



**T.C.
ERCIYES ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI**

**YENİ NESİL METASEZGİSEL YAKLAŞIMLARIN KÜMELEME
PERFORMANSLARININ KARŞILAŞTIRILMASI: RESMİ
İSTATİSTİKLERDE UYGULAMA**

**Hazırlayan
Fatma TURNA**

**Danışman
Doç. Dr. Lale ÖZBAKIR**

Doktora Tezi

**Temmuz 2016
KAYSERİ**

**T.C.
ERCIYES ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI**

**YENİ NESİL METASEZGİSEL YAKLAŞIMLARIN KÜMELEME
PERFORMANSLARININ KARŞILAŞTIRILMASI: RESMİ
İSTATİSTİKLERDE UYGULAMA
(Doktora Tezi)**

**Hazırlayan
Fatma TURNA**

**Danışman
Doç.Dr. Lale ÖZBAKIR**

**Temmuz 2016
KAYSERİ**

BİLİMSEL ETİĞE UYGUNLUK

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

Adı-Soyadı: Fatma TURNA

İmza :



YÖNERGEYE UYGUNLUK

“Yeni Nesil Metasezgisel Yaklaşımların Kümeleme Performanslarının Karşılaştırılması: Resmi İstatistiklerde Uygulama” adlı Yüksek Lisans / Doktora tezi, Erciyes Üniversitesi Lisansüstü Tez Önerisi ve Tez Yazma Yönergesi’ne uygun olarak hazırlanmıştır.

Tezi Hazırlayan

Fatma TURNA

İmza

Danışman

Doç. Dr. Lale ÖZBAKIR

İmza

Endüstri Mühendisliği ABD Başkanı

Doç. Dr. İbrahim DOĞAN

İmza

Doç. Dr. Lale ÖZBAKIR danışmanlığında **Fatma TURNA** tarafından hazırlanan “**Yeni Nesil Metasezgisel Yaklaşımların Kümeleme Performanslarının Karşılaştırılması: Resmi İstatistiklerde Uygulama**” adlı bu çalışma, jürimiz tarafından Erciyes Üniversitesi Fen Bilimleri Enstitüsü **Endüstri Mühendisliği** Anabilim Dalında **Doktora** tezi olarak kabul edilmiştir.

28.07.2016

(Tez Savunma Sınav Tarihi)

JÜRİ:

Başkan: Doç.Dr. Lale ÖZBAKIR(Tez danışmanı)

Üye :Doç. Dr. Sinem KULLUK

Üye Yrd. Doç. Dr. Pınar TAPKAN

Üye :Yrd. Doç. Dr. Hamdi Giray REŞAT

Üye : Yrd. Doç. Dr. Mustafa GÖÇKEN

ONAY:

Bu tezin kabulü, Enstitü Yönetim Kurulunun 09/08/2016 tarih ve 2016/35-02 sayılı kararı ile onaylanmıştır.

09/08/2016
Enstitü Müdürü
Prof. Dr. Mehmet AKKURT

ÖNSÖZ/TEŞEKKÜR

Tez çalışmam boyunca, tez konumun belirlenmesi ve yürütülmesinde değerli görüş ve katkılarını ve kıymetli tecrübelerini esirgemeyen danışman hocam Doç. Dr. Lale ÖZBAKIR'a teşekkürü bir borç bilirim.

Çalışmamın her aşamasında anlayış, destek ve sabrıyla yanımda olan eşim Turgut TURNA'ya ve bu zorlu dönemde hak ettiği ilgiyi gösteremediğim halde yaşından büyük anlayışıyla bana destek ve güç veren biricik oğlum A. Tunahan TURNA'ya sevgi ve teşekkürlerimi sunarım.



YENİ NESİL METASEZGİSEL YAKLAŞIMLARIN KÜMELEME PERFORMANSLARININ KARŞILAŞTIRILMASI: RESMİ İSTATİSTİKLERDE UYGULAMA

Fatma TURNA

Erciyes Üniversitesi, Fen Bilimleri Enstitüsü

Doktora Tezi, Temmuz 2016

Danışman: Doç. Dr. Lale ÖZBAKIR

ÖZET

Veri analizinde kümeleme yöntemlerinin kullanılmasına ilişkin literatürde birçok çalışma bulunmakta ve kümeleme halen araştırmacılar tarafından zor bir problem olarak ele alınmaktadır. Parametre sayısı az ve hesapsal zamanı kısa olan algoritma arayışı nedeniyle yeni nesil yaklaşımların kümeleme problemine uygulanması ilgi çekmektedir.

Bu tez çalışması; özgün olarak literatüre giren ve performanslarını standart test fonksiyonlarında ispatlayan iki yeni nesil metasezgisel yaklaşımın kümeleme problemine uygulanmasını içermektedir. Bu algoritmalarından biri olan İyon Hareketi Algoritması, doğadaki iyonların hareketlerinden esinlenmiştir. Ağırlıklı Süperpozisyon Çekim Algoritması ise temelde süperpozisyon ve ajanların çekim hareketi prensiplerine dayanmaktadır. Her iki algoritma da literatürdeki sürekli, kategorik ve karma 10 veri kümesine uygulanarak, performansları Parçacık Sürü Optimizasyon Algoritması ve Yapay Arı Koloni Algoritması ile karşılaştırılmıştır. Uygun olmayan çözümlerin seçilmemesi için tüm algoritmalara Deb kuralı uygulanmıştır. Elde edilen sonuçlar, her iki algoritmanın da kümeleme probleminde rekabet edebilir olduğunu göstermiştir. Kategorik verilerdeki en iyi performansı gösteren iki algoritma, Türkiye İstatistik Kurumu tarafından düzenli olarak gerçekleştirilen Yaşam Memnuniyet Araştırması'na uygulanmış ve sonuçlar yorumlanmıştır.

Çalışmadaki görüş ve yorumlar hazırlayanın kendisine aittir, Türkiye İstatistik Kurumu'nu bağlamaz. Bu teze dayanarak yapılan tüm çalışmalar için de aynı durum geçerlidir.

Anahtar Kelimeler: Kümeleme; İyon hareketi algoritması; Ağırlıklı süperpozisyon çekimi algoritması; Yaşam memnuniyet araştırması.

COMPARISON OF NEW GENERATION METAHEURISTIC APPROACHES CLUSTERING PERFORMANCES: APPLICATION IN OFFICIAL STATISTICS

Fatma TURNA

Erciyes University, Institute of Science

Ph.D. Thesis, July 2016

Supervisor: Assoc. Prof. Dr. Lale ÖZBAKIR

ABSTRACT

There are many studies in the literature regarding the use of clustering methods and clustering is still considered as a difficult problem by researchers. Due to the searching for an algorithm with less number of parameters and short computational time, application of new generation approaches to the clustering problem draws attention.

This thesis includes the application of two new generation meta-heuristic approaches- originally entered the literature and proved the performances in standard test functions- to the clustering problem. The Ion Motion Optimization Algorithm that one of these algorithms has been inspired by the movements of ions in nature. In the second place the Weighted Superposition Attraction Algorithm is essentially based on superposition and the principles of the attracted movements of agents. Both algorithms applied to 10 continuous, categorical and mixed datasets in the literature and their performances are compared with the Particle Swarm Optimization Algorithm and the Artificial Bee Colony Algorithm. For non-selection of unsuitable solutions, Deb rules were applied to all algorithms. The obtained results showed that both algorithms are competitive in the clustering problem. Two algorithms that show the best performance in categorical data applied to the Life Satisfaction Survey regularly carried out by Turkish Statistical Institute and the results are interpreted.

The opinions and comments in this dissertation belong to writer and do not reflect the opinions of Turkish Statistical Institute. For all the work done on the basis of this thesis is the same as previous.

Keywords: Clustering; Ion motions algorithm; Weighted superposition algorithm; Life satisfaction survey.

KISALTMALAR

VM	:Veri Madenciliği
VTBK	:Veri Tabanlarından Bilgi Keşfi
TÜİK	:Türkiye İstatistik Kurumu
OLS	:Ordinary Least Squares (Sıradan En Küçük Kareler)
YMA	:Yaşam Memnuniyet Araştırması
CFS	:Correlation-based Feature Selection (Korelasyon Tabanlı Özellik Seçimi)
IMO	:Ion Motion Optimization (İyon Hareketi Optimizasyonu)
WSA	:Weighted Superposition Attraction (Ağırlıklı Süperpozisyon Çekimi)
ABC	: Artificial Bee Colony (Yapay Arı Kolonisi)
PSO	: Particle Swarm Optimization (Parçacık Sürü Optimizasyonu)
YSA	: Yapay Sinir Ağları (YSA)
GA	: Genetik Algoritma
EP	: Evrimsel Programlama
ES	: Evrim Stratejileri
GP	: Genetik Programlama
KKO	: Karınca Koloni Optimizasyonu
TB	: Tavlama Benzetimi

TA : Tabu Arama

GRASP : Greedy Randomized Adaptive Search Procedure
(Aç Gözlü Rassal Adaptif Arama Prosedürü)



İÇİNDEKİLER

YENİ NESİL METASEZGİSEL YAKLAŞIMLARIN KÜMELEME PERFORMANSLARININ KARŞILAŞTIRILMASI: RESMİ İSTATİSTİKLERDE UYGULAMA

	Sayfa
BİLİMSEL ETİĞE UYGUNLUK SAYFASI	i
YÖNERGEYE UYGUNLUK SAYFASI.....	ii
KABUL ve ONAY SAYFASI.....	iii
TEŞEKKÜR.....	iv
ÖZET.....	v
ABSTRACT.....	vi
KISALTMALAR	vii
İÇİNDEKİLER	ix
TABLolar LİSTESİ.....	xiv
ŞEKİLLER LİSTESİ	xv

1. BÖLÜM

GİRİŞ

1.1. Çalışmanın Kapsamı ve Önemi	1
1.2. Çalışmanın Amacı ve Yöntemi.....	2
1.3. Çalışmanın Bölümleri.....	4

2. BÖLÜM

VERİ MADENCİLİĞİ

2.1. Giriş	6
2.2. Veri Madenciliği Tanımı	6
2.3. Veri Tabanlarından Bilgi Keşfi Süreci	7

2.4. Veri Madenciliği Süreci	9
2.5. Veri Madenciliğinin Diğer Disiplinlerle İlişkisi.....	11
2.6. Veri Madenciliği Uygulama Alanları	11
2.7. Temel Veri Madenciliği Fonksiyonları	12
2.7.1 Özellik Seçimi	12
2.7.1.1 Korelasyon Tabanlı Özellik Seçimi	13
2.7.2 Sınıflandırma	13
2.7.3 Kümeleme	13
2.7.4 Birliktelik Analizi	13
2.8. Veri Madenciliğinde Karşılaşılan Sorunlar	14
2.9. Resmi İstatistiklerde Veri Madenciliği Uygulamaları	15
2.10. Sonuç	16

3. BÖLÜM KÜMELEME

3.1. Giriş	17
3.2. Kümeleme Tanımı.....	17
3.3. Kümeleme Aşamaları.....	19
3.4. Kümeleme Probleminin Matematiksel İfadesi	21
3.5. Benzerlik / Benzemezlik (Uzaklık) Ölçütleri	22
3.5.1. Sürekli Veriler için.....	23
3.5.2. Kategorik Veriler için	25
3.5.3. Karma Veriler için	26
3.6 Kümeleme Geçerlilik İndeksleri	28
3.7 Kümeleme Yöntemleri	31
3.7.1. Hiyerarşik Kümeleme	32
3.7.2. Hiyerarşik Olmayan Kümeleme	33
3.7.3. Yoğunluk Tabanlı Kümeleme	39
3.7.4. Izgara Tabanlı Kümeleme.....	39
3.7.5. Evrimsel Algoritmalar ile Kümeleme.....	40
3.7.6. Sürü Zekası Algoritmaları ile Kümeleme	40
3.7.7. Fizik Kuralları tabanlı Algoritmalar ile Kümeleme	41
3.8 Metasezgiseller ile Yapılan Kümeleme Çalışmaları.....	42

3.8.1. Sürekli Veri Kümelerinin Kümelenmesinde Metasezgiseller	42
3.8.2. Kategorik ve Karma Veri Kümelerinin Kümelenmesinde Metasezgiseller	44
3.9 Kümeleme Analizinde Karşılaşılan Sorunlar	45
3.10 Sonuç	46

4. BÖLÜM

YENİ NESİL METASEZGİSEL YAKLAŞIMLAR

4.1. Giriş	47
4.2 Yeni Nesil Metasezgisel Algoritmalar	47
4.2.1. İyon Hareketi Algoritması.....	47
4.2.1.1 Likit Aşaması.....	51
4.2.1.2 Kristal Aşaması	52
4.2.2. Ağırlıklı Süperpozisyon Çekim Algoritması.....	54
4.3. Parçacık Sürü Optimizasyon Algoritması	59
4.4. Yapay Arı Kolonisi Algoritması	61
4.5. PSO ve ABC Algoritmalarının Kümeleme Uygulamaları.....	64
4.6. Sonuç	66

5. BÖLÜM

METASEZGİSEL ALGORİTMALARIN KÜMELEMeye UYGULANMASI VE ANALİZ EDİLMESİ

5.1. Giriş	67
5.2. Uygulama	67
5.2.1. Aday Çözümlerin Tanımlanması.....	67
5.2.2. Sürekli Veri Kümelerinde Uygulama.....	68
5.2.2.1 Başlangıç Çözümlerinin Oluşturulması	68
5.2.2.2 Örneklerin Küme Merkezlerine Atanması.....	70
5.2.2.3 Kümelemenin Uygunluk Değerinin Hesaplanması.....	70
5.2.2.4 Orijinal IMO Algoritmasında Yapılan Modifikasyonlar	71
5.2.2.5 Orijinal WSA Algoritmasında Yapılan Modifikasyonlar	75
5.2.3. Kategorik ve Karma Veri Kümelerinde Uygulama.....	76

5.2.3.1 Başlangıç Çözümlerinin Oluşturulması	77
5.2.3.2 Örneklerin Küme Merkezlerine Atanması.....	77
5.2.3.3 Kümelemenin Uygunluk Değerinin Hesaplanması	78
5.2.3.4 Pozisyon Güncellemesi.....	78
5.3. Deneysel Çalışma.....	78
5.3.1 Deney Kurulumu	79
5.3.2 Performans Karşılaştırması	81
5.3.2.1 Sürekli Veri Kümeleri için Karşılaştırma.....	82
5.3.2.2 Kategorik Veri Kümeleri için Karşılaştırma.....	87
5.3.2.3 Karma Veri Kümeleri için Karşılaştırma.....	90
5.3.2.4 Algoritmaların İstatistiksel Olarak Karşılaştırılması	95
5.3.2.4.1 Performans Kriterleri Açısından Algoritmaların İkili Karşılaştırması	95
5.3.2.4.2 Algoritmaların Performans Kriterlerinin Eşzamanlı İstatistiksel Karşılaştırması.....	97
5.4. Sonuç	100

6. BÖLÜM

YAŞAM MEMNUNİYET ARAŞTIRMASI ÜZERİNE VERİ MADENCİLİĞİ UYGULAMALARI

6.1. Giriş	102
6.2 Yaşam Memnuniyet Araştırması.....	102
6.3 Yaşam Memnuniyet Araştırması Verileri Üzerinde Yapılan Veri Madenciliği Uygulamaları	103
6.3.1. Veri Ön İşleme.....	103
6.3.2. Özellik Seçimi	105
6.3.3. Kümeleme	108
6.4. Sonuç	118

7. BÖLÜM

SONUÇ VE TARTIŞMALAR

7.1. Çalışmanın Katkıları	121
7.2. İleriye Yönelik Öneriler.....	122

EKLER.....	123
KAYNAKLAR	140
ÖZGEÇMİŞ.....	150



TABLolar LİSTESİ

Tablo 3.1.	Kümeleme notasyonları.....	21
Tablo 3.2.	Kümeleme yöntemleri	38
Tablo 4.1.	Notasyonlar.....	48
Tablo 5.1.	Veri kümelerinin özellikleri.....	79
Tablo 5.2.	Parametre değerleri.....	81
Tablo 5.3.	Sürekli veri kümelerinde algoritmaların uygunluk değerleri.....	83
Tablo 5.4.	Sürekli veri kümelerinde algoritmaların hata oranları.....	84
Tablo 5.5.	Sürekli veri kümelerinde algoritmaların doğruluk oranları.....	85
Tablo 5.6.	Sürekli veri kümelerinde algoritmaların hesapsal zamanları.....	86
Tablo 5.7.	Kategorik veri kümelerinde algoritmaların uygunluk değerleri	87
Tablo 5.8.	Kategorik veri kümelerinde algoritmaların hata oranları	88
Tablo 5.9.	Kategorik veri kümelerinde algoritmaların doğruluk oranları.....	89
Tablo 5.10.	Kategorik veri kümelerinde algoritmaların hesapsal zamanları	90
Tablo 5.11.	Karma veri kümelerinde algoritmaların uygunluk değerleri	91
Tablo 5.12.	Karma veri kümelerinde algoritmaların hata oranları	91
Tablo 5.13.	Karma veri kümelerinde algoritmaların doğruluk oranları.....	92
Tablo 5.14.	Karma veri kümelerinde algoritmaların hesapsal zamanları	92
Tablo 5.15.	K-ortalamlar algoritması ile metasezgisel algoritmaların uygunluk değerlerinin karşılaştırılması.....	94
Tablo 5.16.	Wilcoxon işaret testinden elde edilen p değerleri.....	95
Tablo 5.17.	Algoritmaların uzlaşma değerleri	98
Tablo 5.18.	Uzlaşma değerleri açısından algoritmaların p değerleri.....	99
Tablo 6.1.	YMA veri kümelerinin özellikleri	103
Tablo 6.2.	Tüm yıllarda özellik seçimi ile seçilen ortak özellikler.....	106
Tablo 6.3.	Yıllara göre ortak olarak belirlenen özellikler.....	107
Tablo 6.4.	YMA kümeleme sonuçları.....	108
Tablo 6.5.	2010 YMA veri kümesi küme merkezleri	109
Tablo 6.6.	2011 YMA veri kümesi küme merkezleri	111
Tablo 6.7.	2012 YMA veri kümesi küme merkezleri	113
Tablo 6.8.	2013 YMA veri kümesi küme merkezleri	115
Tablo 6.9.	2014 YMA veri kümesi küme merkezleri	117

ŞEKİLLER LİSTESİ

Şekil 2.1.	VTBK süreci.....	8
Şekil 2.2.	Veri madenciliği süreci.....	10
Şekil 2.3.	Veri madenciliğinin diğer disiplinlerle ilişkisi	11
Şekil 3.1.	(a) Kümelenmemiş veri (b) Kümelenmiş veri	18
Şekil 3.2.	Küme içi ve kümeler arası uzaklıklar	19
Şekil 3.3.	Kümeleme analizi temel akış şeması	20
Şekil 3.4.	Kümeleme yöntemleri.....	31
Şekil 3.5.	K-ortalamlar tekniği adımları.....	34
Şekil 3.6.	Bir veri kümesinde üç küme merkezinin bulunması durumu için k-ortalamlar tekniğinin kullanılması örneği	34
Şekil 3.7.	CLARA algoritması akışı	36
Şekil 3.8.	CLARANS algoritması akışı	37
Şekil 4.1.	IMO algoritmasının adımları	50
Şekil 4.2.	Likit aşamada anyon ve katyonların en iyi iyonlara hareketi	51
Şekil 4.3.	IMO algoritması kristal aşama kodu.....	53
Şekil 4.4.	Kristal aşamada iyonların yerel optimum etrafında olması	53
Şekil 4.5.	WSA algoritmasının adımları	55
Şekil 4.6.	WSA hedef nokta belirleme prosedürü.....	55
Şekil 4.7.	WSA arama yönü belirleme prosedürü.....	56
Şekil 4.8.	WSA pozisyon güncelleme prosedürü.....	58
Şekil 4.9.	PSO algoritmasının adımları.....	60
Şekil 4.10.	ABC algoritmasının adımları.....	62
Şekil 5.1.	Aday bir çözümün kodlaması	67
Şekil 5.2.	En iyi iyonların her iterasyonda değişmesi.....	73
Şekil 5.3.	En iyi iyonların global olarak korunması	73
Şekil 5.4.	En iyi uygunluk değeri değişmezse en iyi iyonların yeniden üretilmesi	74
Şekil 5.5.	En iyi uygunluk değeri değişmezse en iyi iyonların korunması	74

1. BÖLÜM

GİRİŞ

1.1 Çalışmanın Kapsamı ve Önemi

Verinin elde edilmesi, işlenmesi, saklanması ve değerlendirilmesi bilgiye ulaşmada önemli bir fonksiyona sahiptir. Bilgiye ulaşmak her zaman ve durumda önemli olmakla beraber, verilerin boyutlarının artmasıyla saklanması ve verilerin işlenerek bilgiye dönüşmesi daha çok problem olmaya başlamıştır. Ancak; bilgi teknolojilerinin gelişimi ve bilgisayar sistemlerinin her geçen gün hızlanmasıyla çok büyük boyutlarda verilerin de saklanabilmesi ve daha hızlı işlenebilmesi sağlanmaktadır. Böylece veriden bilgi elde etme süreci hızlanmakta ve karar süreçlerine katkıları daha verimli hale gelmektedir.

Bilgisayar sistemleri ile üretilen veriler, işlenmeden bir anlam ifade etmedikleri için verileri analiz ederek anlamlı bilgiye dönüştürme ihtiyacı doğmuş ve bu da veri madenciliği (VM) tekniklerini ortaya çıkarmıştır. Büyük hacimlerde verinin hemen her alanda toplanabilmesiyle de VM tekniklerinin her alanda kullanılabilmesi mümkün olmuştur.

Küreselleşen ve sürekli gelişen dünyada sosyal problemler de doğru orantılı olarak büyümekte ve problemlerin çözümü de zorlaşmaktadır. Problemlerin nedenleri ve çözüm yöntemleri üzerinde sosyolojik ve psikolojik değerlendirmeler sadece tecrübeye dayalı olarak yapıldığında yeterli olmamakta, çoğu zaman sayısal analiz tekniklerine ihtiyaç duyulmaktadır. Tam da burada; istatistik, VM, optimizasyon, karar destek sistemleri vb. araçlar devreye girmektedir. Sosyal alanlarda son zamanlarda çok sayıda ve çeşitte verinin özellikle anketler aracılığı ile elde edilmesi mümkün olmakla birlikte, anketler üzerinde yapılan basit istatistiksel analizler verideki gizli bilgiyi ve önemli karakteristikleri çıkarmada yeterli olmamaktadır. Bu tür anketler, sosyal problemler hakkında özellikle karar vericilere karar desteği ve keşfedilmemiş bilgiyi

sunacak verileri barındırabilmektedir. Bu anketlerin; toplumların sosyal, kültürel ve ekonomik karakteristikleri, insanların karar verme davranışları, hayata bakış ve mutluluk durumları hakkında madenlenecek gizli ve değerli verileri içermesi ve literatür taramasında resmi istatistiklerin çıkarıldığı anket çalışmaları üzerinde kümeleme kullanan az sayıda çalışmaya rastlanılması, tez çalışmasının motivasyonunu oluşturmaktadır.

Türkiye İstatistik Kurumu (TÜİK) tarafından 2003 yılından itibaren Türkiye’de yaşayan insanların yaşam memnuniyetlerini ölçmek için yapılan Yaşam Memnuniyet Araştırması (YMA) verilerinin önemli bir veri kaynağı olduğu tespit edilmiş ve VM tekniklerinden kümeleme yaklaşımının uygulanmasına karar verilmiştir. Böylece; Türkiye’de yaşayan insanların *umut* düzeyleri açısından kümelendiklerinde anket veri yapısındaki hangi sosyal, kültürel ve ekonomik özellikler açısından farklılaşacakları ortaya konulmaya çalışılmıştır. Bu amaçla; YMA’nın 2010-2014 yılları arasındaki verileri TÜİK’ten temin edilmiştir.

1.2 Çalışmanın Amacı ve Yöntemi

Çalışmanın amacı iki yeni nesil metasezgisel yaklaşımın kümeleme performanslarını karşılaştırarak, performansı yüksek algoritmalar ile YMA gibi önemli bir veri kaynağı üzerinde kümeleme uygulaması gerçekleştirmek ve böylece insanların *umut* düzeylerini etkileyen faktörler açısından gruplama yapısını ve bu yapının yıllar arasındaki değişimini görmektir. Bunun için öncelikle 2010-2014 yıllarına ait veri kümelerinin her birine özellik seçimi uygulanarak her yıl için *umut* değişkenini etkileyen faktörler tespit edilmiştir. Kümeleme öncesi veri kümeleri, tüm yıllarda ortak olan etkin faktörlere indirgenmiştir.

Yeni nesil metasezgisel yaklaşımlardan olan İyon Hareketi Algoritması (IMO) ve Ağırlıklı Süperpozisyon Çekimi Algoritması (WSA), kümeleme problemine uyarlanmış ve sürekli, kategorik ve karma tipteki veri kümelerinde performansları Parçacık Sürü Optimizasyonu (PSO) ve Yapay Arı Kolonisi (ABC) algoritmaları ile karşılaştırılmıştır. İndirgenmiş YMA veri kümesi kategorik tipte veriye sahip olduğu için kategorik veri kümelerinde en iyi performansı gösteren iki algoritma seçilerek YMA’nın eldeki tüm yıllarına ait veri kümelerine uygulanmış ve sonuçları yorumlanmıştır.

Literatürde yaşam memnuniyetine ilişkin karşılaşılan çalışmalardan bazıları şunlardır:

Bonikowska et al. [1] tarafından Kanada'nın Genel Sosyal Anketi ve Kanada Toplum Sağlığı Anketi üzerine yapılan çalışma, yaşam memnuniyetine verilen cevapların durağanlığını ve sosyo-ekonomik karakteristiklerle yaşam memnuniyeti arasındaki memnuniyetin değişkenliğini tutarlı bir analitik çerçevede ölçmeyi amaçlamıştır. Yıldan yıla yaşam memnuniyet skorundaki değişimleri, skor ortalama ve standart sapmaları hesaplayarak, sosyo-ekonomik karakteristiklerle yaşam memnuniyeti arasındaki değişimin büyüklüğünü ise Sıradan En Küçük Kareler Yöntemi (OLS)'ni kullanarak değerlendirmişlerdir.

Chou ve Chi [2], Hong-Kong'daki yaşlı Çinlilerin yaşam memnuniyetleri ile finansal gerilimleri arasındaki ilişkiyi, yaşam yönetim stratejileri kullanarak düzenlemeyi amaçlamışlardır. Hong-Kong'daki 421 yaşlı insana uygulanan anketin analizi için çoklu regresyon kullanmışlardır. Bulgular sonucunda, yaşlıların finansal geriliminin yaşam memnuniyetlerini olumsuz etkilememesi için yaşam yönetim stratejilerinin geliştirilmesi gerektiğini vurgulamışlardır.

Boyce et al. [3], evlilikle beraber yaşam memnuniyetinin arttığı ancak evlendikten kısa bir süre sonra hemen evlilik öncesi seviyesine indiğini iddia eden çalışmalarda kişiliklerin dikkate alınmadığını belirtmişlerdir. 2015 yetişkin Alman vatandaşına yaptıkları ankette kişiliklerin evlilik öncesine göre yaşam memnuniyetindeki katkısını çoklu regresyon kullanarak araştırmışlardır. Elde ettikleri sonuçlar ile evlilik nedeniyle yaşam memnuniyetinde azalma olduğu tezini çürütmüşlerdir.

Fleming et al. [4], Yeni Zelanda mahallelerindeki suç korkusu, yeşil alanların faydaları ve yaşam memnuniyeti arasındaki ilişkiyi regresyon analizi ile analiz etmişlerdir. York ve Bell [5], toplumların hangi karakteristiklerinin o toplumda yaşayan insanların yaşam doyumlarını etkilediğini OLS kullanarak analiz etmişlerdir. Araştırma; kadınların politik temsiliinin geniş olduğu, askeri harcamaların düşük ve sağlık harcamalarının yüksek olduğu yerlerde ulusların daha yüksek bir yaşam memnuniyetine sahip olduğunu ortaya koymuştur.

Chen et al. [6], 2011 yılında Çin'de yaşayan yetişkinler üzerinde yapılan bir ankete dayanarak, yaşam memnuniyeti üzerinde yaşın, medeni durum ve hane halkı

zenginliğinin etkisi olduğunu, çevre ilçe veya şehir bölge özelliklerinin, kentin büyüklüğünün, kişinin ikamet ettiği yerin konumunun ise bağımsız bir etkiye sahip olduğunu ortaya koymuşlardır. Analizlerini, OLS'ye göre yapmışlardır.

Du et al. [7], 1008 Hong Kong ve Macau üniversite öğrencisi arasında yaptıkları çalışmada; kişisel ve ilişkisel özsaygının içsel, aile, eş ve ruhsal olmak üzere dört umut kaynağı ile yaşam memnuniyeti arasındaki ilişkiye aracılık ettiği hipotezini yapısal eşitlik modellemesi kullanarak test etmişlerdir.

Haar et al. [8], Malezyalı, Çinli, Yeni Zelanda Maorili, Yeni Zelanda Avrupalı, İspanyol, Fransız ve İtalyan 1416 işçi üzerinde yapılan anket çalışmasını regresyon ile analiz etmişler ve yedi kültür arasında iş yaşam dengesinin iş ve yaşam memnuniyeti ile pozitif, ankisiyete (kaygılı ve endişeli hissetme) ve depresyon ile negatif ilişki içerisinde olduğu sonucuna varmışlardır.

Samaha ve Hawi [9], yaşam memnuniyeti ile akıllı telefon bağımlılığı, stres ve akademik performans arasındaki ilişkiyi 300 üniversite öğrencisinin katıldığı bir internet anket verisinde Pearson korelasyonu kullanarak analiz etmişlerdir.

Demircan [10], TÜİK tarafından 2012 yılında fert ve hane bazında yapılan YMA'nın ham verileri üzerinde Weka paket programı aracılığıyla sınıflandırma uygulaması yapmıştır. Çalışmada, hedef ve girdi değişkenlerinin belirlenmesinden sonra veri bütünleştirme, kayıp değerlerin doldurulması, tutarsız ve aşırı uç değerlerin veriden çıkarılması gibi veri ön işleme işlemleri gerçekleştirilmiştir. Sonuçta elde edilen veri kümelerine sınıflandırma algoritmaları uygulanmış ve sınıflandırma kuralları belirlenmiştir. Ayrıca; özellik seçimi algoritmaları kullanılarak indirgenmiş veri kümeleri elde edilmiş ve bu veri kümelerine de sınıflandırma uygulaması yapılmıştır.

1.3 Çalışmanın Bölümleri

Tez çalışması şu şekilde organize edilmiştir: İkinci bölümde VM ve resmi istatistiklerde VM çalışmaları hakkında bilgi verilmiş, üçüncü bölümde kümeleme ayrıntılı olarak anlatılmış ve metasezgiseller ile kümeleme literatürü verilmiştir. Dördüncü bölümde, yeni nesil metasezgisel algoritmalar ve çalışma kapsamında karşılaştırma amaçlı kullanılan PSO ve ABC algoritmaları açıklanmıştır. PSO ve ABC algoritmaları ile

yapılan kümeleme literatürü de bu bölümde verilmiştir. Beşinci bölüm, metasezgisel yaklaşımların kümeleme problemine uygulanmasının detaylı olarak açıklandığı bölümdür. Uygulama sonuçları da bu bölümde anlatılmaktadır. Altıncı bölümde, en iyi kümeleme sonuçlarını veren iki metasezgiselin YMA verilerine uygulamasını konu edinmiştir. Uygulama sonuçları ve yorumlar da bu bölümde yer almaktadır. Yedinci bölüm ise çalışmanın literatüre katkısının ve orijinalliğinin ortaya konduğu sonuç ve tartışmalar kısmıdır.



2. BÖLÜM

VERİ MADENCİLİĞİ

2.1 Giriş

Bu bölümde, tezde çalışılan konunun analiz tekniği olan kümelemenin de bir fonksiyonu olduğu VM ile ilgili bilgi verilecektir. VM'nin tanımı, Veri Tabanlarından Bilgi Keşfi Süreci (VTBK), genel VM süreci, disiplinlerle ilişkisi ve uygulama alanları açıklandıktan sonra, temel VM fonksiyonları, VM ile analizde karşılaşılan problemler ele alınacaktır. Temel VM fonksiyonları ele alınırken, tez çalışmasında kullanılan teknikler açıklanacaktır.

Tez çalışmasında kullanılan temel yöntem kümeleme olduğu için kümelemenin detaylı tanıtımı üçüncü bölümde yapılacaktır. Ayrıca; bu bölümün sonunda resmi istatistiklerde VM uygulamalarına dair literatür taraması verilecektir.

2.2 Veri Madenciliği Tanımı

Bu zamana kadar VM'nin birçok tanımı yapılmıştır. Basitçe; büyük ölçekli veriler arasından “değerli olan” bilgiyi elde etme işidir [11]. Sadece öngörülen değil öngörülemeyen bilgilerin de veri yığını içinden madenlenerek bilgisayar programları aracılığıyla çıkarılması şeklinde de tanımlanabilir.

Başka bir tanım da, önceden bilinmeyen ve veri içinde gizli olup potansiyel olarak faydalı olabilecek gizli bilgilerin çıkarılmasıdır [12].

Fayyad et al. [13], VM'yi veri içinde geçerli, yeni, potansiyel olarak faydalı ve anlaşılabilir örüntüleri teşhis etme işlemi olarak açıklamaktadır.

Tanımlardan anlaşılacağı üzere, VM temel olarak çok büyük boyutlardaki verilerin saklanması ve verilerden anlamlı bilgiler çıkarılması işlevlerini kapsamaktadır. VM, VTBK'nın bir alt sürecidir.

2.3 Veri Tabanlarından Bilgi Keşfi Süreci

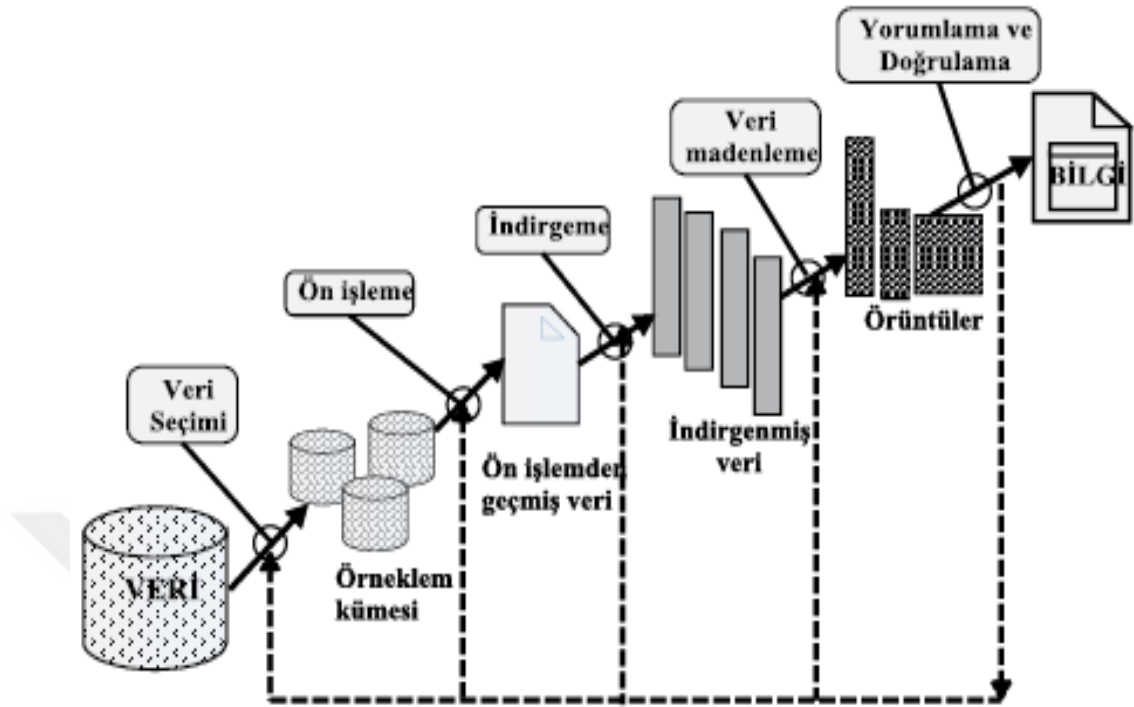
Veri tabanlarından bilgi üretmenin geleneksel yöntemlerle yapılması ve bilgilerin saklandığı veri tabanlarının boyutlarının artması analizin ya yetersiz kalmasına ya da önemli ölçüde zaman kaybına neden olmaktadır. Bu problemlerin üstesinden gelmek için ilk olarak Piatetsky-Shapiro [14] tarafından 1989 yılında gerçekleştirilen ilk VTBK çalışma grubu toplantısında kullanılmış olan VTBK, yaygın olarak hala kullanılmaya devam edilmektedir.

Veriden anlamlı bilgi çıkarmak için kullanılan yöntemler dizisi olarak tanımlanabilen VTBK'da temel problem, karmaşık ve çok büyük boyutlu veriyi özet ve faydalı bir forma getirmektir. VTBK, çok büyük boyutlu ve karmaşık veri yığınları içinden önceden tahmin edilemeyen soruların cevaplarını veren, çekirdeğinde VM bulunan geleceğe yönelik bir tahmin sistemidir [13].

Klasik çözümleme ve yorumlama, veriden anlamlı bilgi çıkarımının geleneksel yöntemini oluşturmaktadır [15]. Hedeflenen bilgiye ulaşmak için uzun bir taramanın yapılması gerekliliğinin yanı sıra, verilerin hacminin büyümesi ve karmaşıklığının artması istatistik, veri yönetimi, makine öğrenmesi, yapay zeka, optimizasyon ve başka bazı alanların kesişiminde bulunan VTBK'lara olan ihtiyacı ortaya çıkarmıştır.

Çevrimiçi veri takibi, ürün kontrol sistemleri, tıp, pazarlama, iş yönetimi, coğrafi bilgi sistemleri, finans, kimya, görüntü tanıma, robot görüş sistemleri, mühendislik, sosyal bilimler, uzay bilimleri, davranış bilimleri ve meteoroloji başta gelen VTBK uygulama alanlarıdır [15].

VTBK süreci içerisinde, VM modelin kurulması ve değerlendirilmesi aşamalarını içermekte ve en önemli kısmını oluşturmaktadır. Bu nedenle, birçok araştırmacı VTBK ile VM terimlerini eş anlamlı olarak da kullanmaktadırlar [16].



