tahminlerde katlanılacak olan maliyetlere ve doğru tahminlerde kazanılacak faydalara ilişkin tahminlere de bu aşamada yer verilmelidir. Bu aşamada mevcut iş probleminin nasıl bir sonuç üretilmesi durumunda çözüleceğinin, üretilecek olan sonucun fayda - maliyet analizinin başka bir deyişle üretilen bilginin işletme için değerinin doğru analiz edilmesi gerekmektedir.

2.1.2. Verilerin Hazırlanması

Veri madenciliğinin en önemli aşamalarından bir tanesi olan verinin hazırlanması aşaması analistin toplam zaman ve enerjisinin %50 - %75 ini harcamasına neden olmaktadır. Bu aşamada firmanın mevcut bilgi sistemleri üzerinde ürettiği sayısal bilginin

iyi analiz edilmesi, veriler ile mevcut iş problemi arasında ilişki olması gerektiği unutulmamalıdır.

Proje kapsamında kullanılacak sayısal verilerin, hangi iş süreçleri ile yaratıldığı da bu veriler kullanılmadan analiz edilmelidir, bu sayede analizci veri kalitesi hakkında fikir sahibi olabilir. Verilerin hazırlanması aşaması kendi içerisinde toplama, birleştirme ve temizleme, dönüştürme adımlarından meydana gelmektedir.

2.1.2.1. Toplama

Tanımlanan problem için gerekli olduğu düşünülen verilerin ve bu verilerin toplanacağı veri kaynaklarının belirlenmesi adıımıdır. Verilerin toplanmasında kuruluşun kendi veri kaynaklarının dışında, nüfus sayımı, hava durumu, merkez bankası kara listesi gibi veri tabanlarından veya veri pazarlayan kuruluşların veri tabanlarından faydalanılabilir.

2.1.2.2. Birleştirme ve Temizleme

Bu adımda toplanan verilerde bulunan farklılıklar giderilmeye çalışılır. Hatalı veya analizin yanlış yönlendirilmesine sebep olabilecek verilerin temizlenmesine çalışılır. Genellikle yanlış veri girişinden veya bir kereye özgü bir olayın gerçekleşmesinden kaynaklanan verilerin, önemli bir uyarıcı enformasyon içerip içermediği kontrol edildikten sonra veri kümesinden atılması tercih edilir. Ancak basit yöntemlerle ve baştan savma olarak yapılacak sorun giderme işlemlerinin, ileriki aşamalarda daha büyük sorunların kaynağı olacağı unutulmamalıdır.

2.1.2.3. Dönüştürme

Kullanılacak model ve algoritma çerçevesinde verilerin tanımlama veya gösterim şeklinin de değiştirilmesi gerekebilir. Örneğin kredi riski uygulamasında iş tiplerinin, gelir seviyesi ve yaş gibi değişkenlerin kodlanarak gruplanmasının faydalı olacağı bilinmelidir.

2.1.3. Modelin Kurulması ve Değerlendirilmesi

Tanımlanan problem için en uygun modelin bulunabilmesi, olabildiğince çok sayıda modelin kurularak denenmesi ile mümkündür. Bu nedenle veri hazırlama ve model kurma aşamaları, en iyi olduğu düşünülen modele varılıncaya kadar yinelenen bir süreçtir.

Model kuruluş süreci, denetimli ve denetimsiz öğrenmenin kullanıldığı modellere göre farklılık göstermektedir. Örnekten öğrenme olarak da isimlendirilen denetimli öğrenmede, bir denetçi tarafından ilgili sınıflar önceden belirlenen bir kritere göre ayrılarak, her sınıf için çeşitli örnekler verilir. Sistemin amacı verilen örneklerden hareket ederek her bir sınıfa ilişkin özelliklerin bulunması ve bu özelliklerin kural cümleleri ile ifade edilmesidir. Öğrenme süreci tamamlandığında, tanımlanan kural cümleleri verilen yeni örneklerle uygulanır ve yeni örneklerin hangi sınıfa ait olduğu kurulan model tarafından belirlenir.

Denetimsiz öğrenmede, kümeleme analizinde olduğu gibi ilgili örneklerin gözlenmesi ve bu örneklerin özellikleri arasındaki benzerliklerden hareket ederek sınıfların tanımlanması amaçlanmaktadır. Denetimli öğrenmede seçilen algoritmaya uygun olarak ilgili veriler hazırlandıktan sonra, ilk aşamada verinin bir kısmı modelin öğrenimi, diğer kısmı ise modelin geçerliliğinin test edilmesi için ayrılır. Modelin öğrenimi, öğrenim kümesi kullanılarak gerçekleştirildikten sonra, test kümesi ile modelin doğruluk derecesi belirlenir. Kurulan modelin değerinin belirlenmesinde kullanılan diğer bir ölçü, model tarafından önerilen uygulamadan elde edilecek kazancın bu uygulamanın gerçekleştirilmesi için katlanılacak maliyete bölünmesi ile edilecek olan yatırımın geri dönüş oranıdır.

Bir modelin doğruluğunun test edilmesinde kullanılan en basit yöntem basit geçerlilik (Simple Validation) testidir. Bu yöntemde tipik olarak verilerin % 5 ile % 33 arasındaki bir kısmı test verileri olarak ayrılır ve kalan kısım üzerinde modelin öğrenimi gerçekleştirildikten sonra, bu veriler üzerinde test işlemi yapılır. Bir sınıflama modelinde yanlış olarak sınıflanan olay sayısının, tüm olay sayısına bölünmesi ile hata oranı, doğru olarak sınıflanan olay sayısının tüm olay sayısına bölünmesi ile ise doğruluk oranı hesaplanır. (Doğruluk Oranı = 1 – Hata Oranı)

Önemli diğer bir değerlendirme kriteri modelin anlaşılabilirliğidir. Bazı uygulamalarda doğruluk oranlarındaki küçük artışlar çok önemli olsa da, birçok işletme uygulamasında ilgili kararın niçin verildiğinin yorumlanabilmesi, çok daha büyük önem taşıyabilir. Çok ender olarak yorumlanamayacak kadar karmaşıklarsalar da, genel olarak karar ağacı ve kural temelli sistemler model tahmininin altında yatan nedenleri çok iyi ortaya koyabilmektedir.

2.1.4. Modelin Kullanılması

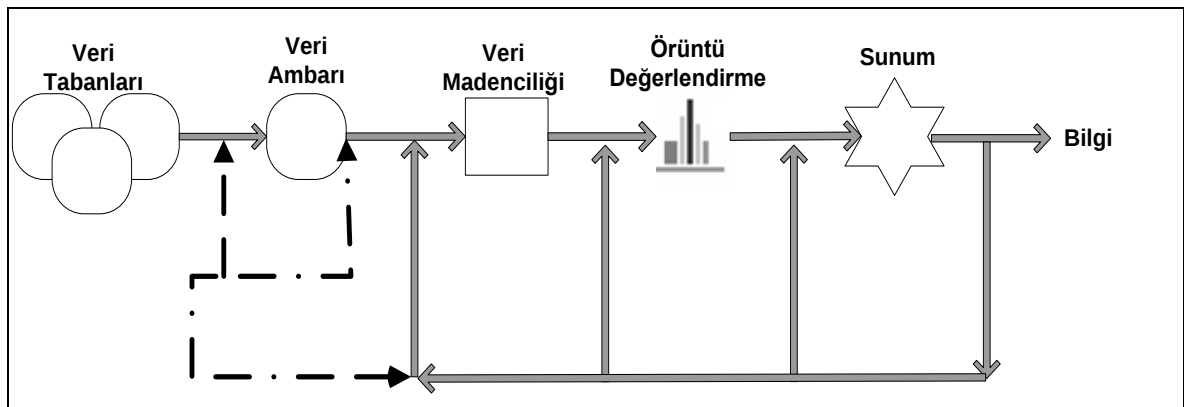
Kurulan ve geçerliliği kabul edilen model doğrudan bir uygulama olabileceği gibi, bir başka uygulamanın alt parçası olarak kullanılabilir. Kurulan modeller risk analizi, kredi değerlendirme, dolandırıcılık tespiti gibi işletme uygulamalarında doğrudan kullanılabilirliği gibi, promosyon planlaması simülasyonuna entegre edilebilir veya tahmin edilen envanter düzeyleri yeniden sipariş noktasının altına düştüğünde, otomatik olarak sipariş verilmesini sağlayacak bir uygulamanın içine gömülebilir.

2.1.5. Modelin İzlenmesi

Zaman içerisinde bütün sistemlerin özelliklerinde ve dolayısıyla ürettikleri verilerde ortaya çıkan değişiklikler, kurulan modellerin sürekli olarak izlenmesini ve ihtiyaç duyuluyorsa yeniden düzenlenmesini gerektirecektir. Tahmin edilen ve gözlenen değişkenler arasındaki farklılığı gösteren grafikler, model sonuçlarının izlenmesinde kullanılan yararlı bir yöntemdir [31].

2.2 Veritabanlarından Bilgi Keşfi Süreci

Veri madenciliği veritabanlarından bilgi keşfi (VTBK) işleminin temel bileşenlerinden birisini oluşturmakla beraber VTBK veri madenciliği dışında kendi içerisinde farklı süreçleri barındırmaktadır. Genel olarak veri tabanından bilgi keşfi işlemi Şekil 2.2’de görüldüğü gibi 5 aşamadan oluşur [32].



Şekil 2.2. Veri tabanından bilgi keşfi aşamaları

VTBK sürecini oluşturan adımlar aşağıdaki şekilde sıralanmaktadır:

a) *Veri Önışlemleri :*

Bu aşamada veri içerisindeki gürültüler, tutarsızlıklar ve düzensizlikler giderilir. Veri tabanında eksik veri varsa bu veriler düzenlenir. Bu işleme veri temizleme (Data cleaning) denir. İkinci aşamada tutarsızlıklar ve gürültüden arındırılmış veriler birleştirilir. Bu işleme ver birleştirme (Data integration) denir. Bu aşamada farklı kaynaklardan toplanan verilerin tek bir veri ambarında toplanabilmesi için gerekli genellemeler ve normalizasyon işlemleri yapılır.

b) *Veri seçme ve Dönüştürme:*

Bu aşamada, veri madenciliği sürecinde kullanılacak veriler ile ilgili bazı ön işlemler yapılır. Bu önışlemler genel olarak veri madenciliği konusu ile ilgili bilgi seçimi, madencilik yapılacak verinin türünün belirlenmesi, veriler arasında hiyerarşik yapının ve genellemelerin belirlenmesi ve veri madenciliği sonunda bulunacak veriler için görselleştirme ve sunum araçlarının belirlenmesi şeklindedir.

c) *Veri madenciliği:*

Anlamlı veri örüntülerini ve veriler arasındaki saklı ilişkileri bulmak için çeşitli analiz yöntemleri ve algoritmaların kullanıldığı aşamadır.

d) *Örüntü Değerlendirme:*

İkinci aşamada belirlenen ölçümlere göre elde edilmiş bilgiyi temsil eden ilginç örüntüleri tanımlamak ve bu örüntülerin ilginçlik seviyelerinin tespit edilmesi aşamasıdır.

e) *Bilgi Sunumu:*

Çeşitli görselleştirme ve raporlama araçları kullanılarak elde edilen verilerin kullanıcılara ulaştırılması aşamasıdır.

Veritabanından Bilgi Keşfi (VTBK) süreci defalarca tekrar ve aşamalar arası atlamalar ve bu aşamalar arasında ileri geri hareketler içerebilmektedir. Günümüzde çoğunlukla veri

madenciliği aşamasına odaklanılmaktaysa da, diğer tüm aşamalar VTBK işleminin bütünlüğü açısından en az veri madenciliği kadar önem taşımaktadır [33].

2.3 Veri Madenciliğinde Kullanılan Modeller

Veri madenciliği modelleri, tahmin edici (predictive) ve tanımlayıcı (descriptive) olmak üzere iki kısımda incelenmektedir.

Tahmin edici modeller, sonuçları bilinen verilerden hareket ederek bir model geliştirir ve kurulan bu modelden yararlanarak sonuçları bilinmeyen veri setleri için sonuç değerlerini tahmin etmeyi hedefler. Örneğin bir banka önceki dönemlerde vermiş olduğu kredilere ilişkin gerekli tüm verilere sahip olabilir. Bu verilerde bağımsız değişkenler kredi alan müşterinin özellikleri, bağımlı değişken ise kredinin geri ödenip ödenmediğidir. Bu verilere uygun olarak kurulan model, daha sonraki kredi taleplerinde müşteri özelliklerine göre verilecek olan kredinin geri ödenip ödenmeyeceğinin tahmininde kullanılmaktadır.

Tanımlayıcı modellerde ise karar vermeye öncülük etmede kullanılabilecek eldeki mevcut verilerdeki örüntülerin tanımlanması sağlanmaktadır. X/Y aralığında geliri, evi ve arabası olan, ayrıca çocukları okul çağında olan aileler ile, çocuğu olmayan ve geliri X/Y aralığından düşük olan ailelerin satın alma örüntülerinin birbirlerine benzerlik gösterdiğinin belirlenmesi tanımlayıcı modellere bir örnektir.

Veri Madenciliği modelleri gördükleri işlevlere göre:

- Sınıflama (Classification) ve Regresyon,
- Kümeleme (Clustering),
- Birliktelik Kuralları(Association Rules) ve Ardışık Zamanlı Örüntüler(Sequential Patterns)

olmak üzere üç ana başlık altında incelemek mümkündür. Sınıflama ve regresyon modelleri tahmin edici, kümeleme, birliktelik kuralları ve ardışık zamanlı örüntü modelleri tanımlayıcı modellerdir.

Sınıflama ve Regresyon modelleri arasındaki temel fark, tahmin edilen bağımlı değişkenin kategorik veya süreklilik gösteren bir değere sahip olmasıdır. Bazı tekniklerde her iki modelin giderek birbirine yaklaşmasından dolayı aynı tekniklerden yararlanılması mümkün olmuştur [34].

2.4 Veri Madenciliğinde Kullanılan Yöntemler

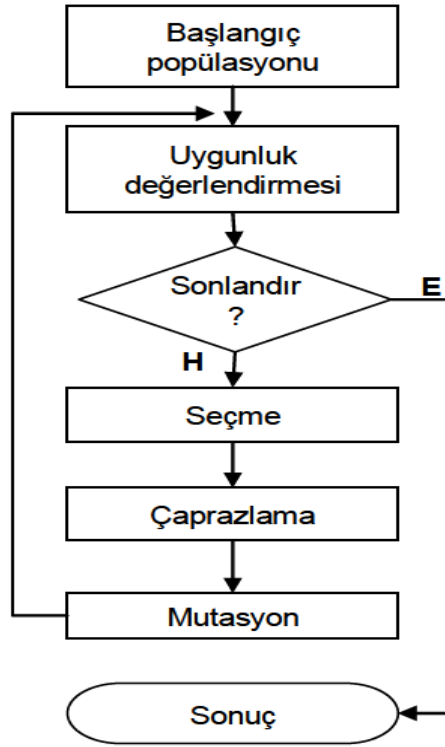
Veri madenciliği algoritmaları genel olarak iki gruba ayrılabilir. Bunlar doğrulamaya ve keşfe dayalı algoritmalarlardır. Doğrulamaya dayalı VM algoritmasında, kullanıcı bir hipotez öne sürer ve sistem bu hipotezi ispatlamaya çalışır. Keşfe dayalı algoritmalar şöyle sıralanabilir:

- Sınıflandırma algoritması: Yeni verileri, belli nitelik değerlerine göre, daha önceden belirlenmiş sınıflara ayırmaya yarar.
- Kümeleme algoritması: Verileri belirlenen niteliklere göre kümeler. Her bir kümede yer alan nesneler dahil oldukları grubu diğer gruplardan ayıran ortak özelliklere sahiptir.
- İlişkilendirme algoritması: Bir veri tabanındaki nesnelerin nitelikleri arasındaki ilişkileri bulur. Bu, bir örüntüde hangi niteliklerin var olmasıyla, başka hangi niteliklerin de onlarla birlikte aynı örüntüde bulunacağını belirler.
- Ardışık örüntüler: Birbirleri ile ilişkisi olan ancak birbirini izleyen dönemlerde gerçekleşen ilişkilerin tanımlanmasında kullanılır.
- Zaman serileri arası bağımlılıklar: Bir veri kümesinde, belirli nitelik değerlerinin belirli zaman aralıklarıyla tekrarlanmasıyla ortaya çıkan örüntüleri bulmak için kullanılır [24].

2.4.1 Genetik Algoritmalar

Genetik algoritmalar öğrenme işleminde evrime dayalı bir yaklaşım kullanan, arama ve optimizasyon problemlerinin çözümünde başarıyla kullanılan ve bilgisayar bilimlerinde pek çok alanda pratik çözümler üretebilen en önemli tekniklerdendir [35]. Genetik Algoritmalar stokastik bir arama yöntemidir. GA, Darwin'in en iyi olan yaşar prensibine dayalı olarak biyolojik sistemlerin gelişim sürecini modeller. Mühendislik ve bilgisayar bilimlerindeki zorlu problemlerin çözümünde yaygın olarak kullanılır [36]. GA tekniği, çözüm uzayının büyüklüğüne rağmen iyi bir çözüme kısa zamanda yakınsamaktadır. Genetik algoritmalarda canlı organizmalardan esinlenilmiştir. Canlı toplumlar nesiller

boyunca, ilk defa Charles Darwin'in "Türlerin Kökeni" adlı kitabında açıkça kaleme aldığı şekilde, doğal seçim süreci ve güçlülerin hayatta kalması prensiplerine göre gelişmişlerdir [37]. Genetik algoritmalar açıklanabilir sonuçlar üretirler. Değişik tipteki verileri işleme özelliğine sahip olan genetik algoritmalar en iyileme (optimizasyon) amacı ile kullanılabilirler. Ayrıca genetik algoritmalar yapay sinir ağları ile ortaklaşa çalışarak başarılı sonuçlar üretmektedir [34]. Genel bir Genetik Algoritma süreci Şekil 2.3'de gösterilmiştir.



Şekil 2.3. Genetik algoritma süreçleri

İlk olarak bireylerin rastgele üretildiği bir başlangıç popülasyonu oluşturulur. Sonraki adımda popülasyon içindeki her bir bireyin belirlenen problem için uygunluk değeri hesaplanır. Uygunluk fonksiyonu probleme özgüdür ve genetik algoritmanın beynini oluşturur. Uygunluk değeri hesaplanan bireyler çaprazlanmak için önceden belirlenen seçme yöntemi ile seçilirler. Seçilen bu bireyler arasında çaprazlama işlemi uygulanır. Çaprazlamadaki amaç baskın bireylerin genlerini bir sonraki kuşağa aktarmasıdır. Çaprazlamadan sonra, birey tekrarını önlemek için mutasyon işlemi uygulanır. Bu işlemde belirlenen mutasyon oranına göre bireylerde belirli sayıda genler mutasyona uğrattılır. Bu adımlar jenerasyon sayısı kadar tekrarlanarak sonuca ulaşılır [38].

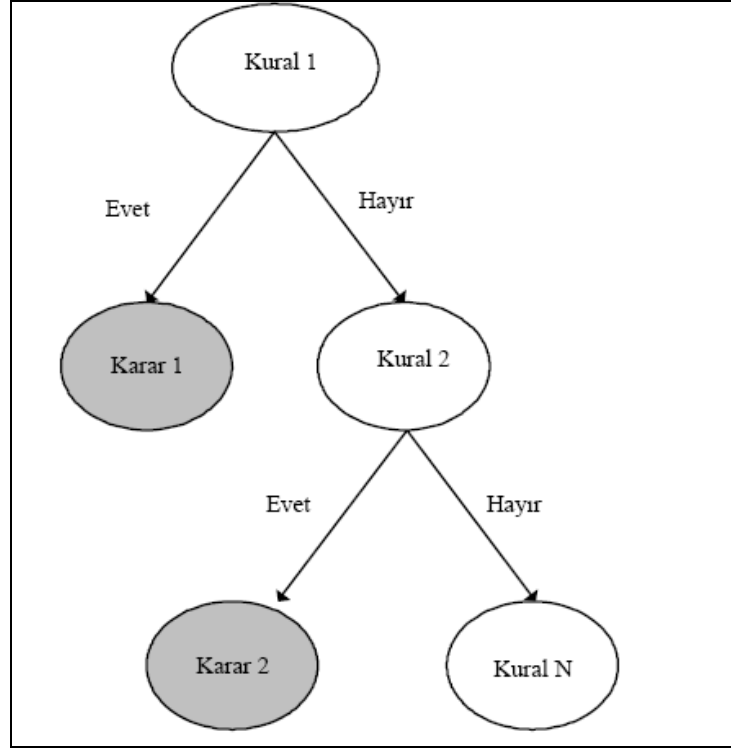
2.4.2 K-En Yakın Komşu

Veri uzayında birbirine yakın olan aynı tip kayıtlar, birbirlerinin komşusu durumundadırlar. Bu anlayış doğrultusunda, çok kolay fakat güçlü olan k-en yakın komşu algoritması geliştirilmiştir. K-en yakın komşu algoritmasının temel felsefesi komşunun yaptığını yaptır. Belirli bir kaydın özelliği tahmin edilmek istenirse, veri uzayında o bireye yakın olan örneğin 10 bireyin davranışına bakılabilir. Bu 10 komşunun davranışının ortalaması hesaplanır ve bu hesaplanan ortalama bireylerin tahmini olarak kabul edilir. K-en yakın komşudaki K harfi araştırdığımız komşu sayısıdır. Örneğin $k=4$ için yeni bir veri sınıflandırılmak istendiğinde eski sınıflandırılmış verilerden en yakın 4 veri alınır. Bu veriler hangi sınıfa dahilse yeni eleman da o sınıfa dahil edilir. Veriler arasındaki mesafe hesabında çeşitli yöntemler mevcuttur ancak çoğunlukla öklit mesafesi kullanılır.

2.4.3 Karar Ağaçları

Tahmin edici ve tanımlayıcı özelliklere sahip olan karar ağaçları, veri madenciliğinde kuruluşlarının ucuz olması, yorumlanmalarının kolay olması, veritabanı sistemleri ile kolayca entegre edilebilmeleri ve güvenilirliklerinin daha iyi olması nedenleri ile sınıflama modelleri içerisinde en yaygın kullanıma sahip tekniklerden birisidir [39].

Karar ağacı kullanarak veriyi sınıflama işlemi iki adımdan oluşur: öğrenme ve sınıflama aşamaları. Öğrenme aşamasında önceden bilinen bir eğitim verisi, model oluşturmak amacı ile sınıflama algoritması tarafından çözümlenir. Öğrenilen model, sınıflama kuralları veya karar ağacı olarak gösterilir. İkinci adım ise sınıflama adımıdır. Sınıflama aşamasında test verisi, sınıflama kurallarının veya karar ağacının doğruluğunu belirtmek amacıyla kullanılır. Eğer doğruluk kabul edilebilir oranda ise, kurallar yeni verilerin sınıflaması amacıyla kullanılır [34]. Şekil 2.4'de karar ağacının yapısı görülmektedir [39].

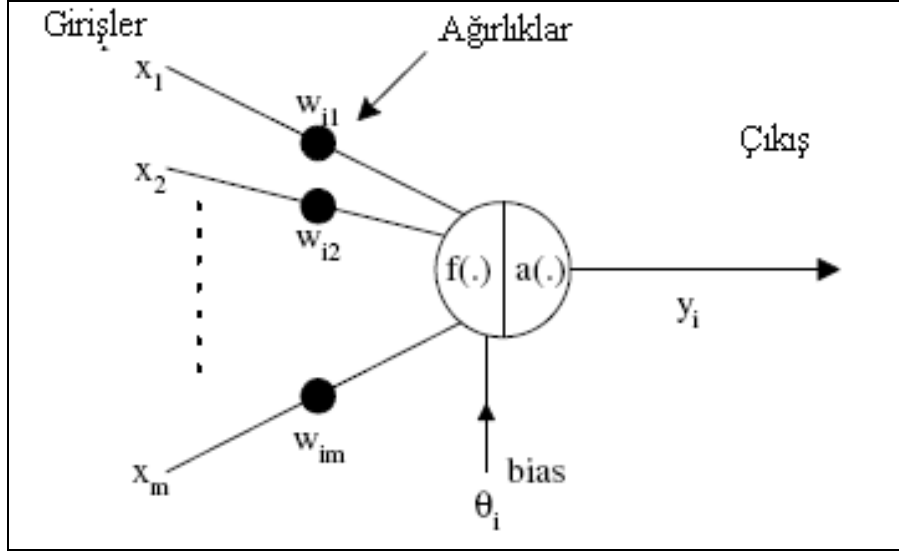


Şekil 2.4. Karar ağacının yapısı

2.4.4 Yapay Sinir Ağları

Yapay sinir ağları biyolojik olarak insan beyninden esinlenmiş ve taklit edilmiştir. İnsan sinir sistemi gibi nöronlardan oluşur. Bu nöronların her biri diğeriyle bağlantılı olup bu bağlantıların ağırlıkları bulunmaktadır [40]. Yapay sinir ağları, insanların deneyimlerinden bir takım bilgiler çıkartması gibi kendisine verilen örneklerden bir takım bilgiler çıkartma yeteneğine sahiptir. Yapay sinir ağları öncelikle sonuçları bilinen belirli bir veri kümesi üzerinde eğitilir. Bu eğitim neticesinde yapay sinir ağının içerisindeki ağırlıklar belirlenir. Bu ağırlıklar kullanılarak yeni gelen veriler işlenir ve bir sonuç üretilir.

Yapay sinir ağları sınıflama, kümeleme ve tahmin amaçları ile kolaylıkla kullanılabilecek genel amaçlı ve güçlü araçlardır [34]. Şekil 2.5’de basit bir yapay sinir ağı modeli gösterilmiştir.



Şekil 2.5. Basit bir yapay sinir ağı modeli

2.4.5 Birliktelik Kuralları

Birliktelik kuralları, işlemlerden oluşan ve her bir işlemin de elemanlarının birlikteliğinden oluştuğu düşünülen bir veri tabanında, bütün birliktelikleri tarayarak, sık tekrarlanan birliktelikleri veritabanından ortaya çıkarmaktır [41]. Birliktelik kuralları, büyük veri kümeleri arasında birliktelik ilişkileri bulurlar. Toplanan ve depolanan verinin her geçen gün artmasından dolayı, şirketler veritabanlarındaki saklı ilişkileri ortaya çıkarmayı hedeflemektedirler. Büyük miktardaki mesleki işlem kayıtlarından ilginç birliktelik ilişkilerini keşfetmek, şirketlerin karar alma işlemlerini daha verimli hale getirebilmektedir.

Birliktelik kurallarının kullanıldığı en tipik örnek Pazar sepeti analizi (Market Basket Analysis) uygulamasıdır. Bu işlem, müşterilerin yaptıkları alışverişlerdeki ürünler arasındaki birliktelikleri bularak müşterilerin satın alma alışkanlıklarını analiz eder. Bu bilgi ışığında market yöneticileri müşterilerin satın alma alışkanlıklarına göre daha etkin satış stratejileri geliştirmesine imkan tanımış olur.

Örneğin bir müşteri süt satın alıyorsa, aynı alışverişte sütün yanında ekmek alma olasılığı nedir? Bu tip bir bilgi ışığında rafları düzenleyen market yöneticileri ürünlerindeki satış oranını arttırabilirler [42]. Birliktelik kuralları eş zamanlı olarak gerçekleşen ilişkilerin tanımlanmasında kullanılır. Örneğin, bira satın alan müşterilerin %75 ihtimalle patates cipsi de almaları veya ekmek ve yağ alanların %90'ının süt de satın almaları birliktelik kuralları kapsamında tespit edilebilir.

Ardışık zamanlı örüntüler ise birbiri ile ilişkisi olan ancak birbirini izleyen dönemlerde gerçekleşen ilişkilerin tanımlanmasında kullanılır. X ameliyatı yapıldığında, 15 gün içerisinde % 45 ihtimalle Y enfeksiyonu oluşması, çekiç alan bir müşterinin ilk üç ay içerisinde % 15, bu dönemi izleyen süre içerisinde ise % 10 ihtimalle çivi alması ardışık zamanlı örüntülerdir.

Sepet analizinde amaç, değişkenler arasındaki ilişkiyi bulmaktır. Bu ilişkilerin bilinmesi şirketin kârını arttırmak için kullanılabilir. Eğer X malını alanların Y malını da çok yüksek bir olasılıkla aldıklarını biliyorsanız veya bir müşteri X malını alıyor ama Y malını almıyorsa o potansiyel bir Y müşterisidir. Sepet analizi günlük işlemler sonucu elde edilen verilerden anlamlı bağıntılar çıkarmada kullanılır. “Eğer A malını alıyorsa % x ihtimalle B malını almaya da meyillidirler” şeklinde bir sonuç A malını satan bir mağaza için çok faydalı bir bilgi olabilir. Sepet analizi uygulamaları; çapraz satış (cross-selling), mağaza raflarının düzenlenmesi (layout), katalog dizaynı ve fiyatlandırma (pricing) gibi alanlarda kullanılmaktadır.

Sepet analizinde mallar arasındaki bağıntı, destek ve güven kriterleri ile hesaplanır. Destek ve güven kriterlerinin tanımları aşağıdaki gibidir.

$$\text{Destek (Support) : } P(X \text{ ve } Y) = \frac{\text{X ve Y mallarını satın alan müşterilerin sayısı}}{\text{Toplam müşteri sayısı}}$$

$$\text{Güven (Confidence) : } P(X/Y) = \frac{P(X \text{ ve } Y) \{X \text{ ve } Y \text{ mallarını satın alanların sayısı}\}}{P(Y) \{Y \text{ malını satın alanların Sayısı}\}}$$

Destek kriteri, veride mallar arasındaki bağıntının ne kadar sık olduğunu, güven kriteri ise Y malını almış olan bir kişinin hangi olasılıkla X malını alacağını söyler. İki ürünün satın alınmasındaki bağıntının önemli olması için hem destek, hem de güven kriterinin olabildiğince yüksek olması gerekmektedir.

İşlemler	Satın Alınan Ürünler	Tekrarlanan Ürün	Destek Değeri
1	Süt, Ekmek, Yumurta	Süt	% 75
2	Süt, Yumurta	Ekmek	% 50
3	Süt, Şeker	Yumurta	% 50
4	Ekmek, Yağ	Şeker	% 25
		Yağ	% 25
		Süt, Yumurta	% 50

Tekrarlanan Ürün	Destek Değeri	Güven Değeri
Süt → Yumurta	% 50	% 66
Yumurta → Süt	% 50	% 100

Şekil 2.6. İşlemler, satın alınan ürünler, destek değerleri ve güven değerleri

Şekil 2.6'nın ilk tablosunda; bir markette farklı zamanlarda yapılan 4 farklı işlem görülmektedir. İkinci tabloda ise her ürünün ve bazı ürün birlikteliklerinin tekrar edilme yüzdeleri verilmiştir. Bu tekrar edilme oranları, bize ürünlerin destek değerini vermektedir. Şeker ve Yağ ürünlerinin destek değerlerinin referans destek değerinden düşük olmaları, bu ürünleri içeren ürün birlikteliklerinin ihmal edilebileceği anlamına gelmektedir. Son tabloda ise destek değeri referans destek değeri olarak kabul edilen % 50 olan ve en az iki üründen oluşan ürün birliktelikleri ele alınarak bunların güven değerleri sonucu görülmektedir. Burada dikkat edilirse Süt → Yumurta ilişkisi % 66 güven değerine sahip iken Yumurta → Süt ilişkisi % 100 güven değerine sahiptir. Çünkü Yumurta satın alınma durumu, Süt satın alınma durumlarının sadece % 66'sında gerçekleşirken Süt satın alınması ise Yumurta alınma durumlarının tümünde gerçekleşmiştir. Bu nedenle Yumurta → Süt birlikteliğinde % 100 güven değeri elde edilmektedir. Bu örnekten de anlaşıldığı gibi veritabanı üzerindeki bir ürün grubunun tekrar edilme miktarı destek değerini, bir ürün grubunu oluşturan herhangi bir ürünün veritabanındaki işlem sayısı ile destek değerine sahip olan ürün birlikteliği içerisindeki yer alış sayısına oranı da güven değerini vermektedir.

sepet analizi tekniğinin güçlü ve zayıf yönlerini sıralamak gerekirse;

Güçlü yönleri;

- Denetimsiz veri madenciliği yöntemini başarı ile uygular
- Anlaşılabilir sonuçlar verir

Zayıf yönleri;

- İleri derecede bilgisayar performansına bağımlıdır
- Arasında bağıntı oluşturacak ürünlerin doğru olarak seçilmesi, analiz süreci için oldukça önemlidir [43].

2.4.5.1 Apriori Algoritması

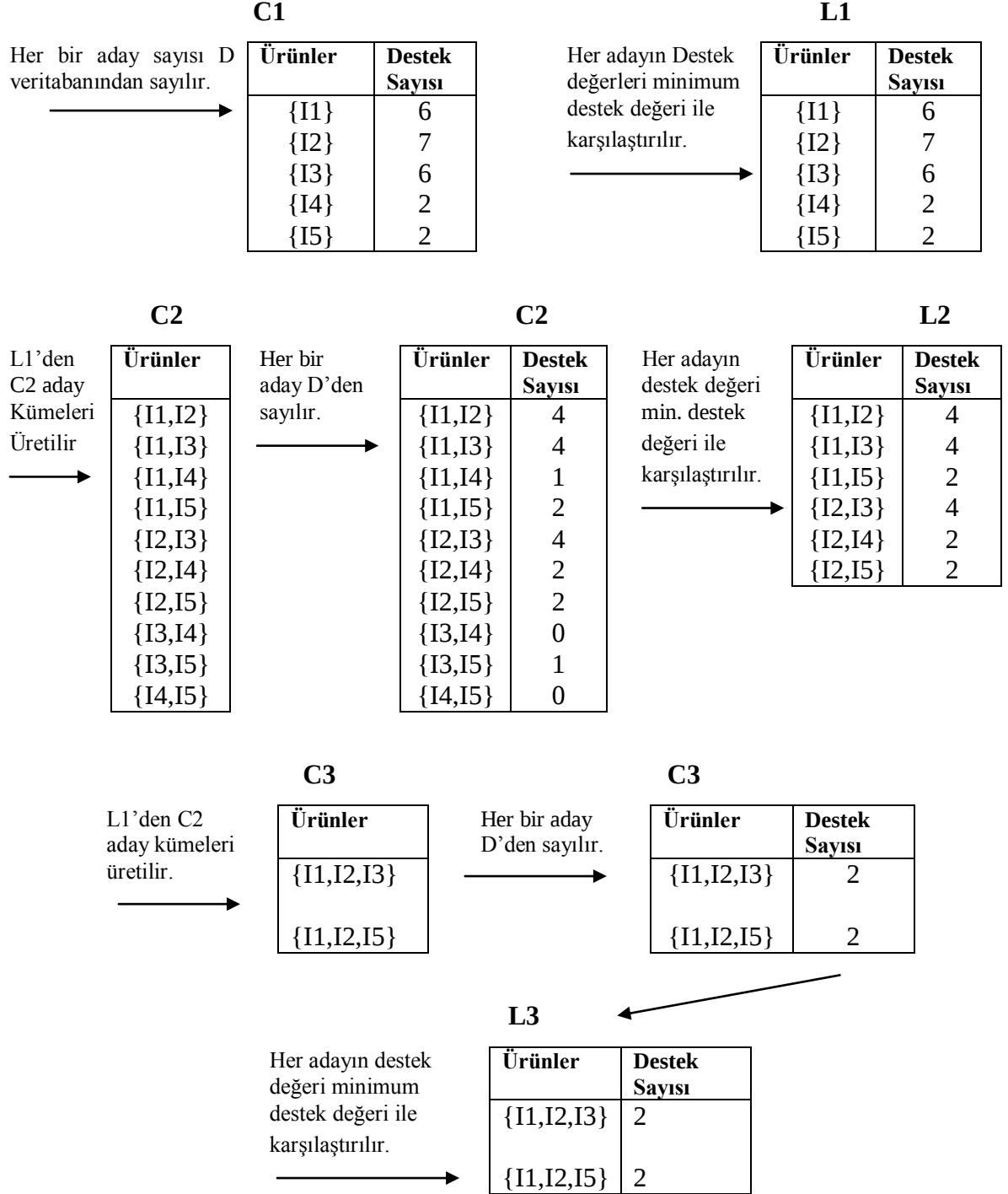
Literatürde birliktelik kuralı çıkaran değişik algoritmalar bulunmaktadır. Apriori Algoritması, birliktelik kuralı çıkarım algoritmaları içerisinde en fazla bilinen algoritmadır [44]. Bu algortmada sık geçen öge kümelerini bulmak için birçok kez veritabanını taramak gerekir. İlk taramada bir elemanlı minimum destek metriğini sağlayan sık geçen öge kümeleri bulunur. İzleyen taramalarda bir önceki taramada bulunan sık geçen öge kümeleri aday kümeler adı verilen yeni potansiyel sık geçen öge kümelerini üretmek için kullanılır. Aday kümelerin destek değerleri tarama sırasında hesaplanır ve aday kümelerinden minimum destek metriğini sağlayan kümeler o geçişte üretilen sık geçen öge kümeleri olur. Sık geçen öge kümeleri bir sonraki geçiş için aday küme olurlar. Bu süreç yeni bir sık geçen öge kümesi bulunmayana kadar devam eder [45].