Şekil 2.1 VTBK süreci [18]

Şekil 2.1’de görüldüğü gibi, çeşitli kaynaklardan toplanan veri içinden örneklem kümesi seçilerek ön işlemde geçirilmekte ve veri, madenleme çalışmasının yapılması için indirgenmektedir. İndirgenen verideki örüntülerin keşfi sonrasında gerekli yorumlamalar yapılarak bilgiye ulaşılmaktadır.

Fayyad et al. [17], VTBK sürecinin kullanıcı tarafından verilen birçok kararla interaktif ve yinelemeli bir süreç olduğunu ve birçok adımı içerdiğini belirtmektedirler. VTBK adımlarını aşağıdaki şekilde özetlemişlerdir:

1. **Uygulama Alanını Tanıma:** Uygulamanın amacını ve ilgili ön bilgiyi içermektedir.
2. **Hedef Bir Veri kümesi Oluşturmak:** Bu aşama, veri kümesinin veya keşfedilecek veri örnekleri veya değişkenlerin bir alt kümesinin seçilmesi işlemlerini kapsamaktadır.
3. **Veri Temizleme ve Ön İşleme:** Veri kümesindeki veya seçilen örnekteki hatalı veya sapan örnek ve niteliklerin çıkarılması gibi temel işlemleri içermektedir. Gürültü faktörü elimine olduğu için keşfedilen bilginin kalitesinin artması mümkün olmaktadır.

4. **Veri İndirgeme ve Projeksiyon:** Amaca uygun bir şekilde veriyi temsil edecek faydalı niteliklerin bulunması aşamasıdır. Verinin etkin temsilini sağlamak için etken değişkenlerin sayısını azaltacak boyut indirgeme veya veri dönüştürme metotları kullanılmaktadır.
5. **VM Fonksiyonu Seçme:** Özetleme, sınıflandırma, regresyon ve kümeleme gibi VM algoritmaları tarafından üretilcek modelin amacına karar verme aşamasıdır.
6. **VM Algoritmasının Seçilmesi:** Verideki örüntüleri araştırarak algoritmanın seçilmesi aşamasıdır. Uygun parametre ve modellere karar verilmektedir.
7. **VM:** Sınıflama, kümeleme, birliktelik analizi, sıralama, modelleme gibi VM sorgularının işletilmesini kapsamaktadır.
8. **Değerlendirme:** Keşfedilen örüntünün yorumlanması ve kullanıcının anlayabileceği şekilde görselleştirilmesi, faydalı bir formata dönüştürülmesi aşamasıdır.
9. **Keşfedilen Bilginin Kullanılması:** Performans sistemine elde edilen bilginin dahil edilmesi, ilgili taraflara rapor edilmesi, önceki bilgi ile potansiyel çelişkilerinin çözümlendirilmesi aşamasıdır.

2.4 Veri Madenciliği Süreci

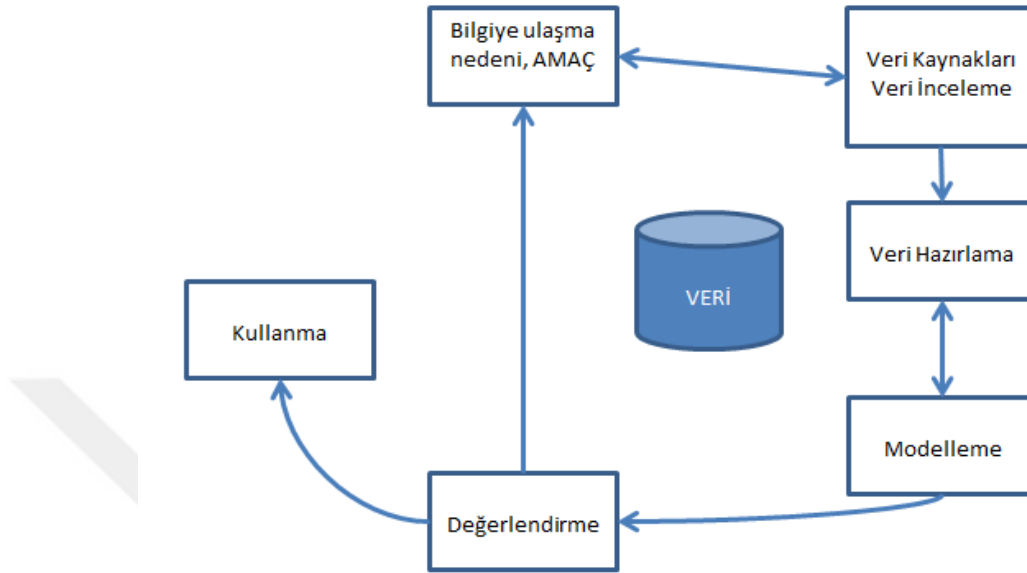
VM’de başarılı olabilmenin en önemli adımı, üzerinde araştırma yapılan işin ve ilgili verilerin içeriğinin tam olarak bilinmesidir. Bu bilgilere sahip olunduktan sonra VM süreci etkin olarak uygulanabilir. VM süreci Şekil 2. 2.’de gösterilmiştir.

Şekil 2.2.’deki VM sürecine ait adımlar ve alt açılımları aşağıda detaylandırılmıştır [19].

- **Problemin Tanımlanması:** Uygulamanın hangi amaçla yapıldığının açık bir şekilde tanımlanmasıdır. Problem açık bir dille anlaşılır bir şekilde ifade edilmiş olmalıdır.

Verilerin Hazırlanması: Gerekli veri kümelerinin toplanması, betimlenmesi ve temizlenmesidir. Veri ön işleme bu aşamada yapılmaktadır. Bu aşama sonrası veri analizi başlayabilmektedir. Veri ön işleme veri kalitesinin sağlanmasında

önemli bir aşamadır. En iyi model belirleninceye kadar yinelenmektedir. Verinin hazırlanması aşamasında aşağıdakiler yapılmaktadır:



Şekil 2.2 Veri madenciliği süreci [18]

- **Verilerin Temizlenmesi:** Aykırı değerlerin, eksik verilerin tespit edilmesi ile gürültünün elimine edilmesi gerçekleştirilmektedir.
- **Verilerin Birleştirilmesi:** Farklı veri tabanlarındaki verilerin bir araya getirilmesi gerçekleştirilmektedir.
- **Verilerin Dönüştürülmesi:** Verilerin madencilik yapılabilmesi için uygun formata dönüştürülmesidir. Bu aşamada veriler; dönüştürme, düzeltme, birleştirme, genelleştirme ve normalleştirme gibi işlemlerden geçirilmektedirler.
- **Verilerin İndirgenmesi:** Büyük hacimli veri kümesinden daha küçük hacimli veri kümesinin elde edilmesi aşamasıdır. Amacı en çok etkileyen özelliklerin seçilmesi olan boyut indirgeme işlemlerini içermektedir.
- **Modelin Kurulması ve Değerlendirilmesi:** Problemin amacına uygun modelin belirlenmesi aşamasıdır. Birçok model denenerek en iyi modelin seçilmesi gerekebilmektedir. Veri hazırlama gibi en iyi model belirleninceye kadar yinelenen bir aşamadır.

- **Modelin Kullanılması:** Kurulan ve belirlenen modelin doğrudan veya bir başka uygulamanın alt parçası olarak uygulanmasıdır.

2.5 Veri Madenciliğinin Diğer Disiplinlerle İlişkisi

VM'nin istatistik, yapay sinir ağları (YSA), yapay öğrenme, veri tabanı sistemleri, veri görselliği gibi birçok disiplinle ilişkisi vardır.



Şekil 2.3 Veri madenciliğinin diğer disiplinlerle ilişkisi [20]

2. 6 Veri Madenciliği Uygulama Alanları

Özellikle büyük hacimli verilerin bulunduğu hemen her alanda VM'nin kullanılması mümkündür. Karar verme ihtiyacını karşılamak adına verilerden anlamlı bilgilerin çıkartılması hemen her alanda talep edilen bir durum olduğu için VM'nin kullanım alanlarının sınırlandırılması söz konusu değildir.

Yine de bu zamana kadar özellikle; pazarlama, satış, müşteri ilişkileri yönetimi, sepet analizi, çapraz satışlar, hedef pazarların belirlenmesi gibi alanlarda kullanımı daha yaygındır. Ayrıca elektronik ticaret, müşteri şikayetlerinin incelenmesi, satış kampanyalarının, verimlilik çalışmalarının yapılması ve indirim kartları ve hediye puanların verilmesi, maliyet analizlerinin yapılarak karlılığın artırılması ve reklamcılık gibi bir çok alanda da ihtiyaç duyulan analizler yapılabilmektedir [21]. Biyoloji, borsa,

perakendecilik, endüstri, bilim ve mühendislik, istihbarat, sağlık, kriminoloji alanlarında da başarılı örnekleri bulunmaktadır [20].

2.7 Temel Veri Madenciliği Fonksiyonları

VM'nin en temel fonksiyonları sınıflandırma, kümeleme ve birliktelik analizidir. Özellik seçimi ise genellikle bu fonksiyonların ilk aşamasını temsil etmektedir.

2.7.1 Özellik Seçimi

Neredeyse tüm sınıflandırma algoritmalarının ilk aşaması özellik seçimidir. İlişkisiz özellikler, sınıf etiketiyle iyi bir ilişkiye sahip olmadığı için modeli zayıflatmaktadırlar. Eğitim verisinin küçük ve bu ilişkisiz özelliklerin eğitim modelinin bir parçası olmasına izin verildiği durumda, aşırı uyum sağlayabildiği için sınıflandırma doğruluğunu kötüleştirebilirler. Örneğin; belirli bir hastalığı tahmin etmek için hastalardan kan örneklerinin alındığı varsayalım. Bilinmektedir ki, kolesterol seviyesi kalp hastalığının bir tahmin edicisi iken, prostatın epitel hücrelerince üretilen protein yapısındaki bir enzim olan PSA'nın seviyesi öyle değildir. Bununla birlikte, eğitim verisi küçük alındığında PSA seviyesi rastgele varyasyonlar nedeniyle kalp hastalığı ile önemli bir korelasyona sahip olabilir. Tek bir değişkenin etkisi küçük olsa da ilişkisiz özelliklerin birikimli etkisi önemli olabilir. Bu durum, görülmeyen test verileri için genelleştirilmiş zayıf bir eğitim modeli sonucunu doğuracaktır. İşte bu nedenle, eğitim süreci esnasında doğru özelliklerin kullanılması kritik bir konudur [22].

Filtreleme ve sarmalama olmak üzere yaygın kullanılan iki özellik seçimi metodu vardır. Tez çalışmasında sarmalama metotlarından biri olan Korelasyon Tabanlı Özellik Seçimi (CFS) kullanılmıştır. Sarmalama yönteminde, seçim aşaması bir sınıflandırma algoritması ile iteratif bir yaklaşım içine gömülmüştür. Her iterasyonda sınıflandırma algoritması özelliklerin bir alt kümesini değerlendirmektedir. Değerlendirme için sınıflandırma algoritması kullanıldığı için bu yaklaşım genel olarak bir özellik kümesi oluşturmaktadır. Kısacası, hedef değişkenini etkileme gücüne göre en etkin değişkenlerden belirli bir yöntemle (örneğin açgözlü) alt küme oluşturularak seçim yapılmaktadır [22].

2.7.1.1 Korelasyon tabanlı Özellik Seçimi

Özellik seçiminin sarmalama tipi algoritmalarındandır. Seçilen özelliklerin bir alt kümesini sunduğu için tez çalışmasında sarmalama yöntemleri özellik seçimi için tercih edilmiştir. CFS'nin temel mantığı, iyi bir özellik alt kümesinin; sınıf değişkeni ile yüksek korelasyonlu ve birbirleriyle düşük korelasyonlu özellikleri içermesidir. Eğer özelliklerin her biri ile sınıf arasındaki korelasyon biliniyorsa ve her özelliğin diğer özelliklerle arasındaki korelasyon biliniyorsa, o zaman özellikler alt kümesi ile sınıf arasındaki korelasyon tahmin edilebilir [23]. Korelasyon katsayısı Eşitlik (2.1) ile hesaplanmaktadır.

$$r_{zc} = \frac{k\bar{r}_{zi}}{\sqrt{k + k(k-1)\bar{r}_{mm}}} \quad (2.1)$$

Burada; r_{zc} , özellikler kümesi ile sınıf arasındaki korelasyonu, k , özellik sayısını, $\bar{r}_{zi}$, özellikler ile sınıflar arasındaki korelasyon ortalamasını, $\bar{r}_{mm}$, özelliklerin birbirleri ile korelasyon ortalamasını ifade etmektedir.

2.7.2 Sınıflandırma

VM fonksiyonları arasında en çok kullanılan ve çalışılan alandır. Veri tabanının bir bölümü eğitim verisi olarak kullanılmakta ve sınıflandırma kuralları oluşturulmaktadır. Böylece yeni bir durum ortaya çıktığında karar vermek mümkün olmaktadır. Sınıflandırma yöntemleri arasında en çok kullanılanlar karar ağaçları, YSA, regresyon, diskriminant analizi ve kural çıkarımıdır.

2.7.3 Kümeleme

Veriler arasındaki benzerliklere bakılarak verilerin gruplandırılmasına kümeleme denmektedir. Üçüncü bölümde ayrıntılı olarak ele alınacaktır.

2.7.4 Birliktelik Analizi

Birliktelik analizi, verideki potansiyel ilişkileri ortaya çıkaran teknikler içerir. Ayrıca, eş zamanlı olarak gerçekleşen ilişkilerin tanımlanmasında kullanılırlar. Birliktelik kurallarının kullanıldığı en yaygın örnek “Sepet Analizi” dir. Bu analiz, müşterilerin alışveriş yaparken aldıkları ürünler arasındaki bağlantıyı, birliktelik kurallarını bularak

müşterilerin satın alma alışkanlıklarını ortaya koymaktadır. Böylece, market sahipleri için karar vermelerine destek amaçlı önemli bilgiler sunulmuş olmaktadır. Apriori algoritması, en çok kullanılan birliktelik analizi algoritmasıdır.

2.8 Veri Madenciliğinde Karşılaşılan Sorunlar

VM'nin büyük hacimli verilerin analizi ihtiyacından da doğduğu göz önüne alındığında büyük hacimli veri tabanlarında karşılaşılan problemler VM analizini de etkilemektedir. Veriler küçük hacimli olduğunda doğru çalışan yöntemler, verilerin büyük hacimli, gürültülü, eksik, boş değerli, artık ve dinamik olması durumunda yanlış çalışabilmektedir [24].

Büyük hacimli veri: Veri kümesinin hacminin büyük olması elde edilen örüntülerin gerçekten var olduğunu göstermesi açısından tercih sebebi iken, elde edilen örüntünün büyüklüğü de bir dezavantaj olarak değerlendirilmektedir. Bu problemin üstesinden gelmek için VM yaklaşımının arama uzayında sezgisel yöntemle tarama yapması, özellik seçimi ve boyut indirgeme gibi yöntemlerin uygulanması söz konusu olmaktadır.

Verinin gürültülü olması: Veri kümesindeki sistem dışı hatalara gürültü denir. Genelde verinin girişi ya da toplanması sırasında meydana gelmektedirler. Quinlan [25], veri kümesindeki gürültünün sınıflandırma üzerindeki etkisi için yaptığı deneyler sonucunda; etiketli öğrenmede etiket üzerindeki gürültünün öğrenme algoritmasının performansını direkt olarak etkilediğini ve performansın düşmesine sebep olduğunu bulmuştur. Yine de eğitim verisindeki özelliklerin gürültü miktarının sadece %10'luk bir kısmının ayıklanabildiğini ifade etmiştir.

Verinin eksik olması: Eksik veriden kastedilen veri kümesindeki olması gereken örneklerin ya da özelliklerin bazılarının bulunmamasıdır.

Boş değerlerin bulunması: Veri kümesindeki bilinmeyen ve uygulanamayan değerlere boş değer denir. Bu değerler için çeşitli çözümler sunulmuştur:

- Boş değerlerin tamamen ihmal edilmesi,
- Boş değerler yerine mümkün bir değer atanması. Böyle bir durumda; o özellikte en çok tekrar eden değer, boş değere en yakın değer veya varsayılan herhangi bir değer tercih edilebilmektedir [25].

Artık verinin olması: Artık veri, veri kümesindeki çıktıyı etkilemeyen gereksiz özelliklere denir. Bu özellikleri veri kümesinden çıkarmak için özellik seçimi algoritmaları geliştirilmiştir.

Dinamik verinin olması: İçeriği sürekli olarak değişen verilere denir. Bu tip veriler VM uygulamalarında problemlere neden olmaktadır. VM’den elde edilen örüntülerin dinamik verinin hangi değeri üzerinden ortaya çıktığının tespiti zorlaşmaktadır. VM algoritmaları çalıştırılırken verilerin değişimine izin verecek bir yapının bulunmaması ya da dinamik veride algoritma ile veri tabanının aynı anda çalıştırılmasının performansı düşürmesi dinamik veri kaynaklı sorunlardır.

2.9 Resmi İstatistiklerde Veri Madenciliği Uygulamaları

Nenonen [26], 2006-2007 yıllarında iş yerinde kayma, tökezleme ve düşme kazaları ile ilgili faktörleri analiz etmek için Fin ulusal iş kazaları ve hastalıkları istatistik verisine VM fonksiyonlarından karar ağaçları ve birliktelik kurallarını uygulamışlardır.

McCharty et al. [27], büyük veri kümelerinin üretildiği ticari uygulamalarda kullanılmak üzere çok sayıda VM tekniğinin geliştirildiğini ve bu tekniklerin geniş hacimli veriyi bilgiye dönüştürerek karar vermede yardımcı olduğunu ifade etmişlerdir. Ancak bu tekniklerin resmi istatistiklere uygulandığı çalışmaların sınırlı olduğunu eklemişlerdir. Federal istatistik ajanslarının VM tekniklerinin uygulanmasına elverişli büyük hacimli veri kümelerine sahip olduğunu ve bazı kurumların yavaş yavaş bu tekniklerin uygulamaya başladığını açıklamışlardır. VM uygulamalarının resmi istatistiklerin kalitesini ve üretim etkinliğini arttırmada yardımcı olacağını ifade etmişlerdir.

McCarthy et al. [28], “2009 Birleşmiş Milletler İstatistik Komisyonu’nun Resmi İstatistiklerde İlerlemeler” adlı toplantısında sundukları makalede, Ulusal Tarım İstatistikleri Servisi’nin VM tekniklerinden hangilerini hangi alanda kullandıklarını ve kullanılmasını önerdikleri bazı teknikleri açıklamışlardır. 2007 Tarım Sayımı çalışmalarında; sayım cevapsızlığının ağırlıklandırılması, sayım posta listesi düzeltilmesi, raporlama hatalarının analizi, anket teşviklerinin dağıtımı ve anket cevapsızlıklarının tahmini için karar ağaçlarını, soru kağıdı tasarımında eyaletlere ait

soru kağıtlarının birleştirilmesi amacıyla hiyerarşik kümeleme yöntemi kullanıldığını belirtmişlerdir.

Miller et al. [29], ABD Ulusal Tarım İstatistikleri Servisi'nin Uydu ile Tarım Sayımı çalışması kapsamında karar ağaçlarını kullandıklarını belirtmişlerdir. Karar ağaçları, uydu ile görünümüleri tespit edilen alanları sınıflandırmakta ve böylece tahmini ekim alanları belirli ekinler için tespit edilebilmektedir.

Hassani et al. [30] makalelerinde, resmi istatistiklere VM uygulamalarının tanıtımını yapmışlar ve VM için yazılım projelerindeki son gelişmeleri de dikkate alarak, resmi veri işleme çalışmasında bir VM uygulama örneği üzerinde akıllı veri kontrol sistem tasarımı önermişlerdir.

Schouten ve Nooji [31], Hollanda İstatistik Ofisi'nin gerçekleştirdiği 1998-2002 yıllarına ait Hanehalkı Yaşam Koşulları Anketi'nin verilerinin cevapsızlık ağırlıklandırmasında karar ağaçlarını kullanmışlardır.

Değirmenci [32], TÜİK tarafından düzenli olarak yapılan Hanehalkı Bütçe Araştırması'nın 2012 verisi üzerinde kümeleme ve birliktelik analizi çalışması yapmıştır.

2.10 Sonuç

Bu bölümde, VM'nin temel fonksiyonlarından olan kümeleme dışındaki özellik seçimi, sınıflandırma ve birliktelik analizi konularına, VTBK ve VM süreçleri ile VM'de karşılaşılan sorunlara yer verilmiş, kümelemenin detaylı anlatımı ise üçüncü bölüme bırakılmıştır. Ayrıca, resmi istatistiklerde VM çalışmalarına ait literatür taraması verilmiştir.

3. BÖLÜM

KÜMELEME

3.1 Giriş

Bu bölümde tezin metot kısmını temsil eden kümeleme üzerinde ayrıntılı durulacaktır. Kümelemenin tanımı, amacı, matematiksel modeli, kümelemede kullanılan benzerlik/benzemezlik ölçütleri ve kümeleme geçerlilik indeksleri, kümeleme yöntemleri ve kümelemede karşılaşılan problemler ele alınacaktır.

Tez çalışmasında kümelemenin sürü tabanlı ve fizik kuralları tabanlı metasezgisel algoritmalar ile yapılması nedeniyle ilgili yöntemlerle kümeleme üzerinde durulacak, kümelemede karşılaşılan problemlerden bahsedilecek ve son olarak da kümelemeye uygulanan metasezgisellerle ilgili literatür verilecektir.

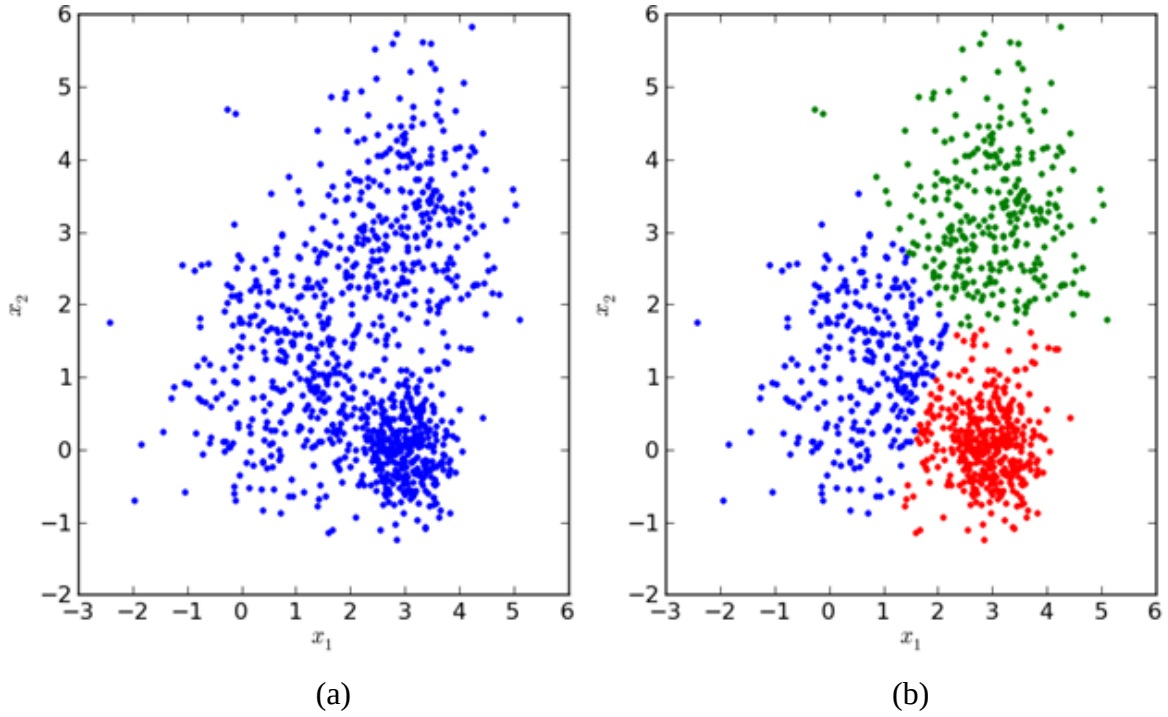
3.2 Kümeleme Tanımı

Kümeleme, veri kümesindeki her bir verinin özelliklerine göre daha hassas tek bir alt küme içerisine atanmasıdır. Veri kümesi, özellikleri olan örnekler olarak temsil edilir [33].

Benzer örnekleri gruplayan veri kümeleme, bir VM ve makine öğrenme işidir [34]. Ölçülen esas karakteristiklerin benzerliğine göre örneklerin doğal gruplandırılmasını keşfetmeyi amaçlayan danışmansız öğrenme tekniğidir [35].

Kümeleme; farklı gruplar için benzeyen örneklerin atanması, daha doğrusu, veri kümesinin tanımlanmış bazı uzaklık ölçülerine göre alt kümelere bölünmesidir. Bir küme, kendi elemanlarının birbirine benzediği ancak diğer kümelerdeki elemanlardan farklı olduğu bir yapıdır [36]. VM’de, genellikle ön işlem olarak kullanılır. Kümeleme problemi zor bir problemdir ve birçok disiplin ve birçok alanda araştırmacılar tarafından çalışılmaktadır [37].

Kümeleme analizi; genellikle çok boyutlu uzayda bulunan nokta ya da bir ölçüm sonucunda elde edilen noktalar olarak ifade edilen örneklerin birbirlerine olan benzerlikleri dikkate alınarak bir araya gelmesi ve gruplanması olarak da tanımlanabilir [38]. Şekil 3.1.(a)'da kümelenmemiş veri ve Şekil 3.1.(b)'de kümelere ayrılmış veriler gösterilmektedir.



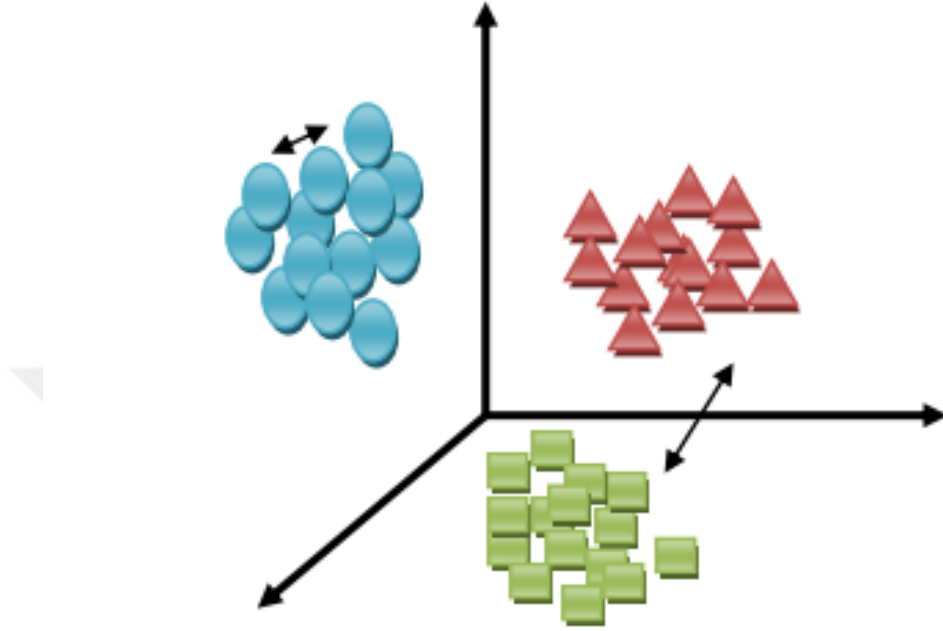
Şekil 3.1 (a) Kümelenmemiş veri (b) Kümelenmiş veri

Bütün bu tanımlamalardan sonra kümelemenin genel özellikleri aşağıdaki şekilde özetlenebilir:

- Verideki sınıf bilgisini kullanmadıkları için danışmansız öğrenme teknikleridir.
- Verileri en iyi temsil edecek vektörleri bulmaya çalışırlar.
- Bu vektörler ile tüm veriler kodlanabilirler ve bilgi farklılığı azalır.
- Bu özelliklerinden dolayı birçok indirgeme algoritmasının temelinde kümeleme bulunmaktadır.

Kümeleme hangi problem için kullanılırsa kullanılsın temel amacı; birey ya da örnekleri sahip oldukları karakteristik özellikleri dikkate alarak gruplamaktır. Kullanıcıya özetleyici bilgi sunmak da başka bir amacıdır. Sadece birbirine benzeyenleri gruplamak değil; aynı zamanda farklı gruplar arasındaki örneklerin de birbirlerine benzerliklerinin

minimize edilmesi de amaçlanmaktadır. Şekil 3.2 küme içi ve kümeler arası uzaklıkları göstermektedir.



Şekil 3.2 Küme içi ve kümeler arası uzaklıklar [39]

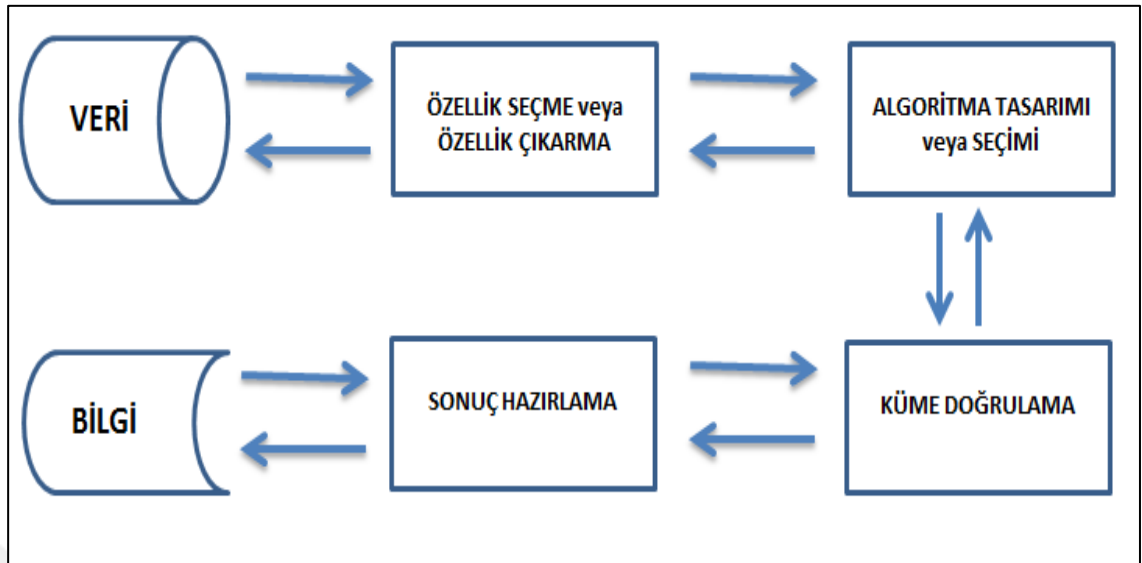
Kümeleme yöntemleri sınıf bilgisi içermeyen ‘etiketsiz’ verilerle ilgilenmektedir. Bu nedenle kümeleme, makine öğrenme tekniklerinin danışmansız öğrenme grubuna girmektedir. Danışmanlı öğrenme tekniği olan sınıflandırmadan en temel farkı da sınıflandırmada sınıf bilgisinin bulunması ve öğrenmenin buna göre yapılmasıdır. Danışmansız öğrenme özelliği nedeniyle veri analizinde kümeleme zor bir problem olarak kabul edilmektedir [40].

3.3 Kümeleme Aşamaları

Şekil 3.3’de kümeleme analizinin temel adımları verilmiştir [40].

Adımlar aşağıdaki şekilde detaylandırılabilir [41,42].

- **Özellik Seçme veya Çıkarma:** Özellik seçme, belirli bir kriter esas alınarak veriye ait özellik kümesinden uygun olanların seçilmesidir. Özellik çıkarma ise veri kümesine ait özellik kümesini dönüştürerek yararlı ve yeni özellikler elde etmektir.



Şekil 3.3 Kümeleme analizi temel akış şeması [40]

- **Algoritma Tasarımı veya Seçimi:** Kümelemede en önemli olgulardan biri, veri elemanları arasındaki benzerlik ölçüsünün belirlenmesidir. Bu aşamada, benzerlik ölçüsünün tespiti yapılarak kümeleme için kullanılacak algoritmanın tasarımı yapılır. Kümeleme yöntemleri ile benzerlik ölçüsü arasında sıkı bir ilişki vardır. Benzerlik ölçüsü kümeleme yönteminin kalitesi ile, kümeleme yöntemi ise kümelemenin kalitesi ile doğrudan bağlantılıdır. Ancak, bir problemi iyi çözen bir yöntemin diğer problemlerde de aynı performansı göstereceği söylenemez. Probleme uygun kümeleme yönteminin seçilmesi bu nedenle çok önemlidir.
- **Küme Doğrulama:** Aynı veri kümesi farklı kümeleme yöntemleri ile kümelendiğinde farklı sonuçlar meydana gelir. Kümeleme algoritmasının parametreleri değiştirildiğinde de sonuçlarda farklılıklar ortaya çıkabilir. Burada kümeleme kalitesini ölçmede doğrulama kriterleri devreye girmektedir. İçsel ve dışsal olmak üzere iki çeşit doğrulama kriteri vardır. İçsel indeksler, veride görünen karakteristik bilgiyi kullanarak küme kalitesini tespit ederken, veri kümesinin yapısına uygun en iyi küme sayısını da belirlerler. Dışsal indeksler ise veride görünmeyen sınıf bilgisi gibi etiketli bilgiyi kullanarak probleme uygun kümeleme yönteminin veya algoritmaya ait parametre değerlerinin saptanmasını sağlarlar. Genelde verilerin önemli bir bölümü etiketsiz olduğundan içsel indeksler daha çok kullanılmaktadır.

- **Sonuç Hazırlama:** Algoritmanın çalıştırılması ile elde edilen kümelenmiş verinin istatistiksel, matematiksel vb. metotlarla analiz edilerek anlamlı bilginin elde edilme aşamasıdır.

3.4 Kümeleme Probleminin Matematiksel İfadesi

Bu çalışmada, kümeleme fonksiyonları için kullanılan notasyonlar Tablo 3.1’de verilmiştir.

Tablo 3.1 Kümeleme notasyonları

D	iki vektör arasındaki benzerlik/benzemezlik
r	Minkowski uzaklık ölçütünde L_p norm
G	kümelerin meydana getirdiği küme
C_g	g . küme
K	küme sayısı
o_i	veri kümesindeki i . örneğin vektörü
o_{ij}	veri kümesindeki i . örneğin j . özelliğinin değeri
O	veri kümesindeki örneklerin kümesi
n	veri kümesinin örnek sayısı
J	veri kümesinin özellik sayısı
w_{ik}	i . örnek k . kümeye ait ise 1, aksi durumda 0
w_{ik}^t	t . iterasyonda i . örnek k . kümeye ait ise 1, aksi durumda 0
z_k	k kümesinin merkezi
z_{kj}^t	k kümesindeki tüm örneklerin j . özelliklerinin t . iterasyondaki ortalaması

Kümeleme probleminde, Eşitlik (3.1)’de belirtildiği gibi, m boyuttaki n tane örnek K sayıda kümeye bölünmektedir [43].

$$G = \{C_1, C_2, \dots, C_K\} \text{ (yani, } C_g \neq \emptyset, \forall g; C_g \cap C_h = \emptyset, \forall g \neq h; \bigcup_{g=1}^K C_g = O) \quad (3.1)$$

Eşitlik 3.1; O veri kümesinin K adet kümeye, hiçbir küme boş kalmayacak, her bir eleman sadece bir kümeye atanacak ve meydana gelen kümelerin birleşimi veri kümesinin tamamını oluşturacak şekilde parçalanmasını ifade etmektedir.

Kümeleme probleminin matematiksel ifadesi Eşitlik (3.2), Eşitlik (3.3) ve Eşitsizlik (3.4)' deki gibidir [44].

$$\text{Min} \sum_{k=1}^K \sum_{i=1}^n w_{ik} D(o_i, z_k) \quad (3.2)$$

$$\sum_{k=1}^K w_{ik} = 1, \quad i = 1, \dots, n \quad (3.3)$$

$$\sum_{i=1}^n w_{ik} \geq 1, \quad k = 1, \dots, K \quad (3.4)$$

$D(o_i, z_k)$, Eşitlik (3.6)'da verilmiştir.

Eşitlik (3.3), bir örneğin sadece bir kümeye atanması kısıtını ifade etmektedir. Eşitlik (3.4) ile de her kümeye en az bir örneğin atanması sağlanmaktadır.

Kümeleme probleminde her örnek bir önceki iterasyonda Eşitlik (3.5) ile güncellenen tüm küme merkezleri içerisinde kendisine en yakın kümeye atanmaktadır [45].

$$z_{kj}^t = \frac{\sum_{i=1}^n w_{ik}^t o_{ij}}{\sum_{i=1}^n w_{ik}^t}, \quad k = 1, \dots, K, j = 1, \dots, m \quad (3.5)$$

3.5 Benzerlik/Benzemezlik (Uzaklık) Ölçütleri

Kümeleme problemi, benzer örnekleri aynı kümede, benzer olmayanları farklı kümelerde toplamayı hedeflediği için ilk olarak iki örnek arasındaki benzerlik/benzemezlik ölçütünün tanımlanması gerekir. Bu ölçütler, verinin sürekli, kategorik ya da karma tipte olmasına bağlı olarak değişmektedir.

3.5.1 Sürekli Veriler İçin

Kullanılan en yaygın ölçütler uzaklık ölçütleridir. Uzaklık ölçütleri arasından da en yaygın olarak bilinen ve kullanılanı, Minkowski uzaklığından türetilen ve Eşitlik (3.6) ile tanımlanan öklid uzaklığıdır [45].

- **Minkowski uzaklığı**

Minkowski uzaklığı Eşitlik (3.6)'nın sol taraf eşitliği ile gösterilmektedir.

$$D_r(o_i, o_l) = \left(\sum_{j=1}^J (o_{ij} - o_{lj})^r \right)^{\frac{1}{r}} \Rightarrow D_2(o_i, o_l) = \sqrt{\sum_{j=1}^J (o_{ij} - o_{lj})^2} \quad (3.6)$$

Buradaki r parametresi metriğin özelleşmesini sağlar. $r=1$ durumu Manhattan City Block uzaklığı, $r=2$ durumu Öklid uzaklığı ve $r \rightarrow \infty$ durumu Chebyshev uzaklığı olarak literatürde yer almaktadır.