Apriori algoritmasında önce aday öge kümeleri oluşturulur. Bu kümeler potansiyel olarak sık geçen öge kümeleridir. C ile gösterilir ve C[1], C[2], C[3],C[k] olarak k-öge kümesini oluştururlar. Her c[k] öge kümesi c[k-1] öge kümesini içerir ve $C[1] < C[2] < C[3] < \dots < C[k]$ şeklinde sıralıdır. Sık geçen k-öge kümeleri ise L ile gösterilir ve minimum destek kısıtlarını sağlarlar.

İşlemler	Satın Alınan Ürün Listesi
T1	I1, I2, I5
T2	I2, I4
T3	I2, I3
T4	I1, I2, I4
T5	I1, I3
T6	I2, I3
T7	I1, I3
T8	I1, I2, I3, I5
T9	I1, I2, I3

Şekil 2.7. Bir marketten satın alınan ürünler listesi

Örnek: Şekil 2.7’de bir firmada satın alınmış ürünlerle ilgili bir veritabanı görülmektedir. Bu veritabanını D olarak adlandıralım. Veritabanında 9 adet işlem görüldüğüne göre $|D| = 9$ denilebilir. Bu durumda D veritabanı üzerinde Apriori algoritması kullanılarak sık kullanılan nesnelerin nasıl bulunabileceğini şekil 2.8’de görebiliriz [46].

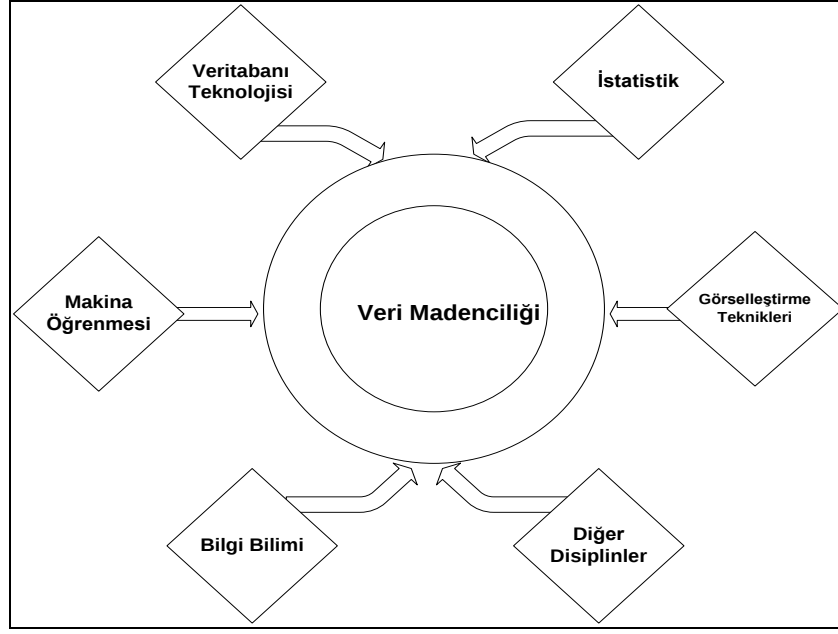


Şekil 2.8. Apriori algoritması ile aday öğelerin ve sık kullanılan öğelerin belirlenmesi

- 1) Algoritmanın ilk iterasyonunda her öge, yani 1 elemanlı ögeler kümesi C1 olarak elde edilir. Algoritma, her ögenin sayısını bulmak için veritabanını basitçe taramaktadır.
- 2) Bu veritabanı için minimum destek sayısı 2 olarak belirlenmiştir. Minimum destek değeri $2/9 = \% 22$ dir. Buradan 1 elemanlı ögeler kümesi olan L1 elde edilebilir. C1 deki tüm ögeler minimum destek değerini sağlamakta ve L1 elde edilmektedir.
- 3) L2'yi bulamak için önce 2 elemanlı aday öge kümesini elde etmek gerekir ve $L1 \times L1$ den C2 elde edilir. L1'in 2'li kombinasyonları C2'nin eleman sayısını vermektedir.
- 4) Daha sonraki işlem D'yi taramak ve C2 deki her aday ögenin minimum destek değerini bulmaktır. Şekil 2.8'de ortadaki 2. tabloda bu görülmektedir.
- 5) C2'de minimum destek değerine sahip olmayan aday ögelerin çıkarılması ile L2 öge kümesi elde edilir.
- 6) 3 elamanlı aday öge kümesinin elde edilmesi daha sonraki bölümde detaylı olarak anlatılacaktır. Normalde $C3 = L2 \times L2 = \{\{I1, I2, I3\}, \{I1, I2, I5\}, \{I1, I3, I5\}, \{I2, I3, I4\}, \{I2, I3, I5\}, \{I2, I4, I5\}\}$ olarak elde edilir. Bunların hepsi sık tekrarlanıyor görülebilir ancak Apriori algoritması özelliğine göre son dört tanesi sık tekrarlanmıyor olabilir. Bu nedenle son dördünü C3'ten çıkarabiliriz. Böylece D'yi taramak ve L3'ü belirlemek için bazı işlemler boşu boşuna yapılmamış olacaktır.
- 7) C3 üzerinde minimum destek değerine sahip olmayan aday ögelerin çıkarılması ile L3 öge kümesi elde edilir.
- 8) Algoritmada $L3 \times L3$ 'ü kullanılarak 4 elemanlı aday öge kümesi elde edilir. Bunun sonucunda $\{\{I1, I2, I3, I5\}\}$ elde edilmesine rağmen bu kümenin alt kümesi olan $\{\{I2, I3, I5\}\}$ L3'te bulunmamaktadır. Bu nedenle $C4 = 0$ dır ve tüm sık kullanılan birliktelikler elde edilerek algoritma sona ermektedir.

2.5 Veri Madenciliğinin Diğer Disiplinlerle İlişkisi

Veri madenciliği, makina öğrenmesi, örüntü tanıma, veritabanı teknolojileri, istatistik, yapay zeka, uzman sistemler, veri görselleştirme (data visualization) alanlarının bir kesişim noktası olarak doğmuş ve bu bağlamda gelişmesini sürdürmektedir. Bu yapı Şekil 2.9'da gösterilmektedir [32].



Şekil 2.9. Veri madenciliğinin ilişkili olduğu diğer disiplinler

2.5.1. Örnek Karşılaştırma

- Metotları farklıdır.
- Verilerin boyutları farklıdır.
- Ancak veri madenciliği, yalnızca heterojen sayılarla değil, heterojen veri alanlarıyla da ilgilenir. Örneğin tıbbi veriler, görüntüler, EKG gibi sinyaller, ateş, kolesterol düzeyi vb. klinik bilgiler ve standart olmayan bir dille yazılmış hekim yorumundan oluşabilir.
- Veri madenciliği, başka birçok disiplinin, özellikle de bilgisayarlı öğrenme/makina öğrenmesi (*machine learning*) ve veritabanı teknolojilerinin de ilgisini çekmektedir [47].

2.5.2. Veri Madenciliği Yaklaşımları

Veri madenciliği yaklaşımları ikiye ayrılır:

- **Model oluşturma:** Makina öğrenmesi ile kural üretme, sinir ağları ve gerileme
- **Örüntü algılama:** Normal EKG dalgalarındaki sapmaların algılanmasından, alışveriş ya da yolculuk örüntülerine kadar değişebilir.

2.6. Veri Madenciliği Türleri

En üst düzeyde, veri madenciliğinin üç türü vardır.

- a. Doğrudan veri madenciliği
- b. Varsayım deneme ve varsayımı daha iyi hale getirme
- c. Dolaylı veri madenciliği

2.6.1. Doğrudan Veri Madenciliği

Bu yöntemde, örneğin hekim, belirli bazı bilgilerin edinilmesiyle (örneğin tıkalı damarların bulunması) ilgilidir. Birçok tıbbi uygulama bu kategoriye girer.

2.6.2. Varsayım Deneme ve Varsayımı Daha İyi Hale Getirme

Kullanıcı bazı varsayımlar üretir ve bunun sistem tarafından doğrulanmasını, değiştirilmesini ve daha uygun olabilecek yenilerinin önerilmesini bekler.

2.6.3. Dolaylı Veri Madenciliği

Sistemde hiçbir kısıtlamanın bulunmadığı ve kullanıcının bulacağı bilgi ya da ne tür bir bilgiyle ilgilenildiğine ilişkin bir beklentinin olmadığı en genel senaryo budur. Bu aynı zamanda en güç yöntemdir [48].

2.7. Bilgi Bulma Süreci

2.7.1. Bilgi Bulma Sürecinin Önemli Özellikleri

- a. İnsan-sistem etkileşimine dayanır. Sistemin alanla ilgili tüm bilgilere sahip olması ve kullanıcının amacını tanıyabilmesi gerektiğinden, bilgi bulma sürecinin tümünü kafamızda canlandırmak çok zordur.
- b. Bilgi bulma işlemi, büyük veritabanlarıyla ilgilidir; bunlara erişmek ve işlemek için gereken yardımı sağlar, ancak büyük dosyalarla çalışmanın etkin yollarını gerektirir.
- c. Bilgi bulma işlemi, disiplinler arası araştırmalara dayanır ve tamamlayıcı nitelikte bilgi teknolojileri gerektirir. Bu disiplinler arası, bilgi bulma yöntemlerini istatistik, makina öğrenmesi ve gerileme çözümlemesi gibi araçlardan ayırır.

2.7.2. Bilgi Bulma Sürecinde Önemli Sorunlar

a. Veritabanlarının Büyük Boyutları

Sırf veritabanlarının boyutu yüzünden, veri madenciliği yöntemlerinden herhangi birinin ham veriyle başarılı olma olasılığı yoktur. Veri madenciliği yöntemleri, bu şekilde elde edilen sonuçların tüm veritabanını temsil edebileceğini umarak, veritabanından bir örneğin çıkarılmasını gerektirebilir. Bir veritabanının boyutlarının küçültülmesi iki yolla olabilir:

- 1. Veri alanında örnekleme;** genellikle rastgele bazı kayıtlar seçilir ve veri madenciliğinin sonraki aşamalarında kullanılır.
- 2. Özellik alanında örnekleme;** her veri kaydının bazı özellikleri seçilir. Birçok özellik varsa, seçim rasgele yapılır.

b. Dinamik Veri Yapısı

Veritabanları, belli aralıklarla güncellenir; yeni kayıtlar eklenir; örneğin, yeni bir SPECT görüntüsü eklenir (aynı hastaya ait yeni bir görüntü ya da yeni bir hastaya ait görüntü) ya da var olanlar yenileriyle değiştirilir.

c. Eksik ya da Kesin Olmayan Veri

Veritabanlarında toplanan bilgiler eksik olabilir ya da kesin olmayabilir. Sorunu çözmek için bu tür veriler geliştirilir.

d. Gürültü

Herhangi bir veri toplama tekniğinin, gürültüden tümüyle arınmış olması çok zordur. Bu nedenle, veri madenciliğinde, gelecekte toplanacak verideki gürültü miktarının yaklaşık olarak o anki veridekiyle aynı olmasına dikkat gösterilmelidir.

e. Eksik Değerler

Bir değer bilinmiyor ya da yanlışlıkla girilmemiş olabilir. Veri madenciliğindeki birçok yöntem, her veri nesnesi için sabit bir boyut (özellik sayısı) gerektirdiğinden, eksik değerler sorun yaratır.

Eksik değerler sorununu çözmek için:

- Eksik değerlerin yerine en olası değerleri,
- Bilinmeyen değerlerin yerine o öznitelik için olası tüm değerleri koymak gerekir.

f. Gerekinden Fazla, Anlamsız ya da Tutarsız Veri

Veri grubunda gerekinden fazla, anlamsız ya da tutarsız veri nesneleri ve/veya öznitelikleri olabilir. Aynı veri ögesi, birden çok kategoriye aitse, tutarsız veri söz konusudur.

2.7.3. Bilgi Bulma Sürecinde Kullanılan Yöntemler

Bilgi bulma sürecinde çeşitli yöntemler önerilmektedir. Bunlardan her biri 4 temel aşamaya ayrılmaktadır:

- Seçim
- İşlem öncesi
- Veri madenciliği
- Yorum

2.7.3.1. Seçim

Hedef bir veri grubu yaratılır. Modern veri grupları çok büyük ve karmaşıktır. Bu nedenle, araştırmanın amacını yansıtan verileri içeren hedef bir veri grubu yaratılır.

2.7.3.2. İşlem Öncesi

Bu aşamada, kullanılacak veri grubu, veri madenciliği yazılımı tarafından çözümleme için hazırlanır. Bunun için; Veri madenciliği yönteminde girdi olarak, veritabanındaki hangi bilgilerin kullanılacağına karar verilmesi gerekir. Sonra, seçilen veriler temizlenir.

2.7.3.2.1. İşlem Öncesi Süreçler (Ön Hazırlık)

Veri, özel veri madenciliği yöntemlerinin gereklerini karşılamak üzere yeniden biçimlendirilir. Bu aşamada, kayıtların ve özniteliklerin yeniden düzenlenmesi gerekebilir. Veri şifrelenir. Eksik veri (tamamlanmamış alanlar), ilgisiz alanlar, değişken olmayan alanlar, eğrilmiş alanlar ve uzak veri noktaları gibi, verideki istenmeyen özelliklerle ilgili sorunların çözülmesi gerekir.

Tamamlanmamış veri alanları, veri grubunda bilgi içermeyen kayıtlardır. Bazı görüşlere göre, bu oran % 70 ya da daha fazla ise bu alan tamamlanmış olarak kabul edilir. Bazı görüşlere göre ise bu oran düşüktür. Bu sorunun çözümü için, göreceli olarak daha fazla alan içeren kayıtlardan oluşan veri grupları önerilmektedir. **İlgisiz alanlar, çözümlemenin** amacı bağlamında önemsiz kabul edilen değerlerin olduğu alanlardır. Bir alanın ilgisiz olduğu kararı dikkatle verilmelidir. **Eğrilmiş alanlar**, değerleri, veri grubunda normal dağılım göstermeyen alanlardır. Değerlerin eğri dağılımını gösterecek bir değişken bulunursa, bu değişken temizlenmiş veride göz ardı edilmelidir ya da eğri, uygun bir matematiksel işlevin uygulanmasıyla düzeltilmelidir. **Uzak veri noktaları**, belirli bir veri alanının normal değer aralığı dışındaki yerlerdir. Bu tür veri noktaları, elde edilecek sonuçları bozabilir. Bu tür kayıtların veri grubundan çıkarılması önerilmektedir. Çıkarılan veriler ilginç veriler içerebilir, ancak bu veriler geleneksel anlamda incelenebilir ve otomatik (veri madenciliği) işlemin bir parçası olmayabilir.

2.7.3.3. Veri Madenciliği Aşaması

Veri modelleme teknikleri seçilir, eğitim ve deneme yordamlarına karar verilir, model oluşturulur ve kalitesi değerlendirilir. Temizlenmiş veri, gizli örüntü ya da eğilimlerin tanımlanması ya da belirli varsayımların denenmesi amacıyla, çözümleme için veri madenciliği yazılımına gönderilir.

2.7.3.4. Yorum

Tüm bilgi bulma süreci çözümlenir. Sonuçların anlaşılmasına çalışılır, yeni bilgilerin ilginç olup olmadığı denetlenir, sonuçlar söz konusu alan açısından (tıbbi açıdan)

yorumlanır ve bunların amaç üzerindeki etkileri araştırılır. Bu aşama, araştırılan alanda uzman olan bir kişi tarafından yürütülür [49].

3. DEPREM KATALOGLARI

Tezin bu bölümünde, bazı deprem katalogları ve tezin uygulama kısmında depremlerle ilgili gerçekleştirilmiş veri madenciliği çalışmalarında kullanılan Türkiye deprem kataloğu ile ilgili bilgiler verilmiştir.

3.1 SCSN (Güney Kaliforniya Sismografik Ağ) Kataloğu

Güney Kaliforniya Sismografik Ağ kataloğu (SCSN) 1932'den günümüze kadar olan depremlerle ilgili çeşitli veriler içermektedir. Bu veriler 1932'den günümüze kadar olan büyüklüğü 3.0 ve daha büyük yerel depremler ve 1981'den şimdiye kadar büyüklüğü 1.8 ve daha fazla olan depremlerle ilgili verilerdir. Ocak 1984'den itibaren ise tüm depremsel hareketlilikler bu katalogda yer almaktadır.

SCSN Kataloğuna Southern California Earthquake Center web sitesinden [50] erişmek mümkündür. Bu sitede yer alan veriler yıllık olarak ASCII formatta kaydedilmiş *.catalog uzantısına sahip dosyalarda tutulmaktadır. Bu verilerin formatı Şekil 3.1'de gösterilmiştir.

Kolon	Uzunluk	Format	Adı	Açıklama
1	4	i4	Yıl	
6	2	i2	Ay	
9	2	i2	Gün	
13	2	i2	Saat	UTC zamanı, yerel saatten 7 yada 8 saat ileri
16	2	i2	Dakika	
19	5	f5.2	Saniye	
26	2	i2	Enlem Derece	
29	5	f5.2	Enlem Dakika	
34	5	i4	Boylam Derece	
39	5	f5.2	Boylam Dakika	
45	1	a1	Kalite	konum kalitesi 'A' +/-1 km yatay mesafe +/-2 km derinlik 'B' +/-2 km yatay mesafe +5 km derinlik 'C' +/-5 km yatay mesafe derinlik sınırlaması yok
47	3	f3.1	Şiddet	
55	5	f5.2	Derinlik	
60	3	i3	Nph	seçilen faz sayısı
67	4	f4.2	Rms	seçilen tüm istasyonların toplam rms değeri
72	7	i7	id	kayıt numarası

Şekil 3.1 SCSN deprem katalog formatı

Şekil 3.2 'de, SCSN deprem kataloğunda tutulan verilerden örnekler gösterilmiştir.

YYY	MM	DD	HH	Mm	SS.ss	Lat	Lon	Q	Mag	Depth	Nph	RMS	Evid
2009	01	01	00	01	09.55	34 49.53	-116 24.33	A	0.4	7.63	5	0.04	10370373
2008	11	12	05	40	15.94	35 58.12	-117 19.50	A	1.5	2.09	35	0.22	10358717
2005	01	06	14	46	31.13	34 7.98	-117 26.45	A	1.7	6.13	24	0.15	14116984
2002	11	09	16	40	19.96	35 16.41	-118 35.78	B	1.9	6.55	20	0.10	13909352
1999	10	17	11	11	00.74	34 27.72	-116 17.04	B	2.5	0.00	17	0.16	9110405
1999	10	17	13	39	41.90	34 38.94	-116 17.40	A	2.7	2.10	33	0.17	9110552
1983	11	25	15	45	07.07	35 54.10	-118 19.21	A	1.6	3.13	15	0.07	25346
1932	12	31	08	20	06.48	33 49.80	-118 6.90	C	1.8	6.00	6	0.10	3359495

Şekil 3.2 SCSN Deprem Kataloğunda Tutulan Veri Örneği

3.2 SCEDC(Güney Kaliforniya Deprem Veri Merkezi) Kataloğu

SCEDC (Southern California Earhquake Data Center- Güney Kaliforniya Deprem Veri Merkezi) kataloğu, 1 Nisan 1981- 31 Ocak 1983 ve 13 Temmuz 1983 ten günümüze kadar olan depremlerin iç merkez verilerinden oluşan bir katalogdur. Bu zaman periyotları SCEDC'den erişilebilen verilerin erişim süreleri ve bu zaman aralıklarındaki dijital dalga formlarını temsil eder. Aynı zamanda bu kayıtlar Güney Kaliforniya Sismik Ağı tarafından işlenmiş verilerdir.

SCEDC kataloğu, 1981 den şimdiye kadar büyüklüğü 1.8 den büyük veya eşit deprem hareketliliklerinden oluşur. Bu katalogda 1983 yılına ait verilerde bir takım zaman farklılıkları bulunmakta bu farklılıkların nedeni ise SCSN ağının (Güney Kaliforniya Sismografik Ağı) bu zaman aralıklarında hareketlilikleri, büyüklükleri ve iç merkez verisini tam olarak işleyememiş olmasından kaynaklanmaktadır.

SCEDC kataloğundaki veriler yıllık olarak ascii *.catalog dosyası şeklinde tutulmaktadır. Şekil 3.3'te SCEDC kataloğunun veri formatı görülmektedir [50].

Kolon	Uzunluk	Format	Adı	Açıklama
1	10	a10	Yıl/ay/gün	
12	10	a10	Zaman	saat:dak:saniye
23	1	a1	Olayın türü	Yerel (L) Bölgesel (R) Ölçüm cihazına uzak (T) Maden patlaması (Q) Şüpheli yer (D)
25	3	f3.1	Şiddet	
29	1	a1	Şiddet tipi	e' enerji şiddeti w' denge şiddeti b' gövde-dalga şiddeti s' yüzey-dalga şiddeti t' yerel şiddet c' son genlik h' helicorder şiddet d' coda sürekli şiddet
32	7	f7.3	Enlem	Derece Dak.
39	8	f8.3	Boylam	Derece Dak.
48	5	f5.1	Derinlik	
54	2	a2	Kalite	olay kalitesi A' +/-1 km yatay mesafe +2 km derinlik B' +/-2 km yatay mesafe +5 km derinlik C' +/-5 km yatay mesafe derinlik sınırlaması yok D' > +/-5 km yatay mesafe
56	7	a7	olayid	
63	4	i4	nph	seçilen faz sayısı
67	4	i4	ngram	
71	4	i4	nterr	
74	3	i3	nport	Portatif araç dosyalannın sayısı

Şekil 3.3 SCEDC katalog formatı

Şekil 3.4'te SCEDC deprem kataloğunda tutulan verilerin örnekleri gösterilmiştir.

YYY/MM/DD	HH:mm:ss.ss	ET	MAG	M	LAT	LOH	DEPTH	Q	EVID	HPH	HGRM
2000/01/01	00:02:46.20	le	0.80	h	34.692	-116.355	12.3	B	9131988	11	0
2000/01/01	00:03:08.00	le	0.70	h	34.439	-116.222	7.0	A	9521550	9	0
2000/01/01	00:03:08.01	le	1.00	n	34.442	-116.228	7.3	A	3321769	9	0
2000/01/01	00:20:00.02	le	2.48	c	32.179	-115.073	6.0	C	9131991	15	0
2000/07/31	16:54:06.27	le	1.10	h	34.511	-116.254	3.0	A	9158937	11	0
2000/07/31	17:18:01.91	le	1.40	h	34.490	-116.269	2.1	A	9158938	19	0

Şekil 3.4 SCEDC deprem kataloğunda tutulan veri örnekleri

3.3 Türkiye Deprem Kataloğu

Türkiye de gerçekleşen depremlerin bir araya getirildiği deprem katalogları değişik araştırmacılar tarafından düzenlenmiştir. Bu çalışmada TUBITAK Marmara Araştırma Merkezi'nden elde ettiğimiz kataloglar kullanılmıştır [51].

Türkiye Deprem Kataloğunda bu araştırmalar bir araya getirilerek derlenmiş ve tek bir katalogda toplanmıştır. Katalogda verilerin kimden alındığını belirtmek için referans kolonunda belirli kısaltmalar kullanılmıştır. Bu kısıtlamalar Tablo 3.1'de gösterilmiştir.

Tablo 3.1 Türkiye deprem kataloğundaki referans kısaltmaları

AJ	: Ambraseys & Jackson (1998)
AJ2	: Ambraseys & Jackson (2000)
EG	: Guidoboni <i>et al.</i> (1994)
EG2	: Guidoboni & Comastri. (2005)
HS	: Soysal <i>et al.</i> (1981)
KE	: Ergin <i>et al.</i> (1967 and 1971)
KU	: Kondorskaya & Ulomov (1999)
PA	: Papazachos <i>et al.</i> (1997)
ST	: Shebalin & Tatevossian (1997)
ISC	: International Seismological Center
INGV	: Istituto Nazionale di Geofisica e Vulcanologia (http://storing.ingv.it/es_web/Data/Es_map.html)

Tablo 3.2 'de, Türkiye Deprem Kataloğunda bulunan magnitüd (büyüklük) bilgileri için kullanılan kısaltmalar ve anlamları gösterilmiştir.

Tablo 3.2 Türkiye deprem kataloğunda yer alan magnitüd bilgileri

Ms	Yüzey dalgasından hesaplanan magnitüd
Mw	Moment magnitüd.
M*	$M^*=0.592.I_0 + 1.63$ (İpek, 1965) hesaplanan magnitüd değeri.
M	Hesaplama konusunda bir bilgi olmayan magnitüd değeri.
Mi	ISC kataloğundan alınan depremlere ait magnitüde değeri.
Me	Eşdeğer magnitüde değeri. Guidoboni ve Comastri (2005)' den alınmıştır.

Türkiye deprem kataloğunda depremlerin olduğu enlem, boylam, tarih, büyüklük, referans derinlik gibi alanlar bulunmaktadır. Şekil 3.5'te Türkiye Deprem Kataloğunda yer alan verilerin formatına dair bir örnek gösterilmiştir[51].

Lat.(deg)	Lon.(deg)	Mag.	Mag. Type	h(km)	Year	Months	Day	HoMinSec	Ref.
37.2000	30.0000	5.50	Mw	20.0	1963	11	22	214131	KU
34.9000	27.4000	4.50	M	33.0	1963	11	26	161948	KE
35.6000	25.7000	4.80	M	33.0	1963	12	19	184338	KE
38.4000	45.3000	4.50	M	33.0	1963	12	31	151808	KE
36.4000	28.9000	4.70	Mi	0.0	1964	01	29	222849	ISC
37.4100	29.8900	5.30	Mi	59.0	1964	01	30	174557	ISC
37.4000	29.7000	4.40	Mi	0.0	1964	01	30	175217	ISC
36.5100	35.8000	4.10	Mi	81.0	1964	02	02	062655	ISC
40.3700	46.1000	4.50	Mi	9.0	1964	02	05	102423	ISC
40.4000	46.0000	4.00	Mi	33.0	1964	02	05	112706	MOS
39.8500	40.3400	4.40	Mi	57.0	1964	02	24	113219	ISC

Şekil 3.5 Türkiye deprem kataloğunda tutulan veri örnekleri

4. TÜRKİYE DEPREM PORTALI (TDEP)

4.1 Giriş

Tezin bu bölümünde, deprem verileri üzerinde veri madenciliği tekniklerinin uygulanması ve bu verilere ait çeşitli istatistiksel bilgilerin sunulması amacıyla tasarlanan TDEP web uygulaması ve bu uygulamada yer alan alt algoritmalar sunulmuştur. Şekil 4.1’de, Türkiye Deprem Portalı yazılımının çalışan web arayüzü gösterilmiştir.

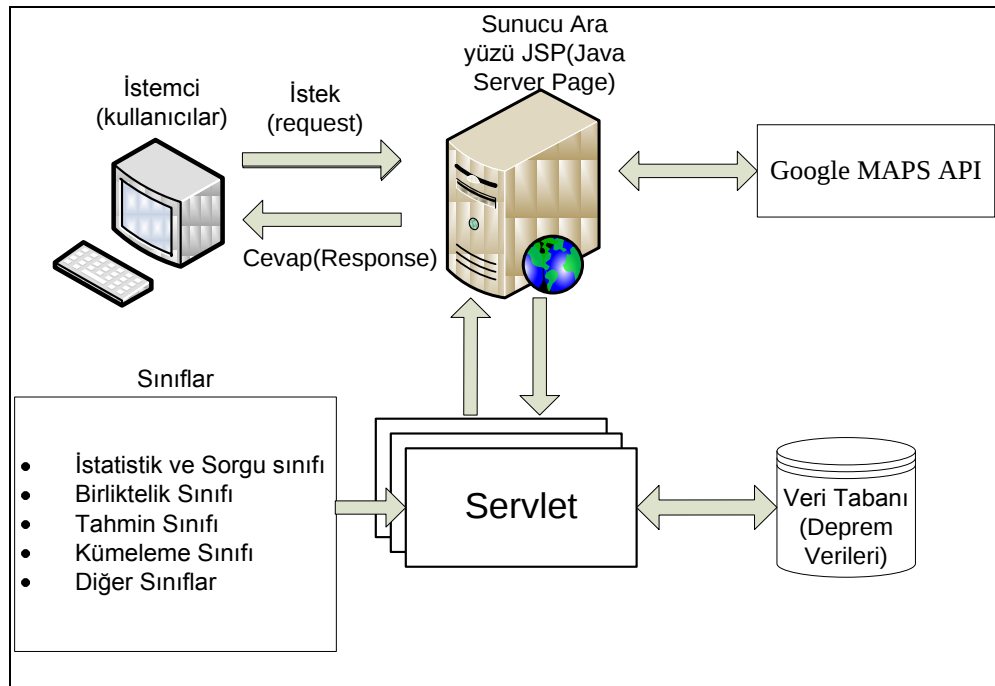


Şekil 4.1. Türkiye Deprem Portalı’nın çalışan web arayüzü

Tasarlanan TDEP uygulamasında, deprem verileri üzerine istatistiksel ve veri madenciliği işlemlerini gerçekleştiren dört farklı alt algoritma bulunmaktadır. Birinci alt algoritma, veri tabanındaki depremlerin sorgulanması ve bu depremlere ait istatistiksel sonuçların bulunması için kullanılan istatistiksel ve sorgu bölümüdür. İkinci alt algoritma, deprem alanları arasındaki hareketlilik ilişkilerinin tespiti için kullanılan birliktelik kuralları algoritmasıdır. Üçüncü alt algoritma, oluşan depremler paralelinde bu bölge üzerinde elde edilen kurallara göre gelecekte deprem oluşabilecek riskli bölgelerin uyarısını veren tahmin kısmıdır. Son olarak, Türkiye üzerindeki depremlerin kümelendirilmesinde kullanılan kümeleme kısımları uygulama içerisinde yer almaktadır.

4.2. Türkiye Deprem Portalı'nın (TDEP) Çalışma Mimarisi

Java tabanlı teknolojiler kullanılarak hazırlanan bu sistemde, birçok sınıf yapısı, veri kaynakları ve bu kaynaklar arasında haberleşmeyi sağlayan servetlar bulunmaktadır. Netbeans 6.0.1 programı ile hazırlanan bu uygulama, Java Server Pages (JSP) teknolojisi ile kullanıcı ve sunucu arasında etkileşimli bir iletişim sağlamaktadır. Şekil 4.2'de, Türkiye Deprem Portalı çalışma mimarisi gösterilmiştir.



Şekil 4.2. Türkiye Deprem Portalı'nın çalışma mimarisi

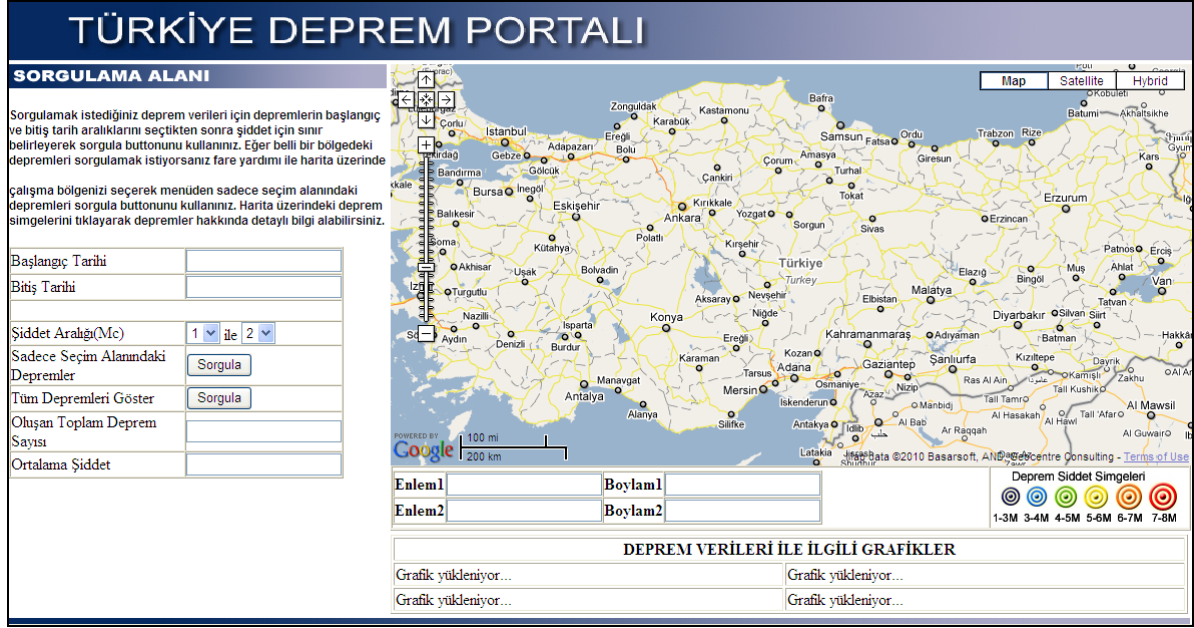
Deprem verilerinin görsel olarak sunulması ve kullanıcıların veriler hakkında daha detaylı bilgi edinmesi için sisteme dinamik haritalar dahil edilmiştir. Google firmasının ücretsiz olarak kullanıma sunduğu Google haritaları uygulaması (Google Maps API) sayesinde kullanıcılar, oluşan depremlerin büyüklüğü ve konumu hakkında haritalar üzerinde detaylı görsel bir içeriğe erişebilmektedir.

Kullanıcılar internet üzerinden sunucu arayüzüne erişerek, deprem verileri için çeşitli isteklerde bulunur. Sunucu kullanıcıdan gelen bu istekleri değerlendirerek ilgili servleti çağırır. Servletler gelen istekler doğrultusunda yürütülecek sınıfları ve kullanılacak diğer veri kaynaklarının yürütümünü gerçekleştirir. Gelen parametreler doğrultusunda sınıflarda ve diğer veri kaynaklarında veriler işlenerek sonuçlar üretilir. Üretilen bu sonuçlar tekrar servletler aracılığı ile sunucuya oradan da kullanıcılara ulaştırılır. Son zamanlarda web teknolojilerinde sıklıkla kullanılan Ajax teknolojisi, sunucu ve servlet arasındaki haberleşmeyi sağlayarak sadece ilgili işlemlerin bulunduğu alanların yenilenmesini sağlar. Bu sayede daha hızlı bir haberleşme sağlanır. Bu sistemde, coğrafi verilerin saklanması ve bu veriler üzerinde geniş bir sorgulama imkânı sunan PostgreSQL veri tabanı yönetim sistemi kullanılmıştır. Türkiye deprem kataloğundaki veriler, hazırlanan veri tabanı içerisindeki tablolara belirlenen formata göre aktarılmıştır.

4.3. Deprem Verileri Üzerinde Sorgulama ve İstatistiksel İşlemler

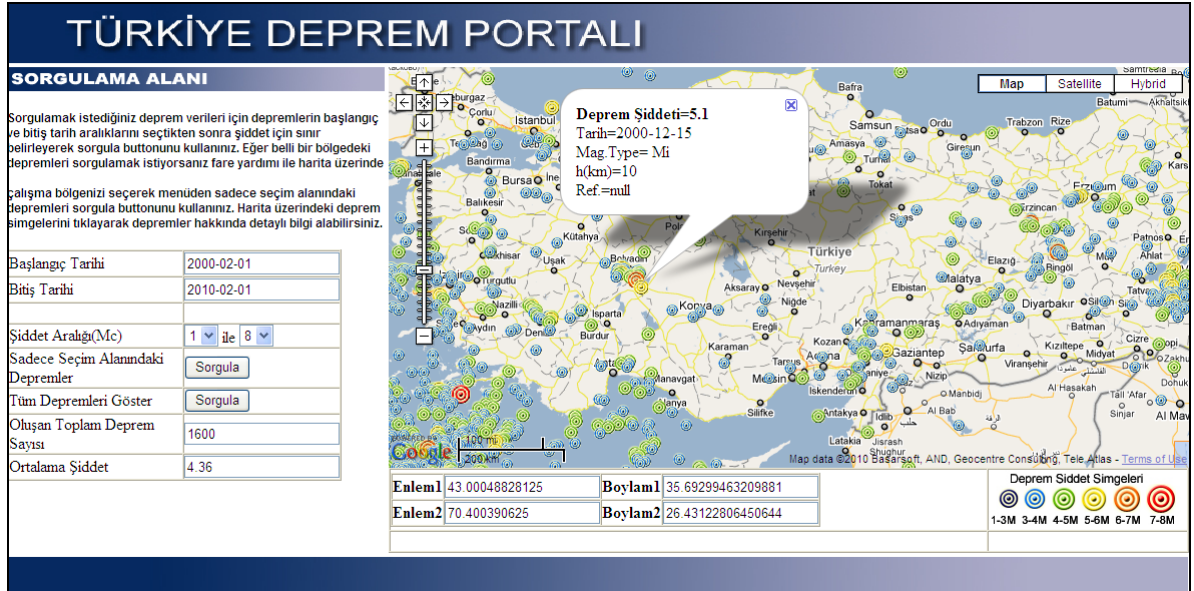
Türkiye Deprem Portalı'nın bu bölümünde kullanıcılar Türkiye üzerindeki depremlerin sorgulamasını ve bu depremler hakkında istatistiksel bilgilere erişim sağlayabilirler. Ayrıca portalın bu kısmında deprem verileri ile ilgili sayısal ve grafiksel veriler sunulmaktadır. Kullanıcılar bu sorgulama işlemlerini portal arayüzde bulunan form bileşenleri yardımıyla kolayca gerçekleştirebilirler.

Şekil 4.3'de Sorgulama ve istatistiksel menü penceresinin genel bir görüntüsü sunulmuştur.



Şekil 4.3. Sorgulama ve istatistik web arayüzü

Şekil 4.4’de Türkiye Deprem Portalı üzerinde yapılan sorgulama sonucunda elde edilen verilerin arayüzde nasıl yerleştirildiğine dair bir örnek gösterilmiştir.

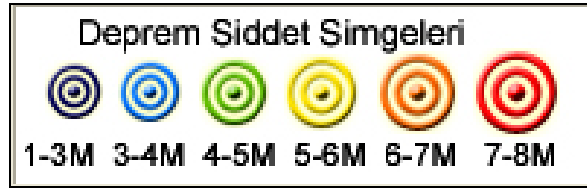


Şekil 4.4. Portal üzerinde deprem verilerinin yerleşimi

Türkiye deprem kataloğundan elde edilen veriler oluştukları konumlara ve büyüklüklere göre görsel bir şekilde dinamik haritalar üzerinde gösterilmiştir. Kullanıcılar Google haritaları uygulamasının sunduğu birçok özelliği bu portal üzerinde kullanabilmektedirler.

Kullanıcılar harita üzerindeki herhangi bir bölgeyi yakınlaştırıp bölgenin detaylı konumunu görebilir. Ayrıca bölgenin hem uydu görüntüsünü hem harita görüntüsünü kullanabilirler.

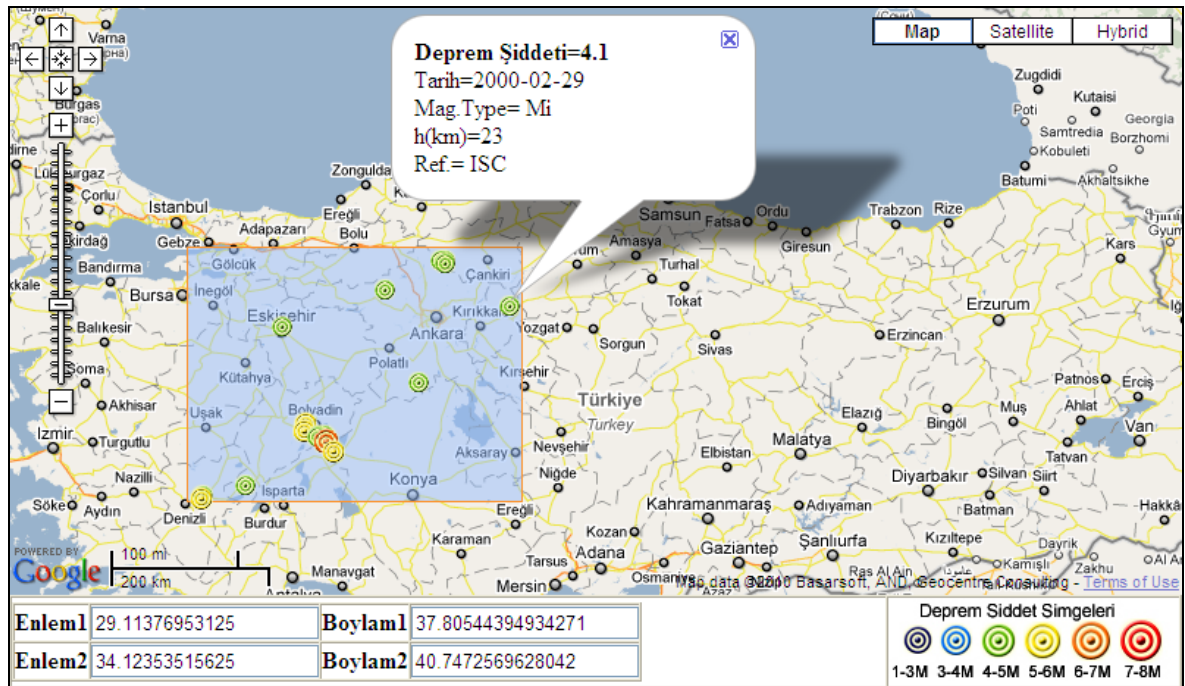
Harita üzerinde depremlerin daha belirgin olarak gösterilmesi için büyüklüklere göre çeşitli simgeler oluşturulmuştur. Şekil 5’de deprem büyüklük simgeleri ve bu simgelerin hangi aralıkları temsil ettiklerini belirten bir şekil gösterilmiştir.



Şekil 4.5. Deprem büyüklüklerini temsil eden simgeler

Bu simgeler belirli bir renk ve büyüklük sıralamasına göre düzenlenmiştir. Böylelikle harita üzerinde depremlerin ayırt edilmesi işlemi görselleştirilerek kullanıcılara kolaylık sağlanmıştır.

Şekil 4.6’da Portalda seçimi yapılan bir bölge üzerindeki depremlerin sorgu sonuçları gösterilmiştir.



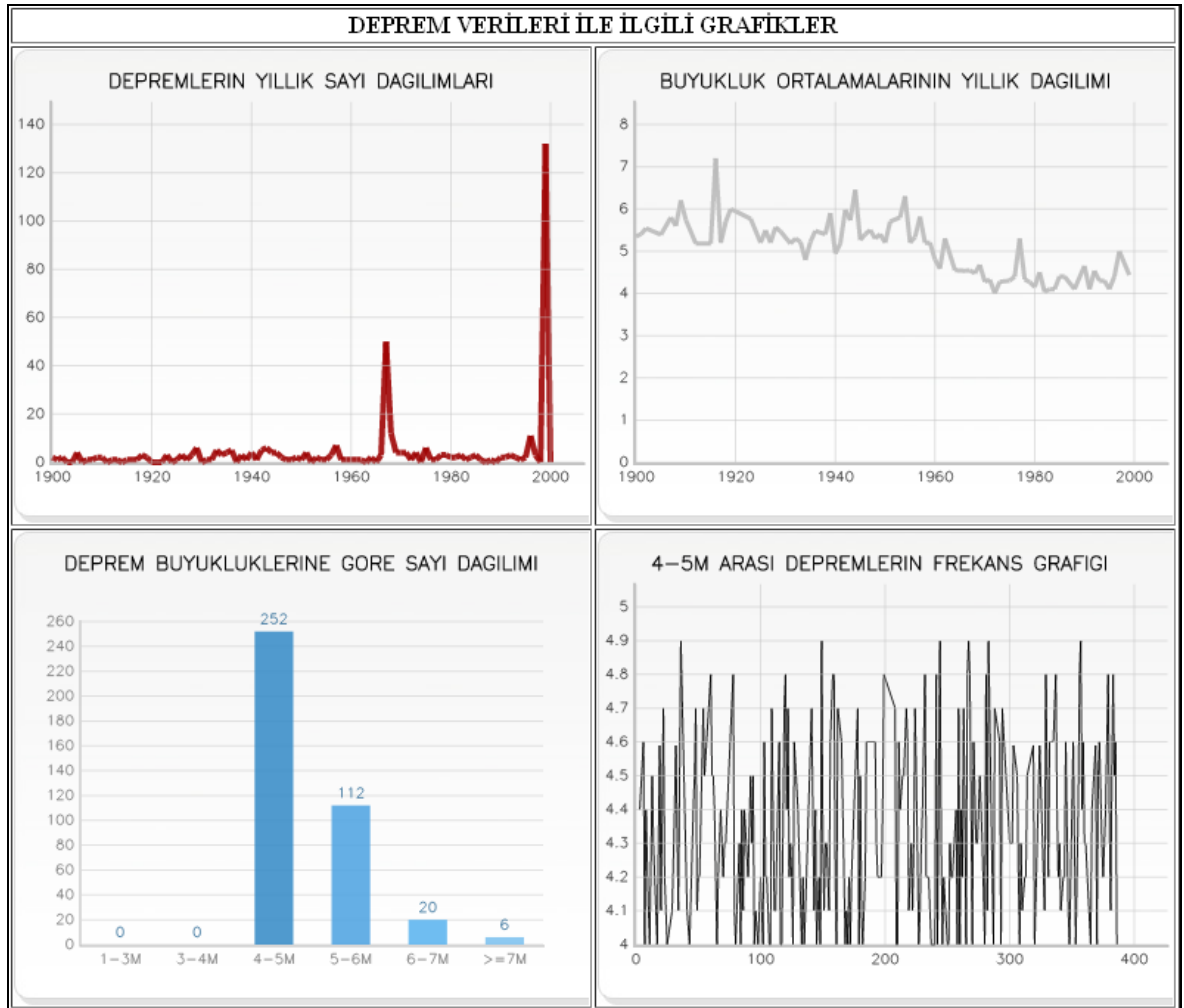
Şekil 4.6. Portalda seçilen bölge üzerindeki depremlerin sorgulanması

4.4. Grafik Ekranı

Portalda, sorgulanan depremlerle ilgili sayısal verilerin yanında istatistiksel grafiklerde yer almaktadır. Bu arayüz içerisinde, sorgulanan yıllar arasındaki deprem sayı dağılımları, oluşan depremlerin büyüklük aralıklarına göre sayıları, yıllık deprem büyüklük ortalamaları gibi grafikler yer almaktadır.

Bu grafikler, kullanıcının sorguladığı tarihler arasında meydana gelen depremler ile ilgili etkili bir görsel veri kaynağı sunmaktadır. Kullanıcı sorgusundan sonra deprem verileri ile birlikte çizdirilen bu grafikler, deprem hareketliliklerinin değerlendirmesinde yardımcı olabilecek niteliktedir.

Şekil 4.7’de, Türkiye Deprem Portalı’nın istatistiksel ve sorgulama kısmında yer alan grafik ekranı gösterilmiştir.



Şekil 4.7. Türkiye Deprem Portalında bulunan istatistiksel grafik ekranı

4.5. Türkiye Deprem Kataloğu Üzerinde Yapılan İstatistiksel Çalışmalar

Tezin bu alt bölümünde, Türkiye deprem kataloğu verileri kullanılarak yapılan bazı istatistiksel ve sorgu çalışmaları yer almaktadır. Bu çalışmalara ait veriler ve grafikler sunularak sonuçlar tartışılmıştır.

Şekil 4.8’de 01.01.1900 ile 01.01.2008 tarihleri arasında Türkiye’de meydana gelen $M_c > 0$ depremlerin yıllara göre sayı dağılım grafiği gösterilmiştir.



Şekil 4.8. 01.01.1900–01.01.2008 tarihleri arasında Türkiye genelinde oluşan $M_c > 0$ büyüklüğündeki depremlerin yıllık sayı dağılımları.

01.01.1900–01.01.2008 tarihleri arasında Türkiye genelinde 11.081 adet deprem meydana gelmiştir. Bu tarihler arasındaki en yüksek depremler 1996 ve 1999 yıllarında meydana gelmiştir. 1970 yılında artan deprem sayısından sonra deprem sayılarının azaldığı, daha sonra 1996 ve 1999 yıllarında tekrar yükseldiği görülmektedir. 1999 yılından sonra tekrar azalışa geçen depremlerin, grafik verisi göz önüne alındığında yaklaşık 2022 yılında tekrar yükselişe geçmesi beklenebilir.

Şekil 4.9’da, 01.01.1900–01.01.2008 tarihleri arasında oluşan depremlerin yıllık ortalama büyüklük grafikleri gösterilmiştir.



Şekil 4.9. 01.01.1900–01.01.2008 tarihleri arasında Türkiye genelinde oluşan $M_c > 0$ büyüklüğündeki depremlerin yıllık büyüklük ortalamaları.

Türkiye genelinde belirtilen tarihler arasındaki ortalama deprem büyüklüğü 4.199 M olarak tespit edilmiştir. Grafikte, deprem büyüklük ortalamalarının gittikçe düşüşe geçtiği görülmektedir. Bunun başlıca nedeni olarak Türkiye üzerinde oluşan depremlerin 1960'lı yıllardan sonra sağlıklı olarak elde edilmesi gösterilebilir. 1900'den 1960'lı yıllara kadar olan deprem verilerinin büyük bir kısmının kaydedilmediği ve bu aralıkta saklanan depremlerin sayıca az fakat büyüklük olarak ortalamanın üzerinde olması grafiğin, gittikçe düşen bir büyüklük ortalaması olduğu görüntüsünü vermektedir. Ayrıca tez çalışmasında referans olarak alınan Türkiye Deprem Katalog verileri 1965 yılına kadar farklı kataloglardan derlenmiştir. 1965 yılından sonra ISC katalog verileri kullanılmıştır. Fakat Türkiye üzerinde oluşan depremlerin büyüklük ortalaması bakımından son yıllara doğru düşüşe geçtiği kesin olarak söylenebilir.

01.01.1900–01.01.2008 tarihleri arasında oluşan 11.081 depremin 4.629 tanesi 3-4M, 5.165 tanesi 4-5M, 1.134 tanesi 5-6M, 130 tanesi 6-7M ve 23 tanesi 7.0 ve 7.0 üzerindeki büyüklüklerde meydana gelmiştir.

Şekil 4.10'da, 01.01.1960–31.12.2003 tarihleri arasında Türkiye genelinde oluşan deprem sayılarının yıllık dağılımları gösterilmiştir.

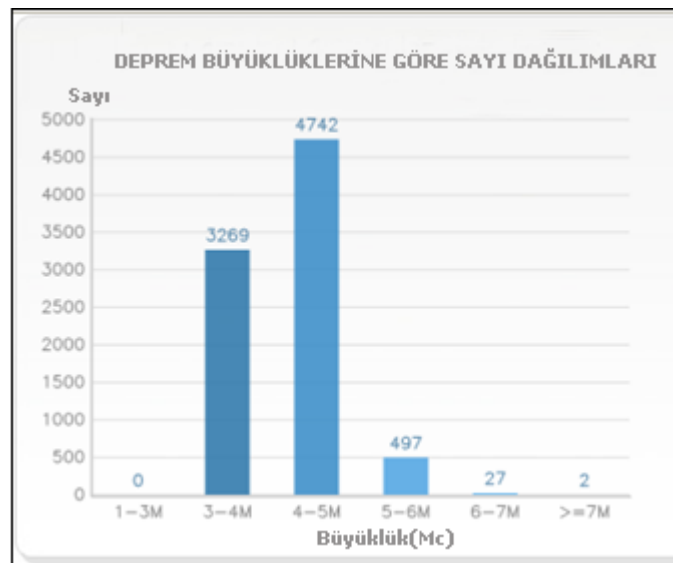


Şekil 4.10. 01.01.1960–31.12.2003 tarihleri arasında Türkiye genelinde oluşan $M_c > 0$ büyüklüğündeki depremlerin yıllık sayı dağılımları.

1970 ve 1996 yıllarında deprem sayılarının yükselişe geçtiği görülmektedir. Bu artışlar arasındaki sürenin yaklaşık olarak 26 yıllık bir süre olduğu gözlemlenmiştir.

Bu tarih aralıklarında, Türkiye genelinde 8.537 deprem meydana gelmiştir. Bu depremlerin 4.742 tanesi 4-5 M büyüklük aralığında gerçekleşmiştir.

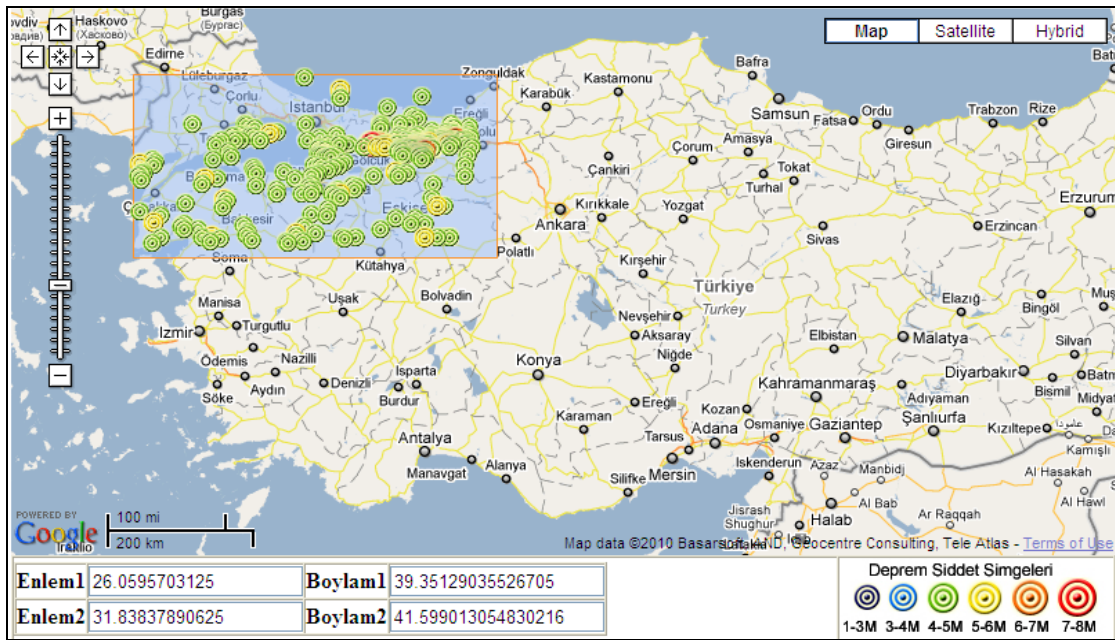
Şekil 4.11’de, 01.01.1960–31.12.2003 tarihleri arasında Türkiye genelindeki depremlerin büyüklüklere göre sayı dağılımları gösterilmiştir.



Şekil 4.11. 01.01.1960–31.12.2003 tarihleri arasında Türkiye genelinde oluşan $M_c > 0$ büyüklüğündeki depremlerin büyüklüklere göre sayısal dağılımları.

Tüm Türkiye genelinde olduğu gibi bu istatistiksel ve grafiksel çalışmalar, kullanıcıların seçim yaptığı herhangi bir bölge üzerinde de yapılabilir. Öncelikle kullanıcı üzerinde çalışacağı bölgeyi harita üzerinden seçer ve gerekli parametreleri girdikten sonra sorgulamalarını ve istatistiksel bilgilerini elde edebilir.

Şekil 4.12’de, Marmara bölgesi ve çevresinde yapılan bir seçim işlemi görülmektedir. Özellikle 17 Ağustos 1999 Gölçükte meydana gelen ve büyük miktarda maddi ve manevi hasara yol açan 7.4 büyüklüğündeki deprem bölgesi seçim alanına dahil edilmiştir.



Şekil 4.12. Marmara bölgesi ve çevresinde yapılan sorgulama ve istatistiksel çalışma

Şekil 4.13’de, Marmara bölgesi ve çevresinde 01.01.1960 ile 01.01.2008 tarihleri arasında meydana gelen deprem sayılarının yıllık dağılım grafiği gösterilmiştir.



Şekil 4.13. 01.01.1960–01.01.2008 tarihleri arasında Marmara bölgesi ve çevresinde meydana gelen $M_c \geq 4,0$ büyüklüğündeki depremlerin yıllık sayı dağılımları

Seçim bölgesinde 01.01.1960–01.01.2008 tarihleri arasında toplam 578 adet deprem meydana gelmiştir. 1967 ve 1999 yıllarında deprem sayılarının bu bölgedeki en yüksek değerlere ulaştığı gözlemlenmiştir. Diğer tarihlerde genel olarak deprem sayısının 10 ve 10'un altında olduğu görülmüştür. Bu bölge üzerindeki deprem hareketliliklerinin yaklaşık 32 yıllık bir süre ile yüksek değerlere çıktığı görülmektedir.

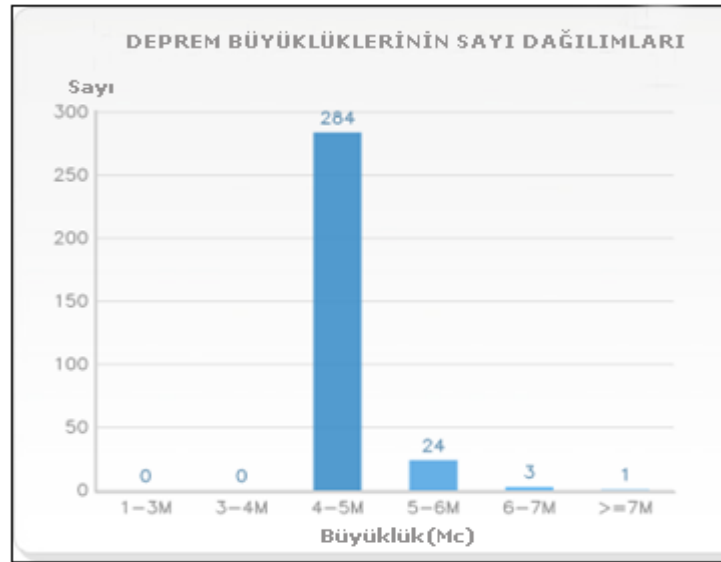
Şekil 4.14'te, Marmara bölgesi ve çevresinde 01.01.1960–01.01.2008 tarihleri arasında $M_c \geq 4.0$ büyüklüğünde meydana gelen depremlerin yıllık ortalama deprem büyüklük grafiği gösterilmiştir.



Şekil 4.14. 01.01.1960–01.01.2008 tarihleri arasında Marmara ve çevresinde oluşan $M_c \geq 4$ büyüklüğündeki depremlerin yıllık ortalama büyüklük grafiği

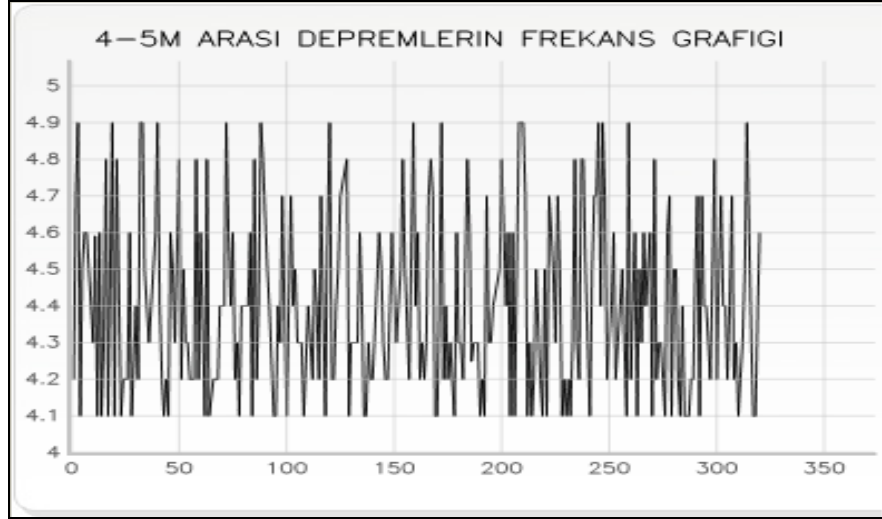
Bu tarih aralıklarında, seçim bölgesindeki depremlerin ortalama büyüklükleri 4.46 olarak tespit edilmiştir. Elde edilen grafikte, Marmara bölgesi ve çevresinde $M_c \geq 4.0$ büyüklüğünde depremlerin 1964'ten 2001 yılına kadar her yıl düzenli olarak meydana geldiği fakat 2001 yılından sonraki yıllarda bu bölgede bazı dönemlerde büyüklüğü 4 ve 4 üstünde hiçbir depremin meydana gelmediği görülmektedir. 1999 yılında bu bölge üzerinde meydana gelen 7.4 büyüklüğündeki Gölcük depremi bu devamlılığın bozulmasında bir etki olarak gösterilebilir.

Şekil 4.15'de, bu bölge üzerinde meydana gelen $M_c \geq 4.0$ büyüklüğündeki depremlerin büyüklüklere göre sayı dağılımları gösterilmiştir.



Şekil 4.15. 01.01.1960–01.01.2008 tarihleri arasında Marmara ve çevresindeki $M_c \geq 4$ büyüklüğündeki depremlerin büyüklük aralıklarına göre sayı dağılımları

Bu bölge üzerinde 01.01.1960 ile 01.01.2008 tarihleri arasında farklı büyüklüklerde toplam 312 adet deprem meydana gelmiştir. Bu depremlerin çoğunluğunu 4–5 arası büyüklükler oluşturmaktadır. 1 adet 7 büyüklüğünde deprem görülmektedir. Bu deprem, Türkiye'nin yakın tarihinde en fazla hasara yol açan 17 Ağustos 1999 Gölcük depremi olarak kayda geçmiştir. Şekil 4.16'da, belirlenen tarihler arasında bölge üzerinde meydana gelmiş olan 4-5M büyüklük aralığındaki depremlerin frekans grafiği gösterilmiştir.



Şekil 4.16. Marmara ve çevresinde meydana gelen 4-5M büyüklüğündeki depremlerin frekans grafiği

Türkiye Deprem Portalı büyük ölçekteki alanlar üzerinde çalıştığı gibi daha ayrıntılı ve alan olarak küçük olan bölgeler üzerindeki hareketliliklerin sorgulanması işleminde de rahatlıkla uygulanmaktadır.

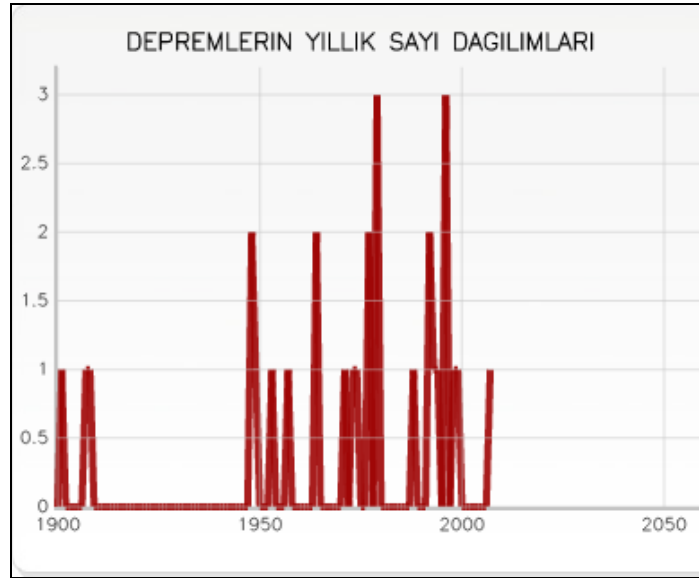
Şekil 4.17’de, 08.03.2010 tarihinde Elazığ Karakoçan’da meydana gelen ve 42 kişinin ölümüne neden olan 6.0 büyüklüğündeki depremin olduğu bölgeyi kapsayan alan üzerinde yapılan seçim işlemi gösterilmiştir.



Şekil 4.17. Sorgulama işlemlerinin daha küçük ölçekteki bölgeler üzerinde uygulaması

01.01.1900–01.01.2008 tarihleri arasında seçilen bu bölge üzerinde farklı büyüklüklerde toplam 99 adet deprem meydana gelmiştir. Bu depremlerden 55 tanesi 3-4M 37 tanesi 4-5M ve 7 tanesi 5-6M büyüklük aralığında oluşmuştur. Bölgedeki ortalama deprem büyüklüğü bu tarihler arasında 4.03 M olarak tespit edilmiştir.

Şekil 4.18’de, seçimi gerçekleştirilen bu bölgede 01.01.1900 ve 01.01.2008 tarih aralıklarında oluşan depremler ile ilgili grafikler gösterilmiştir.



(a)



(b)

Şekil 4.18. 01.01.1900–01.01.2008 tarih aralığında seçimi yapılan bölgeye ait istatistiksel grafikler

- a) Bölgede meydana gelen depremlerin yıllara göre sayı dağılımları
- b) Bölgede meydana gelen depremlerin yıllara göre ortalama büyüklük dağılımları

Şekil 4.18a'da, bu bölge üzerindeki deprem sayısının son 15 yıl içerisinde diğer yıllara oranla arttığı gözlemlenmiştir. 1999 yılı öncesinde bu bölgenin hareketli olduğu gözlemlenmektedir. Şekil 4.18b'de görüldüğü gibi 1996 yılına kadar bu bölgede belirli zaman periyotlarında hiç deprem meydana gelmediği fakat 1996 yılından sonra bu bölgede her yıl farklı sayılarda depremlerin meydana geldiği görülmektedir.

4.6. Bölüm Değerlendirmesi

Tezin bu bölümünde, deprem verileri üzerine gerçekleştirilecek istatistiksel ve veri madenciliği çalışmaları için tasarlanan Türkiye Deprem Portalı uygulaması ile ilgili bilgiler verilmiştir. Hazırlanan portal uygulaması, deprem verilerinin etkili bir şekilde sorgulanabilmesi ve bu veriler üzerinde çeşitli bilimsel çalışmaların yapılabilmesine imkân tanıyan etkili bir araç olarak geliştirilmiştir. Geliştirilen portal bünyesinde, deprem verileri üzerinde gerçekleştirilebilecek istatistiksel çalışmalar için hazırlanan arayüz gösterilmiştir. İstatistiksel bilgilerin sunumunda portal üzerine entegre edilen çeşitli grafiklerden yararlanılmıştır. Bu grafikler sayesinde istatistiksel sonuçların gözlemlenmesi ve değerlendirilmesi işlemi kolaylaştırılmıştır.

Bu bölümde, seçimi yapılan bazı bölgeler üzerinde yapılan çeşitli istatistiksel ve sorgulama çalışmaları sunulmuştur. Elde edilen sonuçlar değerlendirildiğinde, depremler ile ilgili istatistiksel çalışmaların, gelecekte meydana gelebilecek depremler hakkında önemli bilgiler verdiği gözlemlenmiştir. Depremlerin hangi sıklıkta gerçekleştiği ve hangi bölgelerde yoğun olarak olduğu gibi bilgilerin sunulması açısından yapılan bu çalışma başarılı sonuçlar vermiştir.

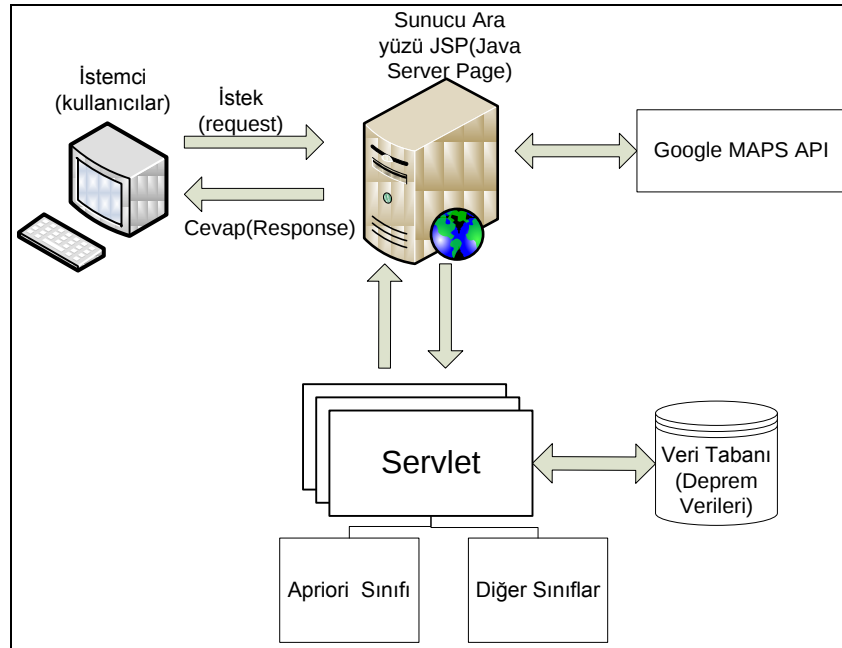
Tasarlanan sistem, sadece tek kullanıcının faydalanabileceği bir yazılım olarak değil bu alanda çalışma yapan bütün kullanıcıların kullanabileceği bir araç olarak dizayn edilmiştir. Kullanıcılar bu sistem üzerinde istedikleri herhangi bir bölgede deprem verilerinin sorgulamasını kolaylıkla gerçekleştirebilir ve bu verilere ait istatistiksel bilgiler elde edebilirler. Gerek kullanımı açısından gerekse de sunduğu bilgi açısından hazırlanan bu sistem, deprem verileri ile ilgilenen kullanıcılar için etkili bir araç niteliği taşımaktadır.