Minkowski metriğini kullanmanın bir dezavantajı, veride büyük ölçeğe sahip bir özelliğin diğer özelliklere baskın gelmesine neden olmasıdır. Özelliklerin belirli bir aralıkta normalize edilmesiyle bu problemin üstesinden gelinebilmektedir [41].

- **Cosinus uzaklığı**

x ve y vektörleri arasındaki benzerlik, aralarındaki açının cosinus'ü ile ölçülebilir. Eşitlik (3.7)'de bu uzaklığın matematiksel formülü bulunmaktadır.

$$\cos(x, y) = \frac{x^T y}{\|x\| \|y\|} \quad (3.7)$$

$\| \cdot \|$, L_2 normu ifade etmektedir. Öklid uzaklığı ile cosinus benzerliği arasında bir ilişki vardır. x ve y birim vektör olduklarında cosinus benzerliği Eşitlik (3.8)'deki gibi benzemezliğe dönüşmektedir [35].

$$D_{\cos}(x, y) = 1 - \cos(x, y) = \frac{1}{2} D_2^2(x, y) \quad (3.8)$$

- **Nokta simetri uzaklığı**

Örnek o_i ve z_k arasındaki nokta simetri uzaklığı Eşitlik (3.9)'daki gibi hesaplanmaktadır [35].

$$D_{ns}(o_i, z_k) = \min_{x_j \in X \setminus x_i} \left\{ \frac{D_2(o_i - z_k, o_j - z_k)}{D_2(o_i, z_k) + D_2(o_j, z_k)} \right\} \quad (3.9)$$

- **Modifiye nokta simetri uzaklığı**

$2z_k - o_i, z_k$ 'nın prototipine göre o_i noktasının simetriğidir. $\sum \min(n)$, n tane en düşük argümanın toplamını hesaplar [35].

$$D_{ns}^*(o_i, z_k) = \frac{1}{2} \sum_{x_j \in z_k} \min(2) \{D_2(2z_k - o_i, o_j)\} \quad (3.10)$$

- **Mahalanobis uzaklığı**

$$D_{Mahalanobis}(o_i, z_k) = \sqrt{(o_i - z_k)^T S^{-1} (o_i - z_k)} \quad (3.11)$$

Burada S , örnek ya da küme içi kovaryans matrisidir. S^{-1} ise kovaryans matrisin tersidir [46].

- **Hotelling T^2 uzaklığı**

İki kümenin ortalama vektörlerinin karşılaştırılmasında kullanılır [39].

$$D_{ht}(z_i, z_k) = T^2 = \frac{n_1 n_2}{n_1 + n_2} (z_i - z_k)^T S^{-1} (z_i - z_k) \quad (3.12)$$

- **Pearson Korelasyon Katsayısı**

Örnekler arasındaki uzaklık Pearson korelasyon katsayısı kullanılarak da hesaplanır [39].

$$r_{ij} = \frac{\sum_{k=1}^d (o_{ik} - \bar{o}_i)(o_{jk} - \bar{o}_j)}{\sqrt{\sum_{k=1}^p (o_{ik} - \bar{o}_i)^2 \sum_{k=1}^p (o_{jk} - \bar{o}_j)^2}} \quad (3.13)$$

Eşitlik (3.13) Pearson katsayısını ifade etmektedir. $\bar{x}_i$, i . örnek üzerinden ölçülen tüm J değişkenlerinin ortalamasıdır ve Eşitlik (3.14) ile hesaplanır.

$$\bar{o}_i = \frac{1}{J} \sum_{k=1}^J o_{ik} \quad (3.14)$$

Korelasyon katsayısı hesaplandıktan sonra iki örnek arasındaki uzaklık Eşitlik (3.15) ile hesaplanır.

$$D(o_i, o_j) = (1 - r_{ij}) \quad (3.15)$$

3.5.2 Kategorik Veriler için

Veri kümesi tamamen kategorik veriden oluştuğu durumlarda k -ortalamlar tekniğinin uyarlanmış hali olan k -modlar en çok kullanılan tekniklerden biridir.

- **K -modlar algoritması:** K -ortalamlar tekniği iki vektör arasındaki Öklid uzaklığını baz aldığı için kategorik verinin kümelenmesinde kullanılamaz. K -ortalamlar tekniğinde aşağıdaki modifikasyonlar yapılarak, k -modlar algoritması oluşturulmuştur.
 - Kategorik veriler için “hamming uzaklığı” veya eşleştirmeli benzemezlilik ölçütünün kullanılması,
 - Kümelerin merkezi olarak ortalamlar yerine, modlarının (en çok tekrar eden değerin) alınması [47].

K -modlar algoritmasının adımları aşağıdaki gibidir:

- 1- Veri örnekleri arasından keyfi olarak K adet küme merkezi üret.
- 2- Eşitlik (3.16)’ya göre en yakın küme merkezine örnekleri ata.
- 3- Örneklerin hepsi atandıktan sonra atanmış örneklere göre küme merkezlerini güncelle.
- 4- Küme merkezi değişmeyinceye kadar veya belirlenmiş bir durdurma kriteri sağlanıncaya kadar Adım 2 ve 3’ü tekrarla [48].

$$D(X, Y) = \sum_{j=1}^{dim} \delta(x_j, y_j) \quad (3.16)$$

$$\delta(x_j, y_j) = \begin{cases} 0 & (x_j = y_j) \\ 1 & (x_j \neq y_j) \end{cases} \quad (3.17)$$

Eşitlik (3.16), X ve Y vektörleri arasındaki benzemezliğin, bu iki vektörün tüm boyutlarındaki benzemezliklerin toplamı olduğunu ifade etmektedir. Eşitlik (3.17) ise bu iki vektörün j . boyutları arasındaki benzemezliğin bu boyutlardaki değerlerin birbirine eşit olması halinde 0, aksi taktirde 1 olduğunu ifade etmektedir.

3.5.3 Karma Veriler için

Karma veriler hem kategorik hem de sürekli tipte özellikler içerirler. Karma veri tipleri için literatürde karşılaşılan uzaklık tipleri aşağıda sıralanmıştır.

- **Heterojen Öklid uzaklığı (Heterojen Euclidean–OverlapMetric)**

x ve y iki vektör olsun. Bu iki vektör arasındaki heterojen öklid uzaklığı Eşitlik (3.18)-Eşitlik (3.22) ile hesaplanmıştır. Kategorik özelliklerin hesaplanmasında çakışma ölçütü (overlap metric), sürekli özelliklerde normalize edilmiş öklid uzaklığı kullanılmıştır [49].

$$HEOM(x, y) = \sqrt{\sum_{a=1}^J d_a(x_a, y_a)^2} \quad (3.18)$$

$$d_a(x, y) = \begin{cases} 1, & \text{eğer } x \text{ ve } y \text{ bilinmiyor ise} \\ \text{overlap}(x, y), & \text{eğer } a \text{ kategorik ise} \\ rn_{diff}_a(x, y), & \text{aksi taktirde} \end{cases} \quad (3.19)$$

$$\text{overlap}(x, y) = \begin{cases} 0, & \text{eğer } x = y \text{ ise} \\ 1, & \text{aksi taktirde} \end{cases} \quad (3.20)$$

$$rn_{diff}_a(x, y) = \frac{|x - y|}{range_a} \quad (3.21)$$

$$range_a = max_a - min_a \quad (3.22)$$

Burada a , örneğin a . boyutunu ifade etmektedir.

- **Değer Fark Ölçütü (Value Difference Metric-VDM)**

Stanfil ve Waltz [50] tarafından kategorik özellikler için uygun bir uzaklık fonksiyonu olarak literatüre tanıtılmıştır. x ve y vektörlerinin a özelliği için basitçe değer fark ölçütü formülü Eşitlik (3.23)'de verilmiştir [50].

$$vdm_a(x, y) = \sum_{c=1}^C \left| \frac{N_{a,x,c}}{N_{a,x}} - \frac{N_{a,y,c}}{N_{a,y}} \right| = \sum_{c=1}^C |P_{a,x,c} - P_{a,y,c}|^q \quad (3.23)$$

Burada;

- $N_{a,x}$, a özelliğinde x değerine sahip olan öğrenme kümesindeki elemanların sayısı,
- $N_{a,x,c}$, c sınıfında a özelliği değeri olarak x 'e eşit olan örneklerin öğrenme kümesindeki sayısı,
- C ise problemdeki çıktı sınıflarının yani kümelerin sayısı,
- q sabit bir sayı, genellikle 1 veya 2 olarak alınır.
- $P_{a,x,c}$, c kümesindeki a özelliğinin x değerine sahip olma olasılığıdır ve Eşitlik (3.24) ile hesaplanmaktadır.

$$P_{a,x,c} = \frac{N_{a,x,c}}{N_{a,x}} \quad (3.24)$$

$$N_{a,x} = \sum_{c=1}^C N_{a,x,c} \quad (3.25)$$

- **Heterojen Değer Fark Ölçütü (Heterogeneous Value Difference Metric-HVDM)**

Karma veri tipleri için önerilen yöntemlerden biri de Heterojen Değer Fark Ölçütü'dür. Eşitlik (3.26) ile hesaplanmaktadır [49].

$$HVDM(x, y) = \sqrt{\sum_{a=1}^J d_a^2(x_a, y_a)} \quad (3.26)$$

$$d_a(x, y) = \begin{cases} 1 & x \text{ ve } y \text{ bilinmiyorsa} \\ \text{normalize edilmiş } vdm_a(x, y), & a \text{ kategorik ise} \\ \text{normalize edilmiş } diff_a, & a \text{ lineer ise} \end{cases} \quad (3.27)$$

$$\text{normalize edilmiş } vdm_a(x, y) = \sum_{c=1}^c \left| \frac{N_{a,x,c}}{N_{a,x}} - \frac{N_{a,y,c}}{N_{a,y}} \right| \quad (3.28)$$

$$\text{normalize edilmiş } vdm2_a(x, y) = \sqrt{\sum_{c=1}^c \left| \frac{N_{a,x,c}}{N_{a,x}} - \frac{N_{a,y,c}}{N_{a,y}} \right|^2} \quad (3.29)$$

$$\text{normalize edilmiş } vdm3_a(x, y) = \sqrt{C * \sum_{c=1}^c \left| \frac{N_{a,x,c}}{N_{a,x}} - \frac{N_{a,y,c}}{N_{a,y}} \right|^2} \quad (3.30)$$

$$\text{normalize edilmiş } diff_a(x, y) = \frac{|x - y|}{4\sigma_a} \quad (3.31)$$

3.6 Kümeleme Geçerlilik İndeksleri (Uygunluk Değerleri)

Bir kümeleme geçerlilik indeksi, verideki doğal bilgiyi kullanarak kümeleme kalitesini ölçer. Bir geçerlilik indeksi aşağıdaki özelliklere sahip olmalıdır:

- Sezgisel bir anlamı olmalı,
- Hesaplaması kolay olmalı,
- Matematiksel olarak hesap verebilir olmalıdır.

Kümeleme algoritmasının türüne göre kümeleme geçerlilik indeksi, bulanık olarak da hesaplanabilir. Literatürdeki küme geçerlilik indeksleri aşağıdaki gibidir [35]. Formüllerde, D_e öklid uzaklığını ifade etmektedir.

- **Grup içi saçılma indeksi (Within-group scatter)**

Kümedeki örneklerin, küme merkezlerine olan öklit uzaklıklarının karelerinin toplamı olarak ifade edilir ve minimize edilmeye çalışılır ve Eşitlik (3.32) ile gösterilmektedir [35].

$$GIYI(G) = \sum_{C_k \in G} \sum_{o_i \in C_k} D_e^2(o_i, z_k) \quad (3.32)$$

- **Gruplar arası saçılma indeksi (Between-groups scatter)**

İki kümenin merkezlerinin arasındaki uzaklık indeksi Eşitlik (3.33)'ye göre hesaplanır ve maksimize edilmeye çalışılır [35].

$$GAUI(G) = \sum_{C_k, C_r \in G} D_e(z_k, z_r) \quad (3.33)$$

- **Bağlantı indeksi (Connectivity index)**

Aynı kümede bulunan örneklerin komşuluk derecelerini değerlendirerek küme bağlantılarını yansıtır. Minimize edilmeye çalışılır ve Eşitlik (3.34)'deki gibi hesaplanır [35].

$$BI(G) = \sum_{i=1}^n \left(\sum_{j=1}^L o_{i,n_{ij}} \right) \quad (3.34)$$

Burada;

$$o_{rs} = \begin{cases} \frac{1}{j}, & \text{eğer } \neg \exists C_k: r \in C_k \wedge s \in C_k; \\ 0, & \text{aksi taktirde} \end{cases} \quad (3.35)$$

n_{ij} , i . örneğin j . en yakın komşusudur. L ise bağlantı ölçümüne katkıda bulunan komşuların sayısıdır.

- **Dunn indeksi**

En yakın komşu uzaklığı ve maksimum küme çapı dağılımı tarafından tahmin edilen seçimde oran tipi indekstir. Maksimize edilmeye çalışılır ve Eşitlik (3.36) ile hesaplanmaktadır [35].

$$DI(G) = \frac{\min_{C_k \in G} \left\{ \min_{C_r \in G \setminus C_k} \{\delta(C_k, C_r)\} \right\}}{\max_{C_k \in G} \{\Delta(C_k)\}} \quad (3.36)$$

Burada;

$$\delta(C_k, C_r) = \min_{o_i \in C_k} \min_{C_r} \{D_e(o_i, o_j)\} \quad (3.37)$$

$$\Delta(C_k) = \max_{o_i, o_j \in C_k} \{D_e(o_i, o_j)\} \quad (3.38)$$

Dunn indeksi küme elemanları arasındaki varyans değerinin küçük olması (Eşitlik (3.38)) ve kümeler arasındaki mesafenin maksimum olmasını (Eşitlik (3.37)) ölçmektedir. Yani, bir kümenin diğer kümelerin elemanlarına olan en küçük uzaklığının, o küme içindeki en uzak iki örneğin uzaklıklarına oranını maksimize etmeye çalışmaktadır.

- **Silhoutte indeksi**

Aynı küme içindeki tüm noktaların birbirlerinden uzaklıkları toplamı ve farklı kümelerdeki noktalar arasındaki en yakın komşu uzaklığına dayalı hesaplanan indekstir. [-1,1] arası değerler alır. Maksimize edilmeye çalışılır ve Eşitlik (3.39) ile hesaplanır [35].

$$SI(G) = \frac{1}{N} \sum_{C_k \in G} \sum_{o_i \in C_k} \frac{b(o_i, C_k) - a(o_i, C_k)}{\max\{b(o_i, C_k), a(o_i, C_k)\}} \quad (3.39)$$

Burada;

$$a(o_i, C_k) = \frac{1}{n_k} \sum_{o_j \in C_k} D_e(o_i, o_j) \quad (3.40)$$

$$b(o_i, C_k) = \min_{C_r \in G \setminus C_k} \left\{ \frac{1}{n_r} \sum_{o_j \in C_r} D_e(o_i, o_j) \right\} \quad (3.41)$$

n_r , r . kümedeki eleman sayısını ifade etmektedir.

- **Turi indeksi**

Küme merkezleri arasındaki minimum uzaklık ve örnek ve ait olduğu küme arasındaki ortalama uzaklık hesaplanarak tahmin edilir. Minimize edilmeye çalışılır ve Eşitlik (3.42) ile hesaplanır [35].

$$TI(G) = (c \times N(\mu, \sigma) + 1) \times \frac{\text{grup içi index}(G)}{\text{gruplar arası index}(G)} \quad (3.42)$$

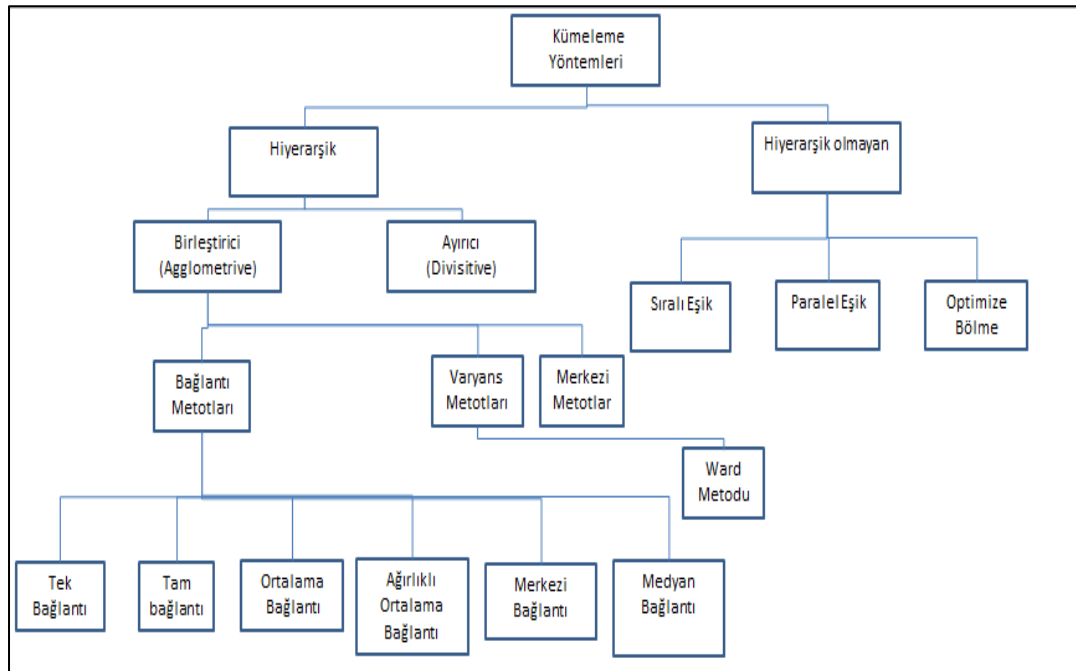
Burada c , kullanıcı tanımlı bir sabit, $N(\mu, \sigma)$, K merkezi aralık μ ortalama ve σ standart sapma ile hesaplanan normal dağılımın değeridir.

$$\text{grup içi index}(G) = \frac{1}{n} \sum_{C_k \in G} \sum_{O_i \in C_k} D_e(o_i, z_k) \quad (3.43)$$

$$\text{gruplar arası index}(G) = \min_{C_r \in G \setminus C_k} \{D_e(z_k, z_r)\} \quad (3.44)$$

3.7 Kümeleme Yöntemleri

Kümeleme yöntemleri hiyerarşik kümeleme ve hiyerarşik olmayan kümeleme şeklinde iki ana grupta incelenmektedir. Şekil 3.4 kümeleme yöntemlerini göstermektedir.



Şekil 3.4 Kümeleme Yöntemleri [51]

3.7.1 Hiyerarşik Kümeleme

Hiyerarşik kümeleme yöntemleri, örnekleri birbirleri ile farklı aşamalarda bir araya getirerek ardışık biçimde kümeler belirlemeyi ve bu kümelere girecek örneklerin hangi uzaklık ya da benzerlik düzeyinde küme elemanı olduğunu belirlemeye yarayan yöntemlerdir.

Hiyerarşik kümeleme yöntemleri, hiyerarşik ağaca benzeyen bir yapı kurarak verinin mümkün gruplandırılma şekillerini araştırır. Böyle bir ağaç yapısı sadece kümelerin basit dizisiyle değil; daha ziyade çok seviyeli hiyerarşik kurulumuyla oluşur [52].

Hiyerarşik yapıda; bir kümede bulunan her bir örnek, bir sonraki kümede de yer almakta, en üst seviyede yer alan küme tüm örneklerin yer aldığı küme olmakta ve alt seviyelerdeki kümelerde daha az örnek yer almaktadır. Hiyerarşik kümelemede küme sayısına görsel olarak karar verilmektedir. Karar verilirken de dendogram adı verilen ağaç diyagramından yararlanılmaktadır. Örnek sayısı $n < 300-400$ olduğu durumlarda tercih edilmektedir. Hiyerarşik kümeleme kendi içinde iki grupta incelenmektedir:

- **Birleştirici (agglomerative):** Her bir örnek başlangıçta ayrı bir küme olarak varsayılır. Bir sonraki adımda en yakın iki küme (örnek) birleştirilerek küme sayısı bir azaltılır. En son aşamada örnekler tek bir kümede birleştirilerek kümeleme sona erer [53]. Birleştirici hiyerarşik kümeleme yöntemlerinin çeşitleri aşağıdaki gibidir:
 - **Tek bağlantı tekniği:** En yakın komşuluk tekniği olarak da bilinen bu teknik, en kısa mesafe esasına dayanır. Uzaklıklar matrisi aracılığıyla birbirine en yakın kümeler birleştirilir ve bu birleştirmeler art arda tekrarlanır.
 - **Tam bağlantı tekniği:** En uzak komşuluk tekniği olarak da bilinen bu tekniğin tek bağlantı yönteminden farkı, iki küme arasında uzaklık olarak her kümedeki eleman çiftlerinin arasındaki uzaklığın en büyüğünün alınmasıdır.
 - **Ortalama bağlantı tekniği:** Farklı gruplarda yer alan gözlem çiftleri arasındaki ortalama uzaklık iki küme arasındaki uzaklık olarak alınır.

- **Ağırlıklı ortalama bağlantı tekniği:** Ortalama bağlantı tekniğine benzer. Farklılığı, yeni oluşan küme ile diğer kümeler arasındaki uzaklığın her kümedeki küme sayısı ile ağırlıklandırılmasıdır.
- **Merkezi bağlantı tekniği:** İki küme arasındaki uzaklık, kümelerin sahip oldukları küme merkezleri arasındaki uzaklık olarak alınır.
- **Medyan bağlantı tekniği:** İki küme arasındaki uzaklık, bu iki kümenin merkezleri arasındaki uzaklığın eşit ağırlıklı olarak hesap edilmesiyle bulunur.
- **Ward'ın bağlantı tekniği:** Kümenin ortasına düşen bir gözlemin, bu kümenin içinde bulunan diğer gözlemlerden ortalama uzaklığını esas alır.
- **Ayırıcı (divisitive):** Tüm örnekler yani veri kümesinin tamamı, tek bir küme olarak varsayılır. En uzak örnekler birbirinden ayrılarak küçük kümeler oluşturulur [53].

3.7.2 Hiyerarşik Olmayan Kümeleme

Küme sayısının belli olmadığı durumlarda hiyerarşik kümeleme kullanılmakta, küme sayısı belli ise hiyerarşik olmayan kümeleme metotlarından faydalanılmaktadır. Bu yöntemde örneklerin kümelere parçalanması rastgele yapılabilir. Kümelere örneklerin atanma kriterleri belirlendikten sonra örnekler kümelere atanır.

Tez çalışması kapsamında ele alınan yöntemlerin de dahil olduğu bu tür kümelemenin en yaygın çeşidi bölünmeli kümelemedir. Hiyerarşik bir yapı olmaksızın, ayırık kümelere örnekleri paylaşmak söz konusudur. En popüler bölünmeli kümeleme algoritmaları; her kümenin bir küme merkezi tarafından temsil edildiği ve kullanılan uygunluk değerinin örnek ile küme merkezi arasındaki uzaklıkların (karesel hata fonksiyonu) toplamı olarak hesaplandığı algoritmalar [54].

Bölünmeli kümeleme metotlarından en çok tanınanı, ilk olarak Macqueen [55] tarafından kullanılan k -ortalamalar kümeleme metodudur. Bölünmeli kümeleme kavramı, teorik olarak k -ortalamalar algoritması ile 1957 yılında başlamış ve günümüze kadar çok sayıda klasik bölünmeli kümeleme algoritması ortaya konmuştur.

- **K -ortalamalar tekniği:** K -ortalamalar tekniği, başlangıç çözümünü k merkezden alır, bu merkezlere en yakın uzaklıktaki veri kümesi örneklerini bu

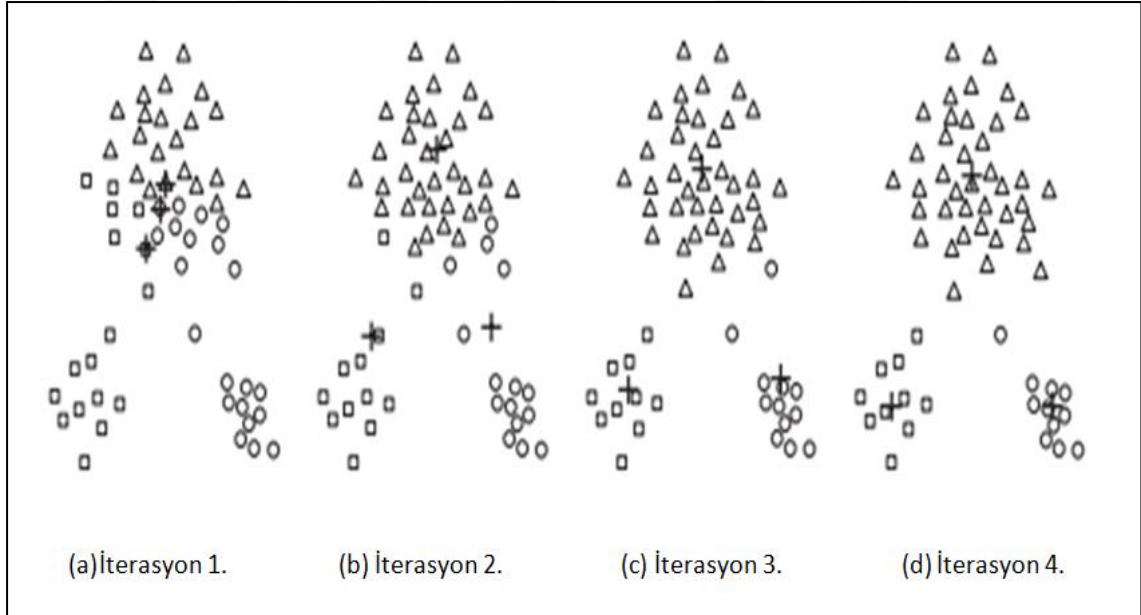
kümelere atar ve kümelerin merkezlerini tekrar bu örneklere göre günceller. Küme merkezlerinde bir değişme olmayıncaya kadar veya belirlenen durdurma kriteri sağlanıncaya kadar bu işlem tekrarlanır. *K*-ortalamlar basit ve hızlı bir algoritmadır; ancak gürültüye duyarlı ve başlangıç çözümüne bağlıdır [56].

Algoritmanın genel aşamaları Şekil 3.5'te verilmiştir.

-
1. *K* tane başlangıç küme merkezi seç.
 2. Her örneği onun en yakınındaki küme merkezine ata.
 3. Her kümenin merkezini güncelle.
 4. Küme merkezleri değişmiyorsa veya durdurma kriteri sağlandıysa dur, yoksa Adım 2'den devam et.
-

Şekil 3.5 *K*-ortalamlar tekniği adımları [57]

Ayrıca algoritma ile üç kümeli bir veri kümesinin dört iterasyon sonucunda küme merkezlerinin bulunması Şekil 3.6'da gösterilmektedir.



Şekil 3.6 Bir veri kümesinde üç küme merkezinin bulunması durumu için *k*-ortalamlar tekniğinin kullanılması örneği [57]

- ***K*-medoidler tekniği:** *K*-medoidler algoritması, *k*-ortalamlar algoritmasının dezavantajlarını gidermek üzere Kaufman ve Rousseeuw [58] tarafından

önerilmiştir. K -ortalamalar algoritmasının dezavantajları aşağıda verilmiştir [59].

- Küme sayısı değerinin bulunmasının zor olması,
- Global kümeler kullanıldığında etkili olmaması,
- Farklı başlangıç çözümleri seçildiğinde kümeleme sonuçlarının değişebilmesi yani başlangıç çözümüne bağlı olması,
- Farklı büyüklük ve farklı yoğunluktaki kümelerde doğal küme yapılarını bulmada algoritmanın başarısız sonuçlar vermesi.

K -medoidler algoritmasında örnekler, k -ortalamalar algoritmasındaki küme merkezleri yerine medoidler ile temsil edilmektedir. Yani kümeyi temsil edecek noktayı bulmak için o kümedeki elemanların ortalamasını almak yerine, o kümenin en merkezindeki noktayı yeni küme merkezi olarak almaktadır. Bu nedenle, k -ortalamalar algoritması ile karşılaştırıldığında sapan değerlere karşı daha az duyarlıdır. Güçlü bir algoritma olarak önerilen önemli bölünmeli algoritmalar arasındadır. Ancak, veri kümesinin hacmi arttıkça k -medoidler algoritması hesapsal zaman açısından etkin olmayan bir şekilde çalışmaktadır [60].

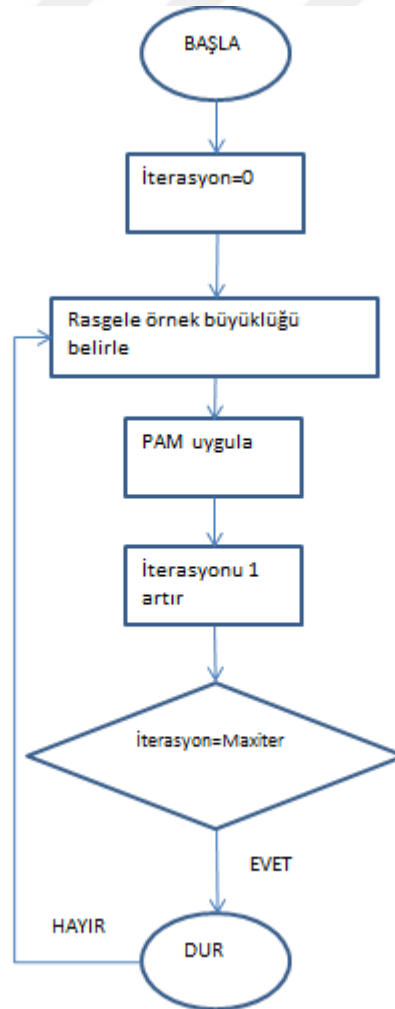
K -medoidler algoritmasının birçok versiyonu bulunmaktadır. Ortaya atılan ilk k -medoidler algoritması, PAM (Partitioning Around Medoids)'dir.

- **PAM:** K -ortalamalar tekniğinde olduğu gibi k adet örneği rastgele seçer ve küme merkezi olarak atar. Her yeni elemanın kümeye katılmasıyla, kümenin elemanlarını dener ve kümenin gelişmesine katkıda bulunacak noktayı tespit edince bulduğu noktayı yeni merkez olarak kabul eder. Eski merkez ise o kümenin herhangi bir elemanı olarak kalır.
- **CLARA:** Büyük veri tabanlarında PAM algoritmasının yetersiz kalmasından dolayı Kaufman ve Rousseeuw [61] tarafından önerilmiştir. CLARA, veri kümesinin tamamını almak yerine, veri kümesinden bir örneklem alıp üzerinde PAM algoritmasını uygular. Örneklem medoidlerini bulur. Buradaki önemli nokta, örneklem yeterince rastgele şekilde alınmış olmasıdır. Eğer rastgelelik yeterince sağlandıysa örneklem medoidleri tüm veri kümesinin yaklaşık olarak medoidleridir. CLARA, veri kümesinin tümünü temsil edecek en iyi medoidleri

bulmak için çoklu örneklem alır ve çıktı olarak en iyi kümelemeyi verir. Doğruluk için kümelemenin kalitesi, sadece örneklemdeki örnekler için değil; veri kümesindeki tüm örneklerin ortalama benzemezliği ile ölçülür [62].

CLARA'nın avantajı, PAM'e göre daha büyük veri kümelerine uygulanabilmesidir. Dezavantajı ise örneklem boyutuna göre performansının değişmesi ve örneklem seçiminin yeterince bağımsız (rastgele) olmaması halinde seçilen örneklem veri kümesini yeterince temsil edememesinden dolayı yanlış sonuçlara ulaşılmasıdır.

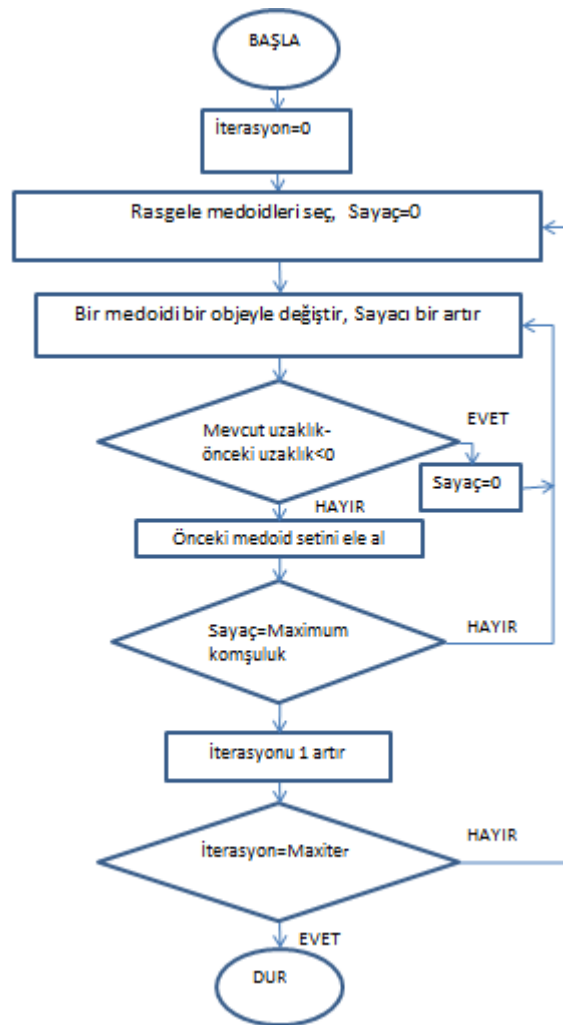
CLARA ile kümelemenin aşamaları Şekil 3.7'de verilmiştir.



Şekil 3.7 CLARA algoritması akışı [63]

- **CLARANS:** CLARANS algoritması, PAM ve veri kümesinin sadece alt kümesinde araştırma yapan CLARA'nın karışımı bir algoritmadır. Verilen bir

zamanda kendini sadece bir örnekleme hapsetmez. CLARANS ve PAM arasındaki fark, CLARANS'ın sadece bir düğümün komşularının bir örneklemini kontrol etmesidir. Ancak, CLARA'dan farklı olarak örneklemler üzerindeki çalışması dinamiktir. Yani, CLARA araştırmanın başında düğümlerin örneklemini düzenlerken, CLARANS ise araştırmanın her adımında komşuların örneklemini düzenler. Bu aramanın yerel olarak kalmasını engellemesi açısından faydalı bir durumdur. Algoritmanın akışı Şekil 3.8'de verilmiştir.



Şekil 3.8 CLARANS algoritması akışı [63]

Han ve Kamber [64], kümeleme yöntemlerini Tablo 3.2'deki gibi sınıflandırmış ve genel özelliklerini açıklamışlardır.

Tablo 3.2 Kümeleme yöntemleri

Yöntem	Genel Özellikleri
Bölünmeli Yöntemler	Küresel şekilde birbirini dışlayan kümeleri bulur.
	Uzaklık temellidir.
	Küme merkezlerinin temsilinde ortalama veya medoid kullanırlar.
	Küçük veya orta hacimli veri kümelerinde etkindirler.
Hiyerarşik Yöntemler	Kümeleme hiyerarşik olarak oluşmaktadır.
	Hatalı birleşimler veya bölünmeler düzeltilemez.
	Mikro kümeleme gibi diğer teknikleri içerebilir veya örnek bağlantılarını dikkate alabilir.
Yoğunluk tabanlı Yöntemler	Keyfi olarak şekillenmiş kümeler bulunabilir.
	Kümeler, düşük yoğunluklu bölgelerle ayrılan, örneklerin yoğun olarak bulunduğu bölgelerdir.
	Küme yoğunluğu: Her nokta komşulukları içinde minimum noktaya sahip olmak zorundadır.
	Sapan değerleri filtreleyebilir.
Izgara tabanlı Yöntemler	Çoklu çözünürlüklü ızgara veri yapısı kullanır.
	Hızlı işlem süresi (örneklerin sayısından bağımsız ancak ızgara sayısına bağlıdır.)

Bölünmeli ve hiyerarşik yöntemler yukarıda açıklanmıştır. Yoğunluk tabanlı ve ızgara tabanlı yöntemlerin açıklamaları aşağıda verilmiştir:

3.7.3 Yoğunluk tabanlı Kümeleme

Bu algoritmalar kümeleri, veri noktalarının yoğun olduğu ve verilen yoğunluk parametresine göre seyrek bölgelerden ayrılmış bölgelerden seçerler. Yoğunluk tabanlı kümeleme, kümeleri keyfi şekilde keşfedebilir [64]. Ester et al. [65] tarafından önerilen DBSCAN (Density Based Spatial Clustering of Applications with Noise) algoritması yoğunluk tabanlı stratejiyi uygulayan ve en yaygın kullanılan algoritmadır. Bu algoritmayı popüler kılan, veri kümesindeki mevcut gruplar hakkında herhangi bir ön bilgiye sahip olmaksızın keyfi şekillerle kümeler keşfedebilme kabiliyetidir [66]. DENCLUE (Density Based CLUstering), OPTICS (Ordering Pointsto Identify the Clustering) diğer yoğunluk tabanlı algoritmalarından bazılarıdır.

3.7.4 Izgara tabanlı Kümeleme

Bu yöntemde algoritmalar, veri noktalarından ziyade hücreleri göz önüne aldıklarından ızgara (grid) tabanlı yöntem ismini almışlardır. Izgara tabanlı algoritmalarda, veri alanı ızgara adı verilen hücrelerin sayısına eşit şekilde sınırlı bir sayıya bölünmektedir. Daha sonra yakın ızgaraların noktalarının oluşturduğu grup için bu ızgaraların üzerinde topolojik olarak komşuluk araştırması yapılmaktadır. Veri noktaları yerine ızgara hücreleri ile çalıştıklarından diğer algoritmalara göre hesapsal zaman açısından daha iyidirler [67]. STING (Statistical Information Grid), STING+, Wave Cluster ve CLIQUE (Clustering In QUEst) ızgara tabanlı algoritmalarındandır.

Bunların yanı sıra, model tabanlı kümeleme içine dahil edilen, iki aşamalı yöntemler de mevcuttur. Bunlardan biri; Kuo et al. [68] tarafından YSA ve çok değişkenli analizi bütünleştirilerek oluşturulan melez yapıdır. Kendini düzenleyen haritalar (self-organizing maps) ve k -ortalamlar tekniği iki aşamalı bir yaklaşım olarak kullanılmaktadırlar. İlk aşamada küme sayısı bulunmakta, k -ortalama tekniği ise son çözüme ulaşılmasında devreye girmektedir.