5. DEPREM VERİLERİ ÜZERİNDE BİRLİKTELİK UYGULAMASI

5.1 Giriş

Bu bölümde, deprem alanlarının birbirleri arasındaki ilişkilerinin tespit edilmesine yönelik olarak veri madenciliği yöntemlerinden birliktelik kuralları, deprem verileri üzerine Türkiye Deprem Portalı (TDEP) aracılığı ile uygulanmıştır. Birliktelik kurallarının bulunması için literatürde sıklıkla kullanılan Apriori algoritması deprem verileri için çalışacak şekilde modellenmiştir. Hazırlanan kural çıkarım arayüzü ile kullanıcılar çalışacakları bölgeyi, büyüklük aralıkları, kural çıkarım tarihleri gibi parametrelerin seçimini kolaylıkla yapabilmekte ve elde edilen sonuçları “txt” formatında alabilmektedir. Geliştirilen uygulama, parametre seçiminin görsel olarak yapılması ve birlikteliklerin internet üzerinden bulunmasında etkili bir araç olarak tasarlanmıştır.

Şekil 5.1’de, birliktelik kural üretimi için hazırlanan sistemin çalışma mimarisi gösterilmiştir.



Şekil 5.1. Birliktelik kural üretimi için hazırlanan sistemin çalışma mimarisi

Kullanıcılar sunucu üzerinden sisteme bağlanarak kural üretimi için çeşitli parametre seçimlerini yapar. Bu parametreler doğrultusunda JSP(Java Server Page) ile hazırlanan

web ara yüzü Ajax teknolojisini kullanarak istekleri Servlete iletir. Servletler gelen istekleri değerlendirerek çalıştırılması gereken sınıflar ve veri tabanları ile iletişimi sağlar. Sınıfların işlediği veriler ve bulunan sonuçlar JSP sunucu arayüzü ile kullanıcılara ulaştırılır. Çalışma bölgesinin görsel olarak görülebilmesi ve deprem verilerinin bulunduğu bölgenin detaylı bir şekilde sunulması için Google haritaları uygulaması web arayüzüne dahil edilmiştir.

Şekil 5.2’de, Türkiye Deprem Portalı içerisinde Apriori uygulamasının çalışan arayüzü görülmektedir.

TÜRKİYE DEPREM PORTALI

SORGULAMA ALANI

Birlikte Kuralın için harita üzerinde ilgili bulmak istediğiniz alan fare yardımı ile seçiniz. Bu alanı belirleyeceğimiz aralıklarda kutulara bölmek için Alanı böl butonunu kullanınız. Başlangıç ve bitiş tarihi, belirlediğiniz alandaki kutular arası kural taraması için algoritmanın referans alacağı tarih aralıklarıdır. Zaman aralığı miktarı ise deprem hareketliliklerinin belirlenen aralıklardaki aylık cinsinden periyot aralığıdır. Tüm parametreleri belirledikten sonra Kuralı üret butonunu yardımı ile bu alandaki kuralı görebilirsiniz. Seçim alanınıza bağlı olarak kural üretimi biraz zaman alabilir. Sayfayı kapatmadan bekleyiniz.

Başlangıç Tarihi: 1900-01-01
 Bitiş Tarihi: 1985-01-01
 Zaman Aralığı (Aylık): 24
 Sıddet Aralığı (Mc): 4 ile 8
 Grid Aralığı Belirleyiniz: 1
 Destek Değeri (Support): 25
 Güven Değeri (Confidence): 75

Kutuları Oluştur Kuralı Üret

Sadece Seçim Alanındaki Depremleri Sorgula

Tüm Depremleri Göster Sorgula

Enlem1: 28 45456964375 Boylam1: 36 96500309285596
 Enlem2: 33 92578125 Boylam2: 40 83043687764923

Deprem Sıddet Simgeleri: 1-3M 3-4M 4-5M 5-6M 6-7M 7-8M

Şekil 5.2. Türkiye Deprem Portalı'nın apriori uygulaması için çalışan web ara yüzü

Hazırlanan bu arayüzün sol kısmında sorgulama paneli bulunmaktadır. Bu panel içerisinde kullanıcıların belirleyebileceği çeşitli parametreler bulunmaktadır. Programın orta kısmı harita uygulamasından oluşmaktadır. Harita üzerinde çeşitli seçim işlemleri ve görsel veriler bulunmaktadır. Depremler enlem ve boylam bilgisine göre programın sağ alt kısmında yer alan renk kodları ile harita üzerinde gösterilmektedir. Ajax teknolojisi kullanılarak hazırlanan mesaj kutusu, çalışma bölgesi üzerinde bulunan sık geçen öğeler ve üretilen kuralları ön bilgi olarak kullanıcıya sunulmaktadır. Daha sonra kullanıcı isterse bu verileri farklı formatlarda kayıt edebilmektedir.

5.2. Kural Üretimi İçin Parametre Seçimi

Birliktelik algoritmasının veriler üzerinde çalışması için çeşitli parametre seçimleri yapılmalıdır. Hazırlanan sistemde tüm bu işlemler için bir panel oluşturulmuştur. Kullanıcılar bu paneli kullanarak algoritmanın gereksinim duyduğu bütün parametreleri bu panel aracılığı ile belirleyebilmektedir.

Şekil 5.3'te, web arayüzünde parametre seçimi için oluşturulan panel yapısı görülmektedir.

Başlangıç Tarihi	1900-01-01
Bitiş Tarihi	2010-01-01
Zaman Aralığı (Aylık)	24
Şiddet Aralığı(Mc)	4 ▼ ile 8 ▼
Grid Aralığı Belirleyiniz	1
Destek Değeri (Support)	25
Güven Değeri (Confidence)	90
Kuruları Oluştur	Kuralları Üret
Sadece Seçim Alanındaki Depremler	Sorgula
Tüm Depremleri Göster	Sorgula

Şekil 5.3. Parametre seçimi için portal üzerinde hazırlanan panel görüntüsü

Başlangıç ve bitiş tarihleri, kural tespiti için algoritmanın hangi aralıktaki deprem verilerini değerlendireceğini belirleyen kısımdır. Algoritma, kural üretimi yaparken bu tarihler arasındaki deprem verilerini veri tabanından alır ve belirlenen formata uygun şekilde giriş verisi olarak değerlendirir.

Zaman aralığı, başlangıç ve bitiş tarihleri arasında bir periyodu temsil eder. Bu kısım aylık olarak girilir. Belirlenen zaman aralığına göre veriler alınırken başlangıç ve bitiş tarihi arasında bu periyoda bağlı deprem verileri alınır. Örneğin 24 aylık bir zaman aralığı seçiminde program 01.01.1900 den başlayarak 2 şer yıllık aralıklarla bitiş tarihine kadar verileri değerlendirir.

Büyükölük aralıđı, depremlerin oluřtukları büyükölüklere göre algoritmaya dahil edilip edilmemesini belirleyen parametredir. 4–8M aralıđında seğıildiđini varsayarsak algoritmaya sadece bu řiddet aralıđındaki deprem verileri girdi olarak sunulacaktır.

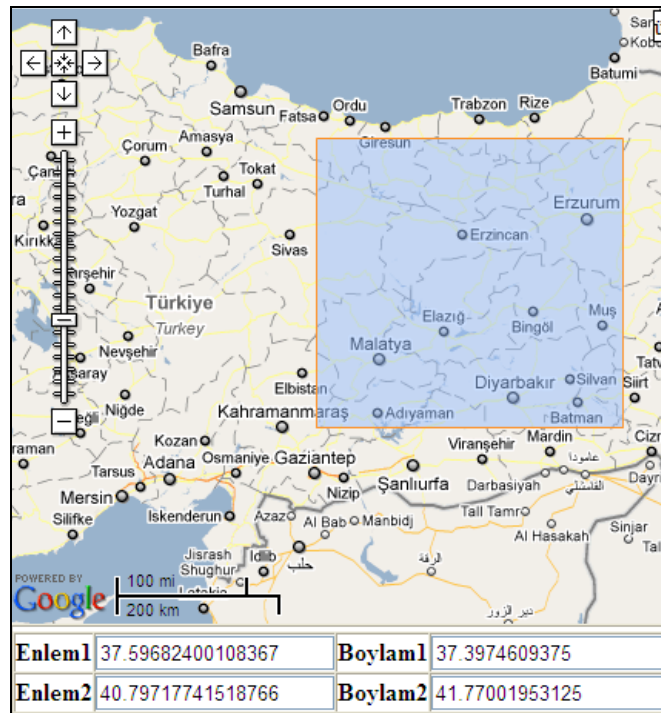
Grid aralıđı kısmında seğıilen bölgenin eřit aralıklı hücrelere bölünmesinde kullanılacak aralık seğıiminin yapıldıđı kısımır. Örneđin grid aralıđı 1° seğıildiđinde 1x1 derecelik hücreler ile seğıim alanı ızgara řeklinde karelere bölünecektir.

Destek ve güven deđerleri apriori algoritmasının kural üretiminde kullandıđı parametrelerdir. Tezin veri madenciliđi bölümünde bu konu ile ilgili ayrıntılı açıklama verilmiřtir.

5.3 alışma Bölgesinin Seğıimi ve Hücrelere Bölünmesi

Kural üretimi yapılacak bölge, kullanıcının isteđine bađlı olarak harita üzerinden fare yardımı ile seğıilebilmektedir. Seğıilen bölgeye ait enlem ve boylam bilgisi oluřturulan “Hucre.java” sınıfına gönderilerek Java Script ve Ajax teknolojileri ile alışma alanı sisteme girdi olarak verilir.

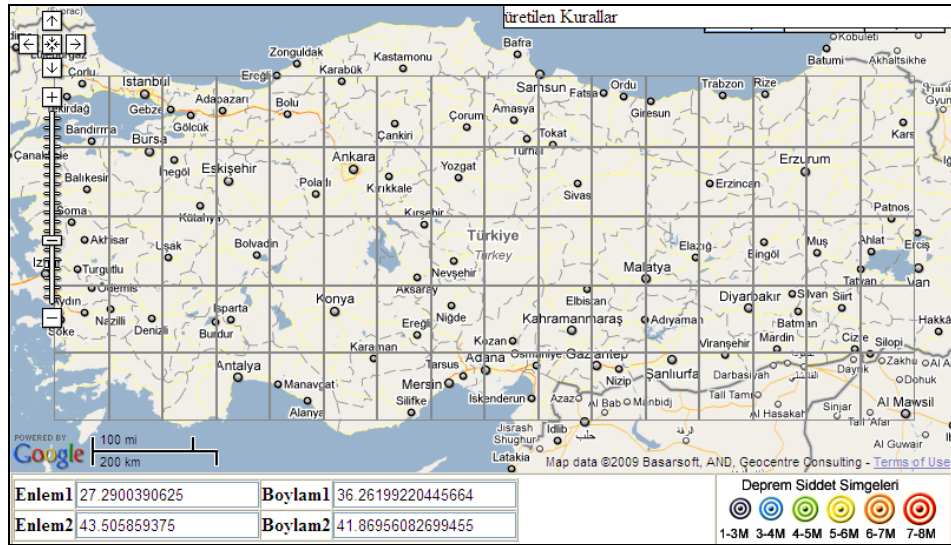
řekil 5.4’de, alışma alanı seğıim işleminin nasıl gerekleřtirildiđini gösteren bir örnek gösterilmiřtir.



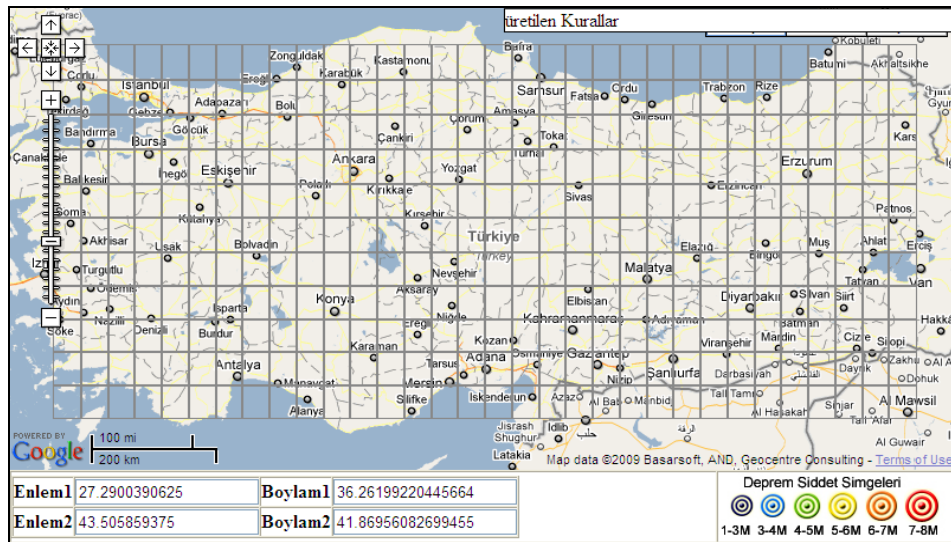
řekil 5.4. alışma alanının portal arayüzü üzerinde seğıimi

Birliktelik algoritmasının uygulanacağı alan seçimi için alt ve üst sınırlar fare yardımı ile belirlenir. Seçimi yapılan bölgeye ait koordinat bilgisi alt kısımda bulunan metin kutuları içerisinde farenin hareketi ile eş zamanlı olarak gösterilir.

Belirlenen bu çalışma alanı daha sonra belli aralıktaki hücelere bölünür. Algoritma verileri değerlendirirken oluşturulan bu hüceler arasındaki birliktelikleri bulmaya çalışır. Şekil 5.5'te, seçimi yapılan bir bölgenin farklı derecelere bölünmüş şekli gösterilmiştir.



(a)



(b)

Şekil 5.5. Çalışma bölgesinin farklı derecelerde hücre aralıklarına bölünmesi

- a) Bölgenin 1x1 derecelik hücelere bölünmesi
- b) Bölgenin 0.5x0.5 derecelik hücelere bölünmesi

5.4. Deprem Verilerin Elde Edilmesi ve Uygun Formata Dönüştürülmesi

Deprem verilerinin kataloglardan alınıp algoritmaya girdi olarak sunulması amacı ile PostgreSQL veri tabanı bağlantısı kurulmuştur. Deprem verileri genel olarak enlem, boylam, büyüklük ve tarih bilgilerini içermektedir. Apriori algoritması boolean bir veri tabanı üzerindeki birliktelikleri tespit etmek için kullanılır [40]. Bu nedenle deprem verilerinin boolean bir tipe dönüştürülmesi gerekir. Tablo 5.1’de, deprem verilerinin algoritmada giriş verisi olarak kullanılabilmesi için boolean tipte gösterimi yapılmıştır.

Tablo 5.1. Deprem verilerinin boolean tipte gösterimi

Zaman	Alan1	Alan2	Alan3	Alan4	Alan5	...
1900	1	0	1	1	0	...
1901	1	1	0	0	1	...
1902	0	0	1	1	1	...
1903	0	1	0	1	0	...
1904	1	1	1	0	1	...
1905	0	0	1	1	0	...
...	...	...	...	...	...	...

Zaman kolonunda, kural üretiminin yapılacağı başlangıç tarihinden başlanarak bitiş tarihine kadar zaman aralığı periyodundaki tarihler sıralanır. Bu tarihlerde çalışma hücrelerinde belirlenen büyüklük aralığındaki deprem verilerinin oluşup oluşmama durumlarına göre veriler boolean formata dönüştürülür. Buradaki alan niteliklerinin her biri çalışma bölgesi üzerindeki hücreleri temsil etmektedir.

5.5. Birlikteliklerin Bulunması İçin Geliştirilen Algoritma Adımları

Bu kısımda apriori algoritmasının deprem verileri üzerine uygulanması ile ilgili yapılan işlemler adımlar şeklinde anlatılmıştır.

Şekil 5.6’da, çalışma alanları arasındaki deprem ilişkilerinin tespitine yönelik geliştirilen kural üretim algoritmasının uygulama adımları gösterilmiştir.

```

Algoritma: Birliktelik Kural Üretimi Algoritması
1   Çalışma alanını belirle: Bolge
2   Çalışma bölgesini hücrelere böl: Nb
3   Başlangıç tarihini belirle: t0
4   Bitiş tarihini belirle: t1
5   Zaman periyodu belirle: tp
6   Büyüklük aralığı belirle: sid
7   Destek ve güven değerlerini belirle: sup, conf
8   Sat=0;
9   while t0 <>t1 do
10      for k:=1 to N do
11          if Deprem(k,t0,t1,sid)=1 then
12              Apdizi(sat,k):=1;
13          else
14              Apdizi(sat,k):=0;
15          end-if
16      end-for
17      sat:=sat+1;
18      t0:=t0+tp;
19  end-while
20  kurallar:=Apriori(sup,conf,Apdizi);
21  kuralları kaydet ve çıktısını ver

```

Şekil 5.6. Hücreler arasındaki birlikteliklerin tespit edilmesi için geliştirilen algoritma

1.Adım: Çalışma alanının seçimi yapılır. Bu kısımda deprem hareketliliklerine göre kural çıkarımı yapılmak istenen alanın seçimi yapılır. Bu seçim işlemi önceki deprem verileri veya uzman görüşü paralelinde yapılabilir.

2.Adım: Çalışma alanı belirli aralıktaki hücrelere bölünür. Bu bölme işleminin amacı her bir alanın veri tabanında farklı bir nitelik olarak temsil edilmesi ve bu hücreler üzerinde algoritmanın değerlendirme yapmasıdır. Bölme işlemi seçilen alan üzerinden yapılmaktadır. Sistem belirtilen aralığa göre çalışma alanındaki hücre sayısını tespit eder ve her bir hücreyi numaralandırarak veri tabanında enlem ve boylam bilgilerini saklar.

3.Adım: Başlangıç tarihi, bitiş tarihi ve zaman periyodu seçimi işlemleri yapılır. Algoritma başlangıç tarihinden başlayarak zaman periyodu kadar tarihi arttırır ve bu tarihte veri tabanından, tüm hücrelerdeki deprem hareketliliğini sorgular. Bu sorgular bir dizi içerisinde 1 ve 0'lar şeklinde tutulur.

4.Adım: Destek ve güven değerlerinin seçimi yapılarak Apriori sınıfı çalıştırılır ve kural üretimi yapılır.

Son olarak üretilen bu kurallar kullanıcılara farklı formatlarda sunulur ve bu alanlar arasındaki ilişkilerin tespit edilmesi işlemi tamamlanır.

Algoritmada kullanılan “Apriori” sınıfı veri madenciliği bölümünde anlatılan algoritma adımlarını içerecek şekilde Java programlama dili ile bir sınıf haline dönüştürülmüştür. Bu sınıfa gönderilen parametreler ile algoritma çalıştırılarak kural üretimi yapar ve elde edilen sonuçları sunar.

5.6. Deprem Verileri Üzerinde Birliktelik Uygulaması

Apriori algoritmasının deprem verilerine uygulanması için belirlenen örnek çalışma alanı Şekil 5.7’de, gösterildiği gibi birer derecelik aralıklara bölünmüş ve her bir hücre veritabanında nitelik olarak kullanılmıştır.

TÜRKİYE DEPREM PORTALI

SORGULAMA ALANI

Birliktelik Kuralları için harita üzerinde ilişki bulmak istediğiniz alanı fare yardımı ile seçiniz. Bu alanı belirleyeceğiniz aralıklarda kutulara bölmek için Alanı böl butonunu kullanınız. Başlangıç ve bitiş tarihi, belirlediğiniz alandaki kutular arası kural taraması için algoritmanın referans alacağı tarih aralıklarıdır. Zaman artış miktarı ise deprem hareketliliklerinin belirlenen aralıklardaki aylık cinsinden periyot aralığıdır. Tüm parametreleri belirledikten sonra Kuralları üret butonunu yardımı ile bu alandaki kuralları görebilirsiniz. Seçim alanınıza bağlı olarak kural üretimi biraz zaman alabilir. Sayfayı kapatmadan bekleyiniz.

Başlangıç Tarihi	1900-01-01
Bitiş Tarihi	1984-01-01
Zaman Aralığı (Aylık)	12
Şiddet Aralığı (Mc)	4 ile 8
Grid Aralığı Belirleyiniz	1
Destek Değeri (Support)	25
Güven Değeri (Confidence)	75
Kutulan Oluştur Kuralları Üret	
Sadece Seçim Alanındaki Depremler	Sorgula
Tüm Depremleri Göster	Sorgula

Enlem1 26.08154296875 Boylam1 35.99578538642032
Enlem2 32.10205078125 Boylam2 42.13082130188811

Şekil 5.7. Kural üretimi için belirlenen örnek çalışma bölgesi

Belirlenen zaman aralıklarında çalışma bölgesinde birer yıllık aralıklarla büyüklüğü $mc \geq 4$ olan depremlerin sorgulanması yapılmıştır. Hücrelerde sorgulanan tarih aralığında

$mc \geq 4$ büyüklüğünde deprem meydana gelmişse bu durum veritabanında “1” olarak temsil edilir aksi durumda “0” olarak temsil edilir. Bu şekilde algoritmanın uygulanması için boolean tipte bir veri tabanı elde edilir. Bu veri tabanı apriori sınıfına giridi olarak sunulmuş ve algoritma bu verileri işleyerek kural üretimi işlemlerini gerçekleştirmiştir.

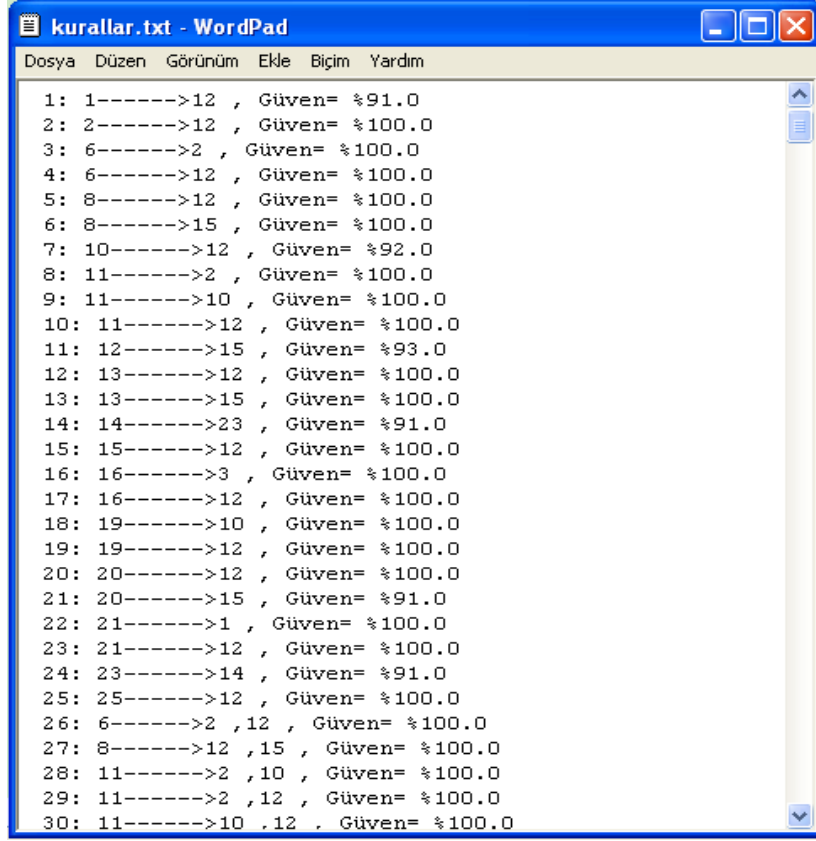
Hedeflenen amaç bu alanlar arasındaki deprem hareketliliğine yönelik ilişkilerini belirleme ve test zamanı olan 01.01.1984–01.01.1995 ve 01.01.1995–01.01.2001 yılları arasında sistemin belirlediği kuralların oluşma oranının gözlemlenerek çalışma bölgesinin karakteristiğini belirlemektir.

Tablo 5.2’de, 01.01.1900–01.01.1984 tarihleri arasında çalışma bölgesindeki alanlar arasında bulunan birliktelik kurallarının bir kısmı gösterilmiştir.

Tablo 5.2. 01.01.1900–01.01.1984 aralığında çalışma bölgesinde bulunan kuralların bir kısmı

No	KURAL	GÜVEN
1	A7 $\longrightarrow$ A2	%99
2	A1 & A7 $\longrightarrow$ A2	%99
3	A1 & A20 $\longrightarrow$ A10	%99
4	A2 & A18 $\longrightarrow$ A3	%99
5	A2 & A16 $\longrightarrow$ A10	%99
6	A3 & A15 $\longrightarrow$ A10	%99
7	A4 & A15 $\longrightarrow$ A6	%99
8	A9 & A21 $\longrightarrow$ A13	%99
9	A1 & A20 $\longrightarrow$ A2 & A10	%99
10	A2 & A3 & A9 $\longrightarrow$ A13	%99
...	...	...

Üretilen bu kurallar tezde bahsedildiği gibi kullanıcıya farklı formatlarda sunulmaktadır. Şekil 5.8’de, üretilen kuralların “txt” formatında kaydedilmiş bir örneği görülmektedir.



```
1: 1----->12 , Güven= %91.0
2: 2----->12 , Güven= %100.0
3: 6----->2 , Güven= %100.0
4: 6----->12 , Güven= %100.0
5: 8----->12 , Güven= %100.0
6: 8----->15 , Güven= %100.0
7: 10----->12 , Güven= %92.0
8: 11----->2 , Güven= %100.0
9: 11----->10 , Güven= %100.0
10: 11----->12 , Güven= %100.0
11: 12----->15 , Güven= %93.0
12: 13----->12 , Güven= %100.0
13: 13----->15 , Güven= %100.0
14: 14----->23 , Güven= %91.0
15: 15----->12 , Güven= %100.0
16: 16----->3 , Güven= %100.0
17: 16----->12 , Güven= %100.0
18: 19----->10 , Güven= %100.0
19: 19----->12 , Güven= %100.0
20: 20----->12 , Güven= %100.0
21: 20----->15 , Güven= %91.0
22: 21----->1 , Güven= %100.0
23: 21----->12 , Güven= %100.0
24: 23----->14 , Güven= %91.0
25: 25----->12 , Güven= %100.0
26: 6----->2 , 12 , Güven= %100.0
27: 8----->12 , 15 , Güven= %100.0
28: 11----->2 , 10 , Güven= %100.0
29: 11----->2 , 12 , Güven= %100.0
30: 11----->10 , 12 , Güven= %100.0
```

Şekil 5.8. Üretilen kuralların txt formatında gösterimi

01.01.1900–01.01.1984 tarihleri arasında çalışma bölgesinde %25 destek ve %75 güven değerleri ile 220 adet kural bulunmuştur.

Bulunan bu kurallar 01.01.1984–01.01.1995 ve 01.01.1995–01.01.2001 tarihleri arasındaki deprem verileri için çalışma bölgesinde ayrı ayrı test edildiğinde Tablo 5.3’de, gösterilen sonuçlar elde edilmiştir.

Tablo 5.3. Üretilen kuralların test sonuçları

Zaman Aralığı	Oluşan Deprem Sayısı ($mc \geq 4$)	Tespit edilen Kural Sayısı	Uygun Kural Sayısı	Tutarlı Kural Sayısı
01.01.1984 01.01.1995	21571	220	162	141
01.01.1995 01.01.2001	21701	220	220	220

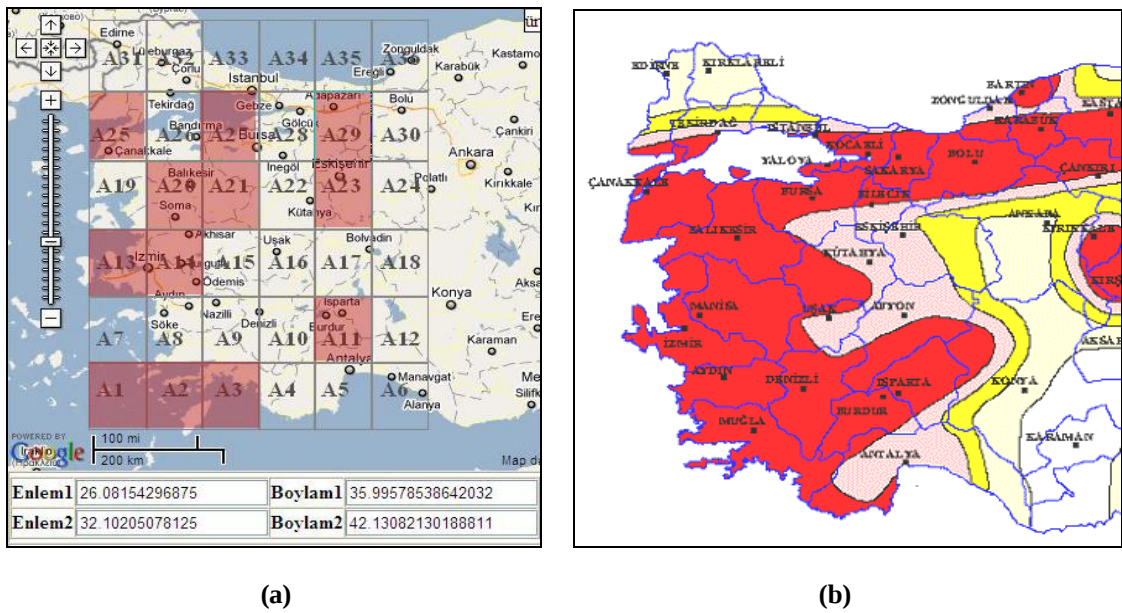
Tespit edilen kural sayısı, 01.01.1900-01.01.1984 tarihleri arasında belirlenen çalışma alanında algoritmanın bulduğu toplam kural sayısını göstermektedir. Uygun kural sayısı alanı, belirtilen test aralığında kural veri tabanındaki test edilebilecek kuralların sayısını temsil etmektedir. Örneğin 1984-1995 yılları arasında test edilecek kural $A1 \& A7 \Rightarrow A2$

kuralı olsun. Bu kuralın uygunluğu şu şekilde test edilmektedir. A1 ve A7 alanında belirtilen tarihler arasında $mc \geq 4$ büyüklüğünde bir deprem meydana gelmişse bu kural test için uygun bir kuraldır. Tutarlı kural alanı ise test edilen bu uygun kuralların doğruluğunu göstermektedir. A1 ve A7 bölgesinde belirtilen test tarihleri aralığında $mc \geq 4$ büyüklüklerinde depremler meydana gelmişse ve bu tarihler arasında A2 bölgesinde de $mc \geq 4$ büyüklüğünde deprem hareketliliği oluşmuşsa bu kural tutarlı kural olarak adlandırılır.

Test verilerine göre çalışma alanında geçmişe dönük keşfedilen kuralların bu bölgede belirli bir zaman diliminde tekrarlandığı ve tüm kuralların belirli bir zaman aralığı içerisinde sağlandığı söylenebilir. Sistemin tespit ettiği tutarlı kural kümesi yardımı ile bölgedeki alanların birbirleri arasındaki sismik hareketlilik ilişkileri belirlenebilir. Bu bölgelerde oluşan bir deprem hareketliliğinin, ilişkili diğer bölgeleri ne derece etkileyeceği hakkında bilgi alınabilir.

Ayrıca bu alan üzerinde sık tekrar eden öğeler bulunarak depremlerin sıklıkla tekrarlandığı ve birbirleri ile ilişkili alanlar, bölgenin karakteristiğini belirlemede uzmana yardımcı olmaktadır.

Şekil 5.9'da, çalışma bölgesi üzerinde apriori algoritmasının tespit ettiği sık geçen öğeler ve çalışma alanının Türkiye deprem haritası üzerindeki gösterimi yapılmıştır.

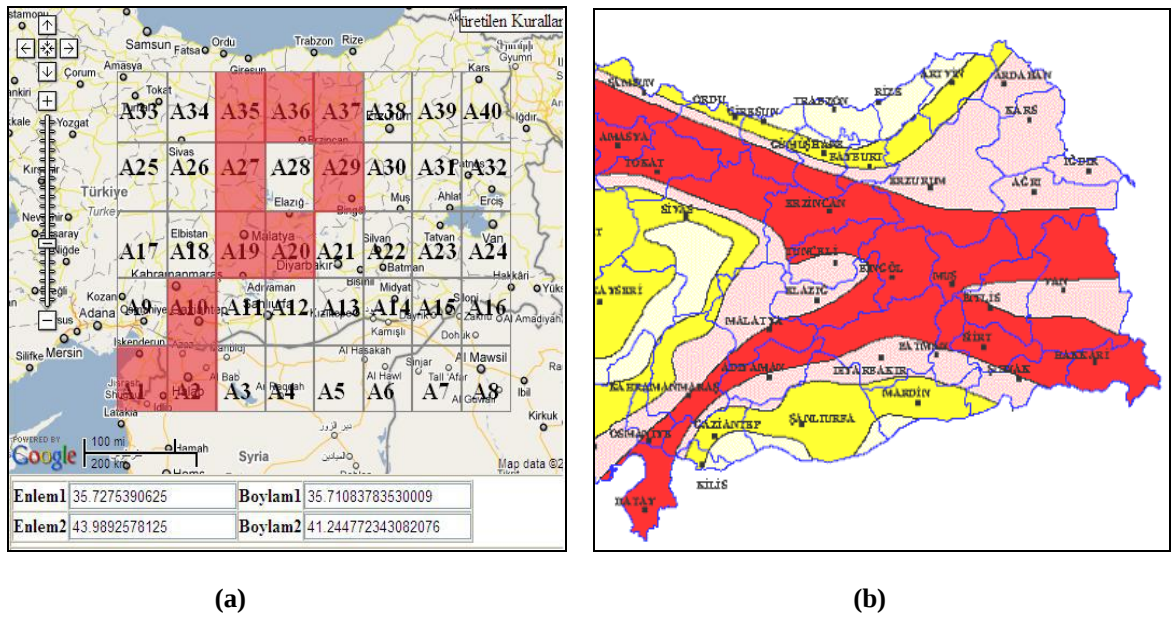


Şekil 5.9. Sık geçen deprem bölgelerinin tespit edilmesi ve Türkiye deprem haritasındaki konumu

- a) Çalışma bölgesi üzerinde sık geçen öğeler
- b) Çalışma bölgesinin Türkiye deprem haritasındaki konumu

Şekil 5.9a’da, çalışma bölgesi üzerinde sık tekrar eden deprem alanları kırmızı renkli hücreler ile gösterilmiştir. Şekil 5.9b’de Türkiye bölgesinin risk derecesine göre deprem alanlarını gösteren haritanın çalışma bölgesini kapsayan kısmı gösterilmiştir. İki şekil karşılaştırıldığında sistemin belirlediği tekrarlı bölgelerin deprem haritası üzerinde 1.dereceden deprem bölgeleri üzerinde olduğu gözlemlenmektedir.

Şekil 5.10’da, Doğu Anadolu bölgesi üzerinde belirlenen çalışma alanındaki sık tekrar eden öğeler ve Türkiye deprem haritasının bu alanı kapsayan kısmı gösterilmiştir.

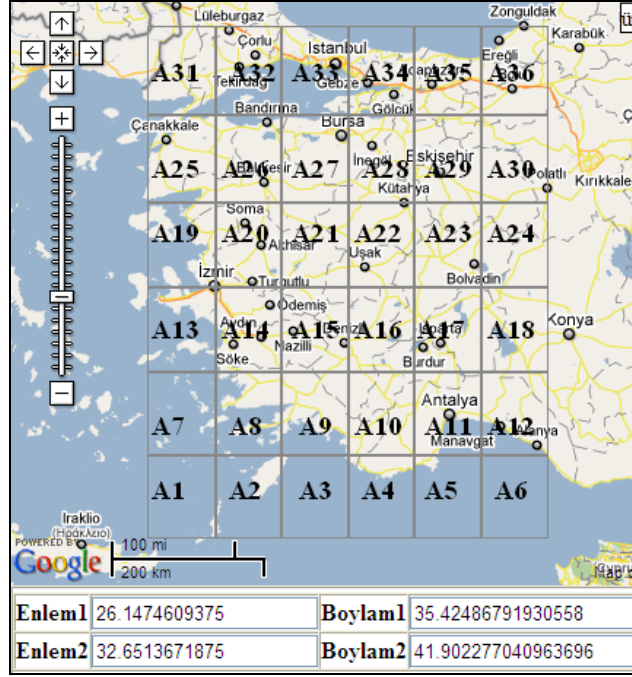


Şekil 5.10. Doğu Anadolu Bölgesi üzerinde belirlenen çalışma alanındaki tekrarlı öğelerin bulunması

- a) Çalışma bölgesi üzerinde bulunan sık geçen öğeler
- b) Türkiye deprem haritasının çalışma alanını kapsayan bölümü

Şekil 5.10a’da, bölge üzerinde bulunan sık tekrar eden deprem alanları Şekil 5.10b’de Türkiye deprem haritası üzerinde 1.dereceden deprem bölgeleri üzerinde olduğu görülmektedir. Türkiye deprem haritası üzerinde gösterilen bazı alanların sık geçen öğe olarak bulunmamasının nedeni bu alanlarda tekrarlı deprem hareketliliğinin olmamasından kaynaklanmaktadır. Fakat bulunan sık geçen alanlar, bu bölge üzerinde en fazla risk taşıyan bölge konumundadır.

Şekil 5.11’de, belirlenen çalışma bölgesi üzerinde 01.01.1964 ve 31.12.2007 tarihleri arasında $m_c \geq 4$ büyüklüğündeki depremler için %80 güven ve %25 destek değerleri ile elde kurallar gösterilmiştir.



(a)

Kural	Güven	Kural	Güven	Kural	Güven
A1→A2	%87	A32→A1,A2	%86	A1,A2,A4→A7	%88
A1→A7	%82	A17→A1,A2,A8	%82	A1,A2,A20→A14	%86
A3→A7	%88	A1,A16→A8	%93	A8,A10,A31→A14,A1	%86
A4→A14	%81	A1,A17→A14	%81	A10,A14,A31→A1,A8	%92
A8→A7	%84	A1,A25→A8	%94	A1,A2,A7,A14→A8	%86
A9→A1	%100	A1,A32→A8	%86	A1,A7,A14,A31→A2,A8	%81
A10→A1	%95	A4,A10→A7	%100	A2,A8,A14,A31→A1,A7	%92
A13→A7	%85	A4,A26→A1	%100	A1,A2,A7,A8,A31→A14	%86
A19→A7	%94	A4,A13→A1,A7	%88	A1,A2,A7,A14,A31→A8	%92
A3→A1,A7	%83	A8,A10→A1,A14	%89	A1,A2,A8,A14,A31→A7	%92
A16→A1,A8	%87	A10,A26→A1,A7	%93	A1,A7,A8,A14,A31→A2	%86
A27→A1,A2	%83	A21,A26→A1,A2	%81	A2,A7,A8,A14,A31→A1	%100

(b)

Şekil 5.11. Çalışma bölgesi üzerinde 01.01.1964 ve 31.12.2007 tarihleri arasında elde edilen kurallar

- c) Çalışma bölgesi seçimi ve hücrelere ayrılması
- d) Çalışma bölgesi üzerinde %80 güven ve %25 destek değeri ile elde edilen kurallar

5.7. Bölüm Değerlendirmesi

Tezin bu bölümünde, deprem alanları arasındaki hareketlilik ilişkilerinin tespit edilmesine yönelik olarak veri madenciliği tekniklerinden birliktelik kuralları deprem verileri üzerine uygulanmıştır. Bu uygulamada, birliktelik kural çıkarım algoritmalarından apriori algoritmasını temel alan bir algoritma geliştirilmiştir. Algoritmanın çeşitli parametre seçimleri ile internet üzerinden deprem verileri için çalışması sağlanmıştır. Dinamik haritalar üzerinde bölge seçim işlemlerinin görsel olarak yapılması ve belirlenen aralıklarda çalışma bölgesinin hücrelere ayrılması uygulamanın büyük avantajlarından. Ayrıca uygulama açısından karmaşık olan algoritmanın, görsel araçlar yardımı ile kolay bir şekilde uygulanabilmesine imkân tanıyan bir arayüz geliştirilmiştir.

Birçok alanda uygulama imkânı bulan birliktelik algoritmalarının, deprem verileri üzerinde de başarılı bir şekilde çalıştığı bu bölüm altında gözlemlenmiştir. Algoritmanın keşfettiği sık geçen öğelerin, Türkiye deprem haritası üzerinde 1.dereceden deprem kuşağındaki alanlara denk geldiği görülmüştür.

Bu bölümde, üretilen kuralların geleceğe yönelik test edilmesi ve tutarlılıklarının gözlemlenmesi için ayrı bir çalışma yapılmıştır. Bu çalışmada elde edilen sonuçlar değerlendirildiğinde, bölge üzerinde elde edilen kuralların belirli bir zaman dilimi içerisinde tamamının gerçekleştiği gözlemlenmiştir. Deprem alanları arasındaki gizli kalmış birlikteliklerin keşfedilmesi, çalışma yapılan bölgenin deprem karakteristiğini belirlemede büyük derecede etkili olduğu gözlemlenmiştir.

Sonuç olarak bu bölüm altında gerçekleştirilen uygulama, birliktelik algoritmasının deprem verileri üzerinde uygulanabilirliğine yönelik bir yaklaşım sunmuştur. Yapılan çalışmalardan elde edilen sonuçlar, birliktelik algoritmasının deprem verileri üzerinde başarılı bir şekilde çalıştığını göstermiştir. Ayrıca bu uygulamanın, tüm kullanıcıların birliktelik algoritmasını deprem verileri üzerinde kolaylıkla kullanabilmesine yönelik geliştirilmiş etkili bir araç olduğu söylenebilir.

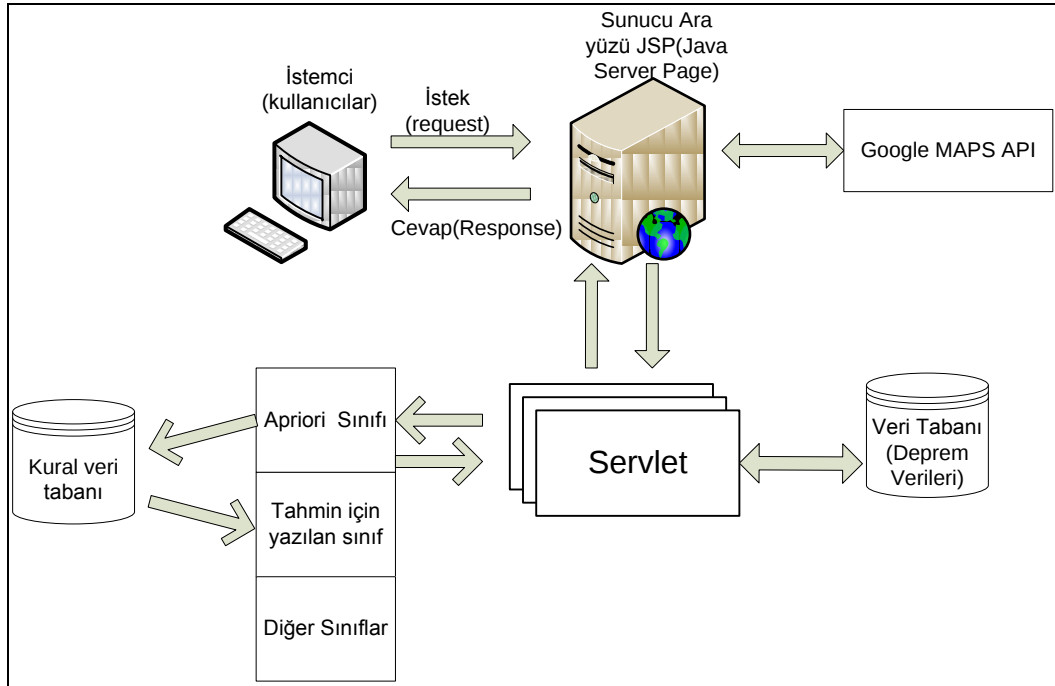
Bu bölüm altında elde edilen sonuçlar ile 1 adet uluslararası konferans yayını ve 1 adet ulusal konferans yayını yapılmıştır[52,53].

6. DEPREM TAHMİNİ İÇİN KURAL TABANLI BİR YAKLAŞIM

6.1 Giriş

Tezin bu bölümünde, deprem bölgelerinde meydana gelebilecek olası sismik hareketlerin tahmini için geliştirilen kural tabanlı bir yaklaşım sunulmuştur. Geliştirilen bu yaklaşım, bölgeler arasındaki birliktelik ilişkilerine dayanmaktadır. Öncelikle algoritma, belirlenen tahmin bölgesi üzerinde kural üretimi yapar. Üretilen bu kurallar, tahmin algoritmasına özellik olarak sunulur. Gerçek zamanlı gelen deprem verileri değerlendirilerek kural veri tabanından bu depremin etki edebileceği ilişkili bölgelerin tespiti yapılır. Depremin olduğu bölge ile ilişkili diğer bölgeler riskli olarak sisteme yansır. Riskli bölgelerde oluşan depremler değerlendirilerek sistemin verdiği uyarıların doğruluğu test edilir. Elde edilen sonuçlar değerlendirildiğinde algoritmanın başarılı tahminler ürettiği gözlemlenmiştir.

Şekil 6.1’de, Türkiye Deprem Portalı’nın kural tabanlı tahmin algoritması için çalışma mimarisi gösterilmiştir.



Şekil 6.1. Portal üzerinde deprem tahmin algoritmasının yerleşimi

6.2 Kural Veritabanının Oluşturulması

Portaldaki apriori sınıfı, seçilen parametreler doğrultusunda çalışma bölgesi üzerindeki kural çıkarımını gerçekleştirmektedir. Üretilen bu kurallar, bir veri tabanında saklanarak tahmin için gerekli örüntü yapısını oluşturur. Algoritma, üretilen bu kurallar yardımı ile gelecekte deprem oluşabilecek riskli alanların tahminini gerçekleştirmektedir.

Şekil 6.2’de, gösterildiği gibi tahmin alanı seçilerek belirlenen aralıklar ile eşit aralıktaki hücelere bölünür. Her bir hücre veri tabanında özellik olarak temsil edilir.



Şekil 6.2 Tahmin bölgesinin seçimi ve hücelere bölünmesi

Kural üretimi için seçilen bölge aynı zamanda sistemin tahmin için kullanacağı bölgeyi temsil etmektedir. Hücre aralıkları, portal ara yüzünde istenilen ölçüde seçilebilmektedir. Bu çalışmada hücre aralıkları 1° olarak seçilmiştir. Algoritmada 4 adet zaman parametresi bulunmaktadır. Bu parametreler, kuralların üretimi için verilerin sıralanmaya başlanacağı tk_0 başlangıç tarihi, kural üretimi için bitiş tarihi olan tk_1 , zaman periyodunu temsil eden t_p tahmin için kullanılacak test verilerinin başlangıç tarihi olan tt_0 ve bitiş tarihi olan tt_1 'dir.

Tablo 6.1’de, kural üretimi için deprem verilerinin düzenlendiği format görülmektedir. Bu tabloda çalışma için ikişer yıllık aralıklarla $mc \geq 4$ büyüklüğündeki sismik veriler veritabanından çekilerek bu depremlerin olduğu bölgeler apriori algoritmasına giriş verisi olarak girilmiştir.

Tablo 6.1. Kural üretimi için deprem verilerinin düzenlenmesi

Tarih ($tk_0=1900.01.01$, $tk_1=1984.01.01$, $t_p=2$)	Deprem Öğeleri
01.01.1900–01.01.1902	{5}{7}{8}{9}{11}{12}{13}{17}{18}{19}{20}{22}
01.01.1902–01.01.1904	{7}{8}{10}{12}{15}
01.01.1904–01.01.1906	{2}{8}{9}{10}{12}{15}{19}{24}{25}
01.01.1906–01.01.1908	{12}{15}{20}
01.01.1908–01.01.1910	{8}{17}{20}
01.01.1910–01.01.1912	{4}{18}
01.01.1912–01.01.1914	{1}{5}{10}{14}{23}
.....	
01.01.1982–01.01.1984	{1}{2}{3}{4}{6}{7}{8}{9}{10}{11}{12}{14}

Çalışma bölgesi üzerinde $tk_0=01.01.1900$, $tk_1=01.01.1984$ ve $mc \geq 4$ büyüklüğündeki sismik veriler $t_p=2$ için oluştukları hücrelere göre apriori algoritmasının çalışabileceği şekilde düzenlenmiştir. Bu tarihler arasındaki sismik veriler, algoritmanın tahminde örüntü olarak kullanacağı kural kümesini temsil etmektedir.

Tablo 6.2’de, $tk_0=01.01.1900$ ile $tk_1=01.01.1984$ tarihleri arasında çalışma bölgesi üzerinde üretilen kuralların bir kısmı gösterilmiştir. Bu örnek çalışmada belirlenen alan üzerinde, minimum güven değeri $minconf=\%75$ ve minimum destek değeri $minsup=\%25$ olmak üzere toplam 98 adet kural bulunmuştur.

Tablo 6.2. Tahmin bölgesi üzerinde üretilen kuralların bir bölümü

Kural ID	Öğeler	Güven (confidence)	Min.Destek(support)
1	{3}→{11}	%76	25
2	{1,3}→{6}	%91.0	25
3	{1,6}→{11}	%85.0	25
4	{6,14}→{11}	%100.0	25
5	{1,11}→{3, 6}	%84.0	25
6	{1, 11, 17}→{6}	%100.0	25
7	{6, 17}→{1, 11 }	%76.0	25
....			
98	{6, 11, 21}→{1}	%100.0	25

Üretilen bu kurallar tahminde kullanılmak üzere hazırlanan bir veri tabanında formatlı bir şekilde saklanmıştır. Tahmin algoritması yürütülürken kuralların kaydedildiği bu veritabanından gerekli bilgileri almaktadır. Yeni bir tahmin bölgesi seçildiğinde veri tabanı içerisindeki kurallar yenilenmektedir.

6.3. Deprem Tahmin Algoritması

Bu kısımda, deprem tahmini için geliştirilen algoritmanın çalışma adımları sunulmuştur. Algoritmada öncelikle geçmiş deprem verileri kullanılarak tahmin bölgesi üzerinde kural üretimleri yapılmaktadır. Üretilen bu kuralların tamamı hazırlanan bir kural veri tabanına kaydedilmektedir. Daha sonra algoritma, gelen deprem verilerini değerlendirerek kural veri tabanından depremin olduğu bölge ile ilişkili diğer bölgelerin tespitini yapmaktadır. Tespit edilen ilişkili bölgeler, gelecekte deprem hareketliliğinin beklendiği bölgeler olarak sisteme yansır. Şekil 6.3'te, tahmin algoritmasında kullanılan kural üretim algoritması gösterilmiştir.

```
Algoritma: Tahmin için kural üretimi algoritması
1 Çalışma bölgesi belirle.
2 Grid aralığı seç ve N adet hücre oluştur.
3 Başlangıç tarihi belirle:  $tk_0$ 
4 Bitiş tarihi belirle:  $tk_1$ 
5 Zaman aralığı belirle:  $t_p$ 
6 Şiddet aralığı belirle:  $sid$ 
7 Minimum destek değeri belirle:  $minsup$ 
8 Minimum güven değeri belirle:  $minconf$ 
9 sat=0;
10 m=0;
11 while  $tk_0 <> tk_1$  do
12   for i=1 to N do
13     if  $tk_0 - t_i$  aralığında i.hücrede  $Mc \geq sid$  deprem varsa
14       apriori_oge[sat][m]=i;
15       m=m+1;
16     end-if
17   end-for
18   sat=sat+1;
19    $tk_0 = tk_0 + t_p$ ;
20 end-while
21 kurallar=Aprioriclass(oge,minsup,minconf)
22 kuralları kural veri tabanına kaydet
```

Şekil 6.3. Tahmin algoritmasının kural üretim bölümü

Şekil 6.4’de, üretilen kurallar paralelinde gelen deprem verileri için gelecek ile ilgili bölgelerin risk tahminini yapan algoritma gösterilmiştir.

Algoritma: Deprem tahmin algoritması

```
1 Tahmine başlanacak zamanı belirle:  $tt_0$ 
2 Tahminin biteceği zamanı belirle:  $tt_1$ 
3 beklenen[],hareketlilik[];
4 while  $tt_0 \neq tt_1$  do
5   for her deprem verisi için do
6     hucre=Depremverisi;
7     if hucre in beklenen[] then
8       write('Riskli Bölge'+hucre+' tahmin edildi');
9     end-if
10    hareketlilik[]=hucre;
11    for her kural için do
12      for kuralın sol kısmı için do
13        if kurallar.sol hareketlilik[] then
14          beklenen[]=kurallar.sag;
15          write(kurallar.sag + 'Riskli Bölgeler');
16        end-if
17      end-for
18    end-for
19  end-for
20   $tt_0 = tt_0 + t_p$ ;
21  clear hareketlilik[], beklenen[];
22 end-while
```

Şekil 6. 4. Deprem tahmini için geliştirilen algoritma yapısı

Riskli bölge tahmininde kullanılan algoritmanın genel uygulama adımlarını basamak şeklinde ifade etmek gerekirse:

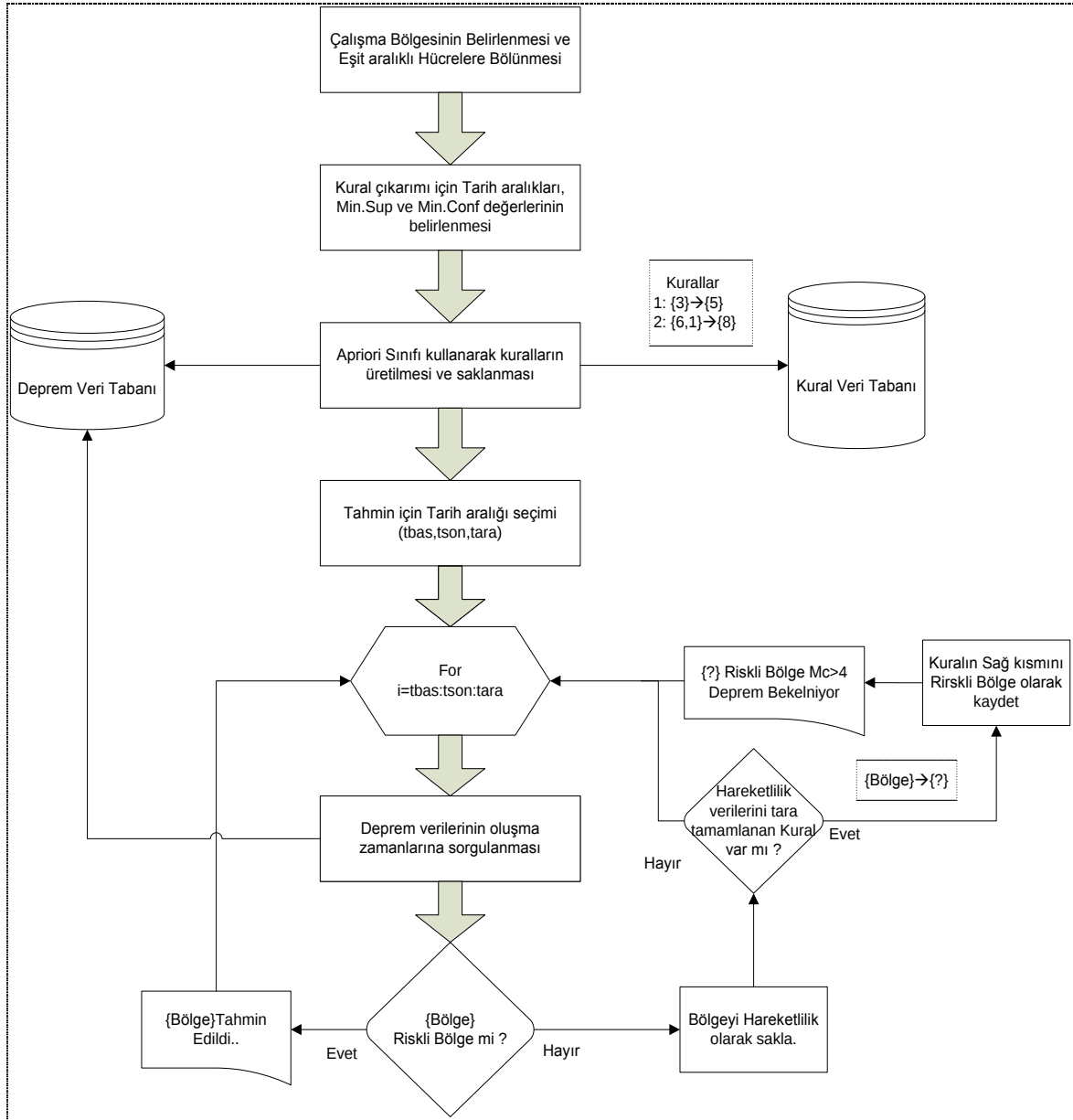
Adım1: Tahmin için sistemin çalışmaya başlayacağı tarih seçilir. Bu tarih aralığından itibaren sistem, veri tabanındaki depremlerin oluşma zamanlarına göre işlenmesini sağlar.

Adım2: Gelen deprem verisi eğer sistemin deprem oluşacağı yönünde uyarı verdiği bölgelerden biriye bu bölgedeki deprem tahmini başarılı olarak kaydedilir.

Adım3: Bu bölge için önceden bir uyarı verilmemiş ise iki yıllık(seçime bağlı) deprem verileri değerlendirilerek kural veri tabanından bu hücrenin tamamladığı bir kural aranır.

Adım4: Bu hücrenin tamamladığı kural için kuralın sağ tarafındaki hücreler riskli bölge olarak sisteme yansır. Bu hücrelerde, seçilen tarih aralığı içerisinde $mc \geq 4$ büyüklüğünde deprem oluşacağı yönünde sistem uyarı verir.

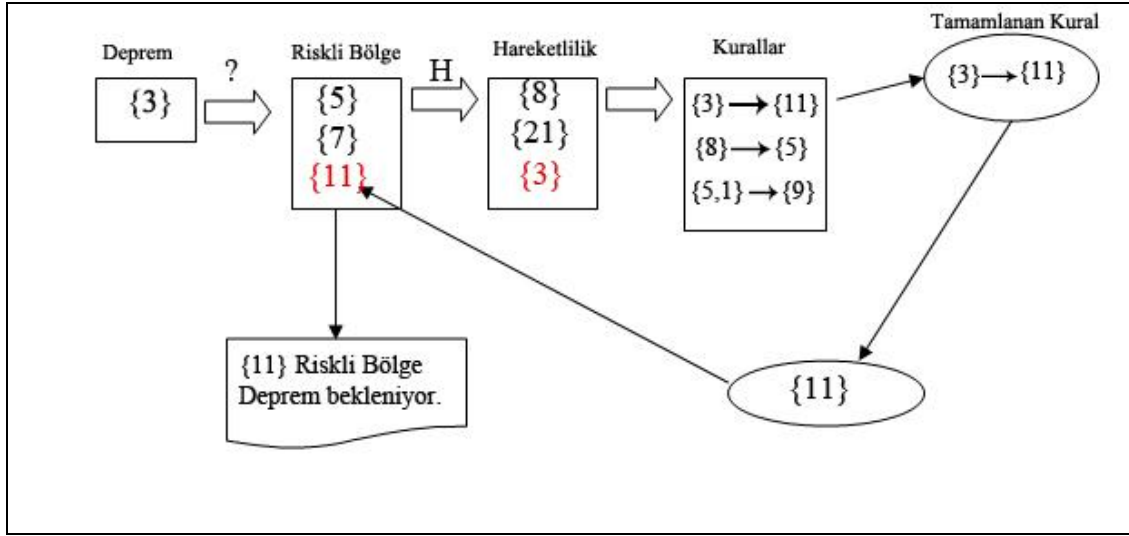
Şekil 6.5'te, deprem tahmin algoritmasının gerçekleştirdiği işlemler akış diyagramı şeklinde gösterilmiştir.



Şekil 6.5. Deprem tahmin algoritması için tasarlanan akış şeması

Öncelikle gelen hücre verisi için algoritma, depremin meydana geldiği hücrenin riskli bölge olarak uyarısının önceden verilip verilmediğine bakar. Eğer bu hücre riskli bölge olarak sistemde işaretlenmişse tahminin önceden yapıldığına dair bir mesaj verir. Eğer bu

hücre riskli bölge olarak işaretlenmemişse hareketlilik verisi olarak kaydedilir. Oluşan yeni hareketlilik verileri değerlendirilerek kural veritabanı taranır ve bu hareketliliklerin tamamladığı kurallar aranır. Bulunan bu kuralların sağ kısmındaki öğeler riskli bölge olarak işaretlenip kaydedilir. Bu bölgelerde, kural çıkarım tarih aralığı içerisinde $mc \geq 4$ büyüklüğünde depremin meydana geleceğinin uyarısı yapılır. Şekil 6.6’da, algoritmanın örnek bir veri için çalışması gösterilmiştir.



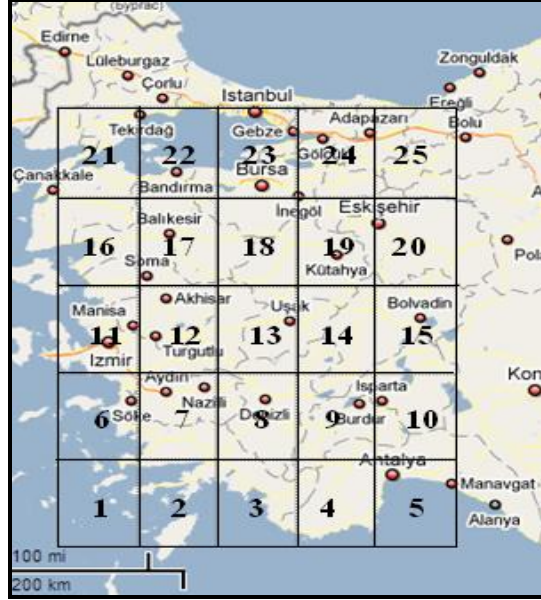
Şekil 6.6. Algoritmanın örnek veri için çalışması

Şekil 6.6’da, tahmin aralığında Sisteme 3. hücrede $mc \geq 4$ büyüklüğünde deprem verisi gelmiştir. Bu hücre riskli bölge içerisinde bulunmadığından dolayı sadece hareketlilik olarak kaydedilir. Daha sonraki aşamada, kural veri tabanı taranarak bu hareketlilik ile birlikte tamamlanan kurallar değerlendirilir. Şekil 6.6’daki örnekte 3.hücredeki deprem verisinin 3 ve 11 numaralı hücreler arasındaki bir kuralı tamamladığı görülmektedir. Bu nedenden dolayı 11 hücrede kural üretimi için kullanılan periyotta $mc \geq 4$ büyüklüğünde bir hareketliliğin oluşacağına dair sistem uyarı vermektedir. Kural üretim için kullanılan periyot aralığında riskli bölgeler ve hareketlilik belleği yenilenmektedir. Her yeni gelen hareketliliğe göre daha önceden oluşmuş depremler göz önüne alınarak kural veri tabanı tekrar taranır ve tamamlanan kural olup olmadığına bakılır.

6.4. Türkiye Deprem Kataloğu Üzerinde Tahmin Çalışması

Bu kısımda, geliştirilen tahmin algoritmasının Türkiye deprem kataloğu üzerinde belirlenen bir çalışma bölgesinde farklı parametreler ile uygulaması yapılmıştır. Bütün

veriler için kural üretim tarih aralığı 01.01.1900–01.01.1985 olarak alınmıştır. Bu tarihler arasında belirlenen büyüklükteki depremler için farklı destek, güven ve zaman aralığına göre kurallar üretilip bu kurallar aracılığı ile gelecekte oluşabilecek depremlerin tahmini yapılmıştır. Şekil 6.7’de, tahmin algoritmasının test edildiği çalışma bölgesi gösterilmiştir.



Şekil 6.7. Deprem tahmin algoritmasının test edildiği çalışma bölgesi

Tablo 6.3’de, $mc \geq 4$ büyüklüğündeki deprem verileri için çalışma alanında farklı parametrelerle yapılan tahmin sonuçları verilmiştir.

Tablo 6.3. 01.01.1985–01.01.1995 test aralığında $mc \geq 4$ büyüklüğündeki depremler için tahmin sonuçları

Min. Support	Min. Confidence	Zaman aralığı	Üretilen Kural Sayısı	Deprem Sayısı ($mc \geq 4$)	Tahmin edilen deprem sayısı	Verilen Uyarı Sayısı	Yanlış Uyarı Sayısı	Uyarı başarıımı
0.25	0.75	24 ay	98	283	164	42	6	%85.71
0.25	0.90	24 ay	24	283	101	24	2	%91.66
0.20	0.75	24 ay	1177	283	199	91	38	%58.24
0.20	0.90	24 ay	383	283	110	43	16	%62.79
0.50	0.90	48 ay	11	283	45	12	2	%83.33
0.50	0.75	48 ay	33	283	75	24	7	%70.83
0.50	0.75	60 ay	746	283	157	40	4	%90.00

Tablo 6.3’de, sistemin zaman aralığının artışına bağlı olarak deprem verileri arasında daha yüksek destek değerlerine sahip kurallar bulunduğu ve bu kuralların paralelinde uyarı başarımının arttığı gözlemlenmektedir. Yüksek güven ve destek değerlerinde bulunan

kuralların güçlü olması, verilen uyarılardaki yanlışlık payını azaltmaktadır. 0.25 destek ve 0.90 güven metriğinde 24 aylık periyotta bulunan kural sayısı 24 ve sistemin başarımı %91.66 iken 0.20 destek ve 0.75 güven değerlerinde bulunan toplam kural sayısı 1177 ve depremler için verilen toplam uyarı sayısı 91’dir. Verilen bu 91 uyarının 38 tanesi gerçekleşmeyen uyarılardır. Sistemin, bu değerler için başarımı %58.24 olarak tespit edilmiştir.

Benzer şekilde, zaman aralığının genişlemesi sistemde sık tekrar eden öğeleri arttırdığından dolayı, geniş zaman aralıklarında kural üretimi daha fazla olmaktadır. Üretilen bu kurallar daha yüksek güven ve destek değerine sahip olduğundan sistemin başarımını yükselmektedir. 0.50 destek ve 0.75 güven değerlerinde 48 aylık periyotta toplam 33 adet kural üretilmiştir. Bu kurallar ile sistem test edildiğinde depremler için toplam 24 uyarı verilmiş ve bu uyarılardan 17 tanesi gerçekleşirken 7 tanesi gerçekleşmeyip sistemin başarımı %70.83 olarak tespit edilmiştir. Aynı destek ve güven değerleri ile zaman aralığı 60 ay olarak genişletildiğinde sistem toplam 746 adet kural üretimi yapmıştır. Üretilen bu kurallar paralelinde geleceğe yönelik depremler için toplam 40 adet uyarı verilmiştir. Bu uyarılardan 36 tanesi gerçekleşirken 4 tanesi gerçekleşmeyip sistemin başarımı %90 olarak belirlenmiştir.

Tablo 6.4’te, $mc \geq 4$ büyüklüğündeki depremler için test aralığı değiştirilerek sonuçlar verilmiştir. Tablo 6.3’ deki destek güven değerleri karşılaştırma için değiştirilmemiştir.

Tablo 6.4. 01.01.1995–01.01.2000 test aralığında $mc \geq 4$ büyüklüğündeki depremler için tahmin sonuçları

Min. Support	Min. Confidence	Zaman aralığı	Üretilen kural sayısı	Deprem Sayısı ($mc \geq 4$)	Tahmin edilen deprem sayısı	Verilen Uyarı Sayısı	Yanlış Uyarı Sayısı	Uyarı başarımı
0.25	0.75	24 ay	98	223	77	21	5	%76.19
0.25	0.90	24 ay	24	223	41	12	4	%66.6
0.20	0.75	24 ay	1177	223	97	41	19	%53.65
0.20	0.90	24 ay	383	223	49	21	7	%66.6
0.50	0.75	48 ay	33	223	89	17	5	%70.58
0.50	0.90	48 ay	11	223	25	5	0	%100
0.50	0.75	60 ay	746	223	145	44	6	%86.36

Tablo 6.4’teki veriler incelendiğinde sistemin Tablo 6.3’deki gibi düşük destek ve güven değerleri ile verilen uyarılardaki yanlışlık oranını arttırdığı gözlemlenmektedir. 0.20 destek ve 0.75 güven değerleri ile 24 aylık periyotta Tablo 6.4’teki başarımlar %58.24 iken benzer şekilde Tablo 6.5’deki başarımlar da %53.65 civarındadır.

Tablo 6.5’de, sistemin $mc \geq 3$ büyüklüğündeki deprem verileri için aynı çalışma bölgesinde farklı parametrelere göre tahmin sonuçları gösterilmektedir.