İki aşamalı yöntemlerde ilk aşamada hiyerarşik teknikler kullanıldığı için ilk aşamada bulunan sonuçlar açısından bir geri alınamama söz konusudur. Yani bir örnek bir

kümeye atandıysa artık başka bir kümeye atanması mümkün değildir. Bu problemin üstesinden gelmek için Kohonen [69] tarafından geliştirilen Kohonen yaklaşımı, öğrenen bir algoritma niteliğinde olduğundan güncellenebilmektedir.

Bunların dışında, kümeleme evrimsel algoritmalar aracılığıyla da yapılmaktadır.

3.7.5 Evrimsel Algoritmalar ile Kümeleme

VM’de yaygın olarak kullanılan evrimsel algoritmalar, doğal evrim sürecinden esinlenerek geliştirilmişlerdir. Gradyan tabanlı tekniklerden tamamen farklı olarak, çok sayıda aday çözümlerin performanslarını eşzamanlı olarak değerlendirerek zeki bir şekilde arama uzayında gezinirler ve küresel en iyiye yaklaşırlar [70].

Evrimsel algoritmaların biyolojik evrimden esinlendikleri; üreme, mutasyon ve doğal seçim gibi mekanizmalardır. Evrimsel algoritmalarla optimizasyonda, aday çözümler bir popülasyondaki bireyleri temsil eder. Popülasyonun evrimi, sayılan mekanizmaların uygulanması sonrasında meydana gelir.

Geleneksel arama metotlarında, probleme bir aday çözüm ortaya konur ve bu aday çözüm değiştirilerek daha iyi çözümler elde edilmeye çalışılır. Evrimsel algoritmalarda ise aday çözümler popülasyonu oluşturulur ve bu popülasyon zamanla evrimleşir. Algoritma, her adayın uygunluk değeri açısından ne kadar güçlü olduğunu hesaplar ve buna göre bir sonraki neslin ebeveynleri olacak ya da yok olacak bireyleri belirler. Bundan sonra, daha iyi bir yeni nesil oluşturmak için ebeveynlere genetik arama (çaprazlama, mutasyon) operatörlerini uygular. Böylece daha güçlü bireyler oluşturularak döngü devam eder.

Evrimsel hesaplamalar, formüle edilişlerine göre farklı isimlerle adlandırılırlar. Genetik algoritmalar (GA), Evrimsel Programlama (EP), Evrim Stratejileri (ES) ve Genetik Programlama (GP) gibi çeşitleri vardır.

3.7.6 Sürü Zekası Algoritmaları ile Kümeleme

Sürü, birbirleriyle etkileşen dağınık yapıda bireylerin oluşturduğu yığına verilen addır. Sürü zekası; bireyler arasındaki dolaylı ve dolaysız etkileşimi, çevre şartlarının değişimine karşı gösterdiği uyum esnekliğini ve kolonideki bazı bireylerin görevlerini

yerine getirememeleri durumunda işlerin devam edebilmesi kabiliyetini temsil eden bir kavramdır. Koloniye oluşturan bireyler çok basit kuralları takip etmektedir. Bireylerin nasıl hareket edeceklerine dair merkezi bir kontrol mekanizması yoktur. Buna rağmen, bireyler arasındaki yerel ve belli bir seviyede olan rastgele etkileşim sayesinde zeki global davranış meydana gelmektedir. Kombinatoriyal optimizasyon, iletişim ağları ve robotik alanlarında sürü zekası algoritmalarının başarıları üssel olarak artmaktadır [71].

Sürü zekası algoritmaları içinden en çok kullanılan PSO [72] ve Karınca Koloni Optimizasyonudur [73]. Tez çalışmada ele alınan algoritmalar; ABC arı kolonisi, PSO kuş sürüsü, WSA ajanların sürüsünü temsil ederler ve sürü zekası algoritmalarıdır. Dördüncü bölümde bu algoritmaların ayrıntılı olarak açıklamaları yapılacaktır.

3.7.7 Fizik Kuralları tabanlı Algoritmalar ile Kümeleme

Popülasyon tabanlı algoritmalar arasında yer alan fizik tabanlı algoritmalar doğadaki fizik kurallarından esinlenerek oluşturulan algoritmalarlardır. Evrimsel ve sürü zekası algoritmalarında olduğu gibi optimizasyon bireylerin ya da ajanların oluşturduğu aday çözümler grubu ile yapılmaktadır. Diğer iki tür algortmadan farkı, arama ajanlarının fiziksel kurallara göre hareket etmesi ya da kombine olmasıdır.

Bu tür algortmalardan biri Manyetik Optimizasyon Algoritmasıdır (MOA). Algoritma, arama uzayında arama ajanlarını hareket ettirmek için elektromanyetik parçacıklar arasındaki elektromanyetik kuvvetleri taklit etmektedir. Elektromanyetik kuvvet parçacıkların uygunluğu ile orantılı olduğundan, arama ajanları en uygun parçacıklara doğru çekilmeye meyillidirler. Bu nedenle, bu algoritmanın arama ajanları en iyi çözümlere doğru hareket ederek gelişmektedirler [74].

Bir diğer fizik tabanlı algoritma, Gravitasyonel Arama Algoritması(Gravitational Search Algorithm-GSA)'dır. Bu algortmada da arama ajanları birbirlerini kitleler halinde, uygunluk fonksiyonları ile orantılı olarak aralarındaki yerçekimi kuvvetine göre çekerler. Algoritma tarafından kitlesel hareketle ilgili olarak Newton'un hareket kanunu kullanılmaktadır [75].

Diğer fizik tabanlı algoritmalar; Işın optimizasyonu (Ray Optimization- RO) [76], Big-Bang-Büyük Patlama (Big-BangBig-Crunch-BBBC) [77], Yapay Kimyasal Reaksiyon

Optimizasyon Algoritması (Artificial Chemical Reaction Optimization Algorithm-ACROA) [78], Kara Delik (Black Hole -BH) [79], Maddenin Halleri Araştırması (States of Matter Search- SMS) [80] ve Kinetik Gaz Molekülleri Optimize Edicisi (Kinetic Gas Molecules Optimizer) [81] algoritmalarıdır.

Tez kapsamında ele alınan IMO algoritması da bu algoritmalar gibi fiziksel kurallardan olan doğadaki iyonların hareketlerinden esinlenerek oluşturulan yeni nesil bir algoritmadır. IMO algoritması, dördüncü bölümde ayrıntılı olarak incelenecektir.

3.8 Metasezgiseller ile Yapılan Kümeleme Çalışmaları

Tez kapsamında; gerek yeni nesil algoritmalar olan IMO ve WSA algoritmaları gerekse literatürde sıkça kullanılan algoritmalar olan PSO ve ABC algoritmaları, kümeleme problemine uygulanmış ve sonuçları karşılaştırılmıştır. Daha önce de bahsedildiği gibi kümelemenin birçok yöntemi vardır. 1990'da doğadan esinlenen metasezgisel yaklaşımlarıyla birlikte kümeleme analizinde yeni bir alan doğmuştur [82]. O günden bugüne birçok metasezgisel kümeleme literatürüne kazandırılmıştır.

Literatür taraması şu şekilde organize edilmiştir: Öncelikle verilerin sürekli olması halinde uygulanan metasezgiseller incelenmiş, ardından kategorik ve karma veri kümelerinde uygulanan metasezgiseller üzerinde literatür taraması gerçekleştirilmiştir.

3.8.1 Sürekli Veri Kümelerinin Kümelemesinde Metasezgiseller

Shelokar et al. [44], karınca kolonisi optimizasyon algoritmasını kümeleme problemine uygulamışlardır. Uzaklık ölçüsü olarak öklid uzaklığı kullanılırken, kümeleme kalitesinin ölçümünde kümeler atanan örneklerin ait oldukları kümelerin merkezlerine olan karesi alınmış öklid uzaklıklarının toplamı kullanılmıştır. Algoritma birkaç veri kümesi üzerinde test edilmiş ve GA, tavlama benzetimi (TB) ve tabu arama (TA) gibi stokastik/sezgisel algoritmalar ile bulunan çözümün kalitesi, değerlendirilen fonksiyon sayısı ve hesapsal zaman açısından karşılaştırılmıştır.

Güngör ve Ünler [83], k -ortalamlar tekniğinin yerel optimuma takılma ve başlangıç çözümüne duyarlı olma gibi iki dezavantajının bulunduğunu ve başlangıç çözümü problemlerinin açgözlü yaklaşım ile halledildiğini ancak; yerel optimuma takılma sorununun devam ettiğini belirterek, çözümü için k -harmonik ortalamlar tekniğini tabu

arama ile melezleyerek önermişlerdir. Örneklerin kümelere atanmasında bulanık bir yapı ve uygunluk değeri hesaplamasında ise harmonik ortalama formülünü kullanmışlardır. İris ve literatürde bilinen diğer veri kümelerinde algoritmayı uygulamış ve tekniğin sağlam bir teknik olduğunu ileri sürmüşlerdir.

Kao et al. [84], k -ortalamalar tekniğinin dezavantajlarını gidermek üzere; k -ortalamalar tekniği, Nelder Mead simpleks yöntemi ve PSO algoritmasından meydana gelen melez bir algoritma önermişlerdir. Modifiye algoritma k -ortalamalar ile başlamakta ve küme merkezleri değişmeyinceye kadar devam etmektedir.

Carpenter ve Grossberg [85] tarafından sinir ağları ile oluşturulan kendini düzenleyen haritalar tekniği ile kümeleme yapılmıştır. Krishna ve Murty [86], GA'nın pahalı çarpazlama operatörü ile uygunluk değerlerinin dezavantajlarını gidermek üzere genetik k -ortalamalar tekniğini önermişlerdir. K -ortalamalar tekniği bir aşamalı olarak uygulandıktan sonra çarpazlama yerine genetik k -ortalamalar kullanılmıştır. Ayrıca, kümeleme için önyargılı mutasyon operatörü tanımlanmış ve Markov zinciri teorisi kullanılarak genetik k -ortalamalar tekniğinin global optimuma yakınsadığı kanıtlanmıştır.

Ruspini [87], mevcut kümeleme metotlarının bazı dezavantajlarından kaçınmak ve orijinal veriyi daha iyi bir yapıya sokmak için bulanık küme teorisini kullanarak kümeleme gerçekleştirmiştir.

Jensi ve Jiji [88], k -ortalamalar tekniğinin iki dezavantajı olan başlangıç çözümüne bağımlı olması ve yerel optimuma takılması problemlerinin üstesinden gelmek için çiçek tozlaşma algoritması (Flower Pollination Algorithm-FPA) olarak bilinen global optimizasyon algoritmasını k -ortalama tekniği ile melezleştirmişlerdir. Önerilen bu algoritmanın sonuçları, k -ortalamalar ve çiçek tozlaşma algoritmalarının sonuçlarıyla karşılaştırılarak iki algoritmadan daha iyi sonuçlar verdiği belirtilmiştir.

Huang et al. [89], sürekli optimizasyon problemleri için geliştirilen karınca kolonisi optimizasyonu (KKO) ile PSO'yu melezleyerek kümeleme problemine uygulamışlardır. Çalışmanın motivasyonu; arama kabiliyetini geliştirmek olmuştur. Yapılan deneysel çalışma sonrasında melez algoritmanın, k -ortalamalar algoritması, PSO ve ACOR (Sürekli KKO Algoritması) ile karşılaştırılarak daha iyi sonuçlar verdiği görülmüştür.

İnkaya vd. [90], kümeleme probleminin çözüm değerlendirme, komşuluk yapısı kurma ve veri kümesini azaltma gibi olgularına yönelmiş ve ilk olarak düzeltilmiş yoğunluk ve göreceli ayırma adını verdikleri iki amaç fonksiyonu belirlemişlerdir. Her amaç fonksiyonu, kümeleme sonuçlarını komşuluklarının yerel karakteristiklerine göre değerlendirmektedir. Bu, bir ön bilgi olmaksızın kümeleme sonuçlarının belirli bir aralıkta kalitesini artırmaktadır. Sonra iki amaç fonksiyonunu kullanarak, KKO önerilmektedir. Bu algoritma, çok amaçlı bir yapıda çalışmakta ve baskın olmayan bir çok çözüm vermektedir. Komşuluk yapısı kurma ve veri kümesini azaltma gibi iki ön işlemi vardır.

Akarsu ve Karahoca [91], çalışmalarında özellik seçimi ve kümeleme için özgün melez bir model sunmuşlardır. Cancer veri kümesini bölümlmek için KKO kullanmış ve veri kümesinden gereksiz ve ilgisiz özellikleri çıkarmak için de “Sıralı Geriye Azalan Özellik Seçimi” tekniğini kullanmışlardır. Önerilen melez yaklaşımın diğer filtreleme yöntemlerinden daha iyi sonuçlar verdiğini belirtmişlerdir.

Shabanzadeh ve Yusof [92], farklı hücrelerin doğal biyocoğrafik dağılımından esinlenerek oluşturulan biyocoğrafya tabanlı olan algoritmanın (Biogeography-Based Optimization-BBO) orijinalinin ana parametrelerinde modifikasyon yaparak kümeleme probleminde uygulamışlardır. Çalışmada yazarlar, çözüm uzayını temsil edecek şekilde başlangıç çözümü önermişlerdir. Uygulama sonuçları algoritmanın iyi sonuçlar verdiğini göstermiştir.

3.8.2 Kategorik ve Karma Veri Kümelerinin Kümelemesinde Metasezgiseller

Wilson ve Martinez [49], hem kategorik hem de sürekli değişkenlerin olduğu veri kümelerinde kullanılabilecek uzaklık ölçütlerini ortaya koymuşlardır. Gao et al. [93], tam sayılı bölünebilir konkav programlama problemlerinin çözümü için dal-sınır PSO melez yaklaşımını sunmuşlardır. Optimum çözümün alt sınırı için doğrusal programlama gevşetmesi çözülerek eldeki en iyi çözüm elde edilmekte ve PSO algoritması yinelenmektedir.

Baş [94], hiyerarşik kümeleme tekniklerinden olan tek bağlantı tekniği, tam bağlantı tekniği, ortalama bağlantı tekniğini ve bölmeli kümeleme tekniklerinden k -modlar algoritmasını kullanarak kümeleme analizi yapmış ve sonuçlarını karşılaştırmıştır.

Izakian ve Ajith [95], kategorik verilerin kümelenmesinde kullanılan en yaygın algoritma olan k -modlar algoritmasının başlangıç küme merkezlerini rassal olarak seçmesi yerel optimum sorununa neden olduğu için sürü tabanlı k -modlar algoritması geliştirmişlerdir.

Cao et al. [96], k -ortalamalar algoritmasının bir uzantısı olarak k -modlar algoritmasını kategorik veri kümelerine uygulamışlardır. Çalışmalarında, basit eşleşme benzemezlik ölçüsü ve Ng benzemezlik ölçüsünü bazı açıklayıcı örnekler kullanarak analiz etmişlerdir.

Qin et al. [97], kategorik veri için hiyerarşik bölünmeli kümeleme algoritmasına dayalı yeni bir bilgi teorisi olan ortalama kazanç oranı (Mean Gain Ratio-MGR) yöntemini önermişlerdir.

3.9 Kümeleme Analizinde Karşılaşılan Sorunlar

Agarwal et al. [98], kümeleme analizindeki sorunları aşağıdaki şekilde özetlemişlerdir:

1. **Benzerlik/benzemezlik ölçüsünün belirlenmesi:** Sürekli değişkenler için Öklid, Manhattan ve Maksimum uzaklık gibi standart ölçüler kullanılmaktadır. Ancak, kategorik özelliklerde uzaklık ölçütünün tanımlanması zordur.
2. **Küme sayısı:** Küme sayısı önceden bilinmiyorsa, tanımlaması zor bir iştir. Doğru sonuçların elde edilmesi için küme sayısının dikkatli bir şekilde belirlenmesi gerekmektedir.
3. **Veri yapısı:** Gerçek hayat verisi her zaman tanımlanabilir kümeleri içermeyebilir. Sınıflandırmada da olduğu gibi; aykırı, eksik, dinamik verilerin olması kümeleme sonuçlarını olumsuz etkilemektedir. Bu gibi durumlarda uygun ön işleme araçlarının kullanılması önerilmektedir.
4. **Veritabanındaki özelliklerin tipi:** Veritabanı, her zaman sürekli veya kategorik şeklinde özelliklerin belirgin olduğu bir yapı sunmaz. Veri tabanları, nominal ve ikili(binary) değişkenler de içerebilmektedirler. Bu tip değişkenler basit bir şekilde hesaplama yapabilmek için kategorik veriye dönüştürülmelidirler.

5. **Başlangıç kümelerinin seçilmesi:** Bölünmeli tip kümelemede, başlangıç küme merkezleri genel olarak rastgele oluşturulmaktadır [96]. Başlangıç küme merkezleri algoritmanın performansı üzerinde önemli bir etkiye sahip olduğundan dikkatli seçilmelidir.

3.10 Sonuç

Bu bölümde tezin ana metotunu oluşturan kümeleme üzerinde detaylı bilgiler verilmiştir. Kümeleme tanımı, aşamaları, yöntemleri, literatürdeki benzerlik/benzemezlik ölçüleri ve uygunluk değerleri, kümeleme kapsamında karşılaşılan problemler ele alınmıştır. Özellikle, uygulamada kullanılan yöntemler üzerinde durulmuş ve sürü zekasına dayalı kümelemeye ilişkin literatüre ayrıntılı yer verilmiştir.

4. BÖLÜM

YENİ NESİL METASEZGİSEL YAKLAŞIMLAR

4.1 Giriş

Bu bölümde, tez çalışmasında kümelemeye uygulanan yeni nesil metasezgisel algoritmalar olan IMO ve WSA ile daha önce kümelemeye uygulanmış olup performanslarını kanıtlamış olan PSO ve ABC algoritmaları hakkında detaylı bilgi verilecektir. Ayrıca bu bölümde, PSO ve ABC algoritmalarının kümeleme uygulamalarına dair literatür taraması yer alacaktır.

4.2 Yeni Nesil Metasezgisel Algoritmalar

IMO ve WSA algoritmalarının ikisi de 2015 yılında yeni nesil optimizasyon algoritması olarak literatüre giren algoritmalar. Her iki algoritma da fonksiyon optimizasyonu problemlerine uygulanmış ve performansları kanıtlanmıştır. Tez çalışması kapsamında bu algoritmalar kümeleme problemine uygulanmış olduğundan, öncelikle bu yeni nesil metasezgisel algoritmaların detaylı tanımlama ve anlatımları yapılacak, daha sonra literatürde mevcut olan PSO ve ABC algoritmaları hakkında bilgi verilecektir.

Algoritmaların notasyon bilgileri Tablo 4.1’de verilmiştir.

4.2.1 İyon Hareketi Algoritması

Javidy ve Hatamlou [99] tarafından 2015 yılında optimizasyon problemlerinin çözümü için önerilen IMO, doğadaki iyonların hareketlerinden esinlenmiştir. Bu nedenle fizik kuralları tabanlı bir algoritmadır. Olabildiğince az parametre, düşük hesapsal zaman, hızlı yakınsama ve yerel optimumdan kaçışı sağlayacak şekilde tasarlanan algoritma, sürekli problemlerin çözümündeki performansı ile rekabet edebilirliğini göstermiştir.

Tablo 4.1 Notasyonlar

$Maxiter$	Maksimum iterasyon sayısı(durdurma kriteri)
$Iteration$	Mevcut iterasyon numarası
J	Problemin boyutu

IMO Parametreleri

A_i	i . anyon vektörü
C_i	i . katyon vektörü
A_{ij}	i . anyonun j . boyutunun değeri
C_{ij}	i . katyonun j . boyutunun değeri
AD_{ij}	i . anyonun en iyi katyon ile arasındaki j .boyuttaki uzaklık değeri
AD_{ij}	i . katyonun en iyi anyon ile arasındaki j .boyuttaki uzaklık değeri
AF_{ij}	i . anyon ile en iyi katyon arasındaki j .boyuttaki çekme kuvvetinin değeri
CF_{ij}	i . katyon ile en iyi anyon arasındaki j .boyuttaki çekme kuvvetinin değeri
$Abest$	En iyi anyon vektörü
$Cbest$	En iyi katyon vektörü
$Abest_j$	En iyi anyonun j . boyutunun değeri
$Cbest_j$	En iyi katyonun j . boyutunun değeri
$AbestFit$	En iyi anyonun uygunluk değeri
$AworstFit$	En kötü anyonun uygunluk değeri
$CbestFit$	En iyi katyonun uygunluk değeri
$CworstFit$	En kötü katyonun uygunluk değeri
$\emptyset_1$	$[-1,1]$ aralığında rastgele üretilen değer

WSA Parametreleri

AA	Yapay ajanların sayısı
$f(i)$	Ajan i 'nin uygunluk değeri
$f(tar)$	Hedef noktanın uygunluk değeri
τ	Kullanıcı tanımlı parametre
λ	Kullanıcı tanımlı parametre
φ	Kullanıcı tanımlı parametre
$weight$	Bir ajanın mevcut noktasının ağırlığı

$weight_i$	i . ajanın ağırlığı
x_{ij}	i . ajanın j . boyutu
tar_j	Hedef noktanın j . boyutu
gap	Bir ajanı hedef noktaya bağlayan vektör
gap_{ij}	i . ajanı hedef noktaya bağlan vektörün j . boyutu
$direct$	Bir ajanın hareket yönünü gösteren vektör
$direct_{ij}$	i . hareket yönünü gösteren vektörün j . boyutu
$sign()$	İşaret fonksiyonu
sl	Adım uzunluğu
UL_j	j . boyutun üst limiti
LL_j	j . boyutun alt limiti

PSO Parametreleri

v_{ij}	i . parçacığın j . boyutunun hız değeri
$pbest_{ij}$	i . parçacığın en iyi pozisyonunun j . boyuttaki değeri
$gbest_j$	Tüm parçacıklar içinde en iyi uygunluk değerine sahip olan parçacığın j . boyutunun değeri
c_1	Yerel öğrenme faktörü
c_2	Global öğrenme faktörü
x_{ij}	i . parçacığın j . boyutunun değeri

ABC Parametreleri

p_i	i . yiyecek kaynağının olasılık değeri
fit_i	i . yiyecek kaynağının nektar miktarı (uygunluk değeri)
SN	Yiyecek kaynağı sayısı (başlangıç popülasyon sayısı)
φ_{ij}	i . yiyecek kaynağının komşusu üretilirken j . (rastgele üretilen) boyut için $[-1,1]$ aralığında üretilen değer
x_{ij}	i . yiyecek kaynağının j . (rastgele üretilen) boyutunun değeri
x_{kj}	k . (rastgele üretilen) yiyecek kaynağının j . (rastgele üretilen) boyutunun değeri
v_{ij}	i . yiyecek kaynağı için üretilen komşu i nin, j . (rastgele üretilen) boyutunun değeri

Bir Yunan terimi olan iyon kelimesi, İngiliz doktor Michael Faraday tarafından 1834 yılında ortaya atılmıştır. Yüklü parçacıklara iyon; negatif yüklü iyonlara anyon, pozitif yüklü iyonlara katyon denmektedir. Doğadaki iyonlardan algoritmanın esinlendiği durum, aynı yüklü iyonların birbirini itme ve farklı yüklü iyonların birbirini çekme eğiliminde olmasıdır [99].

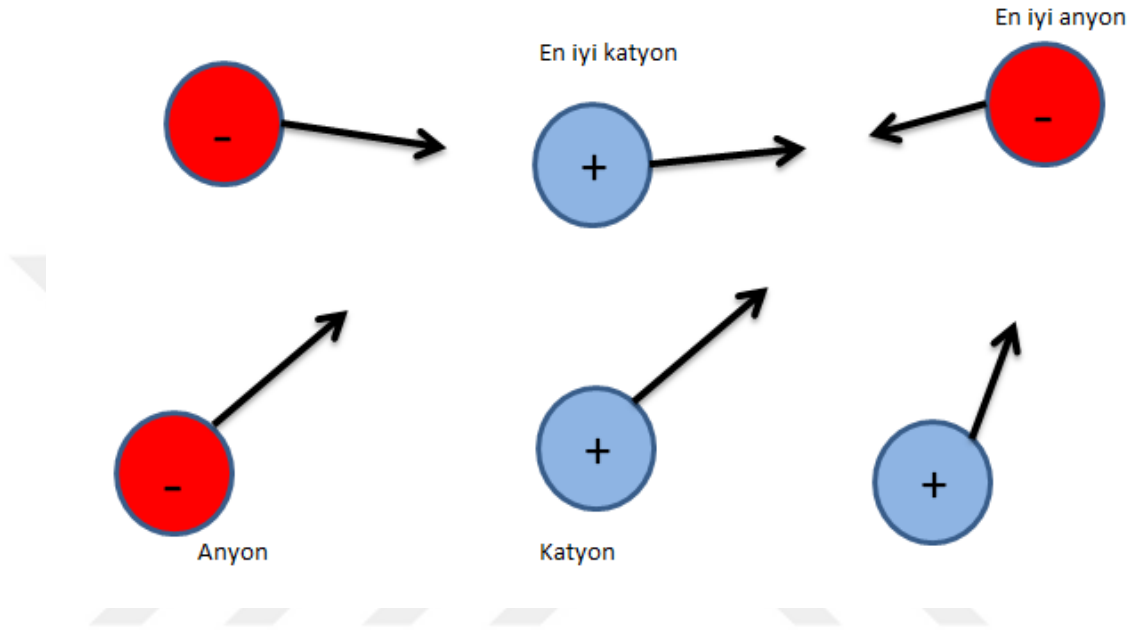
IMO algoritmasında aday çözümler, anyon ve katyonlar tarafından temsil edilirken, aralarındaki itme ve çekme kuvvetleri iyonları arama uzayında hareket ettirmektedirler. Algoritmanın temel mantığı, iyonların zıt yüklü en iyi iyonla doğru hareket etmeleridir. Anyonlar(A) en iyi katyona, katyonlar(C) en iyi anyona doğru çekilirlerken hareket miktarları, aralarındaki çekim kuvvetlerine bağlıdır. En iyi anyon ve katyona olan bu hareketler ile iterasyon yenilenmesi boyunca amaç fonksiyon değerinin geliştirilmesi garantilenmektedir. IMO algoritması; çeşitlenme (diversification) ve yoğunlaşma (intensification) olmak üzere temel iki yapıdan oluşmaktadır [99]. IMO algoritmasının adımları Şekil 4.1’de verilmiştir.

-
1. Popülasyon büyüklüğünü, maksimum iterasyon sayısını, küme sayısını (kümeleme problemi için) belirle, anyon ve katyonları popülasyon büyüklüğünün yarısı olarak rasgtele oluştur.
-
2. **Likit Aşaması**
Iteration=1
Iteration<=Maxiter olduğu sürece gerçeke {
 - Anyon ve katyonların uygunluk değerlerini hesapla.
 - En iyi ve en kötü anyon ve katyonu belirle.
 - Anyonların en iyi katyon tarafından ve katyonların en iyi anyon tarafından çekilme kuvvetlerini hesapla.
 - Anyon ve katyonların pozisyonlarını güncelle.
-
3. **Kristal Aşama**
 En kötü iyonların uygunluk değerlerinin yarısı, en iyi iyonların uygunluk değerlerinden küçük veya eşitse anyon ve katyonları, en iyi iyonlar etrafına dağıt. Küçük bir olasılıkla (<0.05) anyon ve katyonları yeniden oluştur.
-
4. Aksi takdirde, durdurma kriterinin sağlanıp sağlanmadığını kontrol et ve sağlanıyorsa algoritmayı sonlandır, sağlanmıyorsa Adım 2’den devam et.
Iteration=Iteration+1 }
 5. En iyi çözümü rapor et.
-

Şekil 4.1 IMO algoritmasının adımları [99]

4.2.1.1 Likit Aşaması

İyonların hareketlerinin serbest ve aynı yüklü iyonlar arasındaki itme kuvvetinin, zıt yüklü iyonlar arasındaki çekim kuvvetinden az olduğu sıvı aşamasıdır. Likit aşaması Şekil 4.2’de verilmiştir.



Şekil 4.2 Likit aşamada anyon ve katyonların en iyi iyonlara hareketi [99]

Algoritma, bu aşamada arama uzayındaki keşfi daraltmamak için itme kuvvetini ihmal etmektedir. İyonlar arasındaki çekim kuvvetini etkileyen tek faktörün iyonlar arasındaki uzaklık olduğu varsayılmakta ve bu kuvvet Eşitlik (4.1) ve Eşitlik (4.2) ile hesaplanmaktadır.

$$AF_{ij} = \frac{1}{1 + e^{-0.1/AD_{ij}}} \quad (4.1)$$

$$CF_{ij} = \frac{1}{1 + e^{-0.1/CD_{ij}}} \quad (4.2)$$

AF_{ij} , i . anyonun j . boyutta en iyi katyon tarafından çekilme kuvvetini ve CF_{ij} ise, i . katyonun j . boyutta en iyi anyon tarafından çekilme kuvvetini ifade etmektedir. Kuvvetler uzaklık ile ters orantılıdır ve 0,5 ile 1 arasında değişen değerler almaktadırlar.

Burada AD_{ij} , i . anyon ile en iyi katyonun j . boyutları arasındaki ve CD_{ij} , ise i . katyon ile en iyi anyonun j . boyutları arasındaki uzaklığı ifade etmekte ve Eşitlik (4.3) ve Eşitlik (4.4) ile hesaplanmaktadır:

$$AD_{ij} = |A_{ij} - Cbest_j| \quad (4.3)$$

$$CD_{ij} = |C_{ij} - Abest_j| \quad (4.4)$$

$Cbest_j$ ve $Abest_j$ sırasıyla, en iyi katyon ve anyonun j . boyutlarını temsil etmektedir. Anyon ve katyonların güncellenmesi kuvvetlere bağlı olarak Eşitlik (4.5) ve Eşitlik (4.6) ile hesaplanmaktadır [99].

$$A_{ij} = A_{ij} + AF_{ij} \times (Cbest_j - A_{ij}) \quad (4.5)$$

$$C_{ij} = C_{ij} + CF_{ij} \times (Abest_j - C_{ij}) \quad (4.6)$$

Likit aşamada herhangi bir rastgelelik yoktur. İterasyonların sayısı arttıkça, daha çok iyon zıt yüklü iyonlara yakınsayacak ve çeşitlenme/keşif uzayı kademeli olarak azalacaktır. Doğada iyonların sıvı halden katı hale kristalleşmesi gibi, algoritmada da iyonların tek bir çözüme yakınsaması ile yerel optimum riski ortaya çıkacaktır [99].

4.2.1.2 Kristal Aşaması

Arama uzayının şekli bilinmediğinden, likit aşamasında yakınsanan çözüm yerel optimum olabilir. Doğada iyonların çekim kuvvetlerinin maksimum olduğu kristal aşamada, üretilen bir dış itme kuvvetinin kristal parçada kırık oluşturması durumunun algoritma tarafından ifade edilişi Şekil 4.3'te verilmiştir.

```

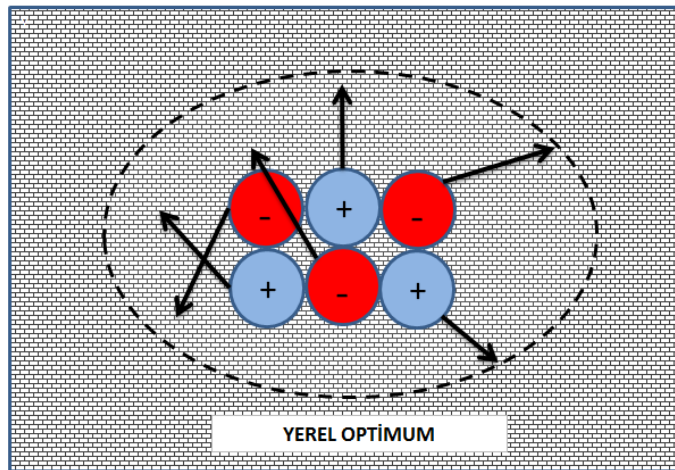
Eğer ( $C_{bestFit} \geq C_{worstFit}/2$  ve  $A_{bestFit} \geq A_{worstFit}/2$ ) {
  Eğer  $rand() > 0.5$ 
     $A_i = A_i + \phi_1 \times (C_{best} - 1)$ 
  Aksi halde
     $A_i = A_i + \phi_1 \times (C_{best})$ 
  Eğer  $rand() > 0.5$ 
     $C_i = C_i + \phi_2 \times (A_{best} - 1)$ 
  Aksi halde
     $C_i = C_i + \phi_2 \times (A_{best})$ 
  Eğer  $rand() < 0.05$ 
     $A_i$  ve  $C_i'$ yi yeniden oluştur
}

```

Şekil 4.3 IMO algoritması kristal aşama kodu [99]

$rand()$, $[0,1]$ aralığında rastgele sayı üretmektedir. Şekil 4.3'deki algoritmaya göre, kristal aşamada her durumda anyon ve katyonlar sırasıyla en iyi katyon ve anyonların etrafında dağılarak yerel optimumdan kaçmaktadırlar. Anyon ve katyonlar çok küçük bir olasılıkla (<0.05), yeniden üretilmektedirler [99].

Kristal aşamada iyonların dağıtımı görsel olarak Şekil 4.4'te gösterilmektedir.



Şekil 4.4 Kristal aşamada iyonların yerel optimum etrafında olması [99]

IMO algoritmasının tezde kullanılan diğer algoritmalarından en önemli farkları;

- Ayarlanması gereken parametresi, maksimum iterasyon ve başlangıç çözüm sayısı (popülasyon büyüklüğü) dışında yoktur. Oldukça az parametre ayarı gerektirmesi algoritmanın üstünlüğü olarak sayılabilir.
- Belirlenen başlangıç popülasyon sayısına göre başlangıç çözümlerini oluşturur. Ancak bunların yarısını anyon yarısını ise katyon olarak belirler. Yani algoritma diğerlerinden farklı olarak farklı iki pozisyon bilgisini eşzamanlı olarak kullanmaktadır.
- Tezde kullanılan diğer üç algoritma sürü zekası tabanlı olup doğadaki sürü halindeki bireylerin davranışlarından esinlenmiştir. IMO ise doğadaki iyonların hareketlerinden esinlenerek ortaya konmuş olan fizik temelli bir algoritmadır.

4.2.2 Ağırlıklı Süperpozisyon Çekim Algoritması

Baykasoğlu ve Akpınar [100,101] tarafından 2015 yılında önerilen WSA algoritması, ajanların çekim hareketleri ile kombine edilmiş süperpozisyonlarına dayalı sürü tabanlı, geliştirici, yeni bir optimizasyon algoritmasıdır. Algoritma, belirli sayıda çözümle başlar ve bazı özel mekanizmalar kullanarak bu çözümleri geliştirmeye çalışır. Arama yönünü ise bir komşu üretme mekanizması aracılığıyla ajanların çekim hareketleri ve süperpozisyonları bütünlüğünde belirler. Algoritmanın adımları Şekil 4.5'te verilmiştir.

WSA, parametre değerlerinin belirlenmesi ve başlangıç çözümünün rastgele oluşturulmasıyla başlar. WSA'nın diğer algoritmalara göre parametre sayısı daha fazladır. Algoritmada kullanıcı tarafından belirlenmesi gereken 6 parametre vardır. Bunlar; maksimum iterasyon sayısı, başlangıç çözümlerinin sayısı, adım uzunluğu ve diğer kullanıcı tanımlı parametreler olan τ , φ , λ parametreleridir. τ , hedef nokta belirleme prosedüründe her çözümün sıralanması ile aldığı sıra değerinin üssü olarak ağırlık hesaplamada kullanılan bir parametredir. φ , pozisyon güncelleme prosedüründe adım uzunluğu hesabında kullanılan ve λ ise aynı prosedürde rastgele üretilen sayı ile karşılaştırılan bir parametredir.

Algoritmada yapay ajanların hareketleri, mevcut keşfedilen noktaların süperpozisyonuna bağlı olarak gerçekleşmektedir. Süperpozisyon kavramı, fizikte ve

diğer ilişkili bilim dallarında birden fazla kaynağın tüm etkilerini belirlemek için kullanılmaktadır.

-
1. Parametrelerin değerlerini belirle.
 2. Başlangıç popülasyonunu rastgele üret.
 3. Başlangıç popülasyonunda yer alan çözümlerin uygunluk değerlerini hesapla ve global en iyi çözümü belirle.
 4. *Iteration=1*
Iteration <= Maxiter olduğu sürece gerçekleştir {
 - Uygunluk değerlerine göre çözümleri sırala.
 - Sıra numaralarına göre her çözümün ağırlığını hesapla.
 - Çözümlerin hareket edeceği hedef noktayı hesapla.
 - Hedef noktanın uygunluk değerini hesapla.
 - Her çözümün, hedef noktanın pozisyon ve uygunluk değerine göre yönünü belirle.
 - Belirlenen yönde her çözümü hareket ettir.
 - Yeni çözümlerin uygunluk değerini hesapla ve global en iyi çözüm ve global en iyi uygunluk değerini güncelle.*Iteration=Iteration+1*
 5. En iyi çözümü rapor et.
-

Şekil 4.5 WSA algoritmasının adımları [100]

Süperpozisyon mekanizması optimizasyon problemine uygulandığında, yapay ajanların tümünün hedef nokta üzerindeki etkisi olarak karşımıza çıkmaktadır. WSA, süperpozisyon prensibine dayalı tüm çözümlerin ağırlıklarından oluşan bir hedef nokta belirlemektedir. Hedef nokta belirleme prosedürü Şekil 4.6’da verilmiştir.

weight = AA boyutunda sıfır vektörü oluştur
tar = J boyutunda sıfır vektörü oluştur
i=1’den AA’ya gerçekleştir {
 $weight(1, i) = i^{(-1)*\tau}$
j=1’den J’ye gerçekleştir
 $tar(1, j) = tar(1, j) + x(i, j) * weight(1, i)$

Şekil 4.6 WSA hedef nokta belirleme prosedürü [100]

WSA, hedef nokta belirlemeden önce her yapay ajana uygunluk değerlerine göre sıralandıklarında elde ettikleri sıra bilgisine göre ağırlık atar. Yani önce yapay ajanları uygunluk değerlerine göre sıralar ve her yapay ajanın elde ettiği sıra bilgisine bağlı olarak hedef noktayı belirler. Hedef nokta üzerinde en yüksek ağırlığı en iyi çözüm alırken, en kötü çözüm en az ağırlığı almaktadır. Ağırlık belirlemede; τ , önemli bir parametredir. τ , büyük bir değer alındığında yapay ajanların etkisi ortaya çıkmayacak, küçük bir değer alındığında ise tüm ajanlar aynı etkiye sahip olacaklardır [100].