Tablo 6.5. 01.01.1985–01.01.2000 test aralığında $mc \geq 3$ büyüklüğündeki depremler için tahmin sonuçları

Min. Support	Min. Confidence	Zaman aralığı	Üretilen kural sayısı	Deprem Sayısı ($mc \geq 3$)	Tahmin edilen deprem sayısı	Verilen Uyarı Sayısı	Yanlış Uyarı Sayısı	Uyarı başarımları
0.25	0.75	24 ay	227	912	480	87	13	%85.05
0.25	0.90	24 ay	24	912	322	40	2	%95.00
0.20	0.75	24 ay	2098	912	573	132	29	%78.03
0.20	0.90	24 ay	573	912	382	68	10	%85.29
0.50	0.90	48 ay	14	912	190	20	1	%95.00
0.50	0.75	48 ay	40	912	394	44	5	%88.63

Tablo 6.5’deki veriler incelendiğinde sistemin $mc \geq 3$ deprem hareketlilikleri için uyarı başarımlarının diğerlerine oranla daha yüksek olduğu görülmektedir.

6.5. Bölüm Değerlendirmesi

Tezin bu bölümünde, depremler için sürekli olarak gündeme gelen tahmin çalışmalarına katkıda bulunması amacıyla yeni bir yaklaşım sunulmuştur. Bu yaklaşım, genel olarak deprem alanlarının birbirleri arasındaki hareketlilik ilişkilerini temel almaktadır. Bir bölgede meydana gelen deprem hareketliliğinin, hangi bölgelere etki edeceği ve hangi bölgelerde deprem oluşabileceği gibi tahminleri kapsayan bu algoritma başarılı bir şekilde bu bölümde uygulanmıştır.

Elde edilen tahmin sonuçları değerlendirildiğinde, sistemin güçlü kurallar için sağlıklı tahminler gerçekleştirdiği ve hata oranının daha düşük olduğu gözlemlenmiştir. Öte yandan düşük destek ve güven metriklerinde üretilen kurallar kullanıldığında, sistemin tahminde hata payının daha yüksek olduğu görülmüştür. Veri tabanında bulunan kuralların düşük destek ve güven değerlerine sahip olması sistemin başarımlarını düşürmektedir.

Geliştirilen bu tahmin algoritması, çevirim içi deprem tahmininde kullanılabilecek etkili bir yaklaşım olarak önerilmiştir. Tasarlanan tahmin algoritmasının, bazı parametrelerde %90’nın üstünde yüksek başarımlarına sahip olduğu görülmüştür. Türkiye Deprem Portalı içerisinde bulunan bu tahmin algoritması ile kullanıcılar belirledikleri herhangi bir

bölge üzerinde kolaylıkla tahmin işlemlerini gerçekleştirebilirler. Çeşitli parametre optimizasyon teknikleri uygulanarak tahmin algoritmasının başarımı artırılabilir.

7. DEPREM VERİLERİNİN KÜMELENDİRİLMESİ

7.1 Giriş

Tezin bu bölümünde, deprem verilerinin kümelendirilmesi için veri madenciliği yöntemlerinden k-Means ve KNN algoritmaları yapılan çeşitli değişikliklerle Türkiye deprem verileri üzerinde başarılı bir şekilde uygulanmıştır. Algoritmalarındaki mesafe hesaplamalarında, depremlerin meydana geldikleri gerçek konum verileri kullanılmıştır. Deprem verileri öncelikle meydana geldikleri coğrafi konuma göre k-Means algoritması ile kümelendirilmiş ve oluşan kümeler içerisindeki hareketlilik grafikleri sunulmuştur. Daha sonra k-Means ile kümelendirilen depremlerin iyileştirilmesine yönelik olarak KNN algoritması ile k-Means algoritması birleştirilerek farklı bir kümelendirme çalışması gerçekleştirilmiştir.

Çalışmada, klasik k-Means algoritmasının performansına büyük ölçüde etki eden rastgele başlangıç küme merkezleri seçme işlemine alternatif yeni bir yaklaşım sunulmuştur. Bu yaklaşımda, başlangıçtaki küme merkezleri rastgele değil veri tabanında bulunan büyük ölçekteki depremler arasından seçilmiştir. Elde edilen kümeleme sonuçları karşılaştırılmış ve algoritmaların farklı k parametreleri ile kümeleme işlemindeki performans analizleri değerlendirilmiştir. Yapılan değerlendirme sonucunda büyük ölçekteki depremleri başlangıç küme merkezleri olarak atayan algoritmanın performansının, klasik k-Means algoritmasının performansına göre oldukça yüksek olduğu gözlemlenmiştir.

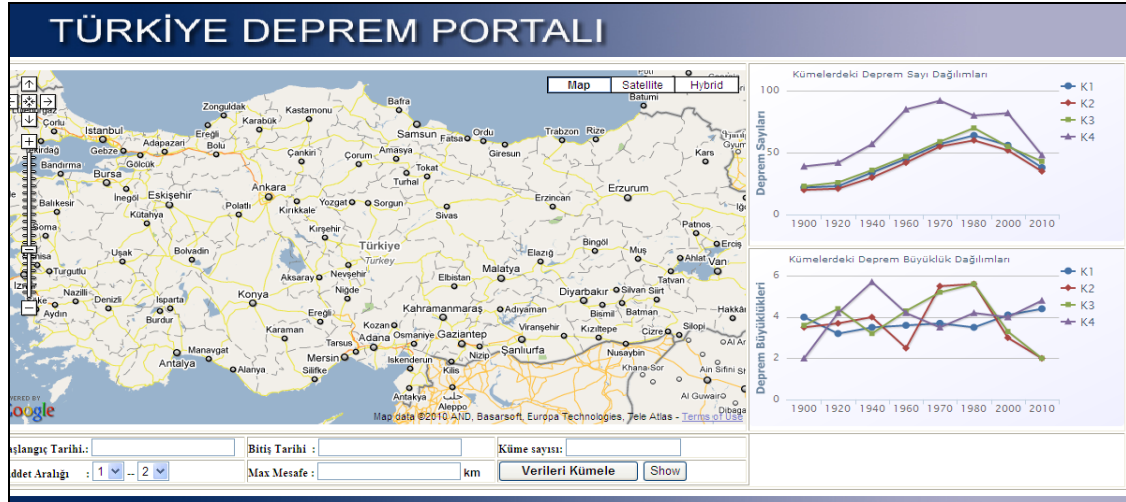
Bu çalışmada, Türkiye Deprem Portalı üzerinde kümeleme işlemlerini gerçekleştirmek için görsel bir arayüz hazırlanmıştır. Depremlerin farklı kümelere ayrılması ile bölgeler içindeki ve bölgeler arasındaki deprem hareketliliklerinin karakteristikleri hakkında bilgi elde edilmesi amaçlanmıştır. Portal üzerindeki grafikler aracılığı ile kümelerdeki deprem hareketliliklerinin daha detaylı gözlemlenmesi sağlanmıştır.

7.2. Türkiye Deprem Portalı Üzerindeki Kümeleme Bölümü

Deprem verileri üzerine uygulanan diğer veri madenciliği yöntemlerinde olduğu gibi kümeleme için de Türkiye Deprem Portalı üzerinde ayrı bir görsel arayüz tasarlanmıştır.

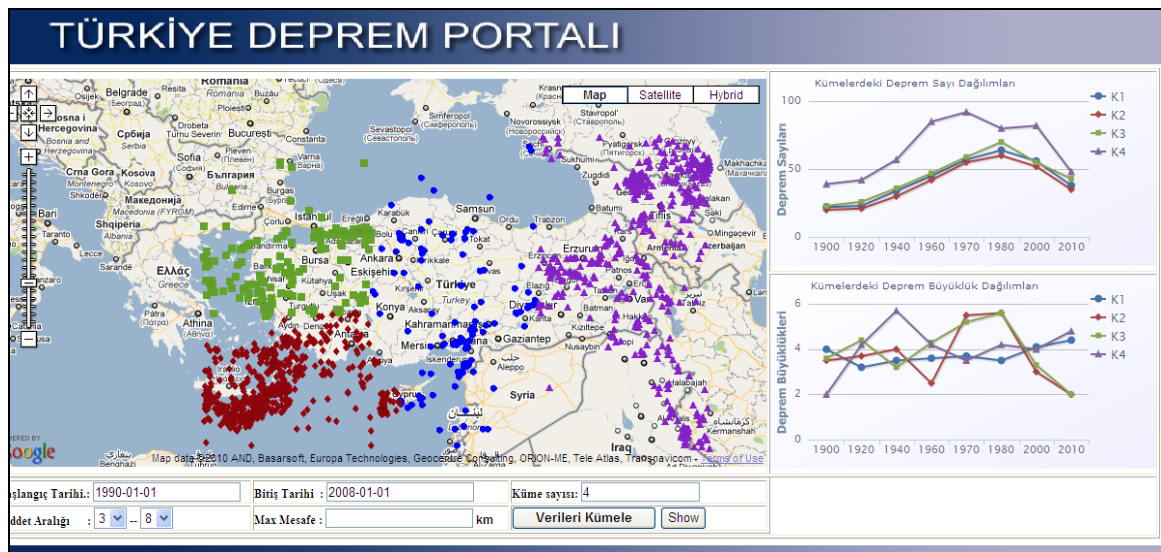
Kullanıcılar bu ara yüzde, seçtikleri tarih ve büyüklük aralığındaki depremleri yine kendi belirledikleri küme sayısına göre kolay bir şekilde kümelerine ayırabilmektedir.

Şekil 7.1’de, Türkiye deprem portalı üzerindeki kümeleme işlemleri için oluşturulan arayüzgörülmektedir.



Şekil 7.1. Türkiye deprem portalında kullanılan kümeleme ara yüzü

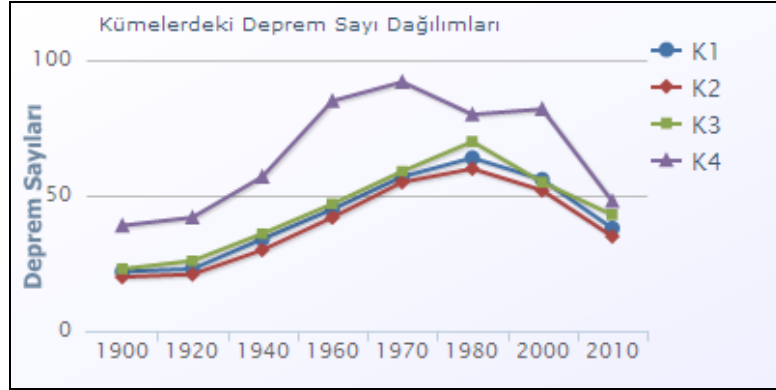
Kullanıcılar bu arayüz üzerinde, belirledikleri parametreler ile kolay bir şekilde kümeleme algoritmasını deprem verileri üzerinde uygulayabilmektedir. Şekil 7.2’de, kümeleme için seçilen çeşitli parametreler ile verilerin arayüzü üzerinde nasıl kümelendiğini gösteren bir şekil gösterilmiştir.



Şekil 7.2. Kümeleme için örnek bir gösterim

Harita üzerinde kümeler farklı simgeler ve renkler ile temsil edilmektedir. Her bir simge o simgenin temsil ettiği kümeye dahil olan deprem verisini göstermektedir. Bu veriler oluştukları konuma göre hangi kümeye yakın ise bu kümenin elemanı olarak kaydedilir.

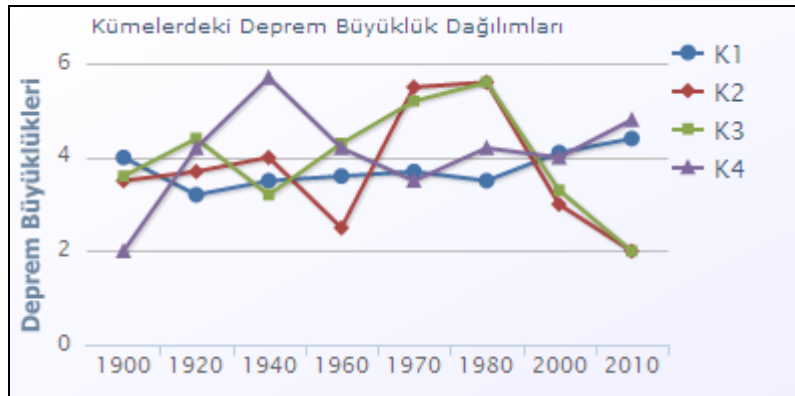
Şekil 7.3’de, kümeleme ara yüzünde bulunan grafik ekranlarından kümelerin yıllık deprem dağılım grafiği görülmektedir.



Şekil 7.3. Kümelerdeki yıllık deprem sayı dağılım grafiği

Yıllık deprem dağılım grafiği ile kullanıcılar kümeler arasındaki bağlantılar hakkında fikir edinebilirler. Ayrıca bu grafik ekranı sayesinde hangi kümelerin bir birlerine deprem dağılımı bakımından benzer veya farklı olduğu gözlemlenebilir.

Şekil 7.4’de, grafik ekranında yer alan, kümelerin yıllık ortalama deprem büyüklük grafiği gösterilmiştir.



Şekil 7.4. Kümelerin yıllık ortalama deprem büyüklük grafiği

Bu grafik yardımı ile kümelerde meydana gelen deprem hareketliliklerinin belirlenen tarihler arasında yıllık ortalama büyüklükleri kümeleri temsil eden renkler ve simgeler ile çizdirilmektedir. Hem görsel açıdan hem de bilgi sunumu açısından bu grafik ekranları önemli birer bileşendir.

Bu bölümün ilerleyen kısımlarında, kümeleme algoritmasının deprem verileri üzerine nasıl uygulandığı ve kümeleme için tasarlanan çeşitli algoritmaların gösterimi yapılmaktadır.

7.3. Deprem Konum Tabanlı K-Means Algoritması

Veri madenciliğinde kullanılan en yaygın kümeleme yöntemlerinden k-means algoritması, bu tez çalışmasında deprem verileri için geliştirilerek kullanımı sağlanmıştır. K-means algoritmasında k adet küme merkezi belirlenerek veriler k adet kümeye ayrılır. Başlangıçta k adet küme merkezinin seçimi eldeki veriler içerisinde rastgele belirlenir. Her niteliğin küme merkezine olan uzaklığı hesaplanır ve en yakın küme merkezi hangisi ise nitelik o kümeye dahil edilir. Daha sonra küme merkezleri eklenen niteliklerin ortalamasına göre yeniden hesaplanır ve nesnelerin kümelenmesinde değişiklik olmayıncaya kadar bu işlemler devam ettirilir. Deprem verileri için tasarlanan kümeleme metodunu aşağıdaki gibi özetleyebiliriz:

1. Kümeleme için çalışılacak zaman aralığı ve deprem büyüklük parametrelerinin seçiminin yapıldığı kısımdır. Bu kısımda belirlenen aralıktaki deprem verilerinin kümelendirilmesi yapılır.
2. Veri tabanından kümelendirilecek veriler süzülerek algoritmada değerlendirilmek üzere farklı bir tabloya kaydedilir. Bu sayede veri tabanı üzerinden gerekli sorguların yapılması esnasında bir performans artırımını sağlar.
3. Deprem verilerinin kaç kümeye ayrılacağı belirlenir. K-Means algoritmasındaki k parametresinin seçimidir. Bu seçim görsel arayüzündeki web form bileşenleri yardımı ile yapılmaktadır.
4. Veriler içerisinde rastgele k tane deprem verisi seçiler ve küme merkezleri olarak konum bilgileri atanır.
5. Her bir nesnenin küme merkezlerine olan uzaklıkları hesaplanır. Mesafe hesaplamasında literatürde sıklıkla kullanılan Öklid mesafe hesabı kullanılmaktadır.

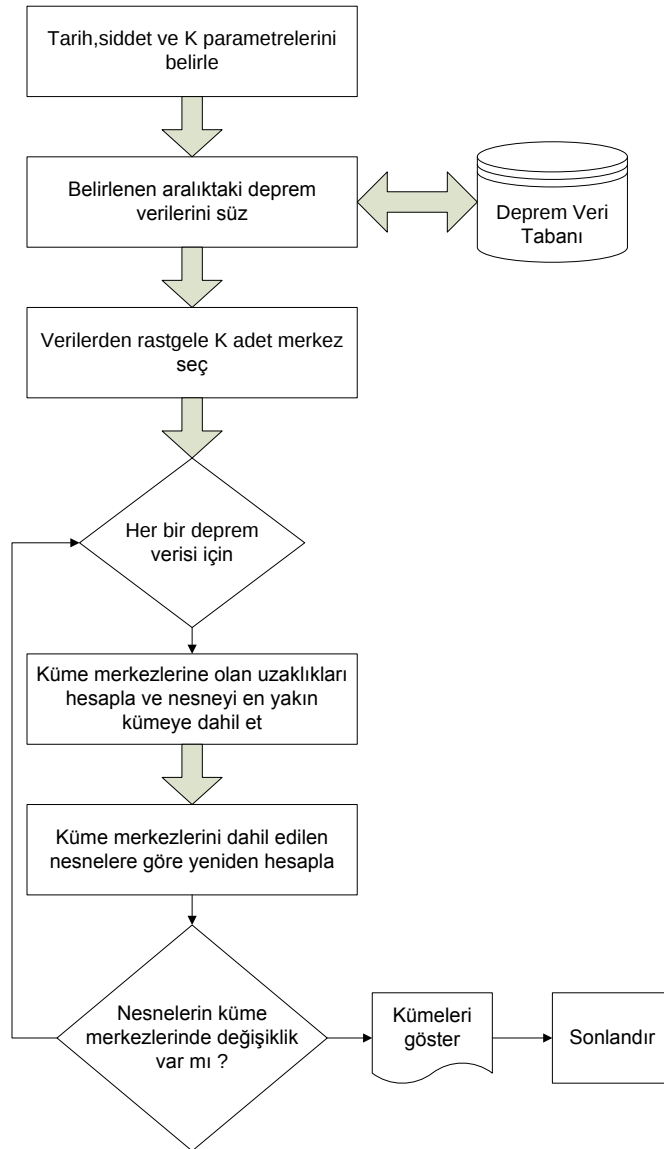
6. Deprem verileri oluşma zamanlarına göre sisteme girilerek en yakın kümeye dahil edilir.

7. Kümelere nesneler dahil olduktan sonra yeni küme merkezlerinin hesaplanması yapılır.

8. Nesnelerin kümelenmesinde değişiklik olmayıncaya kadar 5.adımdan itibaren bu işlemler tekrarlanır.

9. Kümeleme işlemleri tamamlandıktan sonra görsel olarak verilerin dahil oldukları kümeler farklı renklerdeki dairelerle belirtilerek görüntülenir.

Şekil 7.5’de, kümeleme için yapılan tüm bu işlemlerin akış şeması olarak gösterimi yapılmıştır.



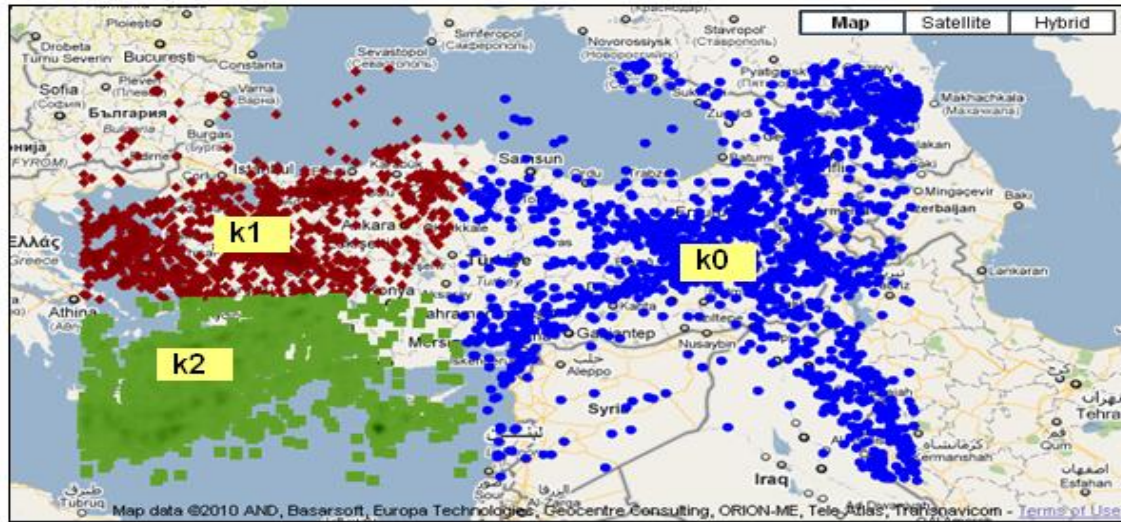
Şekil 7.5. Kümeleme işlemlerinin akış diyagramı

Bu çalışmada, k-Means algoritmasındaki mesafe, deprem verileri için gerçek konum mesafesi olarak modellenmiştir. Veri tabanında bulunan depremlerin olduğu enlem ve boylam gibi coğrafi bilgiler kullanılarak nitelikler arasındaki mesafe hesaplanmaktadır. Postgre SQL veri tabanının coğrafi veriler için bünyesinde barındırdığı çeşitli fonksiyonlar yardımı ile bu işlemler kolaylıkla yapılabilmektedir.

7.4. Türkiye Deprem Verileri Üzerinde Kümeleme Çalışması

Bu kısımda, bir önceki bölümde bahsedilen kümeleme algoritmasının Türkiye deprem verileri üzerinde çeşitli parametre değişiklikleri ile uygulamaları yapılmış ve sonuçlar sunulmuştur.

Şekil 7.6’da, 01.01.1900–01.01.2008 tarihleri arasında $k=3$ ve $Mc \geq 4$ depremlerin kümelendirilmesi sonucu oluşan kümeler gösterilmiştir.



Şekil 7.6. 01.01.1900–01.01.2008 tarih aralığında $k=3$ ve $Mc \geq 4$ için kümeleme sonuçları

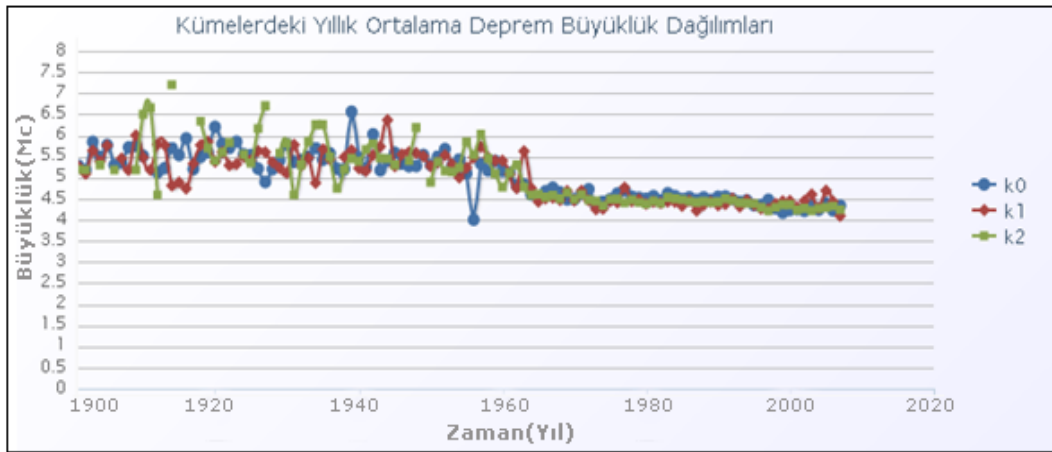
Şekil 7.7’de, oluşan kümelerin yıllık deprem sayılarındaki dağılımlarının grafiği çizdirilmiştir.



Şekil 7.7. 01.01.1900–01.01.2008, $k=3$, $M_c \geq 4$ için oluşturulan kümelerdeki yıllık deprem dağılımı

Kümelerdeki yıllık deprem dağılımları değerlendirildiğinde 1900 den 1964 yılına kadar kümelerdeki $M_c \geq 4$ büyüklüğündeki deprem sayılarının düşük olduğu ve her kümede ortalama aynı seviyede olduğu gözlemlenmiştir. 1964 yılından sonra kümelerdeki deprem hareketliliklerinde farklı artışlar olduğu görülmektedir.

Şekil 7.8’de, oluşturulan bu kümelerdeki $M_c \geq 4$ büyüklüğündeki depremlerin yıllık ortalama deprem büyüklük grafikleri gösterilmiştir.

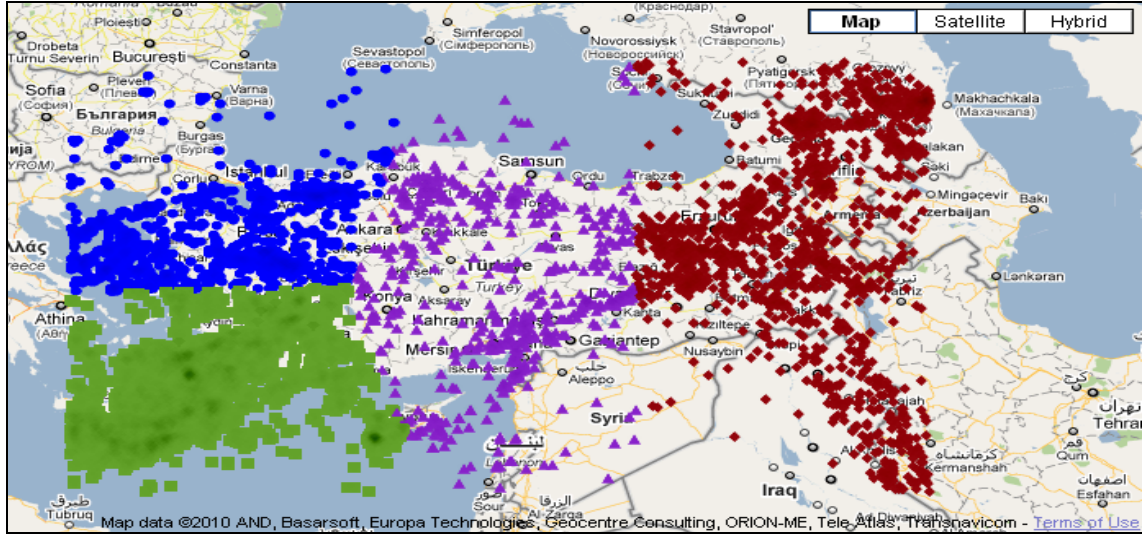


Şekil 7.8. 01.01.1900–01.01.2008 $k=3$, $M_c \geq 4$ için oluşturulan kümelerdeki yıllık ortalama büyüklük

Grafikte elde edilen sonuçlara göre 1900’den 1964 yılına kadar kümelerdeki $M_c \geq 4$ büyüklüğündeki deprem büyüklüklerinin yıllık ortalama değerlerinin farklılıklar gösterdiği görülmektedir. Bu dağılımlar 1964 yılından sonra tüm bölgede tam anlamıyla aralarında bir dengeye kavuşmuştur. 1964 öncesi verilerdeki düzensiz davranışların nedeni Türkiye deprem katalogunda toparlanan verilerin 1964 yılından önceki kısmının çeşitli

kaynaklardan taranması ve 1964 yılından sonraki deprem verilerinin ISC kataloglarından toparlanması olarak söylenebilir.

Şekil 7.9’da, 01.01.1900–01.01.2008 tarihleri arasında $k=4$ ve $M_c \geq 4$ depremlerin kümelenmesi sonucu elde edilen sonuçlar gösterilmiştir.



Şekil 7.9. 01.01.1900–01.01.2008 tarih aralığında $k=4$ ve $M_c \geq 4$ için kümeleme sonuçları

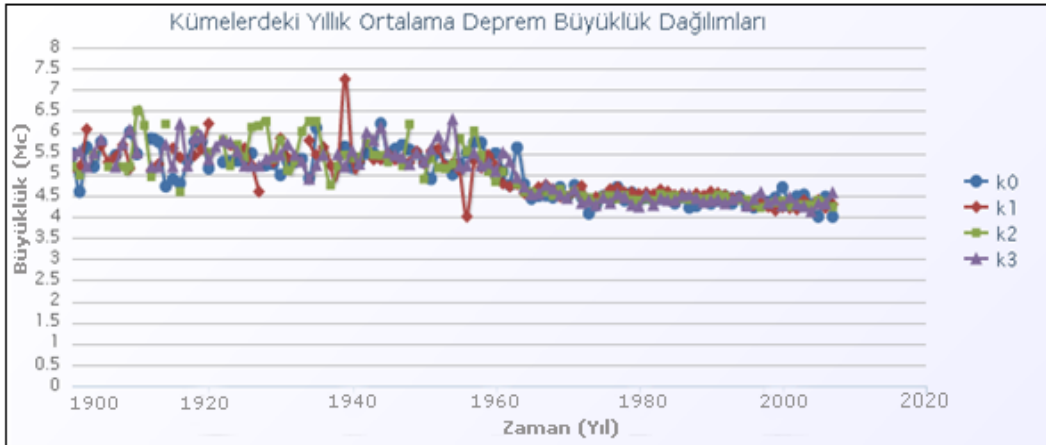
Elde edilen bu sonuç $k=3$ için oluşturulan küme sonuçlarına paralel bir yapıda elde edilmiştir. $K=4$ için elde edilen yeşil alanla Marmara bölgesi ve civarındaki deprem verilerinin farklı bir kümeye ayrıldığı görülmektedir.

Şekil 7.10’da, küme sayısı 4 için elde edilen kümelerdeki yıllık deprem sayı dağılımları gösterilmiştir.



Şekil 7.10. 01.01.1900–01.01.2008, $k=4$, $M_c \geq 4$ için oluşturulan kümelerdeki yıllık deprem dağılımı

“k0” ile gösterilen küme Marmara bölgesi ve civarındaki deprem hareketliliğini temsil etmektedir. Yakın geçmişte Türkiye de büyük ölçüde hasara yol açan 7.4 M büyüklüğündeki 1999 Gölcük depremi bu küme içerisinde bulunmaktadır. Bu kümede hemen hemen her yıl sürekli olarak $M_c \geq 4$ büyüklüğünde depremlerin meydana geldiği gözlemlenmiştir. Grafikte göze çarpan unsurlardan bir tanesi “k2” ile gösterilen kümede 1999 yılı öncesi deprem sayısında hızlı bir artışın olduğu ve bu artışı takiben 1999 yılında Gölcük depremini içeren “k0” kümesinde deprem sayısının aniden arttığıdır. Bu sonuçtan, kümeler arasında bir hareketlilik ilişkisinin tespit edilmesi ile gelecekte meydana gelebilecek büyük ölçekteki depremlerin tahmin edilmesinin mümkün olduğunu söyleyebiliriz. Şekil 7.11’de, k=4 için elde edilen kümelerdeki yıllık deprem büyüklük dağılımları gösterilmiştir.



Şekil 7.11. 01.01.1900–01.01.2008 k=4, $M_c \geq 4$ için oluşturulan kümelerdeki yıllık ortalama deprem büyüklükleri

7.5. Büyük Ölçekli Depremlerin Başlangıç Küme Merkezi Olarak Atanması

K-Means algoritması giriş verilerini k adet kümeye ayırır. Her küme, küme merkezleri ile temsil edilir. K-Means algoritmasının performansı başlangıç kümelerinin merkezlerine bağlıdır. Benzer şekilde üretilen kümelerdeki yapı da başlangıçtaki yapılanmayla doğrudan ilişkilidir. Kümeleme teknikleri genelde örnekler içerisinde rastgele seçilen bir giriş verisi ile başlar ve toplam karesel hatayı azaltmak için bu veriler tekrarlı olarak yinelenir. Rastgele başlatma yöntemi kullanan bölünmeli kümeleme algoritmalarında iyi bir sonuç elde etmek için farklı başlangıçlarla birçok kez algoritmayı yeniden çalıştırmak gerekir.

Bu kısımda, klasik k-Means algoritmasının en büyük performans problemlerinden biri olan küme merkezlerinin rastgele atanması işlemini önlemeye yönelik bir yaklaşım sunulmuştur. Bu yaklaşımda kullanılan temel düşünce deprem verileri üzerinde yapılan diğer çalışmaların sıklıkla dile getirdikleri geçmişte meydana gelmiş büyük ölçekteki depremlerin gelecekte oluşacak diğer depremler üzerindeki etkisi üzerine kurulmuştur. Bu bağlamda Türkiye deprem katalogu üzerinde meydana gelen büyük ölçekli depremler k-Means algoritması için başlangıç küme merkezlerini belirlemede kullanılmıştır. Bu sayede kümelendirme işleminin performansının büyük ölçüde artırılması amaçlanmıştır. Şekil 7.12’de önerilen algoritmanın adımları sunulmuştur.

```
Algoritma: Rasgele olmayan başlangıç küme merkezli k-Means
(NrK-means)
Küme sayısı belirle:  $k$ ;
Tarih aralığı seç: tarih;
Kümelendirilecek min.büüklük: siddet;
for  $i=1$  to  $k$  do
    Center $i$ =Veri tabanından büyük ölçekli depremlerden seç
end-for
For  $j$ =tarih aralığındaki tüm deprem kayıtları için do
    for  $i=1$  to  $k$  do
        Dist $i$ =oklid(Center $i$ ,dep $j$ );
    end-for
    Dep $j$ .kume=Minimum(Dist $i$ );
end-while
Kümelerin çıktılarını ver
```

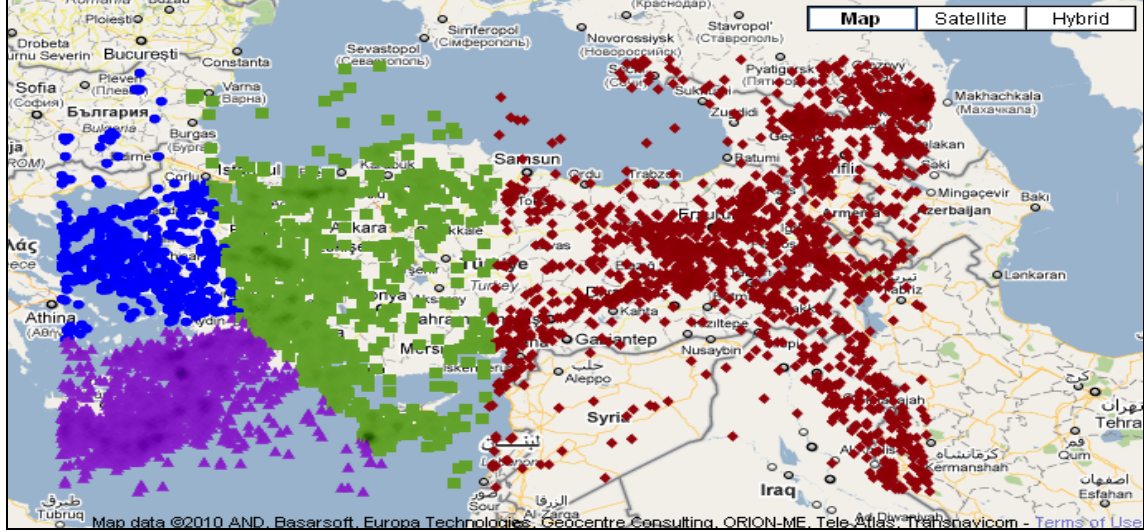
Şekil 7.12. Rasgele olmayan başlangıç küme merkezli k-means algoritması

Algoritmada öncelikle depremlerin ayrılacağı küme sayısı olan k belirlenmektedir. K parametresinin belirlenmesi bu algoritmanın en büyük dezavantajlarından birisidir. Kullanıcıların bu parametreyi bildikleri varsayılsa da maalesef bu çoğu zaman bu doğru olmayan bir yaklaşım olmuştur. Daha sonra kümelenecek istenen deprem büyüklükleri ile tarih aralıkları belirlenir. Veri tabanı içerisindeki büyük ölçekli depremler belirlenen k adet küme için merkez olarak tayin edilir.

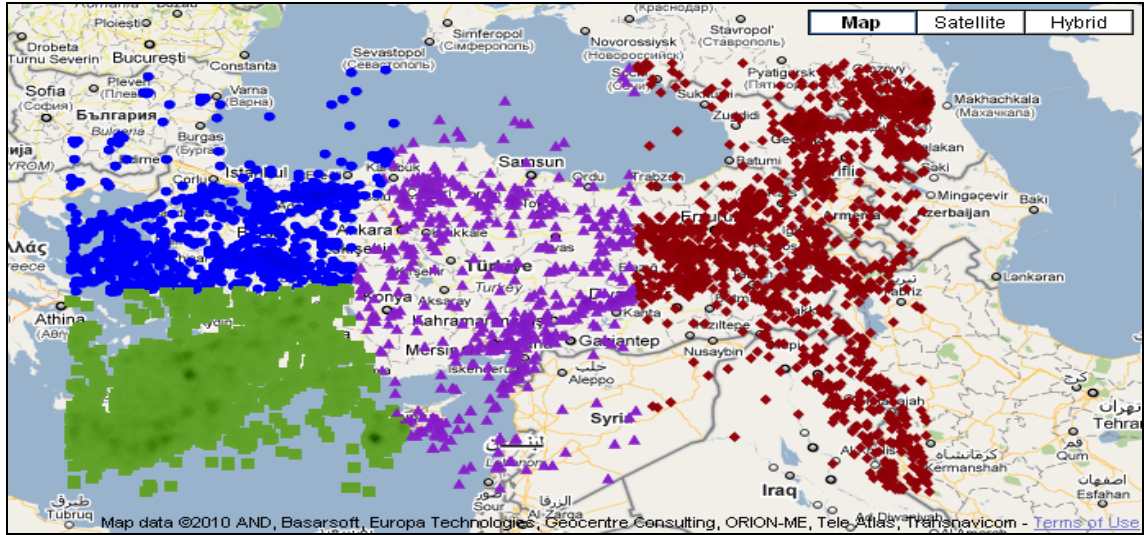
Kümelendirilmek istenen bütün deprem verileri için küme merkezlerine olan uzaklık Öklid bağıntısı ile hesaplanır. Deprem verisinin en yakın olduğu küme merkezi bu

depremin kümesi olarak atanır. Bu işlem tüm deprem verileri için tekrarlanır ve algoritma çalışmasını tamamlar.

Şekil 7.13 'te, aynı parametreler ile klasik k-Means algoritması ile rastgele olmayan k-Means algoritmasının kümeleme sonuçları gösterilmiştir.



(a)



(b)

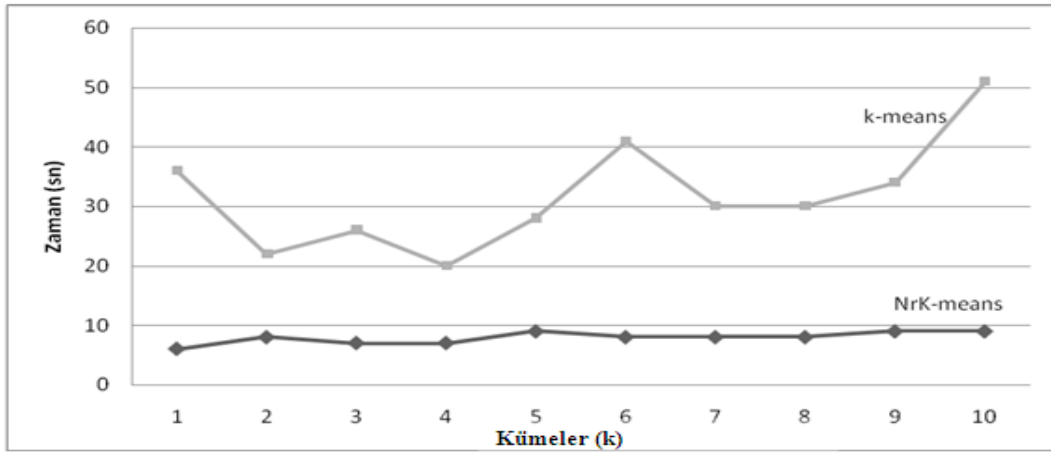
Şekil 7.13. 01.01.1900–01.01.2008 tarih aralığında $k=4$ ve $M_c \geq 4$ deprem verilerinin Nrk-means ve k-means ile kümelendirilmesi

- (a) Nrk-means algoritması ile kümeleme
- (b) K-means algoritması ile kümeleme

Şekil 7.13'te, elde edilen küme sonuçlarına bakıldığında her iki algoritmanın verileri kümelendirmesinde aşırı derecede farklı bir yapının oluşmadığı görülmektedir. Her iki algoritma da verileri benzer şekilde kümelemiştir. Bu algoritmaların küme sonuçlarında az bir farklılık görülse de bu algoritmalar arasında bir performans farklılığı bulunmaktadır.

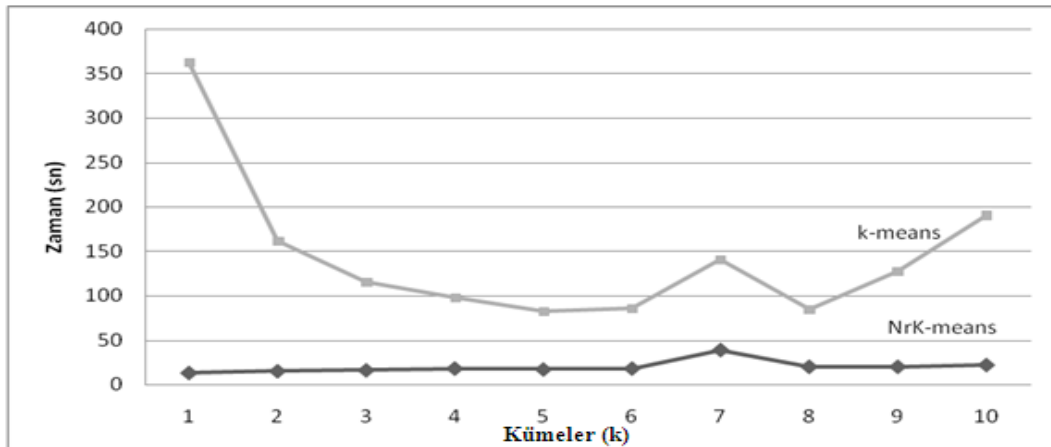
Intel Core 2 Duo P8600 2.40 Mhz işlemci ve 3GB Ram bellek sistemine sahip bilgisayar ile Netbeans 6.0.1 programında yapılan testlerde algoritmaların çalışma süreleri test edilmiştir.

Şekil 7.14'te, k-Means ve NrK-Means algoritmalarının 1263 veriyi farklı k parametreleri ile kümelendirmesindeki performansları gösterilmiştir.



Şekil 7.14. K-means ve NrK-means algoritmalarının 1263 veriyi kümelendirmesindeki hız ölçümleri

Şekil 7.15'de, her iki algoritmanın 3119 deprem verisini kümelendirmesindeki performans grafiği gösterilmiştir.



Şekil 7.15. K-means ve NrK-means algoritmalarının 3119 veriyi kümelendirmesindeki hız ölçümleri

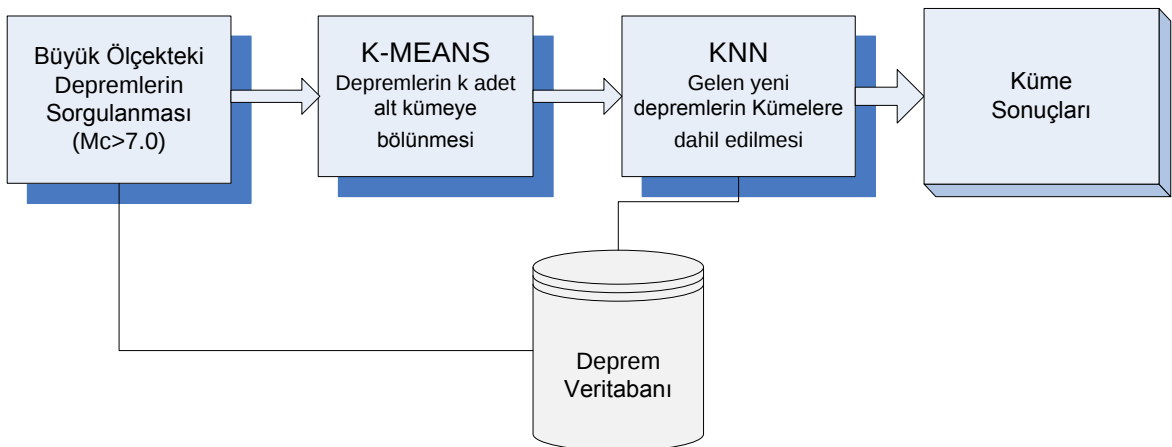
Her iki şekilde NrK-Means algoritmasının çalışma hızının küme sayısı arttıkça büyük farklılıklar göstermediği görülmektedir. Bu grafiklere bakıldığında büyük miktardaki veriler için k-Means algoritmasının performansının düşük olduğu görülmektedir. Bunun başlıca nedeni k-Means algoritmasında veri sayısı arttıkça küme merkezlerinin yeniden hesaplanma işlemleri ve küme merkezleri değiştiğinde tekrardan verilerin kümelerine dahil edilme işlemleridir. NrK-Means algoritmasında küme merkezleri sabit olduğundan dolayı bu işlemler için zaman harcanmamaktadır.

7.6. K-Means -KNN Algoritması ile Deprem Verilerinin Kümelendirilmesi

Bu kısımda, k-Means ve KNN algoritmalarının ortak kullanımı ile deprem verilerinin kümelendirilmesine yönelik bir yaklaşım sunulmuştur. K-Means algoritması ile büyük ölçekteki deprem verileri k adet kümeye ayrılmıştır. Daha sonra gelen deprem verileri önceden oluşturulmuş kümelerle göre en yakın k adet nesnenin ortalaması ile bu kümelerden birisine dahil edilmiştir. K-Means algoritması ile elde edilen kümelerin iyileştirilmesine yönelik yapılan bu uygulamada başarılı kümeler elde edildiği gözlemlenmiştir.

7.6.1 Geliştirilen Yöntem

Deprem verilerinin kümelendirilmesine yönelik geliştirilen uygulamanın çalışma yapısı Şekil 7.16'da gösterilmiştir.



Şekil 7.16. K-Means - KNN ile geliştirilen kümeleme yöntemin çalışma yapısı

Geliştirilen yöntemin çalışma adımları aşağıdaki şekilde sıralanmıştır.

- Başlangıç kümelerini oluşturacak deprem büyüklüğü belirlenir. Türkiye deprem katalogu verileri içerisinde 7.0 ve 7.0 üzerindeki büyüklüklerde meydana gelen depremler alınmıştır.
- Seçimi yapılan depremlerin k-Means yöntemi ile k adet alt kümeye bölünür. Başlangıç merkezleri rastgele atanan küme merkezlerine olan uzaklıklarına göre büyük ölçekli depremler k adet kümeye atanır.
- KNN algoritması ile belirlenen tarih aralığındaki depremler oluşma zamanlarına göre sisteme girilerek en yakın k adet depremin kümelerine göre bu kümelere dahil edilir.
- Kümeleme sonuçları Türkiye deprem portalının kümeleme ekranından görsel olarak sunulur ve bu kümelere ait yıllık deprem sayı dağılımları ve yıllık ortalama şiddet dağılım grafikleri sunulur.

KNN algoritmasının uygulanmasında depremlerin ilgisiz kümelere dahil olmasını engellemek için bir mesafe kısıtlaması yapılmıştır. 666 Km olarak belirlenen bu mesafe, bir nesnenin k adet nesneye olan maksimum uzaklığıdır. Yeni gelen bir deprem verisine 666km mesafe içerisinde k adet deprem verisi yoksa bu nesnenin kümelenmesi işlemi bir sonraki iterasyonda gerçekleştirilir.

Deprem verilerinin kümelendirilmesindeki bir diğer önemli unsur zamandır. Bu bakımdan algoritma içerisinde deprem verileri sıralı olarak sisteme girildiğinden dolayı o zaman dilimine kadar meydana gelmiş deprem verilerine göre yeni deprem verisi kümelere dahil edilir. Şekil 7.17’de, geliştirilen kümeleme yönteminin algoritması gösterilmiştir.

Algoritma: K-Means-KNN kümeleme

```
Küme sayısı belirle:  $k$ ;  
Büyük ölçekli depremler için min.büüklük:  
Tarih aralığı seç: tarih;  
Kümelendirilecek deprem büyüklükleri:  $M_c$ ;  
K adet küme merkezini belirle;  
Dbas="select* from Deprem where mag>=7.0"  
Dbas verilerini K-Means ile k adet alt kümeye ayır;  
Dyeni="select*from Deprem where mah>= $M_c$  and tar=tarih";  
while(tüm nesneler kümelendirilinceye kadar)  
while (Dyeni içerisindeki tüm veriler için)  
Dyeni verisine Max 666km uzaklıktaki k adet deprem verisini getir;  
if(k adet nesne varsa)  
Tüm nesnelerin kümelerini belirle;  
Baskın kümeyi bul;  
Dyeni verisinin kümesini baskın küme olarak güncelle;  
end-if  
end-while;  
end-while;  
Kümelerin çıktılarını ver
```

Şekil 7.17. K-Means - KNN kümeleme algoritması

7.6.2. Türkiye Deprem Verileri Üzerinde Yöntemin Uygulanması

Bu kısımda, geliştirilen K-Means ve KNN algoritmalarının ortak kullanımı ile kümeleme yönteminin Türkiye deprem kataloğu verileri üzerinde çeşitli uygulamaları sunulmuştur.

Şekil 7.18’de, 01.01.1900-01.01.2008 tarihleri arasında Türkiye Deprem Kataloğunda yeralan 7.0 ve üzeri büyüklükteki depremler gösterilmiştir.



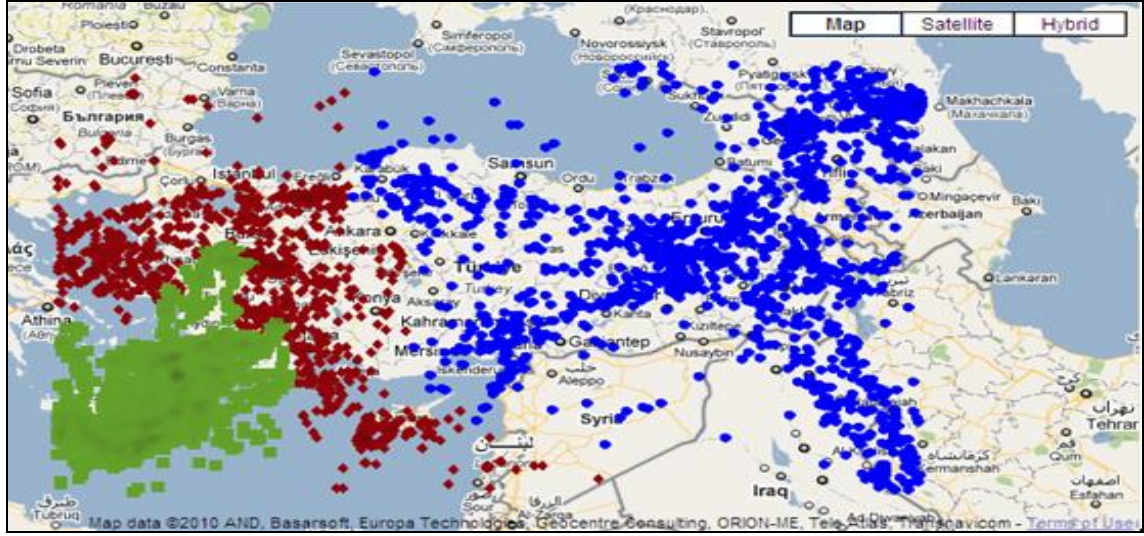
Şekil 7.18. 01.01.1900-01.01.2008 tarih aralığında Türkiye deprem kataloğunda yer alan $M_c \geq 7.0$ büyüklüğündeki depremler

Şekil 7.19’da, 01.01.1900-01.01.2008 tarihleri arasında $k=3$ için $M_c \geq 7.0$ büyüklüğündeki depremlerin k-Means ile oluşturduğu kümeler gösterilmiştir.



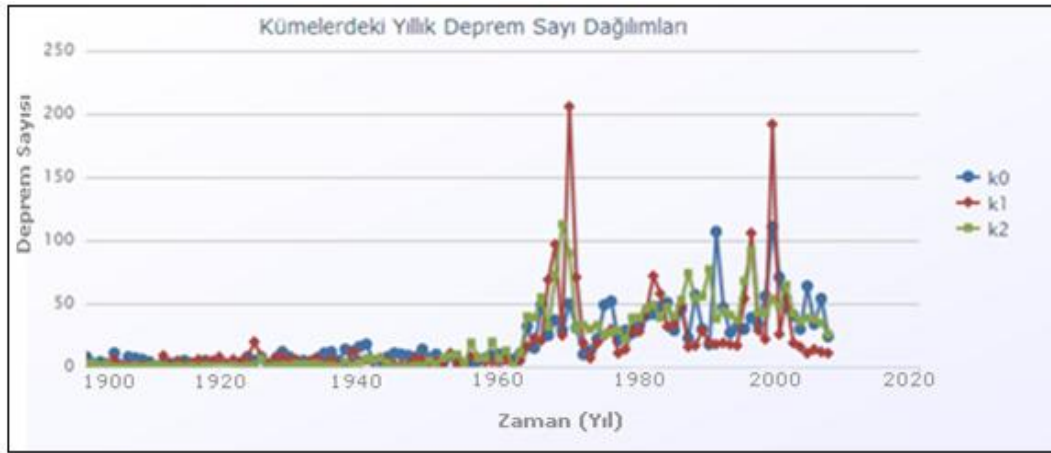
Şekil 7.19. 01.01.1900-01.01.2008 tarihleri arasında $k=3$ için $M_c \geq 7.0$ büyüklüğündeki depremlerin k-Means ile oluşturduğu kümeler

K-Means yöntemi ile büyük ölçekli depremler 3 farklı kümeye bölünmüştür. Şekil 7.20’de, 01.01.1900-01.01.2008 tarih aralığında $M_c \geq 4.0$ büyüklüğündeki depremlerin, k-Means ile oluşturulan kümelere, KNN algoritması kullanılarak dahil edildikten sonra oluşan küme sonuçları gösterilmiştir.

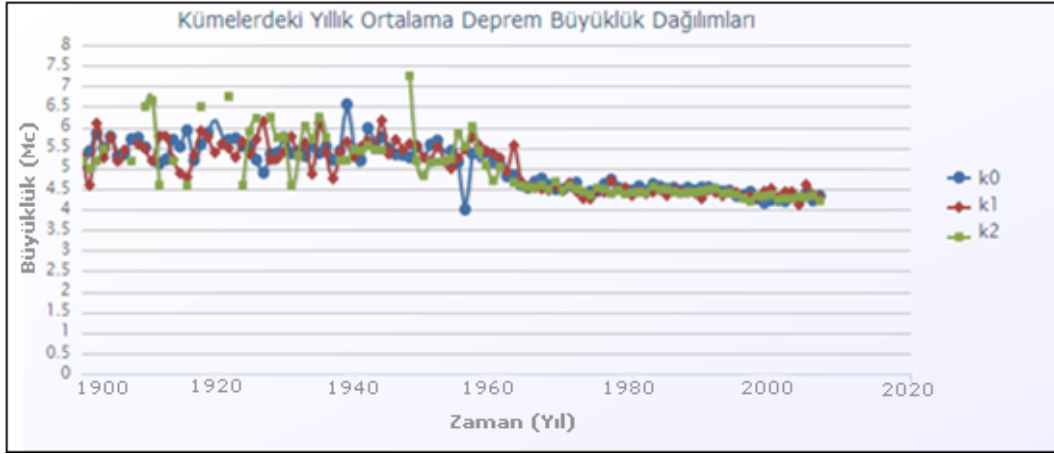


Şekil 7.20. 01.01.1900-01.01.2008 tarih aralığında $M_c \geq 4.0$ büyüklüğündeki depremlerin k-means ile önceden oluşan kümelerle KNN algoritması ile dahil edilmesi sonucu oluşan kümeler

Şekil 7.21’de, K-Means ve KNN ile elde edilen kümelerdeki depremlere ait yıllık sayı dağılımları ile yıllık ortalama büyüklük dağılımları grafikleri gösterilmiştir.



(a)



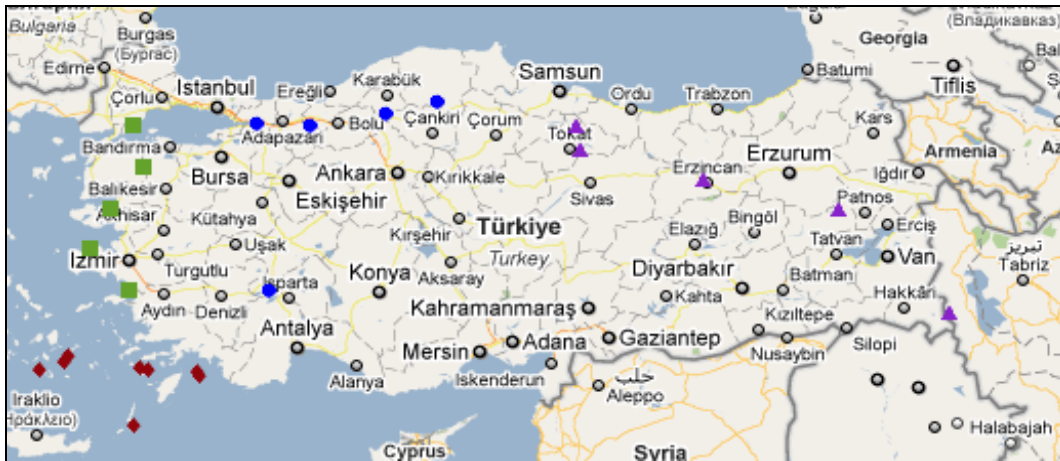
(b)

Şekil 7.21. 01.01.1900-01.01.2008 tarih aralığında $k=3$ için $M_c \geq 4.0$ büyüklüğündeki depremlerin K-Means-KNN algoritması ile oluşan kümelere ait grafikler

- a) Kümelerdeki yıllık deprem sayısı dağılımları
- b) Kümelerdeki yıllık deprem büyüklük ortalamaları

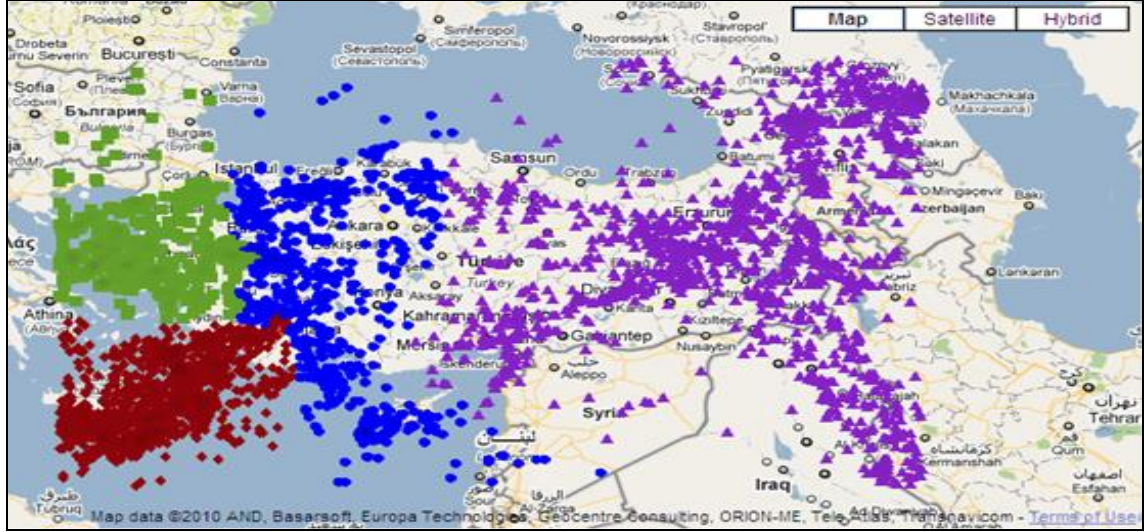
Şekil 7.21a'da, k1 kümesindeki depremlerin belli aralıkta tüm kümelerdeki deprem sayılarının üstüne çıktığı görülmektedir. k0 kümesindeki depremlerin sayı dağılımlarının diğer kümelere göre daha az değiştiği ve k2 deki sayı dağılımlarının genel olarak üst seviyede olduğu görülmektedir.

Şekil 7.22'de, 01.01.1900-01.01.2008 tarih aralığında $k=4$ için Türkiye deprem kataloğundaki büyük ölçekli depremlerin oluşturduğu alt kümeler gösterilmiştir.



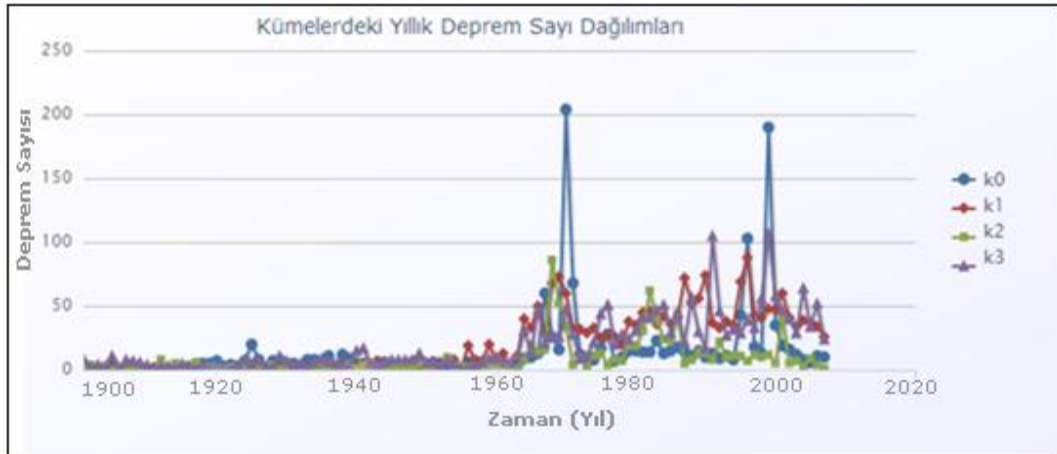
Şekil 7.22. 01.01.1900-01.01.2008 tarih aralığında $k=4$ için $M_c \geq 7.0$ depremlerin k-Means ile kümelendirilmesi

Şekil 7.23'te, 01.01.1900-01.01.2008 tarih aralığında $k=4$ için $M_c \geq 4.0$ büyüklüğündeki depremlerin KNN ile kümelere dahil edilmesi sonucunda oluşan küme yapısı gösterilmiştir.

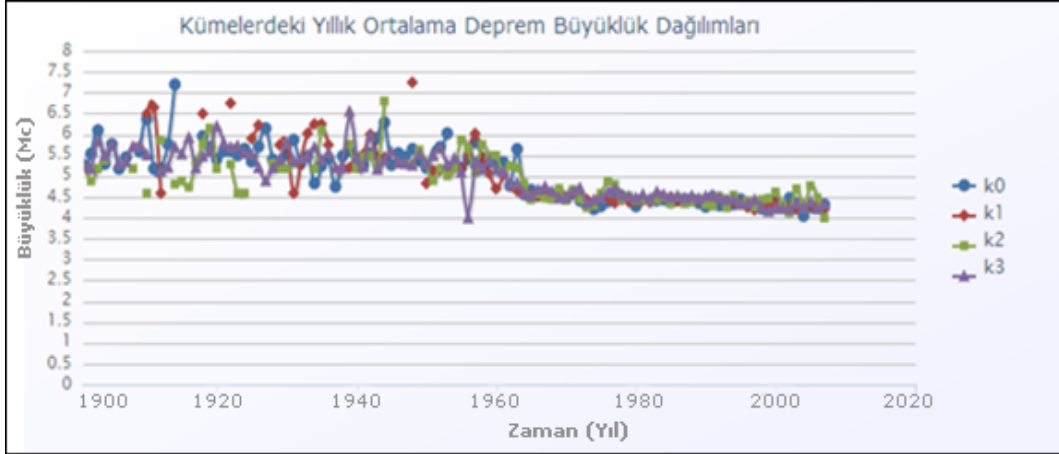


Şekil 7.23. 01.01.1900-01.01.2008 tarih aralığında $k=4$ için $M_c \geq 4.0$ büyüklüğündeki depremlerin KNN ile kümelere dahil edilmesi

Şekil 7.24'te, KNN ile oluşturulan kümelerdeki yıllık deprem sayı dağılımları ile kümelerdeki yıllık ortalama deprem büyüklük dağılım grafikleri sunulmuştur.



(a)



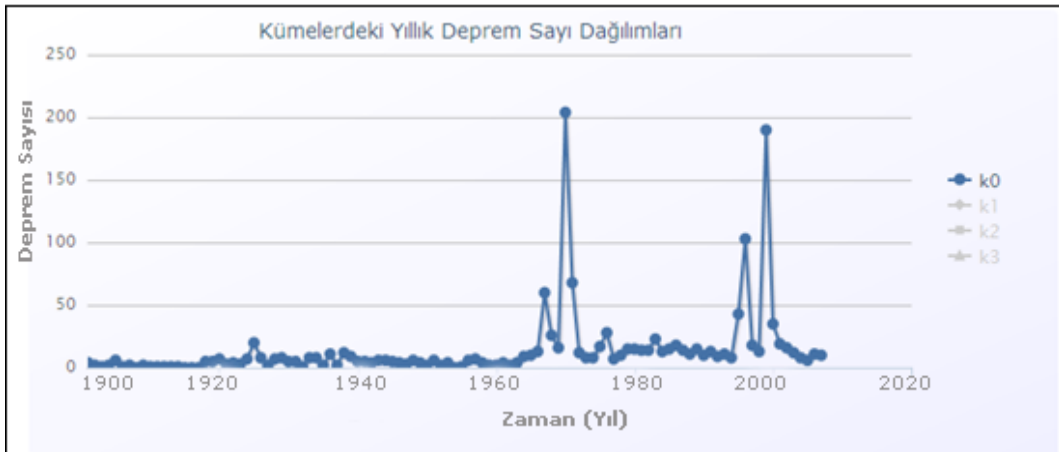
(b)

Şekil 7.24. 01.01.1900-01.01.2008 tarih aralığında $k=4$ için $M_c \geq 4.0$ büyüklüğündeki depremlerin K-Means ve KNN algoritmaları ile oluşan kümelere ait grafikler

- a) Kümelerdeki yıllık deprem sayısı dağılımları
- b) Kümelerdeki yıllık deprem büyüklük ortalamaları

Şekil 7.24a'da k0 kümesindeki deprem sayısı dağılımının diğer kümelere göre iki kez en yüksek değerlere ulaşması bu kümeyi incelenmesi gereken ilginç bir küme haline getirmektedir. Bunun yanında k1 kümesindeki sayı dağılımlarının tüm kümelere göre yüksek değerlerde seyrettiği görülmektedir.

Şekil 7.25'de, "k0" kümesine ait yıllık deprem sayısı dağılım grafiği ayrı olarak gösterilmiştir.

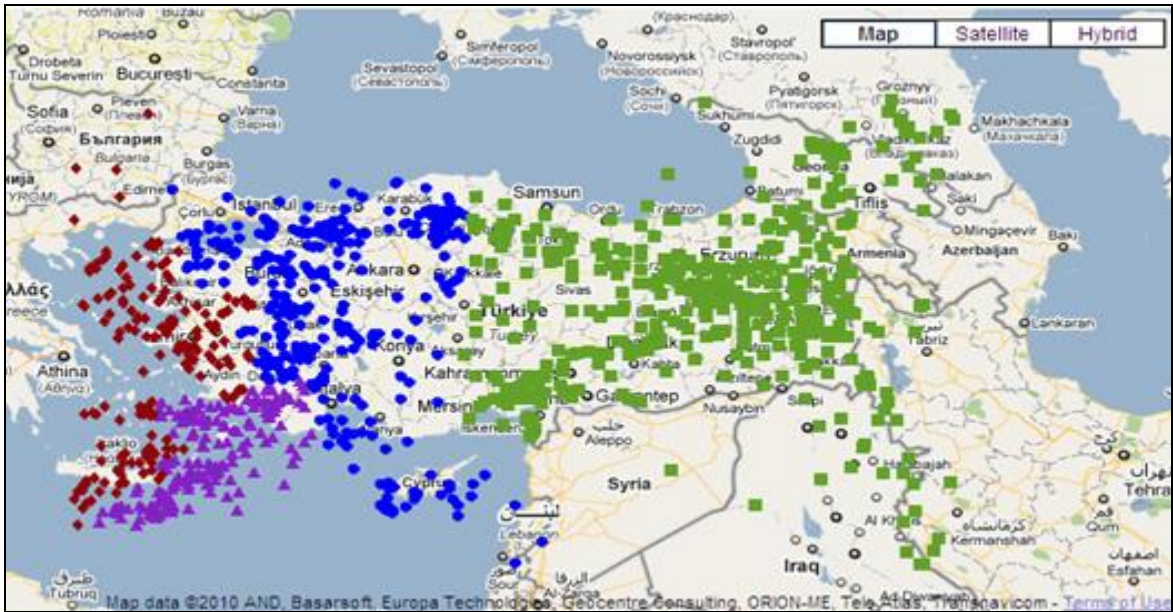


Şekil 7.25. K0 kümesindeki depremlerin yıllık sayısı dağılım grafiği

Şekil 7.25'de, k0 kümesindeki depremlerin 1960 yıllarına kadar düşük sayılarda görüldüğü ve 1970 yılında en yüksek değere ulaştığı görülmüştür. 1970 yılından sonra

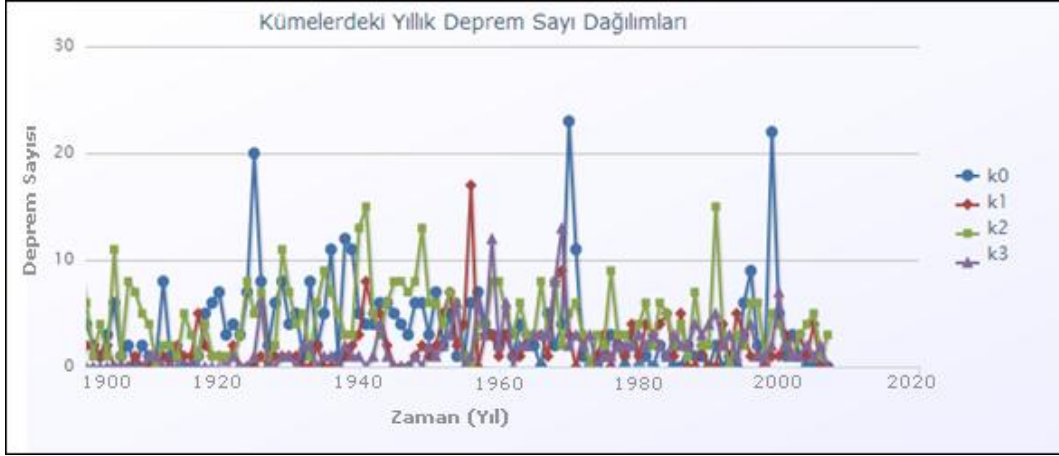
düşüşe geçen deprem sayıları 1996 ve 1999 yıllarında tekrar en yüksek değerine ulaşmıştır. Ayrıca burada göze çarpan bir diğer unsur bu kümedeki deprem sayısı en yüksek değerine ulaşmadan yaklaşık 3 yıl öncesinde bu yükselişlerin belirtisi olan bir yükselişin görülmesidir. 1970 yılındaki tepe değerine ulaşmadan önce 1967 yılında yükselişe geçtiği ve aynı şekilde 1999 yılındaki tepe değerine ulaşmadan önce 1996 yılında belirli bir miktarda yükseldiği görülmektedir. Kümelere ait bu şekilde elde edilecek örüntülerin gelecek depremlerin tahmininde kullanılabilecek önemli bilgiler olduğu söylenebilir.