Hedef nokta algoritmadaki Eşitlik (4.7) ile belirlenmektedir.

$$tar(1, j) = tar(1, j) + x(i, j) * weight(1, i) \quad (4.7)$$

Hedef nokta belirlenir belirlenmez her yapay ajan bu noktaya doğru hareket edip etmeyeceğini planlayacaktır. WSA, tüm çözümleri bu hedef noktaya yaklaştırarak arama uzayında komşuluk üretmeye çalışır (Şekil 4.7).

gap = (AA, J) boyutunda sıfır matrisi oluştur

direct = (AA, J) boyutunda sıfır matrisi oluştur

i=1'den AA'ya gerçeğe {

Eğer $f(i) \geq f(tar)$ ise {

j=1'den J'gerçeğe

gap(i, j) = tar(1, j) - x(i, j)

d=1'den J'ye gerçeğe

direct(i, d) = sign(gap(i, d))

Aksi halde Eğer $f(i) < f(tar)$ ise {

Eğer $rand() < e^{(f(i)-f(tar))}$

j=1'den J'ye gerçeğe

gap(i, j) = tar(1, j) - x(i, j)

d=1'den J'ye gerçeğe

direct(i, d) = sign(gap(i, d))

Aksi halde

d=1'den J'ye gerçeğe

*direct(i, d) = sign(-1 + (1 + 1) * rand()) } }*

}

Şekil 4.7 WSA arama yönü belirleme prosedürü [100]

Yapay ajanların hedef noktaya doğru hareket edip etmeyecekleri, uygunluk değerlerinin hedef noktanın uygunluk değeri karşısındaki durumu tarafından belirlenir. Eğer yapay ajanın uygunluk değeri hedef noktanın uygunluk değerinden daha kötüyse hedef noktaya hareket edecektir. Aksi durumda ise yapay ajan ya hedef noktasına hareket edecek veya rastgele seçilen bir yöne hareket edecektir.

Hareket yönünün tespitinde algoritmadaki Eşitlik (4.8) ve Eşitlik (4.9) kullanılmaktadır.

$$gap(i, j) = tar(1, j) - x(i, j) \quad (4.8)$$

$$direct(i, d) = sign(gap(i, d)) \quad (4.9)$$

Eşitlik (4.8) ile yapay ajan hedef noktaya hareket edecekse, hedef noktanın pozisyonu ile yapay ajanın pozisyonu arasındaki uzaklık bulunur. Eşitlik (4.9) ise yapay ajanın hedef noktaya ilerleme yönünü tespit eder.

Eğer yapay ajanın uygunluk değeri hedef noktanınkinden daha iyiye, Eşitlik (4.10)'da olduğu gibi hesaplanan olasılık (sağ taraf fonksiyonu) ile rastgele oluşturulan olasılık ($rand()$) arasında değerlendirme yapar.

$$Eğer rand() < e^{(f(i)-f(tar))} \quad (4.10)$$

Bu değerlendirmenin sonucu “Evet” ise, yapay ajanın Eşitlik (4.8) ve Eşitlik (4.9) ile hedef noktaya doğru yön belirlemesi yapılır. “Hayır” ise Eşitlik (4.11) ile yapay ajan için rastgele bir yön belirlenir.

$$direct(i, d) = sign(-1 + (1 + 1) * rand()) \quad (4.11)$$

Arama algoritmalarının etkinliğini belirleyen en önemli etken komşu üretme mekanizmasıdır. WSA'da komşuluk üretme, adım uzunluğu ile arama yönünün belirlenmesini içermektedir. Yapay ajanlar, bir pozisyonun diğerine mevcut pozisyonlarının komşuluk alanını keşfederek hareket ederler. WSA algoritmasında yapay ajanların yeni pozisyonlarının güncellenmesini, belirlenen arama yönlerine göre nasıl yaptıkları Şekil 4.8'de verilmektedir.

Eğer $rand() \leq \lambda$

$$sl = sl - e^{-iteration/(iteration+1)} * \varphi * sl$$

Aksi halde

$$sl = sl + e^{-iteration/(iteration+1)} * \varphi * sl$$

$i=1$ 'den AA 'ya gerçekte {

$j=1$ 'den J 'ye gerçekte {

$$x(i,j) = x(i,j) + sl * direct(i,j) * abs(x(i,j))$$

Eğer $x(i,j) < LL(1,j)$

$$x(i,j) = LL(1,j)$$

Aksi halde eğer $x(i,j) > UL(1,j)$

$$x(i,j) = UL(1,j) \} \}$$

Şekil 4.8 WSA pozisyon güncelleme prosedürü [100]

Bir yapay ajanın yeni pozisyonunun güncellenmesi işlemi Eşitlik (4.12) ile yapılmaktadır.

$$x(i,j) = x(i,j) + sl * direct(i,j) * abs(x(i,j)) \quad (4.12)$$

Adım uzunluğu ilk değeri, Tablo 4.1'de de görüleceği gibi WSA'nın belirlenmesi gereken parametrelerindendir.

Baykasoğlu ve Akpınar [100], algoritma için kullandıkları parametrelerin ayarlamasını Taguchi metot kullanarak yapmışlardır. Taguchi yaklaşımı, 1986 yılında Taguchi [102] tarafından ortaya konan bir deney tasarımı yöntemidir. Klasik deney tasarımı yöntemlerinin kullanımları, endüstriyel şartlar altında sağlıklı ve verimli olmamaktadır. Deneyi yapılan sistemin faktör sayısı arttıkça yapılması gereken deney sayısı da hızlı bir şekilde artış göstermektedir. Taguchi getirdiği yaklaşımla, deneylere başlamadan yapılan ayrıntılı analiz ile gerekli deney sayısını önemli şekilde düşürmüştür. Ancak bu yapılırken, faktörler arasındaki etkileşimlerin belirli bir kısmı göz ardı edilmiştir.

Taguchi metodunun amacı, gürültü faktörlerine karşı ürün veya süreci daha sağlam yapacak şekilde kontrol parametrelerinin en iyi kombinasyonunu seçmektir. Metot, az sayıda deney ile çok sayıda değişkeni çalışmak için deneylerin tasarımında ortogonal

(dikey) dizileri kullanmaktadır. Bu diziler, deney sayısını önemli ölçüde azaltmaktadır. Ayrıca, çıkarılan sonuçlar kontrol faktörleri ve onların kurulum değerleri tarafından taranan tüm deney bölgesi üzerinde geçerlidir [103].

WSA çalışmasında yazarlar, adım uzunluğu ve τ , φ , λ kullanıcı tanımlı parametreleri deney tasarımına tâbi tutmuşlardır. Maksimum iterasyon sayısı ve başlangıç çözümlerinin sayısı, diğer algoritmalarla karşılaştırmada esas olduklarından kullanıcı tarafından karar verilmiştir.

Yazarlar, WSA algoritmasını hem kısıtlı optimizasyon hem de kısıtsız optimizasyon problemlerinin her ikisi için de uygulamışlardır.

4.3 Parçacık Sürü Optimizasyon Algoritması

Kennedy ve Eberhart [72] tarafından 1995 yılında geliştirilmiş sürü zekasına dayalı bir metasezgiseldir. Kuşların sürü halinde toplanması, çevrelerine uyum sağlamaları, yiyecek kaynağı bulabilmeleri ve avcılardan kaçabilme gibi bilgi paylaşımı esasına dayanan sosyolojik davranışları taklit edildiği için sosyolojik esinlemeli bir algoritmadır.

Tipik olarak bir hayvan sürüsü, rastgele yiyecek kaynağı bulacak liderlere sahip değillerdir. Onlar, bir yiyecek kaynağına en yakın olan grup üyelerinden birini takip ederler. Sürü zaten daha iyi bir konuma sahip üyeler arasındaki iletişim sayesinde eşzamanlı olarak kendilerini en iyi duruma getirirler. Daha iyi konuma sahip olan birey, sürüsünü bilgilendirecek ve diğerleri de eşzamanlı olarak bilgi verilen bu konuma hareket edeceklerdir. Bu durum, en iyi pozisyon sağlanana ya da bir yiyecek kaynağı keşfedilene kadar devam etmektedir. PSO, optimum sonucu bulmak veya optimum sonuca yaklaşmak için hayvanların bu davranışlarını taklit etmektedir [104].

PSO'nun yaygın uygulamalarında, parçacıklar arama uzayında parçacığın en iyi değeri ve tüm parçacıklar arasında en iyi değere sahip parçacığın değerinin kombinasyonu içinde hareket ederler. PSO'da komşuluk, birbirleri ile iletişim kurabilen parçacıkların bir alt kümesi olarak tanımlanır. İlk PSO modelinde, iletişim kurabilmek için yakın olunması gereken mesafeyi belirlemekte parçacıklar arasındaki gerçek mesafeyi veren öklid komşuluğu kullanılmıştır. Bu, sadece hemen yakınındaki diğer kuşlarla iletişim

kurabilen kuşlarda olduğu gibi biyolojik modellere benzer olarak, kuş sürülerini taklit etmek demektir. Araştırmalar biyolojik modellerden matematiksel modellere doğru kaydığında ve daha az hesapsal zaman amaçlandığında öklid komşuluğu terk edilmiştir. Parçacığın sadece yakınındaki değil, sürüdeki diğer tüm parçacıklardan bilgi alabildiği ve iletişim kurabildiği “gbest model” adı verilen uygulamalar kullanıma girmiştir [105]. PSO algoritmasının pek çok versiyonu vardır. Uygulamada standart olarak kullanılan adımlar Şekil 4.9’da verilmiştir:

-
1. Popülasyon büyüklüğünü, maksimum iterasyon sayısını, küme sayısını (kümeleme problemi için), başlangıç hız değerini belirle.
 2. Başlangıç popülasyonunu rasgele üret.
 3. $Iteration=1$
 $Iteration \leq Maxiter$ olduğu sürece gerçekleştir {
 4. Her parçacık için,
 - a. Uygunluk değerini hesapla.
 - b. Hesaplanan en iyi uygunluk değeri, mevcut değerden iyiye en iyi değeri yenisiyle güncelle ($pbest$).
 5. Global en iyi değer olarak, tüm parçacıklar içinden en iyi uygunluk değerine sahip olanı seç ($gbest$).
 6. Her parçacık için,
 - a. Hız değerlerini güncelle.
 - b. Parçacık pozisyonlarını güncelle. $Iteration=Iteration+1$
 - }
 7. Durdurma kriteri sağlandığında en iyi çözümü rapor et.
-

Şekil 4.9 PSO algoritmasının adımları [105]

Başlangıç popülasyonu rastgele üretildikten sonra, durdurma kriteri sağlanıncaya kadar algoritma çalışmaya devam eder. Bu çalışma sırasında, ilk olarak popülasyondaki çözümlerin uygunluk değerleri hesaplanır ve parçacığın en iyi pozisyon değeri güncellenir. Tüm parçacıkların uygunluk değerleri içinden en iyi değer, global en iyi olarak güncellenir. Bundan sonraki aşama algoritmanın komşuluk üretmesi ile ilgilidir. Hız vektörü Eşitlik (4.13)’e göre ve güncellenen hızlara göre parçacıkların pozisyonları ise Eşitlik (4.14)’e göre güncellenir [106].

$$v_{ij} = v_{ij} + c_1(pbest_{ij} - x_{ij}) + c_2(gbest_j - x_{ij}) \quad (4.13)$$

$$x_{ij} = x_{ij} + v_{ij} \quad (4.14)$$

PSO, kolayca kodlanabilen, yerel optimuma takılmadan çok hızlı bir şekilde global optimuma yakınsayabilen hızlı ve etkili bir optimizasyon algoritması olmasına rağmen bazı zayıf noktaları da vardır. Başlangıç çözümüne bağlı olması ve bazen tutarlılık için farklı parametrelerin daha etkin kombinasyonlarına ihtiyaç duyması bunlardandır [107].

PSO algoritması birçok probleme başarılı bir şekilde uygulanmıştır. Fonksiyon optimizasyonu [108], bulanık sistem kontrolü [109], kümeleme [45], sınıflandırma [110], YSA eğitimi [111] bunlardan sadece bazılarıdır.

4.4 Yapay Arı Kolonisi Algoritması

ABC algoritması, Karaboğa [112] tarafından 2005 yılında bal arılarının arama davranışlarından esinlenerek sürekli problemlerin optimizasyonu için geliştirilen metasezgisel bir algoritmadır. Sürü zekası tabanlı bu algoritma, sürü halinde doğada hareket eden bal arılarının besin bulurken ortaya koydukları davranışları taklit etmektedir. Arıların sosyal bir düzen içinde yaşamaları, bu düzenin kurallarını tüm arıların içgüdüsel olarak bilmesi ve arılar tarafından görevlerin aksatılmadan giderilmesi araştırmacıların dikkatini çekmiş ve bu davranışları modelleyerek problemlere uygulama yoluna gitmişlerdir.

Parametre sayısı az, uygulanması kolay, diğer metasezgisellerle kolayca entegre edilebilen bir algoritma olduğu için hala dünyanın bir çok yerindeki araştırmacılar tarafından cazibesi devam etmektedir [113].

ABC algoritmasında, bir yiyecek kaynağı optimizasyon probleminin mümkün bir çözümünü temsil etmektedir. Algoritmanın adımlarını gösteren Şekil 4.10 aşağıda yer almaktadır:

-
1. Popülasyon büyüklüğünü, maksimum iterasyon sayısını, küme sayısını (kümeleme problemi için), limit değerini belirle.
 2. Başlangıç popülasyonunu rastgele üret.
 3. Popülasyonun uygunluk değerlerini hesapla.
 4. $Iteration=1$
 $Iteration \leq Maxiter$ olduğu sürece gerçekleştir {
 5. Her işçi arı için,
 - a. Yiyecek kaynaklarının komşu çözümünü üret.
 - b. Mevcut çözümler ile komşu çözümler arasında açgözlü yaklaşım uygula.
 - c. Uygunluk değerlerine bağlı olasılık hesapla.
 6. Olasılık değerlerini $[0,1]$ aralığında normalize et.
 7. Her gözcü arı için;
 - a. Yiyecek kaynaklarının komşu çözümünü üret.
 - b. Mevcut çözümler ile komşu çözümler arasında açgözlü yaklaşım uygula.
 8. Limit değerine göre tüketilen kaynağı belirle. Tüketilen kaynak varsa kaşif arı için rastgele bir çözüm üret.
 9. Şimdiye kadarki en iyi çözümü hafızada tut.
 $Iteration=Iteration+1$
 }
 10. Durdurma kriteri sağlandığında en iyi çözümü rapor et.
-

Şekil 4.10 ABC algoritmasının adımları [114]

Bir aday çözümün kalitesi, ilgili yiyecek kaynağının nektar miktarını ifade etmektedir. Arı kolonisinde üç arı tipinden biri olan işçi arılar ile bir diğeri olan gözcü arılar popülasyonda eşit sayıda bulunurlar. Yani başlangıç çözüm sayısı işçi ya da gözcü arıların sayısı miktarınca rastgele üretilir. Daha sonra algoritma içinde işçi, gözcü ve kâşif arıların tümünün aramasında algoritma durdurma kriteri sağlanıncaya kadar devam eder. Bir işçi arı, yerel bilgiye bağlı olarak hafızasındaki pozisyonda bir modifikasyon yapar ve yeni kaynağın (komşu çözümün) nektar miktarını test eder. Eğer yeni kaynağın kalitesi öncekinden daha iyiye işçi arı yenisini hafızasına alarak eskisini unuttur. Aksi halde eskisini hafızasında tutmaya devam eder. Tüm işçi arıların arama süreci tamamlandıktan sonra yiyecek kaynaklarının nektar miktarını ve bu kaynakların pozisyon bilgilerini gözcü arılarla paylaşırlar. Bir gözcü arı tüm işçi arılardan aldığı nektar bilgilerini değerlendirir ve nektar miktarının olasılığına bağlı olarak bir yiyecek

kaynağı seçer. İşçi arılarda olduğu gibi hafızasındaki pozisyonda modifikasyon yaparak komşu üretir ve komşu yiyecek kaynağının nektar miktarı ile mevcut miktarın karşılaştırmasını yaparak en iyisini seçer [115].

Gerek işçi gerekse de gözcü arıların yiyecek kaynakları için komşu üretme mekanizmaları Eşitlik (4.15)'de verilmiştir [116].

$$v_{ij} = x_{ij} + \varphi_{ij}(x_{ij} - x_{kj}) \quad (4.15)$$

Burada, her işçi/gözcü arının yiyecek kaynağı için o işçi/gözcü arının yiyecek kaynağının j . boyutu ve rasgele belirlenmiş k . komşusunun j . boyutuna bağlı olarak bir komşu üretme mekanizması çalıştırılmaktadır. Yani, komşu çözümler mevcut çözümlerin vektörel olarak değişmesinden değil bir boyutunda meydana gelen değişimlerle üretilmektedir.

İşçi arıların komşuluk arama işlemleri tamamlandıktan sonra son yiyecek kaynağı pozisyonları için bu pozisyonlar arasından bir gözcü arının seçim yapmasına yarayacak olasılık değerleri Eşitlik (4.16) ile hesaplanmaktadır [116].

$$p_i = \frac{fit_i}{\sum_{n=1}^{SN} fit_n} \quad (4.16)$$

Gözcü arılar, hesaplanan bu olasılık değerleri içerisinde en yüksek olasılığa sahip yiyecek kaynakları etrafında komşuluk araması yapmaktadırlar. Yani yiyecek açısından daha muhtemel yerlerde algoritma daha derin bir araştırma yapmaktadır.

Gözcü arıların yaptığı detaylı araştırma sırasında parametre olarak kullanıcı tarafından belirlenen *limit* değeri sayısınca kendisinden daha iyi bir komşu çözüm üretememiş yiyecek kaynakları yerine, kaşif arı yeni bir yiyecek kaynağı üretmektedir. Algoritma, bu mekanizmayla limitini dolduran çözümlerde fazla zaman kaybetmemekte ve daha iyi çözümler bulabilme şansını denemektedir. ABC algoritmasının keşfetme üzerine (gerek işçi gerek gözcü arılarda yaptığı ayrıntılı komşu arama) odaklanması yakınsama hızını oldukça yavaşlatmaktadır [117]. ABC algoritması da PSO algoritması gibi birçok probleme uygulanmış ve başarılı sonuçlar alınmıştır.

4.5 PSO ve ABC Algoritmalarının Kümeleme Uygulamaları

Zhang et al. [43], optimum kümelemeyi gerçekleştirmek için ABC algoritmasını kümeleme problemine uygulamışlardır. Örneklerin kümelere atanmasında öklid uzaklığını, oluşan kümelemenin kalitesini hesaplarken de örneklerin ait oldukları kümelerin merkezlerine karesi alınmış öklid uzaklıklarının toplamını kullanmışlardır. Her aday çözümün arama uzayında uygun olmayan çözümlerin üstesinden gelerek arama yapması için Deb kuralı uygulanmıştır. Literatürdeki bilinen veri kümeleri üzerinde algoritma çalıştırılmış ve GA, TB, TA, KKO ve K-NM-PSO algoritmaları ile performansı karşılaştırılmıştır.

Kuo et al. [118], arı koloni optimizasyonu ile otomatik Kernel kümeleme algoritması geliştirmişlerdir. Bu algoritma, uygun küme sayısını belirlemekte ve veri noktalarını doğru kümelere atamaktadır. Bunu kümeleme kabiliyetini arttıran Kernel fonksiyonu ile yapmaktadır. Algoritma karşılaştırma amaçlı veri kümeleri üzerinde denenmiş ve geçerliliği ortaya konmuştur. Sonuçları diğer otomatik kümeleme yöntemleriyle karşılaştırılmıştır.

Karaboğa ve Öztürk [54], ABC algoritmasının kümelemedeki performansını veri kümeleri üzerinde uygulayarak, PSO ve literatürde bulunan dokuz kümeleme tekniği ile karşılaştırmıştır. Örneklerin kümelere atanmasında öklid uzaklığını ve amaç fonksiyon değeri olarak da kümelere atanan örneklerin küme merkezlerine olan karesi alınmış öklid uzaklığını kullanmışlardır. Sınıflandırma hata oranlarına göre yapılan karşılaştırmalarda ABC algoritmasının çok değişkenli kümelemede etkin bir algoritma olduğu sonucuna varmışlardır.

Zou et al. [119], işbirlikçi ABC olarak adlandırdıkları bir algoritma önermişlerdir. Algoritma, elde edilen en iyi boyut bilgisinin kaydedilmesiyle iteratif olarak devam etmektedir. ABC'nin kompleks optimizasyon problemlerinin çözümünde kullanılmak üzere önerilen algoritma, öncelikle 8 kıyaslama fonksiyonu ile karşılaştırılarak performansı kanıtlanmış, daha sonra kümeleme problemine uygulanmıştır. Uzaklık ölçütü olarak öklid uzaklığı, uygunluk değeri olarak ise örneklerin küme merkezlerine olan karesi alınmış öklid uzaklıkları toplamı kullanılmıştır.

Öztürk vd. [56], ABC algoritmasını otomatik kümeleme problemine uygulamışlardır. Öncelikle ABC'nin küresel arama kabiliyetini geliştirmek için PSO algoritmasının komşu aramasını taklit etmişler ve algoritmanın vektörel arama yapmasını sağlamışlardır. Kümeleme geçerlilik indeksi olarak Turi indeks kullanmışlar ve önerdikleri yöntemi en çok bilinen veri ve görüntü kümelerine uygulamışlardır.

Liu et al. [120], kümeleme ve sınıflandırma için eşzamanlı öğrenmeye dayalı PSO önermişlerdir. Geliştirilen algoritma; eğitim örneklerini bölmekte, uygun küme sayısının bulunması için otomatik kümeleme algoritması çalıştırmakta ve sınıflandırma mekanizmasından kümeleme merkezleri elde etmektedir. Daha faydalı yerel bilgi ve daha iyi optimizasyon sonuçları elde etmek için PSO'daki parçacıklara güncelleme stratejisi tanıtılmaktadır. Uygulama sonuçları, algoritmanın sadece hesapsal zamanı kısaltmadığını ayrıca, veri kümelerinde daha iyi bir doğruluğa ulaştığını göstermektedir.

Cura [45], küme sayılarının bilinmesi ve bilinmemesi durumları için kümelemede PSO algoritmasını kullanmıştır. Örneklerin kümelere atanmasında oluşan kümelemenin kalitesini ölçerken de örneklerin ait oldukları küme merkezlerine öklid uzaklığını kullanmışlardır. Küme sayısı bilindiğinde uygulanan PSO sonuçları uygunluk değerleri, hata ve hesapsal zaman açısından ABC, KKO algoritması ve k -ortalama ile melez PSO algoritması ile karşılaştırılmış ve algoritma başarılı sonuçlar elde etmiştir. Küme sayısı bilinmeden yapılan uygulamada aynı kriterler açısından yine üstün sonuçlar elde ettiği görülmüştür. Uygun olmayan çözümlerin üstesinden gelmek için Deb kuralı uygulanmıştır.

Tiejun et al. [121], PSO algoritmasının karma özellikteki verilere uygulanması sırasında, parçacıkların arama uzayındaki yerlerini sınırlamak için bir potansiyel alan sınırlama mekanizması kullanmışlardır.

Matsui et al. [122], kesikli optimizasyon problemlerinde PSO kullanırken, başlangıç noktalarını üretmek için yeni bir yöntem önermişlerdir. Bu yöntem, noktaların hareket planları ve hareket yöntemlerinin revizyonu ile elde edilen değerlerin yuvarlanması esasına dayanmaktadır.

Ji et al. [123], kategorik veri için k -modlar mekanizmasına dayalı ABC adlı bir melez algoritmayı ortaya koymuşlardır. K -modlar algoritmasını örneklerin kümelere atanmasında ve küme merkezlerinin güncellenmesinde kullanmışlardır.

Marinakis et al. [124], iki aşamalı melez kesikli ABC algoritması önermişlerdir. Birinci aşamada özellik seçimi için ABC algoritması kullanılmakta ve ikinci aşamada kümeleme yapmak için Aç Gözlü Rastgele Adaptif Arama Prosedürü (GRASP-Greedy Randomized Adaptive Search Procedure) kullanılmaktadır. Özellik seçimi kesikli bir problem olduğu için ABC algoritmasına bazı modifikasyonlar yaparak kesikli ABC'yi önermişlerdir. Sürekli değerleri kesikleştirmek için sigmoid fonksiyonu kullanmışlardır.

4.6 Sonuç

Bu bölümde, tez çalışmasında kümeleme problemine uygulanan metasezgisel algoritmalar ayrıntılı olarak tanıtılmıştır. Gerek yeni nesil algoritmalar olan IMO ve WSA, gerekse literatürde çokça kullanılan algoritmalar olan PSO ve ABC algoritmalarının çıkış noktaları, adımları ve diğer özellikleri detaylandırılmıştır.

5. BÖLÜM

METASEZGİSEL ALGORİTMALARIN KÜMELEMeye UYGULANMASI ve ANALİZ EDİLMESİ

5.1 Giriş

Bu bölümde, dördüncü bölümde ayrıntılı olarak açıklanan metasezgisel algoritmaların kümeleme problemine uygulanması yer almaktadır. Algoritmalarından PSO ve ABC, literatürde daha önce kümeleme problemine uygulanmıştır. Ancak, IMO ve WSA algoritmaları yeni nesil metasezgisel algoritmalar olup bu tez çalışmasıyla ilk kez kümeleme problemlerine uygulanmış olacaktır.

5.2 Uygulama

5.2.1 Aday Çözümlerin Tanımlanması

Tüm algoritmalar kümeleme problemine uygulanırken aday çözümler Şekil 5.1’de görüldüğü gibi küme merkezi tabanlı temsil ile gerçek kodlama yapılarak tanımlanmıştır. Şekil 5.1’de iki boyutlu bir veri kümesinin üç küme ile temsil edilmesi ortaya konmuştur.

z_{11}	z_{12}	z_{21}	z_{22}	z_{31}	z_{32}
----------	----------	----------	----------	----------	----------

Şekil 5.1 Aday bir çözümün kodlaması [92]
küme sayısı=3, veri kümesi boyutu=2

z_{11} , 1. küme merkezinin 1. boyutunu, z_{12} ise 1. küme merkezinin 2. boyutunu ifade etmektedir.

Tez çalışmasında; sürekli, kategorik ve karma veri kümeleriyle çalışıldığından farklı değişken tiplerine sahip veri kümeleri için algoritmalarda gerekli modifikasyonlar yapılmıştır. Başlangıç çözümleri üretilmesinde, örneklerin kümelere atanmasında, küme merkezlerinin güncellenmesinde ve uygunluk değerlerinin hesaplanmasında kullanılan mekanizmalar değişkenlik gösterdiği için uygulama sırasında algoritmalara yapılan modifikasyonların izah edilişi veri kümelerinin değişken tiplerine göre yapılacaktır.

5.2.2 Sürekli Veri Kümelerinde Uygulama

5.2.2.1 Başlangıç Çözümlerinin Oluşturulması

Başlangıç çözümleri, her özelliğin maksimum ve minimum değerleri arasından rastgele oluşturulmaktadır. Uygun olmayan çözümlerin seçilmemesi için Deb kuralı başlangıç çözümü oluşturma mekanizmasında ve ayrıca uygunluk değerlerinin üretilmesi sonrasında algoritmaların kendilerine özgü akışları içinde bir seçim mekanizması olarak kullanılmaktadır.

Kümeleme probleminde uygun olmayan çözüm;

- Bir örneğin birden fazla kümeye atanmış olması,
- Herhangi bir kümeye hiç bir örneğin atanmamış olması,

durumlarından biri veya ikisinin gerçekleştiği çözümleri ifade etmektedir. Çalışmada her örneğin minimum uzaklıktaki kümeler arasından sadece birine atanması gerçekleştirilerek birinci uygunsuzluk durumunun önüne geçilmiştir. İkinci uygunsuzluk durumu ise kümeleme sonunda gerçekleşebilmekte ve ceza puanı işlemi uygulanmaktadır. Neticede, çalışmada uygun olmayan çözüm sadece boş kalan bir kümenin olmasından ileri gelmekte ve hiçbir örneğin atanmadığı küme sayısı, ceza puanı olarak verilmektedir.

Tez çalışmasında uygun olmayan çözümlerin üstesinden gelmek için kullanılan Deb kısıt işleme kuralına göre, a ve b ile ifade edilen iki çözüm için:

- Eğer a çözümünün uygunsuzluğu $< b$ çözümünün uygunsuzluğu ise o zaman a çözümü, b çözümüne tercih edilir.

- Eğer a çözümünün uygunsuzluğu = b çözümünün uygunsuzluğu ise her iki çözümün uygunluk değerinden minimizasyon probleminde minimum olan ve maksimizasyon probleminde maksimum olan tercih edilir [43].

Başlangıç çözümü oluşturulurken Deb kuralı, uygun olmayan çözümleri, çözümler oluşur oluşmaz tespit ederek yenisini üreten ve tüm çözümler uygun çözümler olana kadar döngüyü devam ettiren bir mekanizma olarak çalışmaktadır. Başlangıç çözümü oluşturma formülü Eşitlik 5.1’de verilmiştir.

$$x_{ik} = attmin_k + rand() * range_k \quad (5.1)$$

Burada x_{ik} , i . çözümün k . boyutunun değerini, $attmin_{1k}$, k . boyutun minimum değerini, $range_{1k}$, k . boyutun maksimum değeri ile minimum değeri arasındaki farkı ifade etmekte ve Eşitlik (5.2) ile hesaplanmaktadır. Eşitlik (5.1)’de $rand()$, $[0,1]$ aralığında rastgele bir değer üretmektedir.

$$range_k = attmax_k - attmin_k \quad (5.2)$$

$attmax_k$, k . boyutun maksimum değeridir.

IMO algoritması dışındaki algoritmalarda başlangıç çözümleri Eşitlik (5.1)’e göre üretilmiştir. IMO’da aday çözümler anyonlar ve katyonlar olarak ikiye ayrılmaktadır. Uygulama sırasında IMO’da hem anyonlar hem de katyonlar için başlangıç çözümleri Eşitlik (5.1)’e göre uygulanmış, ancak arama uzayını daha iyi temsil etmesi açısından Eşitlik (5.3) ve Eşitlik (5.4) ile başlangıç çözümleri oluşturulması denenmiştir.

$$A_{ik} = attmin_k + \emptyset * range_k \quad (5.3)$$

$$C_{ik} = attmax_k - \emptyset * range_k \quad (5.4)$$

Burada $\emptyset$, $[0, 0.5]$ arası rastgele bir değerdir.

Birkaç veri kümesi üzerinde yapılan denemeler Eşitlik (5.3) ve Eşitlik (5.4) ile elde edilen bir başlangıç çözümünün Eşitlik (5.1)’e göre iyi sonuçlara daha hızlı yakınsama sağladığını göstermiştir ve IMO algoritmasının tüm uygulamalarda bu şekilde çalıştırılmasına karar verilmiştir.

5.2.2.2 Örneklerin Küme Merkezlerine Atanması

Veri kümesindeki örneklerin atanacakları kümelerin belirlenmesinde, verilerin küme merkezlerine olan öklid uzaklığı esas alınmıştır. Öklid uzaklığı Eşitlik (5.5) ile hesaplanmaktadır.

$$D(o_i, o_l) = \sqrt{\sum_{j=1}^J (o_{ij} - o_{lj})^2} \quad (5.5)$$

Örnekler, tüm küme merkezlerine öklid uzaklıkları hesaplanarak, en yakındaki küme merkezine atanmaktadırlar. Tüm örnekler atandıktan sonra kümeleme işlemi gerçekleşmekte ve bundan sonraki aşamada küme merkezleri güncellenerek kümelemenin uygunluk değeri hesaplanmaktadır.

5.2.2.3 Kümelemenin Uygunluk Değerinin Hesaplanması

Veri kümesindeki örnekler, öklid uzaklık ölçüsüne göre en yakın kümeye atandıktan sonra elde edilen kümelerin merkezleri Eşitlik (5.6) 'ya göre güncellenmekte ve kümelere atanan örneklerin kendi küme merkezlerinden öklid uzaklıklarının toplamı uygunluk değerini vermektedir. Uygunluk değerinin hesaplanması Eşitlik (5.7) ile yapılmaktadır.

$$z_{kj}^t = \frac{\sum_{i=1}^n w_{ik}^t o_{ij}}{\sum_{i=1}^n w_{ik}^t}, k = \{1, \dots, K\}, j = \{1, \dots, m\} \quad (5.6)$$

$$f^t = \sum_{k=1}^K \sum_{i=1}^n w_{ik}^t Di(o_i, z_k^t) \quad (5.7)$$

Formüllere ilişkin notasyon üçüncü bölümde verilmiştir. Çalışmada seçilen uygunluk değeri hesaplama yöntemi, örneklerin kümelere atanmasında kullanılan uzaklık ölçüsünün, örneklerin atandıkları küme merkezlerine olan uzaklıklarının hesabında da kullanılmasını esas almıştır.

Tüm çözümlerin uygunluk değerleri hesaplandıktan sonra her algoritmada çözümler arasından en iyiyi seçme mekanizması için Deb kısıt işleme metodu uygulanmıştır. Deb kuralı, çözümleri değerlendirmekte ve çözümlere uygunsuzluğun sayısına göre ceza puanı vermektedir. Meydana gelebilecek tek uygunsuz çözüm, hiçbir elemanın atanmadığı kümenin olması durumu olduğu için ceza puanı, hiçbir örneğin atanmadığı küme sayısını ifade etmektedir.

Çalışmada, ABC algoritmasında işçi ve gözcü arıların komşuluk bulmalarının hemen ardından açgözlü yaklaşımla yaptıkları seçim yerine Deb metodu ile seçim yaptırılmıştır.

Kullanılan algoritmalar karşılaştırmaya tabi tutulacağı için tamamında başlangıç çözümü oluşturma, örneklerin kümelere atanması, küme merkezlerini güncelleme ve uygunluk değerlerini oluşturma konusunda ortak bir format takip edilmeye çalışılmıştır. Bunlar dışında tüm algoritmalar kendilerine özgü bir şekilde iteratif olarak devam etmekte ve önceden tanımlanan çalıştırma ve maksimum iterasyon sayısı gerçekleştiğinde kümeleme sonuçları raporlanmaktadır.

5.2.2.4 Orijinal IMO Algoritmasında Yapılan Modifikasyonlar

IMO algoritması üzerinde çalışılırken, kristal aşamada iyonların en iyi anyon ve katyonların etrafına dağılma prensibi üzerinde modifikasyon yapılarak algoritmanın performansının etkilendiği görülmüştür. Orijinal IMO algoritmasında, kristal aşamadaki (C_{best-1}) (Şekil 4.3) mekanizmasının kullanılmasına karşın, kümeleme probleminde -1 yerine *range/iteration* mekanizmasının kullanılmasının yerel aramaya katkı sağlayacağı öngörülerek buna göre algoritma çalıştırılmıştır. Ayrıca; kristal aşamada en iyi anyon ve katyonların etrafına iyonların dağıtılması mekanizmasında diğer metasezgisellerin kullandığı komşuluk yapılarının entegrasyonu ile algoritmanın daha iyi sonuçlar verip vermeyeceği incelenmiştir. Bunun için IMO algoritmasının saçılma mekanizmasının içine vektör farkına dayalı mutasyon entegre edilmiştir. Benzer komşuluk yapısını ABC algoritması da Eşitlik (5.8) ile kullanmaktadır[56].

$$v_{ij} = x_{ij} + \theta_{ij}(x_{ij} - x_{kj}) \quad (5.8)$$

Burada θ_{ij} , -1 ile 1 arasında rastgele üretilmiş bir sayıdır. Eşitlik (5.8) vektörün sadece tek boyutunda değişiklik yapmaktadır. IMO algoritmasının kristal aşamasında iyonlar en iyi anyon ve katyonun etrafında dağılacakları için Eşitlik (5.8)'deki komşuluk yapısı entegre edilirken; bir anyon ile en iyi katyon ve bir katyon ile en iyi anyon arasında tüm boyutlardaki farklar dikkate alınmıştır. Yani vektörün sadece bir boyutunda değil; tüm boyutlarında değişiklik yapılmıştır. Sonuçta; orijinal IMO algoritmasındaki Eşitlik (5.9) yerine, Eşitlik (5.10) ve Eşitlik (5.11) yerine, Eşitlik (5.12) kullanılmıştır.

$$A_i = A_i + \emptyset_1 \times (Cbest - 1) \quad (5.9)$$

$$A_i = A_i + \emptyset_1 \times \left(A_i - \left(Cbest - \frac{range}{iteration} \right) \right) \quad (5.10)$$

$$A_i = A_i + \emptyset_1 \times (Cbest) \quad (5.11)$$

$$A_i = A_i + \emptyset_1 \times (A_i - Cbest) \quad (5.12)$$

Anyonlar için yapılan bu modifikasyon, kristal aşamanın her komşuluk aramasında katyonlar için de Eşitlik (5.13) ve Eşitlik (5.14) ile uygulanmış ve yapılan denemeler sonucunda Orijinal IMO'ya göre daha iyi sonuçlar verdiği görülmüştür.

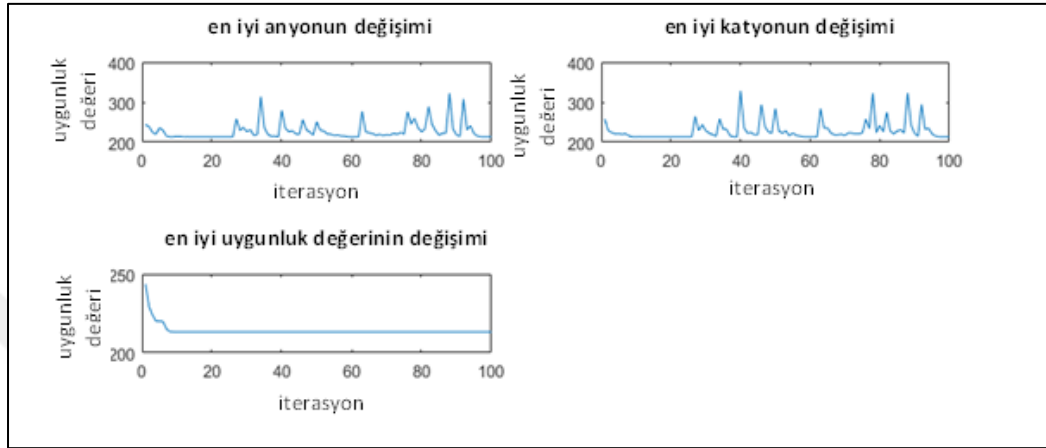
$$C_i = C_i + \emptyset_1 \times \left(C_i - \left(Abest - \frac{range}{iteration} \right) \right) \quad (5.13)$$

$$C_i = C_i + \emptyset_1 \times (C_i - Abest) \quad (5.14)$$

Bu nedenle, bundan sonraki her uygulamada IMO algoritmasının ilk modifikasyonu çalıştırılmıştır.

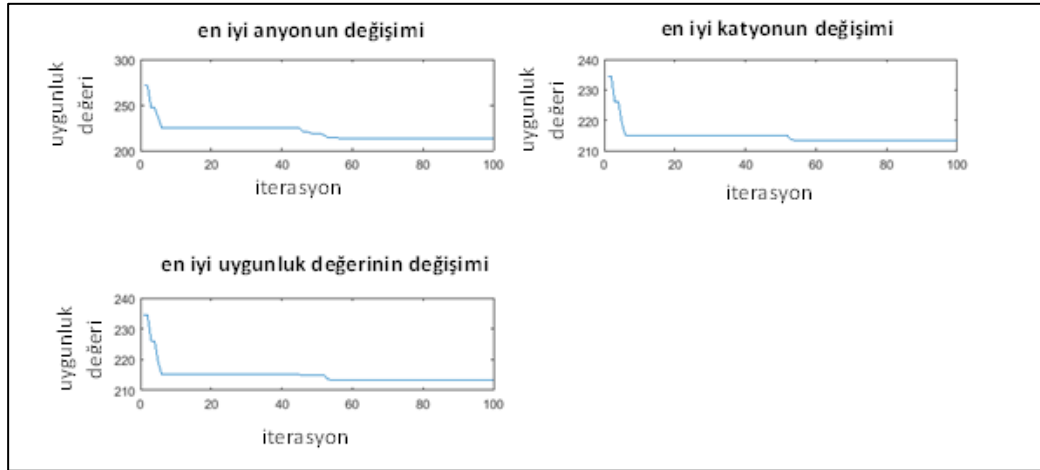
Ayrıca; en iyi anyon ve katyonların iterasyona bağlı olarak değişmesi ve değişmemesi durumunda davranışları gözlemlenmiştir. 4 durum incelemesi yapılmıştır.

- 1. Durumda:** Her iterasyonda elde edilen en iyi anyon ve katyonlar global olarak korunmamış, iterasyon değiştikçe en iyi anyon ve katyonlar da değişmiştir. Bu durum için 100 iterasyon boyunca en iyi anyon, en iyi katyon ve global en iyi çözümün uygunluk değerlerinin değişimleri Şekil 5.2’de görülmektedir.



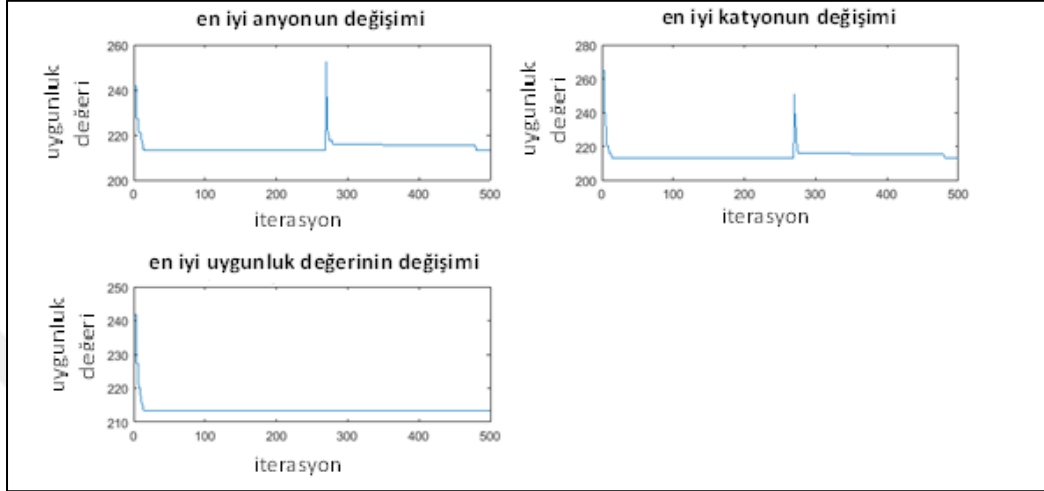
Şekil 5.2 En iyi iyonların her iterasyonda değişmesi

- 2. Durumda:** En iyi anyon ve katyonlar global olarak korunarak her iterasyonda değişmemişlerdir. Şekil 5.3 bu durumu göstermektedir.



Şekil 5.3 En iyi iyonların global olarak korunması

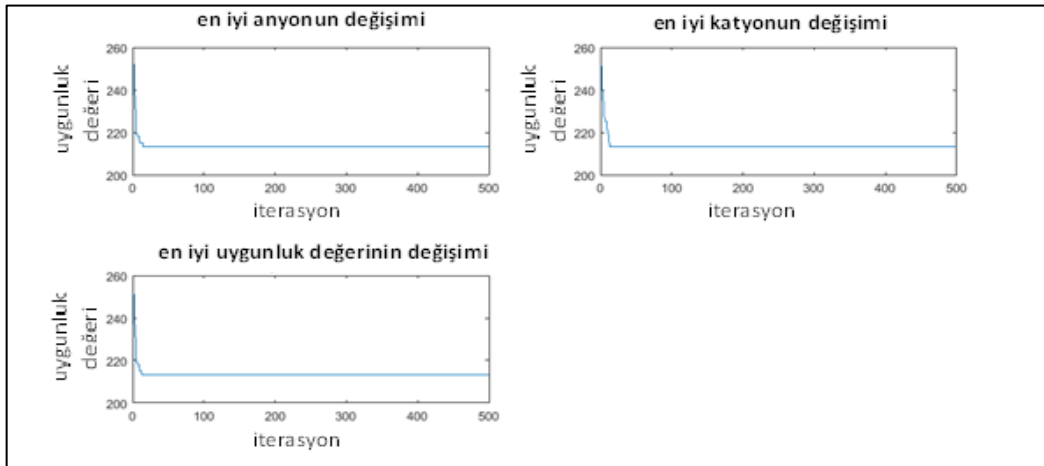
- 3. Durumda:** En iyi anyon ve katyonlar global olarak korunmuş ve 250 iterasyon boyunca en iyi uygunluk değeri değişmezse, en iyi anyon ve katyonlar hafızadan silinerek, tüm anyon ve katyonlar yeniden oluşturulmuştur. Şekil 5.4 bu durumdaki değişimleri göstermektedir.



Şekil 5.4 En iyi uygunluk değeri değişmezse en iyi iyonların yeniden üretilmesi

Şekil 5.4’de en iyi anyon ve en iyi katyondaki ani yükselmelerin sebebi, 250 iterasyon boyunca en iyi uygunluk değerinde bir değişiklik meydana gelmemesinin en iyi anyon ve katyonların yeniden üretilmesine neden olmasıdır.

- 4. Durumda:** 3. durumdan farklı olarak Şekil 5.5’de görüldüğü gibi 250 iterasyon boyunca en iyi uygunluk değeri değişmezse en iyi anyon ve katyonlar hafızadan silinmeden tüm anyon ve katyonlar baştan üretilmiştir.



Şekil 5.5 En iyi uygunluk değeri değişmezse en iyi iyonların korunması

Bu dört durum için de veri kümeleri IMO'da çalıştırılmış ve en iyi sonuçların 2, 3 ve 4. durumlarda bulunduğu ancak; 3 ve 4. durumların hesapsal zamanı arttırdığı tespit edildiği için 2. durum modifikasyonu tercih edilmiştir.

Yapılan modifikasyonlarla Orijinal IMO'dan farklı komşuluk yapısı kullanan ve en iyi anyon ve katyonların iterasyona bağlı olarak değişmediği modifiye bir algoritma meydana gelmiştir. Deneysel çalışma sırasında hem orijinal hem de modifiye algoritma uygulanmıştır. Bundan sonraki bölümde, orijinal IMO algoritması, IMO ve modifiye edilen algoritma ise Modifiye IMO olarak anılacaktır.

5.2.2.5 Orijinal WSA Algoritmasında Yapılan Modifikasyonlar

WSA algoritmasında, hedef nokta belirlenmesinde ve belirlenen hedef noktaya göre yön tayininde modifikasyonlar denenmiştir. Orijinal WSA algoritmasında, hedef nokta belirlenirken kötü çözümlerin de hedef noktaya katkıları olmaktadır. Ayrıca, tüm çözümlerin hedef noktaya doğru yönlerinin tespitinde, iyi bir çözümün kendisinden daha kötü bir çözüm olabilen hedef noktaya yönelmesi mümkün olabilmektedir. WSA algoritmasında yapılan modifikasyon kötü çözümlerin katkısının ortadan kaldırılmasına yönelik olmuştur.

İlk modifikasyonda;

Kötü çözümlerin katkısının elimine edilmesi için o iterasyonda elde edilen en iyi uygunluk değerinin her çözümün uygunluk değerine oranı, o çözümün ağırlığı olarak alınmıştır. Orijinal WSA'daki Eşitlik (5.15) yerine, Eşitlik (5.16) kullanılmıştır.

$$weight(1, i) = i^{(-1)*\tau} \quad (5.15)$$

$$weight(1, i) = \frac{bestf}{f(i)} \quad (5.16)$$

Buradaki *bestf*, o iterasyonda elde edilen en iyi uygunluk değerini ve *f(i)* ise *i*. ajanın mevcut uygunluk değerini ifade etmektedir.

Elde edilen hedef nokta, herhangi bir çözümde kötüye o çözüm korunmuştur, değiştirilmemiştir.

İkinci modifikasyonda;

Hedef nokta direkt olarak iterasyondaki en iyi değer alınmış (yani çözümlerin ağırlıkları hesaplanmamıştır).

Bu modifikasyonlarla beraber her iki modifikasyon ayrı birer algoritma olarak değerlendirilmiş, sonuçta deneysel çalışmada; Orijinal WSA, Modifiye1 WSA ve Modifiye2 WSA şeklinde üç WSA algoritması uygulanarak sonuçları yorumlanmıştır. Orijinal WSA, çalışmanın bundan sonraki kısımlarında WSA olarak anılacaktır.

5.2.3 Kategorik ve Karma Veri Kümelerinde Uygulama

Bu çalışmada; kategorik ve karma veri kümelerinin kümelenmesi için sürekli veri kümeleri için uygulanan algoritmaların her birinde aşağıdaki modifikasyonlar yapılmıştır:

- Öncelikle, veri kümelerinin hangi özelliklerinin kategorik ya da sürekli olduğunu sisteme tanıtmak için *mix_state* adlı bir vektör tanımlaması yapılmıştır.

$$mix_state = [0 \ 1 \ 1 \ 0 \ 1]$$

Mesela, yukarıdaki *mix_state* vektörü; özellik sayısı 5 olan bir veri kümesinin 2,3 ve 5. özelliklerinin sürekli, 1 ve 4. özelliklerinin ise kategorik olduğunu ifade etmektedir.

$$mix_state = [0 \ 0 \ 0 \ 0 \ 0]$$

Yukarıdaki vektör ise, veri kümesinin tamamen kategorik olduğunu göstermektedir.

Gerek başlangıç çözümü üretilirken, gerek küme merkezleri güncellenip uygunluk değeri hesaplanırken ve gerekse algoritma içinde kategorik özelliklerin kesikli alanda kalması sağlanırken *mix_state* vektörüne göre hareket edilmektedir.

5.2.3.1 Başlangıç Çözümlerinin Oluşturulması

Başlangıç çözümü, sürekli değişkenlerde ilgili özelliğin minimum ve maksimum değerleri arasında sürekli alandan rastgele oluşturulurken, kategorik değişkenlerde ilgili özelliğin minimum ve maksimum değerleri arasındaki kesikli alandan rastgele oluşturulmaktadır. *mix_state* vektörüne göre özelliğin kategorik olması durumunda kesikli alandan Eşitlik (5.17)'e göre, özelliğin sürekli olması durumunda ise Eşitlik (5.1)'e göre başlangıç çözümü oluşturulmaktadır.

$$x_{ik} = attmin_k + round(rand() * range_k) \quad (5.17)$$

Burada *round* ifadesi, *rand() * range_k* ifadesinin, $[0, range_k]$ aralığında oluşturduğu sürekli değeri kesikli hale getirmektedir. Yani, Eşitlik (5.17) ile, ilgili özelliğin minimum değeri ile maksimum değeri arasında kesikli bir değer üretilmektedir.

5.2.3.2 Örneklerin Küme Merkezlerine Atanması

Veri kümesindeki örneklerin küme merkezlerine uzaklıklarının hesaplanmasında, heterojen öklid uzaklığı adı verilen ve üçüncü bölümde Eşitlik (3.18) ile hesaplanan bir ölçüt kullanılmıştır. Kategorik özelliklerin uzaklık hesabında çakışma ölçütü, sürekli özelliklerde normalize edilmiş öklid uzaklığı kullanılmıştır [49]. Ancak, burada literatürde rastlanılmayan bir ekleme yapılmıştır. Kategorik özellikler arasındaki uzaklık hesaplamalarında çakışma ölçütü formülü (Eşitlik (3.20)) ile hesaplama yapıldıktan sonra ilgili örneğin birden fazla kümeye uzaklığı minimum ise, o örneğin küme merkezlerine Manhattan uzaklığı hesaplanmakta ve eşit minimumlar arasından Manhattan uzaklığı küçük olan kümeye atama yapılmaktadır. Manhattan uzaklığı Eşitlik (5.18) ile hesaplanmaktadır. Ayrıca, Deb kuralı hem kategorik hem karma veri tipleriyle çalışırken de sürekli veri kümelerinde olduğu gibi uygulanmıştır.

$$D(x, y) = \sum_{i=1}^m |x_i - y_i| \quad (5.18)$$

Tüm örnekler kümelere atandıktan sonra sürekli özelliklerde kümelere atanan örneklerin ortalamaları alınarak yeni küme merkezlerinin bulunmasına karşın, kategorik özelliklerin ortalamaları yerine modları alınmıştır. Mod, verinin o özelliğinde frekansı

yüksek değer demektir. En çok tekrarlanan değer o özelliğin modu olmaktadır. Karma veri kümelerinde; kategorik özelliklerin modları, sürekli özelliklerin ise ortalaması alınarak küme merkezleri güncellenmiştir.

5.2.3.3 Kümelemenin Uygunluk Değerinin Hesaplanması

Sürekli özelliklerde kümeleme gerçekleştikten sonra, her örneğin ait olduğu küme merkezine olan öklid uzaklıklarının toplamı ile uygunluk değeri hesaplanırken, karma verilerde kümeleme sonrası örneklerin ait oldukları küme merkezlerine heterojen öklid uzaklıklarının toplamı (Eşitlik (3.18)) uygunluk değeri olarak alınmaktadır.

5.2.3.4 Pozisyon Güncellemesi

Her algoritmada pozisyon güncellemeleri gerçekleştikten sonra, kategorik yapıda olduğu halde pozisyon güncellemesi sonucu sürekli hale gelen özelliklere basit yuvarlama işlemi uygulanmıştır. Bu işlem ile kategorik özelliklerin kesikli aralıkta kalması sağlanmıştır.

5.3 Deneysel Çalışma

Çalışmada kullanılan veri kümeleri, UCI makine öğrenme deposundan (<http://archive.ics.uci.edu/ml/machine-learning-databases/>) alınan ve Tablo 5.1’de özellikleri verilen veri kümeleridir.

Tablo 5.1’de görüldüğü gibi veri kümeleri içinde cancer, voting ve credit veri kümelerinin eksik değerleri bulunmaktadır. Cancer veri kümesinde az sayıda eksik değer olduğu için bunlar veri kümesinden çıkarılmış, voting veri kümesinde çok sayıda eksik değer olduğundan eksik olan değerler ilgili özellik için yeni bir kategori olarak değerlendirilmiş ve credit veri kümesinde ise eksik değerler sürekli özelliklerin ortalaması, kategorik özelliklerin modu ile doldurulmuştur.

Ayrıca, soybean veri kümesinde 14 özellik tek değerden oluştuğundan bunlar veri kümesinden çıkarılmıştır.

Tablo 5.1 Veri kümelerinin özellikleri

Veri kümesi	Değişken Tipi	Örnek sayısı	Sürekli değişken sayısı	Kategorik değişken sayısı	Sürekli değişkenler	Sınıf sayısı	Eksik değer sayısı	Tek değerli özellik sayısı
iris	sürekli	150	4	0	1- 4	3	0	0
wine	sürekli	178	13	0	1-13	3	0	0
thyroid	sürekli	215	5	0	1- 5	3	0	0
glass	sürekli	214	9	0	1- 9	6	0	0
cancer	kategorik	699	0	9	0	2	16	0
zoo	kategorik	101	0	17	0	7	0	0
soybean	kategorik	47	0	21	0	4	0	14
voting	kategorik	435	0	16	0	2	288	0
credit approval	karma	690	9	6	2,3,8,11, 14, 15	2	67	0
statlog heart	karma	270	5	8	1,4,5,8,10	2	0	0

5.3.1 Deney Kurulumu

Uygunluk değeri dışında, kümeleme kalitesini belirlemek için 3 farklı performans ölçütü kullanılmıştır:

- a) Elde edilen kümelemede hata oranı Eşitlik (5.19) ile hesaplanmıştır [45].

$$ER = \left[\left(\sum_{i=1}^{n-1} \sum_{l=i+1}^n |A_{il} - B_{il}| \right) / \left(\frac{n(n-1)}{2} \right) \right] \times 100 \quad (5.19)$$

Burada A_{il} , kümeleme gerçekleşmeden önce i örneği ile l örneği aynı kümede iseler 1, değilse 0 değerini aldığını ifade eden bir değerdir. B_{il} kümeleme gerçekleşikten sonra eğer i örneği ile l örneği aynı kümede ise 1, değilse 0 değerini almaktadır.

- b) Elde edilen kümelerin doğruluk oranları Eşitlik (5.20) ile hesaplanmıştır [121].

$$u = \frac{\sum_{i=1}^k \alpha_i}{n} \quad (5.20)$$

Burada; k küme sayısı olmak üzere α_i , i . kümede doğru bir şekilde kümelenen örneklerin sayısıdır.

c) Hesapsal zaman.

10 çalıştırma sonrası algoritmaların bir çalıştırmayı (maksimum iterasyon sayısınca) gerçekleştirdiği zamanın saniye olarak hesabıdır. Tüm performans ölçütleri 10 çalıştırma sonrası elde edilen değerlerin ortalaması, en iyisi ve en kötüsü olarak raporlanmıştır. Algoritmalar kodlanırken,

- ABC için; <http://mf.erciyes.edu.tr/abc/form.aspx> web adresindeki,
- WSA için <http://web.deu.edu.tr/baykasoglu/wsa.html> web adresindeki standart fonksiyonlar için MATLAB ile hazırlanmış kod yapıları alınarak, sadece kümeleme problemi için gerekli uyarlamalar yapılmıştır.
- PSO ve IMO algoritmaları ise kümeleme problemi için kodlanmıştır.

Kodlama için MATLAB R2015b programı kullanılmış ve algoritmalar Intel(R) Core(TM) i7-4510U 2.60 GHz CPU 8.00 GB RAM özelliklere sahip bir bilgisayarda çalıştırılmıştır. Algoritmalar için kullanılan parametre değerleri Tablo 5.2’de verilmiştir.

Algoritmalar birbirinden farklı yakınsama hızlarına sahip olsalar da, adil bir ortamda karşılaştırma yapabilmek için başlangıç popülasyon ve maksimum iterasyon sayıları tüm algoritmalar için eşit alınmıştır.

Kümeleme problemi üçüncü bölümde belirtildiği gibi Eşitlik (3.3) ve Eşitlik (3.4) kısıtlarına sahip bir problemdir. Bu nedenle, WSA algoritmasının kısıtlı problemler için önerildiği makalede yazarların çalıştırdığı parametre değerleri ile veri kümelerinde denemeler yapılmıştır. Diğer algoritmalara göre veri kümelerinde gösterdiği düşük performans nedeniyle kısıtsız problemlere uyguladıkları parametreler ile çalıştırma yapılmıştır. Bu durumda çok daha iyi sonuçların ortaya çıkması, WSA algoritması için Tablo 5.2’de verilen parametrelerde karar kılınmasına sebep olmuştur.

PSO algoritmasında c_1 ve c_2 parametreleri için yapılan denemeler sonucunda en iyi uygunluk değerinin $c_1=2$, $c_2=2$ olduğu durumda elde edildiği tespit edilmiştir. Cura [45], PSO’yu kümeleme problemine uygularken $c_1=2$, $c_2=2$ değerlerini kullanmıştır.

Tablo 5.2 Parametre değerleri

Parametreler	ABC	PSO	WSA ve modifikasyonları	IMO ve Hybrid IMO
Popülasyon sayısı	50	50	50	50
Maximum iterasyon	2000	2000	2000	2000
Çalıştırma sayısı	10	10	10	10
<i>Limit</i>	10	-	-	-
Başlangıç hız vektörü (v)	-	0	-	-
Yerel öğrenme faktörü (c_1)	-	2	-	-
Global öğrenme faktörü (c_2)	-	2	-	-
Kullanıcı tanımlı parametre (τ)	-	-	0,8	-
Başlangıç adım uzunluğu (sl)	-	-	0,035	-
Kullanıcı tanımlı parametre (λ)	-	-	1	-
Kullanıcı tanımlı parametre (φ)			0,0001	

ABC algoritmasının *limit* değeri literatürde tercih edilen bir değer olan 10 olarak alınmıştır. Zhang et al. [43], *Limit* değerini 10 olarak kullanmıştır. Tablo 5.2’de görüldüğü gibi IMO ve Modifiye IMO algoritmasının maksimum iterasyon sayısı ve popülasyon sayısı dışında belirlenmesi gereken bir parametresi yoktur. IMO algoritması açısından bu durum bir avantajdır.

5.3.2 Performans Karşılaştırması

Algoritmalar tüm veri kümeleri için çalıştırıldıktan sonra performans kriterlerinin tamamı açısından ortalama, en iyi ve en kötü değerler tablolar halinde raporlanmıştır.

Ayrıca, algoritmaların elde ettikleri sonuçların farklılıklarının istatistiksel olarak anlamlı olup olmadığını belirlemek üzere MINITAB 17 istatistik yazılımında Wilcoxon Eşleştirilmiş İki Örnek Testi kullanılarak algoritmalar ikili karşılaştırmaya tâbi tutulmuştur.

Bağımlı iki grup karşılaştırması için parametrik test varsayımları olan iki grup arasındaki değerlere ilişkin fark değerleri dağılımının normal dağılım göstermemesi ve eş varyanslığın sağlanmaması durumunda parametrik olmayan Wilcoxon Eşleştirilmiş İki Örnek Testi kullanılır. Her bir karşılaştırmada iki algoritma aynı veri kümeleri açısından karşılaştırılacağından, normalitenin ve eş varyanslığın her karşılaştırmada sağlanamamasından dolayı ve karşılaştırılan örnek sayıları eşit olduğundan parametrik olmayan Wilcoxon işaret testinin kullanılması uygun bulunmuştur.

Algoritmaların sonuçları karşılaştırılırken öncelikle ortalama, en iyi ve en kötü değerlerin yer aldığı tablolar sürekli, kategorik ve karma veri kümeleri için tüm performans ölçütleri ayrı ayrı tablolar halinde verilmiş, daha sonra Wilcoxon testi sonuçlarına göre yorumlar yapılmıştır.

5.3.2.1 Sürekli Veri Kümeleri için Karşılaştırma

Algoritmaların veri kümelerinde uygunluk değeri sonuçlarına göre 10 çalıştırma içinde ortalama, en iyi ve en kötü değerleri Tablo 5.3’de verilmiştir.

Tablo 5.3’e bakıldığında, IMO ve WSA dışında her algoritmanın 4 veri kümesi için de aynı en iyi çözümleri bulunduğu, Modifiye IMO algoritmasının ise her çalıştırmada en iyi çözümü bulunduğu görülmektedir. Yani Modifiye IMO algoritması; her 10 çalıştırmada da sürekli veri kümelerinin 4’ü için de en iyi değeri bulmaktadır. Modifiye IMO algoritması için sürekli veri kümelerinde bu durum bir üstünlük olarak ifade edilebilir. Her veri kümesinde her çalıştırmada en iyi sonucu bulmak algoritmanın sağlamlığına dair önemli bir kanıttır.

Modifiye2 WSA algoritmasının PSO algoritması ile rekabet edebilir bir algoritma olduğu görülmektedir. Gerek PSO, gerekse Modifiye2 WSA glass veri kümesi haricindeki diğer veri kümelerinde en iyi sonuçları her çalıştırmada bulmuştur.

Tablo 5.3 Sürekli veri kümelerinde algoritmaların uygunluk değerleri

Veri kümesi	Kriter	PSO	ABC	IMO	Modifiye IMO	WSA	Modifiye1 WSA	Modifiye2 WSA
iris	Ortalama	97,22	97,22	97,22	97,22	97,22	97,23	97,22
	En iyi	97,22	97,22	97,22	97,22	97,22	97,22	97,22
	En kötü	97,22	97,22	97,22	97,22	97,22	97,23	97,22
wine	Ortalama	16530,50	16530,50	16530,50	16530,50	16530,50	16530,50	16530,50
	En iyi	16530,50	16530,50	16530,50	16530,50	16530,50	16530,50	16530,50
	En kötü	16530,50	16530,50	16530,50	16530,50	16530,50	16530,50	16530,50
thyroid	Ortalama	1960,59	1961,11	1961,29	1960,59	1961,65	1961,25	1960,59
	En iyi	1960,59	1960,59	1960,59	1960,59	1960,59	1960,59	1960,59
	En kötü	1960,59	1961,75	1962,98	1960,59	1962,78	1962,57	1960,59
glass	Ortalama	213,71	213,22	213,58	213,20	214,27	213,24	213,52
	En iyi	213,20	213,20	213,38	213,20	213,22	213,20	213,20
	En kötü	218,23	213,22	214,16	213,20	216,11	213,42	215,48

IMO ve WSA algoritmalarının glass veri kümesinde diğer algoritmalar tarafından bulunan en iyi sonuç olan 213,20'yi bulamaması diğer algoritmaların bu algoritmalarla olan üstünlüğü olarak değerlendirilebilir. IMO'da yapılan modifikasyon ve WSA'da yapılan ikinci modifikasyonun sürekli veri kümelerinde bu algoritmaların uygunluk değerlerine göre performanslarını arttırdığı ortadadır.

Hata oranlarına göre kümeleme kalitesi Tablo 5.4'de verilmiştir.

Tablo 5.4'e bakıldığında, en iyi değerler açısından PSO 2 veri kümesinde, ABC 3 veri kümesinde, IMO 3 veri kümesinde, Modifiye IMO 2 veri kümesinde, WSA 2 veri kümesinde, Modifiye1 WSA 2 veri kümesinde ve Modifiye2 WSA ise 3 veri kümesinde en iyi sonucu bulmuştur.

Tablo 5.4 Sürekli veri kümelerinde algoritmaların hata oranları

Veri kümesi	Kriter	PSO	ABC	IMO	Modifiye IMO	WSA	Modifiye1 WSA	Modifiye2 WSA
iris	Ortalama	11,41	11,41	11,41	11,41	11,78	11,78	11,41
	En iyi	11,41	11,41	11,41	11,41	11,50	11,50	11,41
	En kötü	11,41	11,41	11,41	11,41	12,03	12,03	11,41
wine	Ortalama	28,09	28,09	28,09	28,09	28,09	28,09	28,09
	En iyi	28,09	28,09	28,09	28,09	28,09	28,09	28,09
	En kötü	28,09	28,09	28,09	28,09	28,09	28,09	28,09
thyroid	Ortalama	36,45	36,54	36,50	36,45	36,63	36,53	36,45
	En iyi	36,45	36,14	36,14	36,45	36,14	36,14	36,45
	En kötü	36,45	36,76	36,76	36,45	37,26	36,97	36,45
glass	Ortalama	32,77	32,54	32,57	32,90	32,89	32,70	32,56
	En iyi	31,56	32,37	31,75	32,90	31,77	32,37	31,00
	En kötü	32,90	32,65	33,62	32,90	33,74	33,97	33,66

Ortalamalar açısından PSO 3 veri kümesinde, ABC 3 veri kümesinde, IMO 2 veri kümesinde, Modifiye IMO 3 veri kümesinde, WSA ve Modifiye1 WSA 1 veri kümesinde, Modifiye2 WSA ise 3 veri kümesinde en iyi sonucu bulmuştur. ABC ve Modifiye2 WSA'nın diğer algoritmalara en iyi ve ortalama değerler açısından üstünlük sağladığı görülmektedir.

Doğruluk oranlarına göre kümeleme kalitesi Tablo 5.5'de verilmiştir.

Tablo 5.5'e bakıldığında en iyi değerler açısından 2 veri kümesinde ABC, ortalama değerler açısından ise ABC ve IMO dışındaki algoritmalar 3 veri kümesinde en iyi performansı göstermiştir.

Tablo 5.5 Sürekli veri kümelerinde algoritmaların doğruluk oranları

Veri kümesi	Kriter	PSO	ABC	IMO	Modifiye IMO	WSA	Modifiye1 WSA	Modifiye2 WSA
iris	Ortalama	90,00	89,87	90,00	90,00	90,00	89,80	90,00
	En iyi	90,00	90,67	90,00	90,00	90,00	90,00	90,00
	En kötü	90,00	89,33	90,00	90,00	90,00	89,33	90,00
wine	Ortalama	70,79	70,62	70,79	70,79	70,79	70,79	70,79
	En iyi	70,79	71,35	70,79	70,79	70,79	70,79	70,79
	En kötü	70,79	70,22	70,79	70,79	70,79	70,79	70,79
thyroid	Ortalama	80,47	80,28	80,42	80,47	80,47	80,47	80,47
	En iyi	80,47	80,47	80,47	80,47	80,47	80,47	80,47
	En kötü	80,47	78,60	80,00	80,47	80,47	80,47	80,47
glass	Ortalama	57,29	57,49	58,18	57,48	58,32	58,93	58,41
	En iyi	57,29	57,29	58,88	57,48	59,35	59,35	58,88
	En kötü	57,29	58,66	57,48	57,48	56,08	58,41	57,93

Tablo 5.6, 10 çalıştırma sonunda hesapsal zaman açısından elde edilen değerlerin ortalaması, en iyisi ve en kötüsünü içermektedir.

En iyi ve ortalama kriterleri açısından bakılınca, iris veri kümesi haricinde tüm algoritmalarda hesapsal zaman performansını en iyi PSO algoritmasının gösterdiği görülmektedir.

Tablo 5.6 Sürekli veri kümelerinde algoritmaların hesapsal zamanları

Veri kümesi	Kriter	PSO	ABC	IMO	Modifiye IMO	WSA	Modifiye1 WSA	Modifiye2 WSA
iris	Ortalama	44,52	490,71	48,61	50,04	54,09	44,08	46,08
	En iyi	43,17	486,23	45,95	49,79	40,73	41,22	41,95
	En kötü	47,87	493,60	52,56	50,89	63,97	58,09	59,00
wine	Ortalama	46,74	520,08	48,70	62,05	61,72	65,92	52,17
	En iyi	45,83	514,09	46,29	61,72	46,71	51,48	50,62
	En kötü	47,72	524,35	52,60	63,05	68,47	71,14	54,23
thyroid	Ortalama	45,87	511,21	47,77	55,63	63,64	62,61	49,81
	En iyi	44,90	506,94	46,48	55,31	48,08	46,96	48,22
	En kötü	47,50	514,04	51,64	56,11	68,63	70,08	51,56
glass	Ortalama	71,46	812,13	119,33	146,88	97,55	109,06	107,58
	En iyi	67,28	797,08	109,29	136,32	74,90	90,05	99,86
	En kötü	73,17	829,45	137,22	170,92	104,75	113,55	112,01

Sürekli veri kümelerinde yeni nesil metasezgisel algoritmaların üstünlükleri aşağıdaki şekilde sıralanabilir:

- 1- Kümeleme için en önemli performans göstergesi olan uygunluk değerlerine göre Modifiye IMO her çalıştırmada en iyi değeri bulabilen tek algoritma olmuştur.
- 2- PSO en iyi olmakla birlikte, IMO ve WSA algoritması hesapsal zaman açısından PSO ile yarışabilir durumdadır.
- 3- Özellikle Modifiye2 WSA uygunluk değerleri açısından PSO ile yarışabilmektedir.

5.3.2.2 Kategorik Veri Kümeleri için Karşılaştırma

Kategorik veri kümelerinin kümelenmesinde uygunluk değerlerine göre sonuçlar Tablo 5.7’de verilmiştir.

Tablo 5.7 Kategorik veri kümelerinde algoritmaların uygunluk değerleri

Veri kümesi	Kriter	PSO	ABC	IMO	Modifiye IMO	WSA	Modifiye1 WSA	Modifiye2 WSA
cancer	Ortalama	2497	2497	2497	2497	2497	2497	2497
	En iyi	2497	2497	2497	2497	2497	2497	2497
	En kötü	2497	2497	2497	2497	2497	2497	2497
zoo	Ortalama	146,6	140	140	140	141,4	140	140,1
	En iyi	140	140	140	140	140	140	140
	En kötü	152	140	140	140	148	140	141
soybean	Ortalama	199	199	199	199	199	199	199
	En iyi	199	199	199	199	199	199	199
	En kötü	199	199	199	199	199	199	199
voting	Ortalama	1702	1701	1701	1701	1701	1701	1701
	En iyi	1701	1701	1701	1701	1701	1701	1701
	En kötü	1706	1701	1701	1701	1701	1701	1701

Tabloya bakıldığında algoritmaların birbirlerine belirgin bir üstünlük sağlamadığı görülmekle birlikte; PSO, WSA ve Modifiye2 WSA algoritmalarının her veri kümesi için her çalıştırmada en iyi sonucu bulamadığı görülmektedir.

Tablo 5.8’de kategorik veri kümelerinde algoritmaların hata oranlarına göre sonuçlar verilmiştir.

Tablo 5.8 Kategorik veri kümelerinde algoritmaların hata oranları

Veri kümesi	Kriter	PSO	ABC	IMO	Modifiye IMO	WSA	Modifiye1 WSA	Modifiye2 WSA
cancer	Ortalama	11,79	11,82	11,82	11,82	11,82	11,82	11,74
	En iyi	11,56	11,82	11,82	11,82	11,82	11,82	11,04
	En kötü	11,82	11,82	11,82	11,82	11,82	11,82	11,82
zoo	Ortalama	3,53	9,40	9,20	9,20	8,82	9,03	9,03
	En iyi	0,00	8,93	8,93	8,93	4,95	8,93	8,67
	En kötü	9,64	9,86	9,86	9,86	10,53	9,56	9,82
soybean	Ortalama	0,94	2,33	0,70	2,57	1,43	1,89	2,55
	En iyi	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	En kötü	4,63	4,63	4,63	4,63	4,63	4,63	4,63
voting	Ortalama	24,03	24,29	24,29	24,49	24,49	24,49	24,29
	En iyi	23,16	23,50	23,50	23,50	23,50	23,50	23,50
	En kötü	25,48	25,48	25,48	25,48	25,48	25,48	25,48

Tabloya bakıldığında, PSO algoritmasının iris veri kümesi dışında en iyi ve ortalama kriterler açısından diğer algoritmalarından daha az hata ile kümeleme yaptığı görülmektedir. Diğer algoritmaların ise birbirlerine belirgin bir üstünlükleri görülmemektedir.

Tablo 5.9’da doğruluk oranlarına göre kümeleme sonuçları verilmektedir. Tabloya bakıldığında, doğruluk oranlarında ortalamalar ve en iyi değerler açısından iris veri kümesi dışında en iyi performansı PSO algoritmasının gösterdiği ve iris veri kümesinde en iyi performansı Modifiye2 WSA’nın gösterdiği görülmektedir.

Tablo 5.9 Kategorik veri kümelerinde algoritmaların doğruluk oranları

Veri kümesi	Kriter	PSO	ABC	IMO	Modifiye IMO	WSA	Modifiye1 WSA	Modifiye2 WSA
cancer	Ortalama	93,72	93,70	93,70	93,70	93,70	93,70	93,75
	En iyi	93,85	93,70	93,70	93,70	93,70	93,70	94,14
	En kötü	93,70	93,70	93,70	93,70	93,70	93,70	93,70
zoo	Ortalama	96,83	94,46	95,05	94,75	94,36	94,85	93,70
	En iyi	100,00	95,05	95,05	95,05	95,05	95,05	93,70
	En kötü	94,06	94,06	95,05	92,08	92,08	93,07	92,08
soybean	Ortalama	99,15	98,72	99,36	97,66	98,72	98,30	97,66
	En iyi	100,00	100,00	100,00	100,00	100,00	100,00	100,00
	En kötü	95,75	95,75	95,75	95,75	95,75	95,75	95,75
voting	Ortalama	86,07	85,89	85,89	85,75	85,75	85,75	85,89
	En iyi	86,67	86,44	86,44	86,44	86,44	86,44	86,44
	En kötü	85,06	85,06	85,06	85,06	85,06	85,06	85,06

Tablo 5.10’da hesapsal zaman açısından kümeleme sonuçları yer almaktadır. Tabloya bakıldığında, en iyi değerler açısından cancer veri kümesinde IMO, zoo veri kümesinde Modifiye1 WSA, soybean veri kümesinde Modifiye2 WSA ve voting veri kümesinde PSO en iyi performansı göstermiştir. Ortalamalar açısından ise cancer veri kümesi dışında diğer üç veri kümesinde Modifiye2 WSA en iyi sonucu vermiştir. Cancer veri kümesinde ortalama açısından en iyi performansı IMO algoritmasının gösterdiği görülmektedir.