Şekil 7.26'da, 01.01.1900–01.01.2008 tarih aralığında $k=4$ için $M_c \geq 5.0$ büyüklüğündeki depremlerin kümelendirilmesine ait sonuçlar gösterilmiştir.

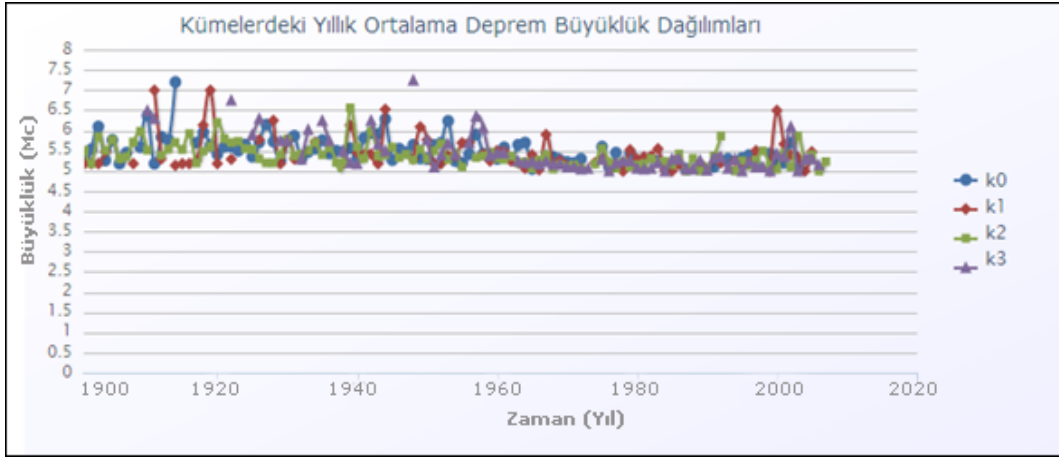


Şekil 7.26. 01.01.1900–01.01.2008 tarih aralığında $k=4$ için $M_c \geq 5.0$ büyüklüğündeki depremlerin kümelendirilmesine ait sonuçlar

Şekil 7.27'de, 01.01.1900-01.01.2008 tarih aralığında $k=4$ için $M_c \geq 5.0$ büyüklüğünde oluşan kümelere ait yıllık deprem sayı dağılımı ve yıllık ortalama büyüklük dağılım grafikleri sunulmuştur.



(a)

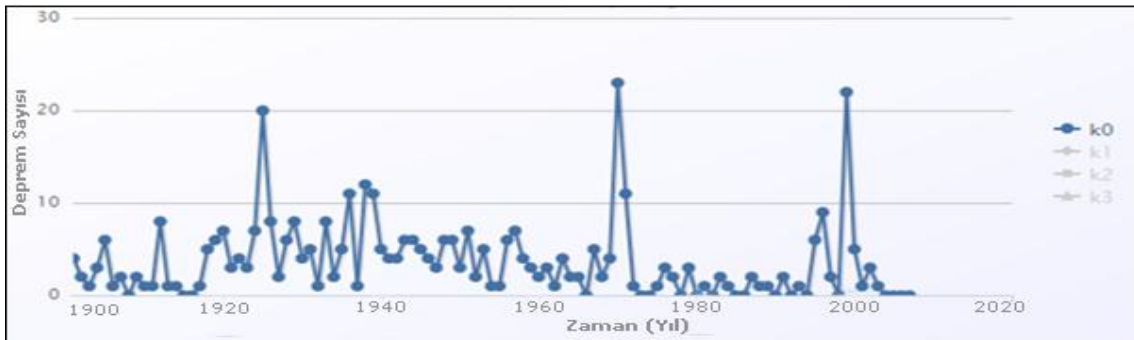


(b)

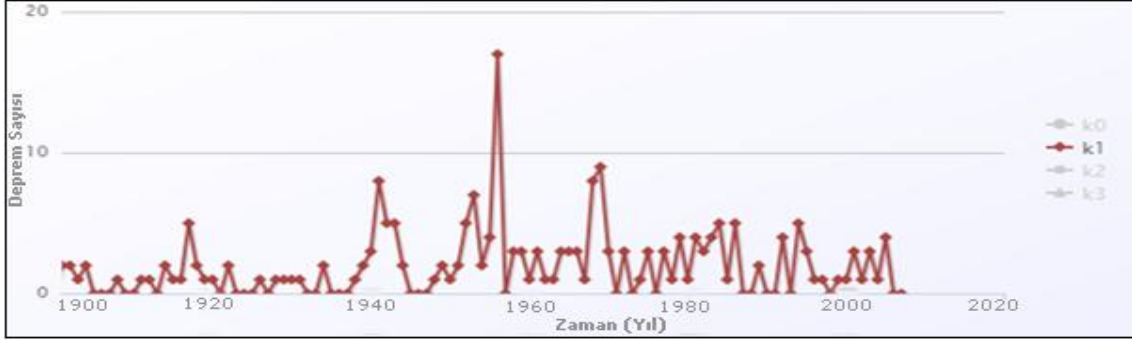
Şekil 7.27. 01.01.1900-01.01.2008 tarih aralığında $k=4$ için $M_c \geq 5.0$ büyüklüğündeki depremlerin K-Means ve KNN algoritmaları ile oluşan kümelerle ait grafikler

- a) Kümelerdeki yıllık deprem sayı dağılımları
- b) Kümelerdeki yıllık deprem büyüklük ortalamaları

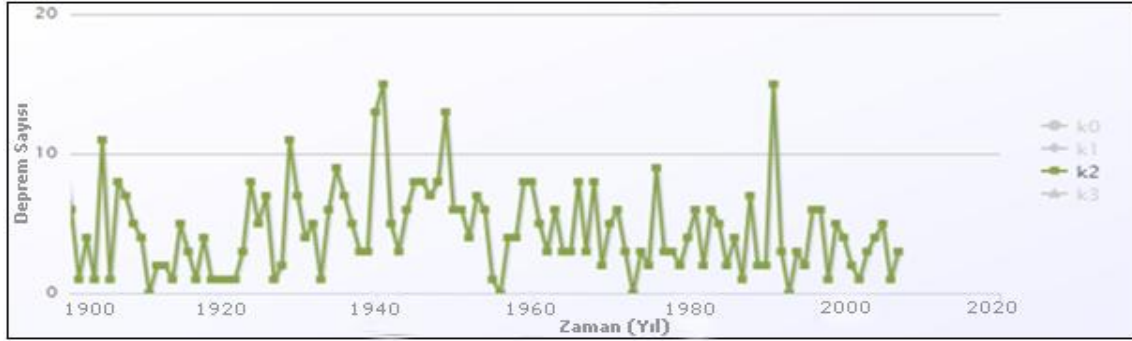
Şekil 7.28’de, kümelerdeki yıllık deprem sayı dağılımlarının her bir küme için ayrı ayrı grafikleri gösterilmiştir.



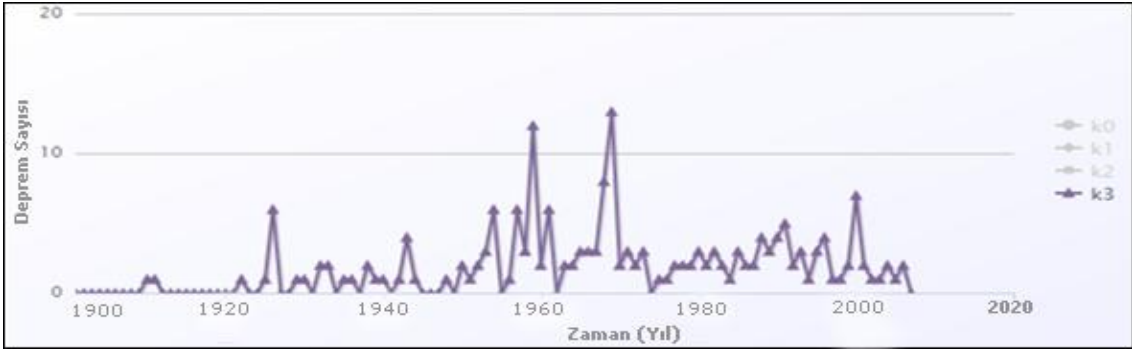
(a)



(b)



(c)



(d)

Şekil 7.28. 01.01.1900–01.01.2008 tarih aralığında $k=4$ için $M_c \geq 5.0$ büyüklüğündeki depremlerin yıllık sayı dağılımlarının her küme için ayrı ayrı gösterimi

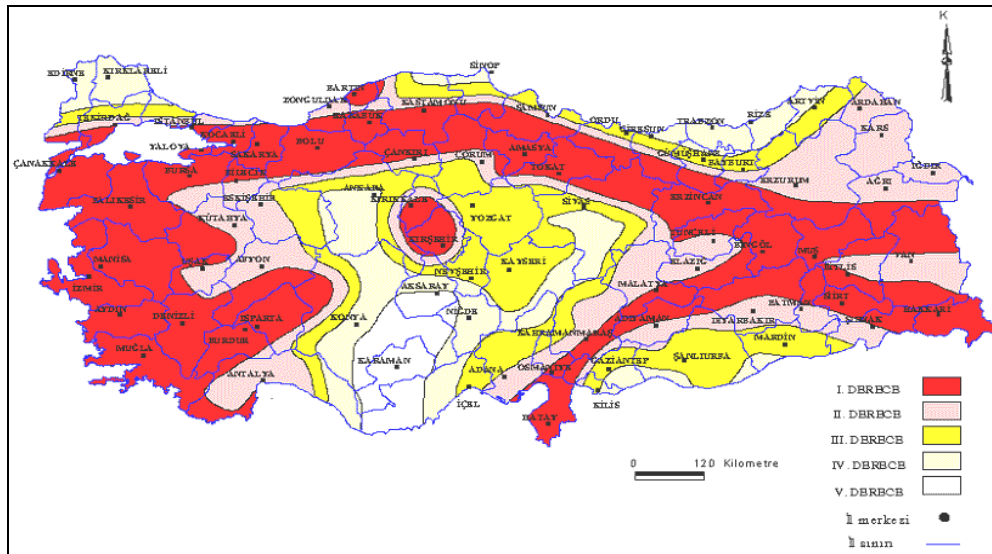
- a) k0 kümesindeki depremlerin yıllık sayı dağılım grafiği
- b) k1 kümesindeki depremlerin yıllık sayı dağılım grafiği
- c) k2 kümesindeki depremlerin yıllık sayı dağılım grafiği
- d) k3 kümesindeki depremlerin yıllık sayı dağılım grafiği

$M_c \geq 5.0$ büyüklüğündeki depremlerin sayı dağılım grafiklerinin daha düşük büyüklükteki depremlerin sayı dağılım grafiklerine göre farklı bir şekilde oluştuğu

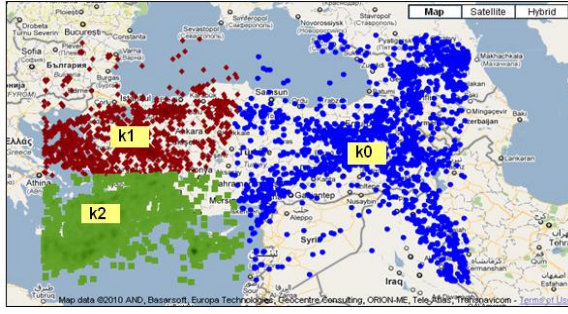
gözlemlenmektedir. 1960 yılı öncesindeki verilerin genellikle büyük ölçekli depremlerden oluşması ve daha küçük ölçekli depremlerin tam olarak 1960'lı yıllardan sonra tutulması bu grafiklerin farklı oluşmasında büyük etkindir. Şekil 7.28 den de görüleceği gibi her küme içerisindeki deprem hareketliliklerinin bir birlerine göre farklılıklar gösterdiği görülmektedir. Bazı kümelerde, $M_c \geq 5.0$ büyüklüğünde depremlerin seyrek bir şekilde oluştuğu, bazı kümelerde sıklıkla tekrarlandığı görülmüştür. k0 kümesindeki sayı dağılımına bakıldığında belirli aralıklarla bu kümedeki $M_c \geq 5.0$ büyüklüğündeki depremlerin sayısında yüksek değerlere ulaşıldığı ve sonrasında belirli bir dönem neredeyse bu büyüklüklerde depremlerin oluşmadığı gözlemlenmiştir. Türkiye'nin doğusunu temsil eden k2 kümesindeki sayı dağılımına bakıldığında, hareketliliklerin düzenli bir yapıda ilerlediği göze çarpmaktadır.

Depremler ile ilgili mantıklı kümelerin elde edilmesi ve bu kümeler içerisindeki deprem aktivitelerinin incelenmesi, gelecek depremler ile ilgili daha iyi tahminler gerçekleştirmede büyük rol oynayabileceği söylenebilir.

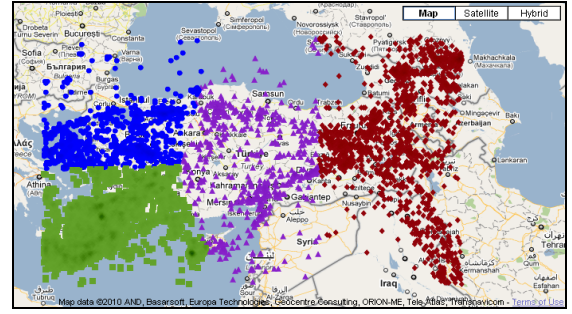
Şekil 7.29'da, kümeleme sonuçlarının karşılaştırılmasına yönelik olarak k-Menas ile elde edilen küme sonuçları ile k-Means KNN algoritmaları ile elde edilen kümeleme sonuçları verilmiştir. Bu kümelerin sonuçlarının gözlemlenmesi için Türkiye deprem haritası sunulmuştur.



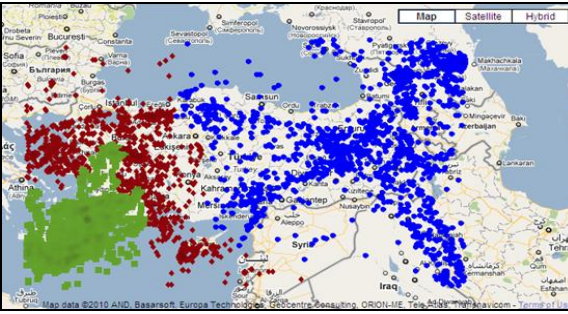
(a)



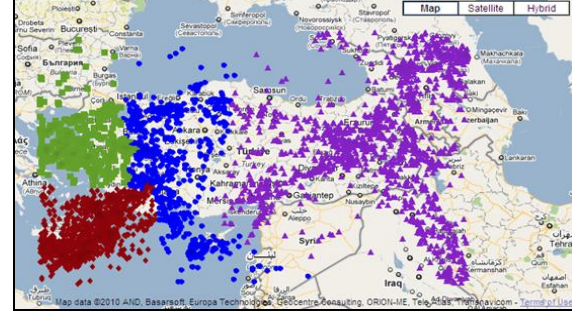
(b)



(c)



(d)



(e)

Şekil 7.29. Kümeleme sonuçlarının karşılaştırılması için küme sonuçlarının gösterimi

- a) Türkiye deprem haritası
- b) $k=3$ için k-Means algoritması ile elde edilen kümeleme sonuçları
- c) $k=4$ için k-Means algoritması ile elde edilen küme sonuçları
- d) $k=3$ için k-Means-KNN algoritması ile elde edilen küme sonuçları
- e) $k=4$ için k-Means-KNN algoritması ile elde edilen küme sonuçları

7.7. Bölüm Değerlendirmesi

Tezin bu bölümünde, depremlerin kümelendirilmesi için veri madenciliğinde sıklıkla kullanılan kümeleme algoritmalarından k-means ve KNN algoritmaları deprem verileri üzerine başarılı bir şekilde uygulanmıştır. Türkiye üzerindeki depremler farklı kümeler bölünerek deprem hareketlilikleri arasındaki ilişkiler tespit edilmeye çalışılmıştır. Elde edilen sonuçlar değerlendirildiğinde, büyük ölçekteki depremler öncesinde kümeler arasında bir hareketlilik ilişkisinin olduğu görülmüştür. Bazı kümelerin, diğer kümelerde oluşacak büyük çaptaki depremlerden önce, daha fazla hareketli olduğu tespit edilmiştir.

Klasik k-Means algoritmasındaki başlangıç küme merkezi belirleme problemine alternatif yeni bir yaklaşım olarak NrK-Means algoritması sunulmuştur. Bu yaklaşımda, başlangıç küme merkezleri rastgele değil, büyük ölçekli depremler arasından seçilen merkezler olarak atanmıştır. Klasik k-Means ve NrK-Means algoritmalarının farklı

sayıdaki depremleri kümelendirmesindeki performansları karşılaştırılmıştır. Yapılan karşılaştırmada NrK-Means algoritmasının verileri kümelendirmede daha hızlı olduğu görülmüştür.

K-Means algoritması kullanılarak yapılan kümeleme işlemlerinin iyileştirilmesine yönelik olarak K-Means ve KNN algoritmalarının birlikte yürütüldüğü bir kümeleme algoritması tasarlanmıştır. Bu algorithmada, büyük ölçekli depremler k-Means algoritması ile kümelendirilerek başlangıç sınıfları oluşturulmuştur. Daha sonra kümelendirilmek istenen deprem verileri KNN algoritması ile önceden oluşturulmuş kümelere dahil edilmiştir. Bu yöntemle Türkiye üzerindeki deprem haritası göz önüne alındığında yalnız k-Means algoritmasının kullanıldığı küme sonuçlarına göre daha mantıklı kümelerin oluştuğu görülmüştür.

8. SONUÇLAR

Bu tez çalışmasında, öncelikle bilgisayar bilimlerinin popüler konularından biri olan veri madenciliği konusuna değinilmiş ve bu alanda kullanılan çeşitli teknikler irdelenmiştir. Daha sonra, dünya genelinde deprem verileri için kullanılan çeşitli katalogların tanıtımı yapılmıştır. Bu kataloglardan, tez çalışmasında kullanılan Türkiye Deprem Kataloğu'nun veri formatı hakkında detaylı bilgi verilmiştir. Bölüm 4'te deprem verileri üzerine gerçekleştirilen çalışmalar için geliştirilen Türkiye Deprem Portalı uygulamasının tanıtımı yapılmıştır. Geliştirilen bu uygulama, deprem verilerinin dinamik haritalar üzerinde görsel olarak sorgulanabilmesi, birçok algoritmanın kolayca entegre edilebilmesi ve internet üzerinden hizmet verebilmesi gibi büyük avantajlara sahiptir. Ayrıca Bölüm 4'te, bazı bölgeler üzerinde depremler ile ilgili yapılan çeşitli istatistiksel ve sorgulama çalışmalarına yer verilmiştir. Yapılan çalışmalar sonucunda, depremler ile ilgili istatistiksel çalışmaların, gelecekte meydana gelebilecek depremler hakkında önemli bilgiler verdiği gözlemlenmiştir. Depremlerin, ne sıklıkta gerçekleştiği ve en çok hangi bölgelerde olduğu gibi bilgilerin sunulması açısından yapılan bu çalışma başarılı sonuçlar vermiştir.

Bölüm 5, veri madenciliği tekniklerinden birliktelik kurallarının deprem verileri üzerinde kullanımına yönelik yapılan çalışmayı kapsamaktadır. Birliktelik kuralları, deprem bölgeleri arasındaki hareketlilik ilişkilerinin tespit edilmesinde kullanılmıştır. Elde edilen kurallar, bu algoritmanın deprem verileri üzerinde başarılı bir şekilde çalıştığını göstermiştir. Algoritmanın yürütümü sonucu tespit edilen tekrarlı öğelerin önemli bir kısmının, Türkiye deprem haritası üzerinde 1.dereceden deprem bölgelerine denk geldiği görülmüştür.

Bölüm 6'da, gelecekte meydana gelebilecek deprem hareketlerinin tespitine yönelik bir yaklaşım sunulmuştur. Bu yaklaşım, deprem hareketliliklerine göre kural veritabanından yararlanarak gelecekte deprem oluşabilecek bölgeler ile ilgili çeşitli tahminler gerçekleştirmektedir. Türkiye deprem kataloğu üzerinde gerçekleştirilen çeşitli tahmin çalışmaları, algoritmanın başarılı sonuçlar ürettiğini göstermiştir.

Bölüm 7'de, depremlerin oluşukları konuma göre kümelendirilmesine yönelik yapılan çalışmayı kapsamaktadır. Deprem verilerinin kümelendirilmesi ile hareketliliklerin daha detaylı incelenebilmesi sağlanmıştır. Bu kapsamda, veri madenciliğinde kümeleme için sıklıkla kullanılan k-Means ve KNN algoritmaları başarılı bir şekilde deprem verileri

üzerinde çalışacak şekilde modellenmiştir. K-Means algoritması kullanılarak yapılan kümeleme işlemini iyileştirmek amacıyla k-Means ve KNN algoritmaları hibrit bir şekilde kullanılarak depremlerin kümelendirilmesi işlemleri gerçekleştirilmiştir. Ayrıca k-Means algoritmasının büyük performans problemlerinden biri olan başlangıç küme merkezi atama işlemine alternatif bir yaklaşım olan NrK-means algoritması bu bölüm altında sunulmuştur. Bu algoritmada, başlangıç kümelerinin merkezleri, katalogda yer alan büyük çaplı depremler arasından seçilmiştir. Klasik k-Means ve NrK-Means algoritmalarının performansları karşılaştırıldığında, önerilen yeni yaklaşımın kümeleme işlemini daha hızlı gerçekleştirdiği gözlemlenmiştir.

Deprem verileri üzerine birçok uygulamayı bünyesinde bulunduran bu tez çalışması, deprem hareketliliklerinin izlenmesi ve bu veriler üzerinde çeşitli veri madenciliği teknikleri ile analizlerin yapılmasına yönelik geliştirilmiş başarılı bir çalışma olarak sunulmuştur. Ayrıca tez çalışmasında kullanılan yöntemlerin uygulanması için geliştirilen Türkiye Deprem Portalı, deprem alanında çalışan uzmanların yararlanabilecekleri bir araç olarak önerilmiştir.

Bu Tez çalışması, 1 adet uluslararası konferans yayını [52] ve 1 adet ulusal konferans yayını [53] ile tamamlanmıştır. Ayrıca 1 adet bildirinin özeti kabul edilmiştir.

KAYNAKLAR

- [1] **Url-1.** <http://www.belgenet.com/deprem/depremt.html>, Türkiye'deki Büyük depremler. Nisan 2010.
- [2] **Url-2.** <http://www.elazig.gov.tr/p997-rakamlarla-elazig-depremi.html>. Elazığ depremi Mart 2010.
- [3] **Holliday, J.R., Nanjo K. Z., Tiampo K. F., Rundle, J. B., and Turcotte, D. L.** 2008 Earthquake forecasting and its verification, *Nonlinear Processes in Geophysics* , 0000:0001–10 SRef-ID: 1607-7946/npg/2008-0000-0001.
- [4] **Kafka, A. L.** 2007. Does seismicity delineate zones where future large earthquakes are likely to occur in intraplate environments? *The Geological Society of America Special Paper* 425.
- [5] **Kanamori, H.** 2003. Earthquake prediction: An overview. In *International Handbook of Earthquake & Engineering Seismology*, ed. W. H. K. Lee, H. Kanamori, P. Jennings, and C. Kisslinger, 1,205–1,216. Amsterdam and Boston: Academic Press.
- [6] **Turcotte, D. L.,** 1991. Earthquake prediction. *Annual Review of Earth and Planetary Sciences* .19,263–281
- [7] **Holliday, J. R., Chien-Chih C., Tiampo, K. F., Rundle, J. B., Turcotte, D. L., and Donnellan,** 2007. A RELM earthquake forecast based on pattern informatics *Seismological Research Letters*. 78, 87-93.
- [8] **Nanjo, K. Z., Rundle, J. B., Holliday, J. R., and Turcotte, D. L.** 2005. Pattern Informatics and Its Application for Optimal Forecasting of Large Earthquakes in Japan, *Pure Appl. Geophys.*
- [9] **Kossobokov, V. G., Keilis-Borok, V. I., Turcotte, D. L., and Malamud, B. D,** 2000. Implications of a statistical physics approach for earthquake hazard assessment and forecasting. *Pure Appl. Geophys.*, 157, 2323–2349.
- [10] **Suratgar, A.A, Setoudeh, F., Salemi, A.H., Negarestani, A.,** 2008. Magnitude of Earthquake Prediction Using Neural Network. *Fourth International Conference on Natural Computation*, 448-452.
- [11] **Colombo, D., Gitis, V., Franco R.,** 1997. Application of pattern recognition techniques to long-term earthquake prediction in central Costa Rica. *Engineering Geology*, 48, 7-18.

- [12] **Zhang, Q., Wang C.**, 2008. Using Genetic Algorithm to Optimize Artificial Neural Network: A Case Study on Earthquake Prediction. *Second International Conference on Genetic and Evolutionary Computing*, 128-131.
- [13] **Ansari, A., Noorzad, A., Zafarani, H.**, 2009. Clustering analysis of the seismic catalog of Iran. *Computers & Geosciences*, **35**, 475–486.
- [14] **Pei, T., Zhu, A., Zhou, C., Li, B. and Qin, C.**, 2006. A new approach to the nearest-neighbour method to discover cluster features in overlaid spatial point processes. *International Journal of Geographical Information Science*, **Vol. 20, No. 2**, 153–168.
- [15] **Bodri, B.**, 2001. A neural-network model for earthquake occurrence. *Journal of Geodynamics*, **32**, 289–310.
- [16] **Adeli, H., Panakkat, A.**, 2009. A probabilistic neural network for earthquake magnitude prediction. *Neural Networks*, **22**, 1018-1024.
- [17] **Guo, W., Li, D., Watson, A.**, 2001. Application of Fuzzy Logic in Seismic Zonation. *Internatonial Fuzzy System Conference*, 829-832.
- [18] **Bayrak, Y., Öztürk, S., Çınar, H., Kalafat, D., Tsapanos, T. M., Koravos, G.Ch., Leventakis, G.A.**, 2009. Estimating earthquake hazard parameters from instrumental data for different regions in and around Turkey. *Engineering Geology*, **105**, 200–210.
- [19] **Yamashina, K.**, 2006. Trial of earthquake prediction in Japan and a statistical test of time-shift. *Tectonophysics*, **417**, 169–182.
- [20] **Jafari, M.A.**, 2009. Statistical prediction of the next great earthquake around Tehran, Iran. *Journal of Geodynamics*, Volume **49**, Issue1, 14-18.
- [21] **Url-3.** <http://quakesim.jpl.nasa.gov/scorecard.html>. QuakeSim Projesi Nisan 2010.
- [22] **Sıramkaya, E.**, 2005. Veri Madenciliğinde Bulanık Mantık Uygulaması, *Yüksek Lisans Tezi*, Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Konya.
- [23] **Two Crows Corporation.** Introduction to Data Mining and Knowledge Discovery, Third Edition ISBN: 1-892095-02-5.
- [24] **Ergüneş, H.F.**, 2004. Genetik Algoritmaların Veri Madenciliğinde Kullanılmasıyla İlginç Kuralların Bulunması, *Yüksek Lisans Tezi*, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- [25] **Zhang, X., Edwards, J., etc**, 2007. Personalised online sales using web usage data mining, *Computers in Industry*, **58**, 772–782.
- [26] **Akpınar, H.**, 2000, Veri tabanlarında bilgi keşfi ve veri madenciliği, *İ.Ü. İşletme Fakültesi Dergisi*, **29**, 1-22.

- [27] **Bojarczuk, C. C., Lopes, H. S., etc.,** 2004. A Constrained-Syntax Genetic Programming System for Discovering Classification Rules: Application to Medical Data Sets, *Artificial Intelligence in Medicine*, **30**, 27–48.
- [28] **Agrawal, R., Mehta, M., etc.,** 1996. The Quest Data Mining System, *IBM Almaden Research Center San Jose 6s California*, U.S.A.
- [29] **Anonim,** Information discovery in databases and data mining, 2000. I. U. *The school of business administration magazine*, **C:29**, 1, April.
- [30] **Allahverdi, N.,** 2002. An artifical intelligence application with expert systems, *Satin Publication Distribution*, Istanbul.
- [31] **Akbulut, S.,** 2006. Veri Madenciliği Teknikleri İle Bir Kozmetik Markanın Ayrılan Müşteri Analizi ve Müşteri Segmentasyonu, *Yüksek Lisans Tezi*, Gazi Üniversitesi, Ankara.
- [32] **Dinçer, E.,** 2006. Veri Madenciliğinde K-Means Algoritması ve Tıp Alanında Uygulanması, *Yüksek Lisans Tezi*, Kocaeli Üniversitesi, Kocaeli.
- [33] **Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R.,** 1994. Advances in data mining and Knowledge Discovery, *AAAI Pres*, USA.
- [34] **Dolgun., M. Ö.,** 2006. Büyük Alışveriş Merkezleri İçin Veri Madenciliği Uygulamaları, *Yüksek Lisans Tezi*, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- [35] **Mitchell, M.,** 1999. Introduction to Genetic Algorithms, A Bradford Book The MIT Press, England.
- [36] **Kaya, M.,** 2009. MOGAMOD: Multi-objective genetic algorithm for motif discovery, *Expert System with Applications*, **36**, 1039-1047.
- [37] **Beasley, D., Bull, D.R., Martin, R.R.,** 1993. An Overview of Genetic Algorithms:Part 1.Fundamentals, *Univ. Computing*, **15(2)**, 58–69.
- [38] **Gülten, A., Doğan, Ş.,** 2008. Genetik Algoritmalar Yönteminin Biyomedikal Verileri Üzerindeki Uygulamaları, *DAUM*, **7 (1)**, 12-16.
- [39] **Türkoğlu, İ., ve Toraman, S.,** 2007. Karar Ağaçları ve Fraktal Analiz Kullanarak Histopatolojik İmgelerin Sınıflandırılması, *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, **22 (4)**, 753 - 758.
- [40] **Karabatak, M., Ince, M. C.,** 2009. An expert system for detection of breast cancer based on association rules and neural network, *Expert Systems with Applications*, Volume **36**, Issue 2, Part 2, 3465-3469
- [41] **Yıldırım, Ö., Doğan Ş., ve Karabatak, M.,** 2008. Öğrenci Davranışlarının Tespitine Yönelik Geliştirilen Karar Destek Sistemi, *Bilimde Modern Yöntemler*

Sempozyumu (BMYS 2008), Eskişehir Osmangazi Üniversitesi, Eskişehir, Bildiriler Kitabı, 859-866.

- [42] **Altıntaş, T.**, 2006. Veri Madenciliği Metotlarından Olan Kümeleme Algoritmalarının Uygulamalı Etkinlik Analizi, *Yüksek Lisans Tezi*, Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Sakarya.
- [43] **Karabatak, M.**, 2008. Özellik Seçimi, Sınıflama Ve Öngörü Uygulamalarına Yönelik Birlikte Kuralı Çıkarımı Ve Yazılım Geliştirilmesi, *Doktora Tezi*, Fırat Üniversitesi, Fen Bilimleri Enstitüsü, Elazığ.
- [44] **Agrawal, R., Imielinski, T., Swami, A.**, 1993, Mining association rules between sets of items in large databases, *In ACM SIGMOD Conf. Management of Data*.
- [45] **Karabatak, M., İnce, M.C.**, 2004. Apriori Algoritması ile Öğrenci Başarısı Analizi, *Elektrik Elektronik Bilgisayar Mühendisliği Sempozyumu, ELECO 2004*, Elektronik-Bilgisayar, 348-352, Bursa.
- [46] **Han, J., Kamber, M.**, 2006. Data Mining: Concepts and Techniques, Second Edition.
- [47] **David, J. H.**, 1998. The American Statistician.
- [48] **Dogan, S.**, 2007. Veri madenciliği kullanarak biyokimya verilerinden hastalık teşhisi, *Yüksek Lisans Tezi*, Fırat Üniversitesi, Fen Bilimleri Enstitüsü, Elazığ.
- [49] **Nariman, A.**, 2005. Designing a Framework Of Intelligent Information Process On Dentistry Administration Data.:FASS.
- [50] **Url-4.** <http://www.data.scec.org/> . Güney Kaliforniya Deprem Katalogları, Mayıs 2010.
- [51] **Url-5.** <http://www.mam.gov.tr/YDBE/index.html>. Türkiye Deprem Kataloğu, Mayıs 2010.
- [52] **Aydın, G., Yıldırım, Ö.**, 2010, “Detecting Relationships Between of Earthquake Zones Using Association Rules”, *International Symposium on INnovations in Intelligent SysTems and Applications, INISTA 2010*, 21-24 June, Kayseri & Cappadocia Turkey, 297-301.
- [53] **Aydın, G., Yıldırım, Ö.**, 2010, “Pattern Informatics Metodu İle Deprem Tahmini ve Türkiye Deprem Verileri Üzerinde Uygulanması”, *Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu, ASYU 2010*, 21-24 Haziran, Kayseri & Kapadokya, Bildiriler Kitabı, 1-4.

ÖZGEÇMİŞ

Özal YILDIRIM

Tunceli Üniversitesi
Tunceli Meslek Yüksekokulu
Bilgisayar Programcılığı Programı

E-Posta: oyildirim@tunceli.edu.tr

- | | |
|------------------|--|
| 1983 | Tunceli/Çemişgezek'te doğdu |
| 1999–2001 | Elazığ Atatürk Lisesini tamamladı |
| 2002–2006 | Fırat Üniversitesi Teknik Eğitim Fakültesi Elektronik ve Bilgisayar Eğitimi Bölümünden mezun oldu. |
| 2006–2007 | Muğla/Bodrum Gündoğın Fahriye Ilıcak A.M.L ve E.M.L de sözleşmeli bilgisayar öğretmeni olarak görev yaptı. |
| 2007–2009 | Hakkari Anadolu Meslek ve Ticaret Meslek Lisesinde kadrolu bilgisayar öğretmeni olarak görev yaptı. |
| 2008- | Fırat Üniversitesi Bilgisayar Mühendisliği Yazılım A.B.D da yüksek lisansa başladı. |
| 2009- | Tunceli Üniversitesi Tunceli Meslek Yüksekokulu Bilgisayar Programcılığı Programında öğretim görevlisi kadrosunda çalışmaktadır. |