Tablo 5.10 Kategorik veri kümelerinde algoritmaların hesapsal zamanları

Veri kümesi	Kriter	PSO	ABC	IMO	Modifiye IMO	WSA	Modifiye1 WSA	Modifiye2 WSA
cancer	Ortalama	605,63	1560,99	535,25	631,43	613,81	603,01	611,41
	En iyi	594,71	1000,45	520,15	619,54	610,61	598,04	585,86
	En kötü	616,07	1678,56	600,79	657,88	631,22	634,87	655,85
zoo	Ortalama	589,72	4107,65	535,87	610,36	521,22	504,75	491,89
	En iyi	501,55	3856,88	478,44	570,73	479,03	474,53	482,02
	En kötü	625,36	4500,85	568,51	632,24	746,60	526,84	503,81
soybean	Ortalama	305,49	4094,38	303,42	351,30	291,92	282,54	279,54
	En iyi	298,94	3982,00	300,55	298,75	279,41	276,52	267,27
	En kötü	310,61	4127,08	308,95	392,75	301,47	299,48	289,66
voting	Ortalama	490,15	8500,25	519,03	509,65	497,59	485,27	483,67
	En iyi	478,62	5260,15	508,91	507,51	488,94	481,17	480,56
	En kötü	512,60	10250,65	541,02	513,76	518,81	496,87	486,88

Kategorik veri kümelerinde yeni nesil metasezgisel algoritmaların üstünlükleri aşağıdaki şekilde sıralanabilir:

- 1- WSA'nın modifiye algoritmaları hesapsal zaman açısından oldukça iyi performans göstermektedirler.
- 2- IMO hesapsal zaman açısından PSO ile yarışabilmektedir.
- 3- Uygunluk değerlerine göre her çalıştırmada en iyi değerleri veren dört algortmadan üçü IMO, Modifiye IMO ve Modifiye1 WSA algoritmalarıdır.

5.3.2.3 Karma Veri Kümeleri için Karşılaştırma

Karma veri kümeleri için yapılan kümelemede uygunluk değerleri Tablo 5.11'de verilmiştir.

Tablo 5.11 Karma veri kümelerinde algoritmaların uygunluk değerleri

Veri kümesi	Kriter	PSO	ABC	IMO	Modifiye IMO	WSA	Modifiye1 WSA	Modifiye2 WSA
Credit approval	Ortalama	1123,48	1212,09	1209,48	1210,36	1209,96	1210,37	1178,16
	En iyi	1109,21	1211,63	1208,13	1209,23	1206,83	1209,70	1166,78
	En kötü	1132,05	1212,22	1210,77	1212,07	1211,42	1210,67	1209,77
Statlog heart	Ortalama	416,579	419,41	405,29	416,50	418,31	427,86	419,66
	En iyi	416,381	418,14	392,24	416,41	417,56	422,84	417,47
	En kötü	416,389	422,61	415,75	416,89	419,17	434,69	422,91

Credit veri kümesinde en iyi ve ortalama değerler açısından en iyi sonucu PSO algoritması verirken, statlog heart veri kümesinde en iyi sonucu IMO algoritması vermiştir.

Hata oranlarına göre kümeleme sonuçları, Tablo 5.12’de verilmiştir.

Tablo 5.12 Karma veri kümelerinde algoritmaların hata oranları

Veri kümesi	Kriter	PSO	ABC	IMO	Modifiye IMO	WSA	Modifiye1 WSA	Modifiye2 WSA
Credit approval	Ortalama	30,45	44,00	42,76	42,61	43,30	42,96	39,17
	En iyi	27,23	42,74	42,18	42,29	42,74	42,85	38,06
	En kötü	33,42	44,36	43,29	42,74	44,12	43,07	41,25
Statlog heart	Ortalama	40,50	41,47	39,29	41,30	42,09	45,00	41,78
	En iyi	39,94	40,60	37,83	41,24	41,24	43,02	40,60
	En kötü	41,24	42,45	41,55	41,55	42,74	48,03	42,45

Uygunluk değerlerinde olduğu gibi, credit veri kümesinde en iyi ve ortalama değerler açısından en iyi sonucu PSO algoritması verirken, statlog heart veri kümesinde en iyi sonucu IMO algoritması vermiştir.

Tablo 5.13 doğruluk oranlarına göre kümeleme sonuçlarını göstermektedir.

Tablo 5.13 Karma veri kümelerinde algoritmaların doğruluk oranları

Veri kümesi	Kriter	PSO	ABC	IMO	Modifiye IMO	WSA	Modifiye1 WSA	Modifiye2 WSA
Credit approval	Ortalama	81,25	63,55	69,10	69,30	68,38	68,84	73,22
	En iyi	83,77	68,99	69,86	69,71	69,13	68,99	74,49
	En kötü	78,84	56,23	68,41	69,13	67,25	68,70	68,55
Statlog heart	Ortalama	71,93	62,74	73,26	71,04	70,07	65,78	70,44
	En iyi	72,59	69,26	74,82	71,11	71,11	68,89	71,85
	En kötü	71,11	60,00	70,74	70,74	69,26	60,37	69,69

Aynı şekilde, credit veri kümesinde en iyi ve ortalama değerler açısından en iyi sonucu PSO algoritması verirken, statlog heart veri kümesinde en iyi sonucu IMO algoritması vermiştir.

Tablo 5.14 hesapsal zaman açısından kümeleme kalitesini göstermektedir.

Tablo 5.14 Karma veri kümelerinde algoritmaların hesapsal zamanları

Veri kümesi	Kriter	PSO	ABC	IMO	Modifiye IMO	WSA	Modifiye1 WSA	Modifiye2 WSA
Credit approval	Ortalama	676,11	9032,51	711,71	698,60	694,64	694,69	665,53
	En iyi	667,88	8908,16	702,86	692,67	689,36	691,33	661,35
	En kötü	703,98	9214,90	742,82	703,26	706,66	700,61	667,55
Statlog heart	Ortalama	282,52	4978,20	298,88	320,14	284,62	292,07	289,90
	En iyi	280,82	3734,09	294,84	301,89	282,02	290,29	284,83
	En kötü	287,03	14993,90	309,09	333,56	290,52	293,25	295,73

Tabloya bakıldığında en yavaş algoritmanın ABC algoritması olduğu, credit veri kümesinde en iyi ve ortalamalar açısından en hızlı algoritmanın Modifiye2 WSA ve statlog heart veri kümesinde ise PSO olduğu görülmektedir.

Ayrıca, kümelemede hızlı ve etkin bir algoritma olan k -ortalamlar algoritmasının; sürekli, kategorik ve karma veri kümeleri için çalışmada kullanılan uzaklık ölçüleri kullanılarak MATLAB kodu yazılmış ve küme içi uzaklık değerlerinin Tablo 5.3, Tablo 5.7 ve Tablo 5.11 sonuçları ile karşılaştırılması Tablo 5.15’de verilmiştir.

Tablo 5.15’e bakıldığında k -ortalamlar ile elde edilen kümelemenin küme içi uzaklıkları toplamalarının statlog heart veri kümesi dışında metasezgisel algoritmalarından daha iyi olmadığı görülmektedir. Her ne kadar hızlı ve etkin bir algoritma olsa da metasezgiseller kümelemede k -ortalamlar tekniğinden küme içi uzaklıklar toplamı açısından daha iyi sonuçlar vermektedir.



Tablo 5.15 K-ortalamlar algoritması ile metasezgisel algoritmaların uygunluk değerlerinin karşılaştırılması

		K-ortalamlar	PSO	ABC	IMO	Modifiye IMO	WSA	Modifiye1 WSA	Modifiye2 WSA
iris	Ortalama	103,52	97,22	97,22	97,22	97,22	97,22	97,23	97,22
	En iyi	97,39	97,22	97,22	97,22	97,22	97,22	97,22	97,22
	En kötü	124,19	97,22	97,22	97,22	97,22	97,22	97,23	97,22
wine	Ortalama	23703,76	16530,50	16530,50	16530,50	16530,50	16530,50	16530,50	16530,50
	En iyi	22999,06	16530,50	16530,50	16530,50	16530,50	16530,50	16530,50	16530,50
	En kötü	24590,54	16530,50	16530,50	16530,50	16530,50	16530,50	16530,50	16530,50
thyroid	Ortalama	2086,5	1960,59	1961,11	1961,29	1960,59	1961,65	1961,25	1960,59
	En iyi	2086,5	1960,59	1960,59	1960,59	1960,59	1960,59	1960,59	1960,59
	En kötü	2086,5	1960,59	1961,75	1962,98	1960,59	1962,78	1962,57	1960,59
glass	Ortalama	249,95	213,71	213,22	213,58	213,2	214,27	213,24	213,52
	En iyi	233,35	213,2	213,2	213,38	213,2	213,22	213,2	213,2
	En kötü	275,37	218,23	213,22	214,16	213,2	216,11	213,42	215,48
cancer	Ortalama	2577	2497	2497	2497	2497	2497	2497	2497
	En iyi	2497	2497	2497	2497	2497	2497	2497	2497
	En kötü	2772	2497	2497	2497	2497	2497	2497	2497
zoo	Ortalama	164	146,6	140	140	140	141,4	140	140,1
	En iyi	141	140	140	140	140	140	140	140
	En kötü	178	152	140	140	140	148	140	141
soybean	Ortalama	224	199	199	199	199	199	199	199
	En iyi	199	199	199	199	199	199	199	199
	En kötü	251	199	199	199	199	199	199	199
voting	Ortalama	1704,5	1702	1701	1701	1701	1701	1701	1701
	En iyi	1701	1701	1701	1701	1701	1701	1701	1701
	En kötü	1706	1706	1701	1701	1701	1701	1701	1701
credit	Ortalama	1145,95	1123,48	1212,09	1209,48	1210,36	1209,96	1210,37	1178,16
	En iyi	1109,21	1109,21	1211,63	1208,13	1209,23	1206,83	1209,7	1166,78
	En kötü	1192,66	1132,05	1212,22	1210,77	1212,07	1211,42	1210,67	1209,77
heart	Ortalama	404,54	416,579	419,41	405,29	416,5	418,31	427,86	419,66
	En iyi	388,37	416,381	418,14	392,24	416,41	417,56	422,84	417,47
	En kötü	424,34	416,389	422,61	415,75	416,89	419,17	434,69	422,91

5.3.2.4 Algoritmaların İstatistiksel Olarak Karşılaştırılması

5.3.2.4.1 Performans Kriterleri Açısından Algoritmaların İkili Karşılaştırması

Wilcoxon işaret testinden elde edilen p değerleri Tablo 5.16’da verilmiştir. $\alpha = 0,05$ anlamlılık düzeyinde istatistiksel analizin yorumları aşağıdaki gibi yapılabilir:

Tablo 5.16 Wilcoxon işaret testinden elde edilen p değerleri

Algoritmalar	Uygunluk değeri	Hata oranı	Doğruluk oranı	Hesapsal zaman
PSO-ABC	0,830	0,042	0,032	0,006
PSO-IMO	0,675	0,529	0,726	0,610
PSO-Hybrid IMO	0,590	0,022	0,052	0,006
PSO-WSA	0,529	0,009	0,108	0,221
PSO-Modified1 WSA	0,673	0,018	0,080	0,760
PSO-Modified2 WSA	1,000	0,076	0,108	0,838
ABC-IMO	0,584	0,059	0,014	0,006
ABC-Hybrid IMO	0,100	0,933	0,193	0,006
ABC-WSA	1,000	0,834	0,059	0,006
ABC-Modified1 WSA	0,418	1,000	0,124	0,006
ABC-Modified2 WSA	0,787	0,554	0,076	0,006
IMO-Hybrid IMO	0,584	0,142	0,108	0,041
IMO-WSA	0,059	0,08	0,108	0,476
IMO-Modified1 WSA	0,59	0,042	0,141	0,476
IMO-Modified2 WSA	0,787	0,673	1,000	0,185
Hybrid IMO-WSA	0,106	0,800	1,000	0,041
Hybrid IMO-Modified1 WSA	0,059	0,673	1,000	0,025
Hybrid IMO-Modified2 WSA	0,855	0,205	0,208	0,006
WSA-Modified1 WSA	1,000	0,529	0,834	0,262
WSA-Modified2 WSA	0,418	0,236	0,205	0,041
Modified1 WSA-Modified 2 WSA	0,295	0,107	0,441	0,041

- Uygunluk değerleri açısından algoritmalar arasında anlamlı bir fark olmadığı görülmektedir.
- PSO ve ABC arasında PSO lehine; hata oranı, doğruluk ve hesapsal zaman açısından anlamlı bir fark olduğu görülmektedir.
- PSO ve Modifiye IMO arasında PSO lehine; hata oranı ve hesapsal zaman açısından anlamlı bir fark olduğu görülmektedir.
- PSO ile WSA, Modifiye1 WSA arasında PSO lehine; hata oranı açısından anlamlı bir fark görülmektedir.
- ABC ve IMO arasında IMO lehine, doğruluk ve hesapsal zaman açısından anlamlı bir fark görülmektedir.
- Modifiye IMO ve WSA arasında WSA lehine, hesapsal zaman açısından anlamlı bir fark görülmektedir.
- Modifiye IMO ve Modifiye1 WSA arasında Modifiye1 WSA lehine, hesapsal zaman açısından anlamlı bir fark görülmektedir.
- Modifiye IMO ve Modifiye2 WSA arasında Modifiye2 WSA lehine, hesapsal zaman açısından anlamlı bir fark görülmektedir.
- WSA ve Modifiye2 WSA arasında Modifiye2 WSA lehine, hesapsal zaman açısından anlamlı bir fark görülmektedir.
- ABC ile Modifiye IMO, WSA, Modifiye1 WSA ve Modifiye2 WSA arasında hesapsal zaman açısından anlamlı bir fark olduğu görülmektedir.
- Modifiye1 WSA ile Modifiye2 WSA arasında Modifiye2 WSA lehine, hesapsal zaman açısından anlamlı bir fark görülmektedir.

- IMO ile Modifiye IMO arasında ise IMO lehine, hesapsal zaman açısından anlamlı bir fark olduğu görülmektedir.

Bütün bu değerlendirmeler ışığında;

- PSO algoritmasının Modifiye2 WSA ve IMO algoritmalarına herhangi bir üstünlüğü bulunmadığı görülmektedir. Yani kümelemede bu yeni nesil metasezgiseller PSO ile yarışabilmektedir.
- PSO, Modifiye IMO algoritmasına hata oranı ve hesapsal zaman açısından üstünlük sağlasa da uygunluk değerleri açısından Modifiye IMO'nun sürekli ve kategorik veri kümelerinde her çalıştırmada en iyi sonucu bulmasına karşın PSO'nun bulamadığı görülmektedir.
- PSO algoritmasının WSA ve Modifiye1 WSA algoritmasına hata oranı dışında herhangi bir üstünlüğü görülmemektedir. WSA algoritmalarının PSO ile kümeleme probleminde rekabet edebilir bir algoritma olduğu söylenebilir.
- PSO, ABC algoritmasına genel olarak üstünlük sağlayan bir algoritmadır.
- ABC algoritmasının yeni nesil metasezgisel algoritmalarla herhangi bir üstünlüğü bulunmamakla birlikte, yeni nesil metasezgisel algoritmaların ABC algoritmasına özellikle hesapsal zaman açısından üstünlükleri olduğu görülmektedir.

5.3.2.4.2 Algoritmaların Performans Kriterlerinin Eşzamanlı İstatistiksel Karşılaştırması

Kümeleme geçerlilik iç indekslerinden küme merkezlerine uzaklıkların toplamını ifade eden uygunluk değeri ve algoritmaların hesapsal zamanlarının ortalama değerlerini aynı amaç doğrultusunda birleştirerek daha tutarlı bir karşılaştırma yapmak için minimum sapma metodu (MDM) kullanılmıştır. MDM metodu, tekil amaçların kısmi sapmalarının toplamını minimize eden en iyi uzlaşlı çözümünü bulmayı amaçlar ve ele

alınan ve minimize edilen uygunluk değeri ve hesapsal zamanı olmak üzere iki kriter açısından Eşitlik (5.21) ile hesaplanır.

$$Uzlaş\ değeri = \frac{f_1^{min}(UD) - f_1(UD)}{f_1^{min}(UD) - f_1^{max}(UD)} + \frac{f_2^{min}(HZ) - f_2(HZ)}{f_2^{min}(HZ) - f_2^{max}(HZ)} \quad (5.21)$$

Burada;

f_1^{min} , 1.kriter (uygunluk değeri) açısından ilgili veri kümesinde algoritmalarından elde edilen değerlerin minimumu.

f_1^{max} , 1. kriter (uygunluk değeri) açısından ilgili veri kümesinde algoritmalarından elde edilen değerlerin maksimumu.

$f_1(UD)$ 1. kriter (uygunluk değeri) açısından ilgili algoritmadan elde edilen değer.

UD uygunluk değerini ve HZ hesapsal zamanı ifade etmektedir.

Elde edilen uzlaş\ değerleri Tablo 5.17’de verilmiştir.

Tablo 5.17 Algoritmaların uzlaş\ değerleri

	PSO	ABC	IMO	Modifiye IMO	WSA	Modifiye WSA1	Modifiye WSA2
iris	0,0010	1,0000	0,0101	0,0133	0,0224	1,0000	0,0045
wine	0,0000	1,0000	0,0041	0,0323	0,0317	0,0405	0,0115
thyroid	0,0000	1,4906	0,6645	0,0210	1,0382	0,6586	0,0085
glass	0,4772	1,0152	0,4249	0,1018	1,0352	0,0850	0,3453
cancer	0,0686	1,0000	0,0000	0,0938	0,0766	0,0661	0,0743
zoo	1,0271	1,0000	0,0122	0,0328	0,2202	0,0036	0,0152
soybean	0,0068	1,0000	0,0063	0,0188	0,0032	0,0008	0,0000
voting	1,0008	1,0000	0,0044	0,0032	0,0017	0,0002	0,0000
credit approval	0,0013	2,0000	0,9761	0,9844	0,9794	0,9841	0,6171
statlog heart	0,5002	1,6256	0,0035	0,5049	0,5772	1,0020	0,6382
TOPLAM	3,0830	12,1314	2,1060	1,8064	3,9859	3,8409	1,7145

Wine, cancer ve soybean veri kümelerinde tüm algoritmalar aynı en iyi sonucu bulduğu için Eşitlik (5.21)’nin ilk bölümünde payda kısmı “0” değerini alacaktır. Bu nedenle, ilgili veri kümesinin minimum ve maksimum uygunluk değerleri arasındaki farkı ifade eden payda kısmının bu üç veri kümesinde hesaplanmasında k-ortalamlar algoritmasının bu üç veri kümesinde bulunduğu ortalama değer kullanılmıştır.

Uzlaş ı değ erleri, algoritmaların performansları açısından uygunluk değ erlerinin ve hesapsal zamanlarının eş zamanlı dikkate alınmasını ifade etmekte olup, Tablo 5.17’deki toplam değ erlere göre performans sıralaması aşağıdaki şekilde ortaya çı kmaktadır;

- 1- Modifiye WSA2
- 2- Modifiye IMO
- 3- IMO
- 4- PSO
- 5- Modifiye WSA1
- 6- WSA
- 7- ABC

Algoritmalar uzlaş ı değ erleri açısından ikili karşı laşt ırmaya tabi tutulmuş ve Wilcoxon iş aret testi uygulanmı ş tır. Elde edilen p değ erleri Tablo 5.18’de verilmi ş tir.

Tablo 5.18 Uzlaş ı değ erleri açısından algoritmaların p değ erleri

Algoritmalar	p değ eri
PSO-ABC	0,014
PSO-IMO	0,476
PSO-Modifiye IMO	0,919
PSO-WSA	0,308
PSO-Modifiye WSA1	0,919
PSO-Modifiye WSA2	1,000
ABC-IMO	0,006
ABC-Modifiye IMO	0,006
ABC-WSA	0,008
ABC-Modifiye WSA1	0,009
ABC-Modifiye WSA2	0,006
IMO-Modifiye IMO	0,415
IMO-WSA	0,014
IMO-Modifiye WSA1	0,415
IMO-Modifiye WSA2	0,541
Modifiye IMO-WSA	0,308
Modifiye IMO-Modified WSA1	0,838
Modifiye IMO-Modifiye WSA2	0,308
WSA-Modifiye WSA1	0,838
WSA-Modifiye WSA2	0,032
Modifiye WSA1-Modifiye WSA2	0,154

$\alpha = 0,05$ anlamlılık düzeyinde istatistiksel analizin yorumları aşağıdaki gibi yapılabilir:

- ABC ve diğer algoritmalar arasında diğer algoritmaların lehine anlamlı bir fark vardır.
- IMO ile WSA arasında, IMO lehine anlamlı bir fark vardır.
- WSA ve Modifiye WSA2 arasında, Modifiye WSA2 lehine anlamlı bir fark vardır.

Kategorik verilerden edilen sonuçlara göre, gerçek hayat verisi olan Yaşam Memnuniyet Anketi verileri üzerinde kümeleme yapmak için uygunluk değerlerine ve hesapsal zamana göre en iyi algoritmalar olan Modifiye IMO ve Modifiye1 WSA algoritmaları seçilmiştir. Gerçek hayat verisi kategorik ve örnek hacmi büyük bir veri kümesi olduğundan kategorik verilerde hızlı ve etkin uygunluk değeri veren algoritmalar seçilmiştir. Modifiye2 WSA, Modifiye1 WSA'ya göre daha hızlı sonuçlar vermesine rağmen Modifiye1 WSA'nın uygunluk değerlerine göre önde olduğu görülmektedir. Modifiye IMO ise gerek sürekli veri kümelerinde gerekse kategorik veri kümelerinde hep en iyi değerleri her çalıştırmada bulmuştur. Sağlam bir algoritma olduğu öngörüldüğü için kategorik veriye sahip gerçek hayat verisinin kümelemesinde tercih edilmiştir.

5.4 Sonuç

Bu bölümde, tezde kullanılan metasezgisel algoritmaların sürekli, kategorik ve karma veri kümelerinde kümelemeye uygulama şekilleri açıklanmıştır. Yapılan deneysel çalışmadan ve uygulama sonuçlarından bahsedilmiştir. Gerçek hayat verisinin kümelemesinde kullanılmak üzere algoritmalar arasından Modifiye IMO ve Modifiye1 WSA seçilmiştir.

Metasezgisel algoritmaların kümeleme problemindeki performansları yaygın olarak literatürde incelenmektedir. Ancak halen parametre sayısı az, hesapsal zamanı kısa olan algoritma arayışı devam etmektedir. Bu çalışma ile IMO gibi diğer birçok metasezgiselle göre daha az parametreye sahip olan bir algoritmanın ve hesapsal zaman açısından

önemli ölçüde hızlı olduğu görülen WSA algoritmasının kümeleme problemindeki performansları değerlendirilmiştir. Çalışmanın orjinalliği bu algoritmaların sadece kümeleme problemine uygulanmasında değil, ayrıca sürekli, kategorik ve karma veri kümelerinin tamamındaki kümeleme performanslarının ele alınmasındadır.



6. BÖLÜM

YAŞAM MEMNUNİYET ARAŞTIRMASI ÜZERİNE VERİ MADENCİLİĞİ UYGULAMALARI

6.1 Giriş

Bu bölümde; gerçek hayat verisi olan YMA üzerinde gerçekleştirilen VM uygulamaları üzerinde durulacaktır. YMA üzerinde ayrıntılı bilgi verilecek, özellikle YMA üzerinde gerçekleştirilen kümeleme uygulamaları ve sonuçlarından bahsedilecektir.

6.2 Yaşam Memnuniyet Araştırması

Türkiye’de yaşayan insanların yaşam memnuniyetlerini ölçmek için 2003 yılından itibaren TÜİK tarafından düzenli olarak yapılan Yaşam Memnuniyeti Araştırması, Türkiye’de yaşayan insanların sosyal, kültürel ve ekonomik karakteristikleri açısından zengin bir bilgi kaynağıdır. Memnuniyet araştırması iki ayrı anketten oluşmaktadır. Hane halkı anketinde, hanenin genel geçim koşulları, eğitim hizmetlerinden memnuniyeti, sahip oldukları işyerleri ve yaşadıkları asayiş sorunları ile ilgili sorular vardır. Bireysel mutluluğu ölçen ankette ise bireylerin mutluluk, umut, kamu hizmetlerinden memnuniyet, beklenti, kişisel gelişim, değerler ve Avrupa Birliği’ne bakışını değerlendiren sorular mevcuttur.

Bu çalışmada, bireysel mutluluk ve umut düzeyini ölçen anket değerlendirilmiştir. YMA’nın 2010-2014 yılları arasındaki verileri TÜİK’ten temin edilmiştir. Veri kümelerinin özellikleri Tablo 6.1’de verilmiştir.

Tablo 6.1 YMA veri kümelerinin özellikleri

Anket yılı	Özellik sayısı	Sürekli özellik sayısı	Kategorik özellik sayısı	Örnek sayısı
2010	225	2	223	7027
2011	237	2	235	7368
2012	223	2	221	7956
2013	239	2	237	196203
2014	238	2	236	7984

Mevcut verilerdeki en son yıla ait veri 2014 verisidir. 2014 yılına ait ankette yer alan özellikler, özelliklerin değer aralığı ve değişken tipi Ekler bölümünde Ek-1 olarak verilmiştir.

6.3 Yaşam Memnuniyet Araştırması Üzerinde Yapılan Veri Madenciliği Uygulamaları

YMA veri kümelerinde sırasıyla aşağıdaki VM teknikleri uygulanmıştır:

6.3.1 Veri Ön İşleme

Yapılan veri ön işleme çalışmaları aşağıda sıralanmıştır:

- a. Hane anketinden bireylerin ailelerine sorulan sorulardan aşağıda verilenler bireysel mutluluğu ölçen anketin veri kümesine eklenmiştir. Bunlar;
 - i. HH BÜYÜKLÜĞÜ: Hanede kaç kişi yaşadığını ifade etmektedir. 1-17 arasında kategorik değerler almaktadır.
 - ii. GELİR GRUBU: Ailenin gelir durumunu ifade etmektedir. 0-1179 TL, 1180-1692 TL, 1693-2369 TL, 2370-3471 TL ve 3471’den fazla olmak üzere 5 kategorik değer içermektedir.
 - iii. OLCEK_GELİR-KARŞILAMA: Hane gelirinin temel ihtiyaçları karşılamadaki durumu ile ilgilidir. Çok kolay, Kolay, Orta, Zor ve Çok zor olmak üzere 5 kategorik değer içermektedir.
- b. Cevapsız bırakılan özellikler ve sorulara verilen cevaplar arasındaki tezat ve aykırılıkların olduğu örnekler Excel programı yardımıyla tespit edilerek veri kümesinden çıkarılmıştır. Mesela, 2010 yılı veri kümesinde “çalışmıyor” olduğu halde çalışılan işle ilgili sorulara cevap veren 14 örnek, veri

kümesinden çıkarılmıştır. 2012 yılında 13 örnek, geliri olduğu halde bazı gelir sorularına cevap vermediği için veriden çıkarılmıştır. Aynı veri kümesinde 3 kişi ücretli çalışan olduğu halde kazançtan memnuniyet ile ilgili soruya “parasal kazancı yok” olarak cevap verdiği için veriden çıkarılmıştır. 2013 yılında bir okul bitirdiği halde, “şimdiye kadar alınan eğitimden memnuniyet” sorusuna “eğitim almadı” olarak cevap veren 675 kişi veriden çıkarılmıştır. Aynı yıl “Ülkemizin önümüzdeki 5 yılda hangi konularda değişim göstereceği” sorusuna 135.327 kişi cevap vermediği için bu özellik veriden çıkarılmıştır.

- c. BİRİMNO, REFERANS YILI, FERTNO ve FERT FAKTÖRÜ özellikleri çalışmanın amacı açısından anlamsız özellikler olduğundan veriden çıkarılmıştır.
- d. CALISMA DURUMU, CALISMAMA NEDENI ve ÖZEL-KAMU sorularını temsil eden özellikler ortak bilgi içerdiğinden CALISMA DURUMU özelliği adı altında birleştirilmiştir. OKUMA-YAZMA sorusu ile OKUL BITEN sorusu ortak bilgi içerdiğinden OKUL BITEN başlığı altında birleştirilmiştir. Belediye veya İl Özel İdare Hizmetleri’nden memnuniyet kategorilerindeki sorular ortak bilgi içerdiğinden birleştirilmiştir.
- e. Herhangi bir soruda cevapsız bırakılma sayısı toplam kişi sayısının önemli bir kısmını oluşturuyorsa, o soru veri kümesinden çıkarılmıştır. Mesela, 2010 yılında “üniversiteye giriş sistemi ile ilgili düşünceler” kategorisindeki sorularda 7027 kişinin 5045’i soru sadece lise ve üzeri eğitim alanlara sorulduğu için bu soruya ya muhatap olmamış ya da cevap vermemiştir. Ayrıca; KAPKAC, GASP, DARP, FERT KOTU, TEHDIT, MAGDURIYET, DOLANDIRICILIK, BASKASUC sorularına verilen cevaplar çoğunlukla “Hayır” şeklinde olduğu ve UMUT değişkenini doğrudan etkileyen sorular olmadığı için tüm yıllarda bu sorular veri kümesinden çıkarılmıştır.
- f. Verilerdeki bazı boşluklar anketin mantığına uygun bir şekilde doldurulmuştur. Mesela; OLCEK MEMNUNİYET KAZANC sorusundaki boşluklara, ilgili kişilerin CALISMA DURUMU’na “çalışmıyor” olarak cevap vermesi nedeniyle “çalışmıyor” değeri atanmıştır.

g. Sonuçta;

- i. 2010 verisi; 168 özellik ve 6967 adet örneğe,
- ii. 2011 verisi; 174 özellik ve 7346 adet örneğe,
- iii. 2012 verisi; 173 özellik ve 7938 adet örneğe,
- iv. 2013 verisi; 193 özellik ve 195528 örneğe düşürülmüştür.
- v. 2014 verisi ise 190 özelliğe indirgenmiş ve örnek sayısı değişmeyerek 7984 adet örnek olarak kalmıştır.

6.3.2 Özellik Seçimi

Özellik seçimi metodu, insanların *umut* düzeylerini etkileyen faktörlerin tespiti için her yıla ait veri kümesine ayrı ayrı uygulanmıştır. WEKA 3.7.13 paket programı kullanılarak hedef değişkenle korelasyonu yüksek özellikler arasından bir alt küme ortaya koyduğu için CFS özellik seçimi algoritması, her bir veri kümesinde kullanılarak her yıl *umut* düzeyini etkileyen faktörler tespit edilmiştir. CFS, sınıf değişkeni ile yüksek korelasyonlu ve birbirleriyle düşük korelasyonlu özellikleri içeren alt kümeler oluşturmaktadır. Özellikler ile sınıf arasındaki korelasyon hesaplanır, her özelliğin diğer özellikler arasındaki korelasyonu hesaplanır ve böylece elde edilen özellik alt kümelerinin sınıf ile arasındaki korelasyon elde edilir. En yüksek korelasyona sahip özellik alt kümesi seçilir.

Yapılan özellik seçimi ile tüm yıllarda ortak olarak seçilen özellikler Tablo 6.2’de gösterilmiştir.

Tablo 6.2 Tüm yıllarda özellik seçimi ile seçilen ortak özellikler

SEÇİLEN ÖZELLİKLERİN İSİMLERİ	
1	OLCEK GELİR KARŞILAMA
2	OLCEK_MEMNUNİYET_SAGLIK
3	OLCEK_MEMNUNİYET_EVLİLİK
4	OLCEK_MEMNUNİYET_KONUT
5	OLCEK_MEMNUNİYET_GELİR
6	OLCEK_MEMNUNİYET_GN_SAG
7	GUVEN EV
8	GUVEN CEVRE
9	UMUT_BASAMAK
10	GECEKIS_KARSILASTIRMA
11	GELECEK_KARSILASTIRMA
12	GELECEK_HAYAT
13	GELECEK_MALI
14	GELIRI_AZALDI

Tablo 6.2'deki ortak özellikler dışında özellik seçimi ile seçilmiş olup sadece bazı yıllarda ortak olan ve kaç yıl için ortak özellik olarak seçildikleri veya sadece tek bir yıl seçilen özelliklerin bilgisi Tablo 6.3'de verilmiştir.

Tablo 6.3 Yıllara göre ortak olarak belirlenen özellikler

Bazı Yıllarda Ortak Olan Özellikler	Özelliklerin Açıklamaları (Ek-1’de açıklaması olanlar boş bırakılmıştır)	Seçildiği yıl sayısı
OLCEK_MEMNUNİYET_AKRABA		4
OLCEK_MEMNUNİYET_EGİTİ		3
OLCEK_MEMNUNİYET_EĞİTİM		3
OLCEK_MEMNUNİYET_IS		2
PARA	Toplumda itibarlı olmak için para önemli midir?	2
OLCEK_MEMNUNİYET_SEMT		2
OLCEK_MEMNUNİYET_SGK		2
ULASIM_KARAYOLU		2
OLCEK_MEMNUNİYET_SOS_HAYAT		2
OLCEK_MEMNUNİYET_KİSİSEL_BKM		2
GELİR DÜZEY DURUM	Gelir düzeyinden dolayı hangi sıklıkla baskı hissedildiği	2
DUNYA SAYGINLIĞI		1
EVDEKİLERLE İLİSKİLERDEN MEMNUNİYET	Ev fertleriyle ilişkiden memnuniyet	1
OLCEK_MEMNUNİYET_KAZANC		1
OLCEK_MEMNUNİYET_ARKADAS		1
ULASTIRMA MEMNUNİYET	Ulaştırma hizmetlerinden memnuniyet	1
TÜRKİYE İŞ ÇALIŞMA	Gelecek yıl Türkiye'deki iş/çalışma durumu nasıl olacak?	1
ASAYİŞ MEMNUNİYET	Asayiş hizmetlerinden memnuniyet	1
İLAC ALIMI	İlaç alımı ile ilgili işlemlerde sorun var mı?	1
SIRA BEKLEME	Muayene ve tahlil için sıra beklemede sorun var mı?	1
2012 TÜRKİYE EKONOMİK	Gelecek yıl Türkiye'nin ekonomik durumu nasıl olacak?	1
UCUZ URUN	Son bir yılda daha ucuz ürünlere yönelme durumu	1
BORÇLANDI	Son bir yılda borçlanma durumu	1
EKONOMİK ACIDAN	Türkiye önümüzdeki 5 yıl ekonomik açıdan nasıl olacak	1

6.3.3 Kümeleme

Tezin beşinci bölümünde yeni nesil metasezgisel algoritmalarından IMO, WSA, Modifiye IMO, Modifiye1 WSA ve Modifiye2 WSA ile diğer iki algoritma olan PSO ve ABC'nin kümeleme problemine uygulanması ve sonuçları tartışılmıştır. Elde edilen sonuçlara göre beşinci bölümde, Modifiye IMO ve Modifiye1 WSA algoritmalarının gerçek hayat verisi olan YMA'nın indirgenmiş verisine uygulanmasına karar verilmiştir. Tüm yıllarda ortak 14 özellik açısından tüm yılların indirgenmiş veri kümeleri üzerinde bu iki algoritma ile kümeleme gerçekleştirilmiş ve sonuçları Tablo 6.4'de verilmiştir.

Tablo 6.4 YMA kümeleme sonuçları

Veri kümeleri	Kriter	Modifiye IMO				Modifiye1 WSA			
		Uygunluk değeri	Hata oranı	Doğruluk oranı	Hesapsal zaman (sn)	Uygunluk değeri	Hata oranı	Doğruluk oranı	Hesapsal zaman
2010 veri kümesi	Ortalama	36913,10	46,91	66,32	8596,40	36987,40	46,97	66,35	8350,40
	En iyi	36844,00	46,23	66,65	8453,40	36844,00	46,60	67,12	8320,60
	En kötü	37032,00	47,68	66,16	9306,50	37033,00	47,27	66,16	8439,60
2011 veri kümesi	Ortalama	38902,20	47,69	67,85	9203,20	38900,40	47,79	67,85	9006,10
	En iyi	38772,00	47,23	67,85	9165,90	38853,00	47,00	67,85	8980,50
	En kötü	39488,00	48,75	67,85	9260,60	38953,00	48,35	67,85	9049,90
2012 veri kümesi	Ortalama	41269,50	48,64	70,38	9912,40	41207,60	48,45	70,34	9787,90
	En iyi	41089,00	48,37	70,55	9850,50	41100,00	47,96	70,47	9762,00
	En kötü	41569,00	49,18	70,31	9972,10	41343,00	49,01	70,31	9862,60
2013 veri kümesi	Ortalama	1015064,0	40,65	71,35	13590,0	1015755,0	40,56	71,35	120354,0
	En iyi	1013581,0	40,65	71,35	12528,0	1015229,0	40,56	71,35	118237,0
	En kötü	1015229,0	40,65	71,35	15550,0	1018469,0	40,56	71,35	125468,0
2014 veri kümesi	Ortalama	41433,40	48,30	68,56	9981,10	41348,80	48,29	68,56	10199,30
	En iyi	41267,00	47,86	68,56	9926,41	41304,00	47,93	68,56	9750,11
	En kötü	41727,00	48,69	68,56	10032,40	41455,00	48,56	68,56	14096,40

Kümeleme danışmansız bir öğrenme tekniği olduğu için sınıf etiket bilgisi kümeleme yaparken kullanılmamaktadır. Ancak çalışmada kümeleme yaparken amaçlanan, hedef değişken açısından kümelerin ayırım noktalarını bulabilmek olduğu için, 4 kategorik değere (çok umutlu, umutlu, umutlu değil, hiç umutlu değil) sahip olan *umut* hedef değişkeninin sınıf sayısı, algoritmalara küme sayısı parametre değeri olarak verilmiştir. Veri kümeleri 4 kümeye ayrılırsa meydana gelecek gruplamanın özelliklerinin ortaya

çıkarılması amaçlanmıştır. Tablo 6.4'te görüldüğü üzere uygunluk değerlerinin en iyileri açısından performansı daha yüksek olan algoritmanın Modifiye IMO olduğu görülmektedir. Bu nedenle, bu algoritmanın kümeleme sonuçlarının yorumu yapılmıştır. Modifiye IMO algoritmasının tüm yıllara ait veri kümeleri için oluşturduğu küme merkezleri sırasıyla; Tablo 6.5, Tablo 6.6, Tablo 6.7, Tablo 6.8 ve Tablo 6.9'da verilmiştir.

Tablo 6.5 2010 YMA veri kümesi küme merkezleri

Özellik ismi	1. Küme Merkezi	2. Küme Merkezi	3. Küme Merkezi	4. Küme Merkezi
OLCEK GELİR KARŞILAMA	Orta	Orta	Zor	Zor
OLCEK_MEMNUNİYET_SAGLIK	Memnun	Memnun	Memnun	Memnun
OLCEK_MEMNUNİYET_EVLİLİK	Memnun	Memnun	Memnun	Memnun
OLCEK_MEMNUNİYET_KONUT	Memnun	Memnun	Memnun	Memnun
OLCEK_MEMNUNİYET_GELİR	Orta	Memnun	Memnun Değil	Memnun Değil
OLCEK_MEMNUNİYET_GN_SAG	Memnun	Memnun	Memnun	Memnun
GUVEN EV	Güvenli	Güvenli	Güvenli	Güvenli
GUVEN CEVRE	Güvenli	Güvenli	Güvenli	Güvenli
UMUT_BASAMAK	Duzey 6	Duzey 6	Duzey 5	Duzey 4
GECMIS_KARSILASTIRMA	Aynı Seviyede Kaldı	Gelişti	Aynı Seviyede Kaldı	Geriledi
GELECEK_KARSILASTIRMA	Aynı Seviyede Kalacak	Gelişecek	Aynı Seviyede Kalacak	Gerileyecek
GELECEK_HAYAT	Aynı Olacak	Daha iyi olacak	Aynı Olacak	Daha Kötü Olacak
GELECEK_MALI	Aynı Olacak	Daha iyi olacak	Aynı Olacak	Daha Kötü Olacak
GELIRI_AZALDI	Hayır	Hayır	Hayır	Evet

Tablolarda kalın (bold) olarak işaretlenen değerler veri kümesinin hangi özellikler açısından kümeler arasında farklılıkların olduğunu göstermektedir. Tablo 6.5

incelendiğinde, kümelerin birbirinden özellikle; “Umut Basamağı (refah düzeyi)” özelliği ile “Gelecek Mali” özelliği arasındaki özelliklere ek olarak “Ölçek Gelir Karşılaması”, “Ölçek Memnuniyet Gelir” ve “Geliri Azaldı” özellikleri açısından ayrıldığı görülmektedir.

Kümelerin tamamında insanların, sağlık, evlilik, konut ve sağlık hizmetleri açısından memnuniyet duydukları ve evde ve bulundukları çevrede kendilerini güvende hissettikleri görülmektedir. Kümeleri birbirinden ayıran esas özellikler açısından kümelerin tarifi şöyle yapılabilir:

1.kümedeki insanlar, gelirlerinin ihtiyaçlarını orta düzeyde karşıladığını ve gelirlerinden orta derecede memnun olduklarını söyleyen, Türkiye’de yaşayan insanların refah düzeylerinin 1-11 arasında olduğu düşünüldüğünde, kendilerinin 6. düzeyde olduklarını, geçmiş 5 yıla göre bugünkü durumlarının maddi-manevi açıdan aynı kaldığını ve gelecek 5 yıl ve gelecek yıl da durumlarının aynı kalacağını düşünen ve geçmiş yıla göre gelirlerinin azalmadığını ifade eden insanlardır.

2.kümedeki insanlar, birinci kümeden farklı olarak; gelirlerinden memnun, geçmiş 5 yıla göre durumlarının daha iyi olduğunu ve gelecek 5 yıl ve gelecek yıl da daha iyi olacağını düşünen insanlardır.

3.kümedeki insanlar; gelirlerinin ihtiyaçlarını zor karşıladığını ve gelirlerinden memnun olmadıklarını belirten, 5. refah düzeyinde olduklarını, geçmiş 5 yıla göre bugün aynı durumda olduklarını ve gelecek 5 yıl ve gelecek yıl da aynı olacaklarını ve geçmiş yıla göre gelirlerinin azalmadığını ifade eden insanlardır.

4.kümedeki insanlar, üçüncü kümeden farklı olarak 4. refah düzeyinde olduklarını, geçmiş 5 yıla göre bugün daha kötü durumda olduklarını ve gelecek 5 yıl ve gelecek yıl da daha kötü durumda olacaklarını ve geçmiş yıla göre gelirlerinde azalma olduğunu ifade eden insanlardır.

Buna göre, 1. kümedeki insanların “Umutlu”, 2. kümedeki insanların “Çok umutlu”, 3. kümedeki insanların “Umutlu değil” ve 4. kümedeki insanların ise “Hiç umutlu değil” kategorilerine dahil oldukları söylenebilir.

Tablo 6.6 2011 YMA veri kümesi küme merkezleri

Özellik ismi	1. Küme Merkezi	2. Küme Merkezi	3. Küme Merkezi	4. Küme Merkezi
OLCEK_GELİR_KARŞILAMA	Orta	Orta	Zor	Zor
OLCEK_MEMNUNİYET_SAGLIK	Memnun	Memnun	Memnun	Memnun
OLCEK_MEMNUNİYET_EVLİLİK	Memnun	Memnun	Memnun	Memnun
OLCEK_MEMNUNİYET_KONUT	Memnun	Memnun	Memnun	Memnun
OLCEK_MEMNUNİYET_GELİR	Orta	Memnun	Memnun Değil	Memnun Değil
OLCEK_MEMNUNİYET_GN_SAG	Memnun	Memnun	Memnun	Memnun
GUVEN_EV	Güvenli	Güvenli	Güvenli	Güvenli
GUVEN_CEVRE	Güvenli	Güvenli	Güvenli	Güvenli
UMUT_BASAMAK	Düzey 6	Düzey 6	Düzey 4	Düzey 5
GECEMIS_KARSILASTIRMA	Aynı Seviyede Kaldı	Gelişti	Geriledi	Aynı Seviyede Kaldı
GELECEK_KARSILASTIRMA	Aynı Seviyede Olacak	Gelişecek	Gerileyecek	Aynı Seviyede Olacak
GELECEK_HAYAT	Aynı Olacak	Daha İyi Olacak	Daha Kötü Olacak	Aynı Olacak
GELECEK_MALI	Aynı Olacak	Daha İyi Olacak	Daha Kötü Olacak	Aynı Olacak
GELİRİ_AZALDI	Hayır	Hayır	Hayır	Hayır

Tablo 6.6 incelendiğinde; 2011 veri kümesinin yine aynı özellikler açısından kümelerinin birbirinden ayrıldığı görülmektedir. Sadece önceki yıla göre gelirlerinde azalma meydana gelip gelmemesi özelliği tüm kümelerde azalmanın meydana gelmediği yönünde olduğundan, 2011 verisinde ayırt edici bir nitelik taşımamaktadır.

Kümelerin tamamında insanların, sağlık, evlilik, konut ve sağlık hizmetleri açısından memnuniyet duydukları ve evde ve bulundukları çevrede kendilerini güvende hissettikleri ve gelirlerinde önceki yıla göre azalma olmadığını ifade ettikleri

görülmektedir. Kümeleri birbirinden ayıran esas özellikler açısından kümelerin tarifi şöyle yapılabilir:

1.kümedeki insanlar; gelirlerinin ihtiyaçlarını orta düzeyde karşıladığını ve gelirlerinden orta derecede memnun olduklarını söyleyen, 6. refah düzeyinde olduklarını, geçmiş 5 yıla göre durumlarının aynı kaldığını ve gelecek 5 yıl ve gelecek yıl da durumlarının aynı kalacağını düşünen insanlardır.

2.kümedeki insanlar, birinci kümeden farklı olarak; gelirlerinden memnun, geçmiş 5 yıla göre durumlarının daha iyi olduğunu ve gelecek 5 yıl ve gelecek yıl da daha iyi olacağını düşünen insanlardır.

3.kümedeki insanlar; gelirlerinin ihtiyaçlarını zor karşıladığını ve gelirlerinden memnun olmadıklarını belirten, 4. refah düzeyinde olduklarını, geçmiş 5 yıla göre bugün daha kötü durumda olduklarını ve gelecek 5 yıl ve gelecek yıl da daha kötü olacaklarını ifade eden insanlardır.

4.kümedeki insanlar, üçüncü kümeden farklı olarak 5. refah düzeyinde olduklarını, geçmiş 5 yıla göre bugün aynı durumda olduklarını ve gelecek 5 yıl ve gelecek yıl da aynı durumda olacaklarını ifade eden insanlardır.

Buna göre, 1. kümedeki insanların “Umutlu”, 2. kümedeki insanların “Çok umutlu”, 3. kümedeki insanların “Hiç umutlu değil” ve 4. kümedeki insanların ise “Umutlu değil” kategorilerine dahil oldukları söylenebilir.

Tablo 6.7 2012 YMA veri kümesi küme merkezleri

Özellik ismi	1. Küme Merkezi	2. Küme Merkezi	3. Küme Merkezi	4. Küme Merkezi
OLCEK GELİR KARŞILAMA	Orta	Orta	Zor	Zor
OLCEK_MEMNUNİYET_SAGLIK	Memnun	Memnun	Memnun	Memnun
OLCEK_MEMNUNİYET_EVLİLİK	Memnun	Memnun	Memnun	Memnun
OLCEK_MEMNUNİYET_KONUT	Memnun	Memnun	Memnun	Memnun
OLCEK_MEMNUNİYET_GELİR	Orta	Memnun	Memnun Değil	Memnun Değil
OLCEK_MEMNUNİYET_GN_SAG	Memnun	Memnun	Memnun	Memnun
GUVEN EV	Güvenli	Güvenli	Güvenli	Güvenli
GUVEN CEVRE	Güvenli	Güvenli	Güvenli	Güvenli
UMUT_BASAMAK	Düzey 6	Düzey 6	Düzey 5	Düzey 4
GECEMIS_KARSILASTIRMA	Aynı Seviyede Kaldı	Gelişti	Aynı Seviyede Kaldı	Geriledi
GELECEK_KARSILASTIRMA	Aynı Seviyede Olacak	Gelişecek	Aynı Seviyede Olacak	Gerileyecek
GELECEK_HAYAT	Aynı Olacak	Daha İyi Olacak	Aynı Olacak	Daha Kötü Olacak
GELECEK_MALI	Aynı Olacak	Daha İyi Olacak	Aynı Olacak	Daha Kötü Olacak
GELİRİ_AZALDI	Hayır	Hayır	Hayır	Evet

Tablo 6.7 incelendiğinde 2012 yılı verisinin kümelenmesinde, kümelerin 2010 verisindeki gibi aynı özellikler açısından farklılık gösterdiği görülmektedir.

Kümelerin tamamında insanların, sağlık, evlilik, konut ve sağlık hizmetleri açısından memnuniyet duydukları ve evde ve bulundukları çevrede kendilerini güvende hissettikleri görülmektedir. Kümeleri birbirinden ayıran esas özellikler açısından kümelerin tarifi şöyle yapılabilir:

1.kümedeki insanlar; ihtiyaçlarının gelirleri tarafından orta düzeyde karşılandığını ve gelirlerinden memnuniyetlerinin orta derecede olduğunu söyleyen, refah düzeylerinin 6. düzey olduğunu, geçmiş 5 yıla göre bugünkü durumlarının aynı ve gelecek 5 yıl ve gelecek yıl da aynı kalacağını düşünen ve önceki yıla göre gelirinde azalma olmadığını ifade eden insanlardır.

2.kümedeki insanlar; birinci kümeden farklı olarak gelirlerinden memnun olduklarını, geçmiş 5 yıla göre durumlarının geliştiğini ve gelecek 5 yıl ve gelecek yılda da gelişeceğini düşünen insanlardır.

3.kümedeki insanlar; ihtiyaçlarının gelirleri tarafından zor karşılandığını ve gelirlerinden memnun olmadıklarını belirten, refah düzeyi açısından kendilerini 5. düzeyde gören, geçmiş 5 yıla göre bugün aynı olduklarını ve gelecek 5 yıl ve gelecek yılda da aynı durumda olacaklarını düşünen, önceki yıla göre gelir durumlarında azalma olmadığını ifade eden insanlardır.

4.kümedeki insanlar; üçüncü kümeden farklı olarak, refah düzeylerinin 4. düzey olduğunu, geçmiş 5 yıla göre durumlarının gerilediğini ve gelecek 5 yıl ve gelecek yılda da gerileyeceğini düşünen ve önceki yıla göre gelirinde azalma meydana geldiğini ifade eden insanlardır.

Buna göre, 1. kümedeki insanların “Umutlu”, 2. kümedeki insanların “Çok umutlu”, 3. kümedeki insanların “Umutlu değil” ve 4. kümedeki insanların “Hiç umutlu değil” kategorilerine girdikleri söylenebilir.

Tablo 6. 8 2013 YMA veri kümesi küme merkezleri

Özellik ismi	1. Küme Merkezi	2. Küme Merkezi	3. Küme Merkezi	4. Küme Merkezi
OLCEK GELİR KARŞILAMA	Orta	Zor	Orta	Zor
OLCEK_MEMNUNİYET_SAGLIK	Memnun	Memnun	Memnun	Memnun
OLCEK_MEMNUNİYET_EVLİLİK	Memnun	Memnun	Memnun	Memnun
OLCEK_MEMNUNİYET_KONUT	Memnun	Memnun	Memnun	Memnun
OLCEK_MEMNUNİYET_GELİR	Memnun	Memnun değil	Memnun	Memnun Değil
OLCEK_MEMNUNİYET_GN_SAG	Memnun	Memnun	Memnun	Memnun
GUVEN EV	Güvenli	Güvenli	Güvenli	Güvenli
GUVEN CEVRE	Güvenli	Güvenli	Güvenli	Güvenli
UMUT_BASAMAK	Duzey 6	Duzey 6	Duzey 6	Duzey 4
GECEMIS_KARSILASTIRMA	Aynı Seviyede Kaldı	Aynı Seviyede Kaldı	Gelişti	Geriledi
GELECEK_KARSILASTIRMA	Aynı Seviyede Olacak	Fikri Yok	Gelişecek	Gerileyecek
GELECEK_HAYAT	Aynı Olacak	Fikri Yok	Daha İyi Olacak	Daha Kötü Olacak
GELECEK_MALI	Aynı Olacak	Fikri Yok	Daha İyi Olacak	Daha Kötü Olacak
GELİRİ_AZALDI	Hayır	Hayır	Hayır	Evet

Tablo 6.8 incelendiğinde; 2013 yılı verisinin kümelenmesinde, kümelerin 2010 verisindeki gibi aynı özellikler açısından farklılık gösterdiği görülmektedir.

Kümelerin tamamında insanların, sağlık, evlilik, konut ve sağlık hizmetleri açısından memnuniyet duydukları ve evde ve bulundukları çevrede kendilerini güvende hissettikleri görülmektedir. Kümeleri birbirinden ayıran esas özellikler açısından kümelerin tarifi şöyle yapılabilir:

1.kümedeki insanlar; ihtiyaçlarının gelirleri tarafından orta düzeyde karşılandığını ve gelirlerinden memnun olduklarını söyleyen, refah düzeylerinin 6. düzey olduğunu, geçmiş 5 yıla göre bugünkü durumlarının aynı ve gelecek 5 yıl ve gelecek yıl da aynı kalacağını düşünen ve önceki yıla göre gelirinde azalma olmadığını ifade eden insanlardır.

2.kümedeki insanlar; ihtiyaçlarının gelirleri tarafından zor karşılandığını ve gelirlerinden memnun olmadıklarını belirten, refah düzeyi açısından kendilerini 6. düzeyde gören, geçmiş 5 yıla göre bugün aynı olduklarını ve gelecek 5 yıl ve gelecek yılda durumları hakkında fikirleri olmadığını ifade eden, önceki yıla göre gelir durumlarında azalma olmadığını belirten insanlardır.

3.kümedeki insanlar; birinci kümeden farklı olarak geçmiş 5 yıla göre durumlarının geliştiğini ve gelecek 5 yıl ve gelecek yılda da gelişeceğini düşünen insanlardır.

4.kümedeki insanlar; ikinci kümeden farklı olarak, refah düzeylerinin 4. düzey olduğunu, geçmiş 5 yıla göre durumlarının gerilediğini ve gelecek 5 yıl ve gelecek yılda da gerileyeceğini düşünen ve önceki yıla göre gelirinde azalma meydana geldiğini ifade eden insanlardır.

Buna göre, 1. kümedeki insanların “Umutlu”, 2. kümedeki insanların “Umutlu değil”, 3. kümedeki insanların “Çok umutlu” ve 4. kümedeki insanların “Hiç umutlu değil” kategorilerine girdikleri söylenebilir.

Tablo 6.9 2014 YMA veri kümesi küme merkezleri

Özellik ismi	1. Küme Merkezi	2. Küme Merkezi	3. Küme Merkezi	4. Küme Merkezi
OLCEK GELİR KARŞILAMA	Orta	Zor	Orta	Zor
OLCEK_MEMNUNİYET_SAGLIK	Memnun	Memnun	Memnun	Memnun
OLCEK_MEMNUNİYET_EVLİLİK	Memnun	Memnun	Memnun	Memnun
OLCEK_MEMNUNİYET_KONUT	Memnun	Memnun	Memnun	Memnun
OLCEK_MEMNUNİYET_GELİR	Memnun	Memnun Değil	Memnun	Memnun Değil
OLCEK_MEMNUNİYET_GN_SAG	Memnun	Memnun	Memnun	Memnun
GUVEN EV	Güvenli	Güvenli	Güvenli	Güvenli
GUVEN CEVRE	Güvenli	Güvenli	Güvenli	Güvenli
UMUT_BASAMAK	Düzey 5	Düzey 4	Düzey 5	Düzey 3
GECEK_KARSILASTIRMA	Gelişti	Aynı Seviyede Kaldı	Aynı Seviyede Kaldı	Geriledi
GELECEK_KARSILASTIRMA	Gelişecek	Aynı Seviyede Kalacak	Aynı Seviyede Kalacak	Gerileyecek
GELECEK_HAYAT	Daha İyi Olacak	Aynı Olacak	Aynı Olacak	Daha Kötü Olacak
GELECEK_MALI	Daha İyi Olacak	Aynı Olacak	Aynı Olacak	Daha Kötü Olacak
GELİRİ_AZALDI	Hayır	Hayır	Hayır	Evet

Tablo 6.9 incelendiğinde, 2014 yılı verisinin kümelenmesinde, kümelerin 2010 verisindeki gibi aynı özellikler açısından farklılık gösterdiği görülmektedir.

Kümelerin tamamında insanların; sağlık, evlilik, konut ve sağlık hizmetleri açısından memnuniyet duydukları, evde ve bulundukları çevrede kendilerini güvende hissettikleri görülmektedir. Kümeleri birbirinden ayıran esas özellikler açısından kümelerin tarifi şöyle yapılabilir:

1.kümedeki insanlar; gelirlerinin ihtiyaçlarını orta düzeyde karşıladığını ve gelirlerinden memnun olduğunu söyleyen, refah düzeyi açısından kendilerini 5. düzeyde gören, geçmiş 5 yıla göre bugünkü durumlarının geliştiğini, gelecek yıl ve gelecek 5 yılda da daha iyi duruma geleceklerini düşünen ve gelir düzeyinde önceki yıla göre bir azalma meydana gelmediğini ifade eden insanlardır.

2.kümedeki insanlar; gelirlerinin ihtiyaçlarını zor karşıladığını ve gelirlerinden memnun olmadığını söyleyen, refah düzeyi açısından kendilerini 4. düzeyde gören, geçmiş 5 yıla göre bugünkü durumlarının aynı düzeyde kaldığını, gelecek yıl ve gelecek 5 yılda da aynı seviyede kalacağını düşünen ve gelir düzeyinde önceki yıla göre bir azalma meydana gelmediğini ifade eden insanlardır.

3.kümedeki insanlar; birinci kümeden farklı olarak, geçmiş 5 yıla göre bugünkü durumlarının aynı düzeyde kaldığını, gelecek yıl ve gelecek 5 yılda da aynı seviyede kalacağını düşünen insanlardır.

4.kümedeki insanlar; ikinci kümeden farklı olarak, kendilerini 3. refah düzeyinde gören, geçmiş 5 yıla göre bugünkü durumlarının daha kötü olduğunu ve gelecek yıl ve gelecek 5 yılda da daha kötü olacağını ifade eden ve gelirlerinde geçen yıla göre azalma meydana geldiğini belirten insanlardır.

Kümeler incelendiğinde; 1. kümedeki insanların “Çok umutlu”, 2. kümedeki insanların “Umutlu değil”, 3. kümedeki insanların “Umutlu” ve 4. kümedeki insanların “Hiç umutlu değil” kategorilerine karşılık geldiği görülebilmektedir.

6.4 Sonuç

Kümeleme sonuçları, tüm yıllarda veri kümelerinde kümeleri birbirinden ayıran temel özelliklerin bulunduğunu ve kümelerin rahatlıkla birbirinden ayrıldığını göstermektedir. Tüm yıllarda kümeler; -“GELİRİ_AZALDI” özelliğinin 2011 yılında gösterdiği farklılık dışında- aynı özellikler açısından farklılık arz etmektedir.

Kümelerde ortak özellikler; “ÖLÇEK MEMNUNİYET SAĞLIK, ÖLÇEK MEMNUNİYET EVLİLİK, ÖLÇEK MEMNUNİYET KONUT, ÖLÇEK MEMNUNİYET GN SAĞ, GÜVEN EV, GÜVEN ÇEVRE ve 2011’e has olmak üzere GELİRİ AZALDI” özellikleridir.

Kümeleri birbirinden ayıran temel özellikler ise; “ÖLÇEK GELİR KARŞILAMA, ÖLÇEK MEMNUNİYET GELİR, UMUT BASAMAK, GEÇMİŞ KARŞILAŞTIRMA, GELECEK KARŞILAŞTIRMA, GELECEK HAYAT, GELECEK MALİ ve GELİRİ AZALDI özellikleri olmuştur.

Kümelemede gizli hedef değişkeni olan *umut* değişkeninin kategorilerine göre kümeleme işlemi belirgin bir şekilde yapılabilmiş ve insanların umut düzeylerinin yukarıda sayılan 8 ayırt edici özellik tarafından farklılık arz ettiği tespit edilmiştir. Ayrıca, yıllar açısından kümeleme sonuçları arasında çok az bir farklılık olduğu söylenebilmektedir.

2010 ve 2012 yıllarının kümelenmesi tamamen aynı olup, 2011 yılının bu yıllardan tek farkı “GELİRİ AZALDI” özelliğidir. 2010 ve 2012 yıllarında “Hiç umutlu değil” kategorisindeki kişilerin gelirlerinde azalma olduğunu ifade etmelerine karşın, 2011 yılında aynı kategoridekilerin gelirlerinde azalma olmadığını ifade etmeleri söz konusudur.

2013 yılı veri kümelenmesinin 2010 ve 2012 yıllarındakinden farkı, “OLCEK MEMNUNİYET GELİR”, “UMUT BASAMAĞI” ve “GELECEK KARSILASTIRMA”, “GELECEK HAYAT” ve “GELECEK MALİ” özelliklerindedir. 2010 ve 2012 yıllarında, “Umutlu değil” kategorisindeki kişilerin refah düzeyleri 5. düzey iken, 2013 yılında 6. düzeydir. Aynı kategorideki kişilerin geçmiş 5 yıla göre bugünkü durumlarının, gelecek yıl ve gelecek 5 yıl durumlarının 2010 ve 2012 yılında aynı olacağını ifade etmelerine karşın, 2013 yılında bir fikirlerinin olmadığını beyan etmeleri de diğer bir farklılıktır. Ayrıca, “Umutlu” kategorisindeki kişilerin 2010 ve 2012 yıllarında gelirlerinden orta düzeyde memnun olmalarına karşın 2013’de memnun olduklarını belirtmeleri diğer bir farktır.

2014 yılı veri kümelenmesinin 2010 ve 2012 yıllarındakinden farkı; “UMUT BASAMAĞI” ve “OLCEK_MEMNUNİYET_GELİR” özelliklerindedir. 2010 ve 2012 yıllarında “Umutlu” kategorisindeki kişilerin refah düzeyleri 6 iken 2014’de 5, “Çok umutlu” kategorisindekilerin 6 iken 2014’de 5, “Hiç umutlu değil” kategorisindekilerin 4 iken 2014’de 3’tür. Ayrıca, “Umutlu” kategorisindeki kişilerin 2010 ve 2012 yıllarında gelirlerinden orta düzeyde memnun olmalarına karşın 2014’de memnun olduklarını belirtmeleri diğer bir farktır.

Bu farklılıklar dışında genel olarak, yıllar arasında çok önemli farklılıklar bulunmamaktadır. Sonuç olarak, tüm yıllarda kümeleme sonuçları, insanların aynı özellikler açısından farklı kümelere ayrıldığını göstermiştir.

YMA gibi önemli sosyal karakteristikleri içeren bir sosyal anketin verilerinin kümeleme problemiyle ifade edilişi literatürde yaygın olan bir çalışma değildir. Bu çalışma ile bu boşluğun doldurulması amaçlanmıştır.



7. BÖLÜM

SONUÇ ve TARTIŞMALAR

7.1 Çalışmanın Katkıları

Bu bölümde yapılan tez çalışmasının literatüre katkısı ve orijinallliği açıklanmıştır. Çalışmada, yeni nesil metasezgisel yaklaşımların kümeleme problemindeki performansları karşılaştırılarak etkin performansa sahip olan yaklaşımların gerçek hayat verisi olan YMA üzerinde uygulaması gerçekleştirilmiştir.

Kümeleme probleminde metasezgisel algoritmaların kullanımı yaygın olmakla beraber; parametre sayısı az ve hesapsal zamanı kısa olan algoritma arayışı halen devam etmektedir. Çalışmada kullanılan yeni nesil metasezgisellerden biri olan IMO, parametre sayısı açısından birçok algoritmaya üstünlük sağlamaktadır. Maksimum iterasyon sayısı ve popülasyon büyüklüğü haricinde hiçbir parametreye ihtiyaç duymayan IMO algoritması ve hesapsal zaman açısından bu çalışmada da üstün performans sağlayan WSA algoritmasının kümelemeye uygulanması çalışmanın orijinallliğini ortaya koymaktadır.

Kümeleme problemi literatürde daha çok sürekli değişkenler için çalışılmakta, kategorik ve karma veri kümelerindeki çalışmalara daha az rastlanılmaktadır. Bu çalışma ile sadece sürekli değil, kategorik ve karma veri kümelerinde de kümeleme yeni nesil algoritmalar ile gerçekleştirilmiş ve etkin sonuçlar elde edilmiştir.

YMA, sosyal içerikli zengin bir veri kaynağı özelliğini taşımaktadır. Bu çalışma ile ayrıca, bu tür anketlerde kümeleme uygulaması konusunda literatürdeki mevcut boşluğun doldurulması amaçlanmıştır.

7.2 İleriye Yönelik Öneriler

Kümeleme, üzerinde birçok yaklaşım uygulanmış ve uygulanmaya devam edilen zor bir problemdir. Yapılan çalışmada, farklı tipte veri kümeleri için kümeleme yapılırken küme sayısının belli olduğu varsayımı altında algoritmaların performansları değerlendirilmiştir. Ancak kümelemenin danışmansız bir öğrenme tekniği olduğu dikkate alındığında, küme sayısını tespit edebilen yaklaşımların kullanılması da gerekebilmektedir. Bu nedenle; çalışmada kullanılan metasezgisellerin otomatik kümeleme için de uygulaması gerçekleştirilebilir.

Kümeleme probleminde kullanılan benzerlik/benzemezlik ölçütlerinin veri kümeleme kalitesini etkilemesi literatürde çalışılan konulardan biridir. Öklid uzaklığı yerine cosinüs uzaklığı kullanılarak algoritmaların kümelemede hata oranları veya doğruluk açısından performansları karşılaştırılabilir.

Kümeleme geçerliliğinde sadece küme içi örneklerin birbirine yakınlıklarının en küçüklenmesi değil, Turi indeks, Dunn indeks, Silhoutte indeks ve üçüncü bölümde açıklanan diğer geçerlilik indeksleri açısından da kümeleme sonuçları karşılaştırılabilir.

EKLER

EK-1 YMA ÖZELLİKLERİN İSİMLERİ ve AÇIKLAMALARI

SIR A NO	ÖZELLİKLER	ÖZELLİKLERİN AÇIKLAMASI	ÖZELLİK DEĞER ARALIĞI	ÖZELLİK VERİ TİPİ	KATEGORİ SAYISI
1	BİRİM NO	Anket soru kağıdında belirtilen ve haneyi tanımlayan form numarası	100101-145610	Sürekli	-
2	REFERANS_YIL	Referans yıl	2014 (tek değer)	Kategorik	1
3	FERT_NO	Fert Sıra numarası	1-13	Kategorik	13
4	BITİRİLEN_YAS	Ferdin bitirdiği yaş	18-98	Sürekli	-
5	CINSİYET	Görüşülen kişinin cinsiyeti	1-2	Kategorik	2
6	MEDENİ_DURUM	B1 Medeni durumunuz nedir?	1-4	Kategorik	4
7	OKUL_BITEN	B2 En son hangi okulu bitirdiniz?	0-32	Kategorik	33
8	OKUMA_YAZMA	B3 Okuma yazma biliyor musunuz?	1-2	Kategorik	2
9	CALISMA_DURUM	B4 Son bir hafta içinde ücretli ya da ücretsiz olarak bir işte çalıştınız mı?	1-3	Kategorik	3
10	CALISMAMA_NEDENİ	B5 Çalışmama nedeniniz nedir?	1-9, 98	Kategorik	10
11	OZEL_KAMU	B6 İşyeriniz/işiniz kamuya mı yoksa özel sektöre mi ait?	1-2	Kategorik	2
12	ISTEKİ_DURUM	B7 Çalıştığınız işteki durumunuz nedir?	1-4	Kategorik	4
13	ISTEKİ_SORUN1	B8.1 2014 yılı içinde, çalıştığınız yerde idari konularda problem yaşıyor musunuz?	1-2	Kategorik	2
14	ISTEKİ_SORUN2	B8.2 2014 yılı içinde, çalıştığınız yerde, ücretler arası farklılık problemi yaşıyor musunuz?	1-2	Kategorik	2
15	ISTEKİ_SORUN3	B8.3 2014 yılı içinde, çalıştığınız yerde, ücret miktarlarında problem yaşıyor musunuz?	1-2	Kategorik	2
16	ISTEKİ_SORUN4	B8.4 2014 yılı içinde, çalıştığınız yerde, çalışma koşullarında problem yaşıyor musunuz?	1-2	Kategorik	2
17	ISTEKİ_SORUN5	B8.5 2014 yılı içinde, çalıştığınız yerde, ücretlerin zamanında ödenmesi ile ilgili problem yaşıyor musunuz?	1-2	Kategorik	2

18	ISTEKI_SORUN6	B8.6 2014 yılı içinde, çalıştığınız yerde, ücretlerin eksik ödenmesi ile ilgili problem yaşıyor musunuz?	1-2	Kategorik	2
19	MUTLULUK	B9 Yaşamınızı bir bütün olarak düşündüğünüzde ne kadar mutlusunuz?	1-5	Kategorik	5
20	KIM_MUTLU	B.10 Sizi hayatta en çok kim mutlu eder?	1-8, 98	Kategorik	9
21	NE_MUTLU	B11 Sizi hayatta en çok ne mutlu eder?	1-6, 98	Kategorik	7
22	OLCEK_MEMNUNİYET_SAGLIK	B12.1 Sağlığınızdan memnun musunuz?	1-5	Kategorik	5
23	OLCEK_MEMNUNİYET_EVLILIK	B12.2 Evliliğinizden memnun musunuz?	1-5	Kategorik	5
24	OLCEK_MEMNUNİYET_EGITIM	B12.3 Şimdiye kadar almış olduğunuz eğitimden memnun musunuz?	1-6	Kategorik	6
25	OLCEK_MEMNUNİYET_KONUT	B12.4 Oturduğunuz konuttan memnun musunuz?	1-5	Kategorik	5
26	OLCEK_MEMNUNİYET_SEMT	B12.5 Oturduğunuz semtten veya mahalleden memnun musunuz?	1-5	Kategorik	5
27	OLCEK_MEMNUNİYET_IS	B12.6 İşinizden memnun musunuz?	1-5	Kategorik	5
28	OLCEK_MEMNUNİYET_KAZANC	B12.7 İşinizden elde ettiğiniz kazançtan memnun musunuz?	1-6	Kategorik	6
29	OLCEK_MEMNUNİYET_GELIR	B12.8 Aylık hanehalkı gelirinizden memnun musunuz?	1-5	Kategorik	5
30	OLCEK_MEMNUNİYET_SOS_HAYAT	B12.9 Sosyal hayatınızdan (eğlence, kültürel ve sportif faaliyetler gibi) memnun musunuz?	1-5	Kategorik	5
31	OLCEK_MEMNUNİYET_KISISEL_BKM	B12.10 Kendinize ayırdığınız zamandan memnun musunuz?	1-5	Kategorik	5
32	OLCEK_MEMNUNİYET_IS_TRFK_ZMN	B12.11 İşe geliş gidiş için harcadığınız (ayırdığınız) zamandan memnun musunuz?	1-5	Kategorik	5
33	OLCEK_MEMNUNİYET_AKRABA	B13.1 Akrabalarınızla ilişkilerinizden memnun musunuz?	1-5	Kategorik	5
34	OLCEK_MEMNUNİYET_ARKADAS	B13.2 Arkadaşlarınızla ilişkilerinizden memnun musunuz?	1-5	Kategorik	5

35	OLCEK_ MEMNUNİYET_ KOMSU	B13.3 Komşularınızla ilişkilerinizden memnun musunuz?	1-5	Kategorik	5
36	OLCEK_ MEMNUNİYET_ ISILISKI	B13.4 İşinizle ilgili kişilerle ilişkilerinizden memnun musunuz?	1-5	Kategorik	5
37	OLCEK_ MEMNUNİYET_ GN_SAG	B14.1 Sağlık hizmetlerinden memnun musunuz?	1-5	Kategorik	5
38	OLCEK_ MEMNUNİYET_ ASAYIS	B14.2 Asayiş hizmetlerinden memnun musunuz?	1-6	Kategorik	6
39	OLCEK_ MEMNUNİYET_ ADLI	B14.3 Adli hizmetlerden memnun musunuz?	1-6	Kategorik	6
40	OLCEK_ MEMNUNİYET_ EGITI	B14.4 Eğitim hizmetlerinden memnun musunuz?	1-6	Kategorik	6
41	OLCEK_ MEMNUNİYET_ SGK	B14.5 Sosyal Güvenlik Kurumu'nun hizmetlerinden memnun musunuz?	1-6	Kategorik	6
42	OLCEK_ MEMNUNİYET_ ULASIM	B14.6 Ulaştırma hizmetlerinden memnun musunuz?	1-6	Kategorik	6
43	ELEKTRO_HIZ	B15 Elektronik ortamda sunulan kamu hizmetlerinden yararlanıyor musunuz?	1-3	Kategorik	3
44	OLCEK_ MEMNUNİYET_ ELKNIK_KAM	B16 Elektronik ortamda sunulan kamu hizmetlerinden memnuniyet derecenizi söyler misiniz?	1-5	Kategorik	5
45	BELEDIYE	B17 Yaşadığınız yerleşim yeri belediye sınırları içinde mi?	1-2	Kategorik	2
46	MEMNUNİYET_COP	B18.1 Belediyenin çöp ve çevresel atık toplama hizmetinden memnuniyet durumunuzu söyler misiniz?	1-7	Kategorik	7
47	MEMNUNİYET_KAN	B18.2 Belediyenin kanalizasyon hizmetinden memnuniyet durumunuzu söyler misiniz?	1-7	Kategorik	7
48	MEMNUNİYET_SU	B18.3 Belediyenin şebeke suyu hizmetlerinden memnuniyet durumunuzu söyler misiniz?	1-7	Kategorik	7
49	MEMNUNİYET_ TASIMA	B18.4 Belediyenin toplu taşıma hizmetlerinden memnuniyet durumunuzu söyler misiniz?	1-7	Kategorik	7
50	MEMNUNİYET_ ZABITA	B18.5 Belediyenin zabıta hizmetlerinden memnuniyet durumunuzu söyler misiniz?	1-7	Kategorik	7

51	MEMNUNİYET_YOL	B18.6 Belediyenin yol/kaldırım yapımı hizmetlerinden memnuniyet durumunuzu söyler misiniz?	1-7	Kategorik	7
52	MEMNUNİYET_YESİL	B18.7 Belediyenin yeşil alan ve ağaçlandırma hizmetinden memnuniyet durumunuzu söyler misiniz?	1-7	Kategorik	7
53	MEMNUNİYET_KIRLILIK	B18.8 Belediyenin hava kirliliği ile mücadeleden memnuniyet durumunuzu söyler misiniz?	1-7	Kategorik	7
54	MEMNUNİYET_SPOR_MRK	B18.9 Belediyenin sağlık, spor merkezi olanaklarından memnuniyet durumunuzu söyler misiniz?	1-7	Kategorik	7
55	MEMNUNİYET_IMAR	B18.10 Belediyenin imar/iskan/ruhsat işlemlerinden memnuniyet durumunuzu söyler misiniz?	1-7	Kategorik	7
56	MEMNUNİYET_ENGELLI	B18.11 Belediyenin engellilere yönelik düzenlemelerden memnuniyet durumunuzu söyler misiniz?	1-7	Kategorik	7
57	MEMNUNİYET_YARDIM	B18.12 Belediyenin hasta ve yoksullara yardımlarından memnuniyet durumunuzu söyler misiniz?	1-7	Kategorik	7
58	MEMNUNİYET_SERGI_FEST	B18.13 Belediyenin sergi, festival, kermes, konser faaliyetlerinden memnuniyet durumunuzu söyler misiniz?	1-7	Kategorik	7
59	MEMNUNİYET_KURS	B18.14 Belediyenin meslek edindirme/el becerisi geliştirme kurslarından memnuniyet durumunuzu söyler misiniz?	1-7	Kategorik	7
60	MEMNUNİYET_ISIK	B18.15 Belediyenin ışıklandırma hizmetlerinden memnuniyet durumunuzu söyler misiniz?	1-7	Kategorik	7
61	MEMNUNİYET_TEMIZLIK	B18.16 Belediyenin temizlik hizmetlerinden memnuniyet durumunuzu söyler misiniz?	1-7	Kategorik	7
62	MEMNUNİYET_ITFAIYE	B18.17 Belediyenin itfaiye hizmetlerinden memnuniyet durumunuzu söyler misiniz?	1-7	Kategorik	7

63	MEMNUNİYET_ CENAZE	B18.18 Belediyenin cenaze hizmetlerinden memnuniyet durumunuzu söyler misiniz?	1-7	Kategorik	7
64	MEMNUNİYET_ LEVHA	B18.19 Belediyenin sokak levhaları ve dış kapı numaralandırma hizmetinden memnuniyet durumunuzu söyler misiniz?	1-7	Kategorik	7
65	MEMNUNİYET_GIDA_ TESİS	B18.20 Belediyenin gıda üreten tesislerin denetimi hizmetlerinden memnuniyet durumunuzu söyler misiniz?	1-7	Kategorik	7
66	MEMNUNİYET_IL_ KANAL	B19.1 İl özel idarenin, kanalizasyon hizmetinden memnun musunuz?	1-7	Kategorik	7
67	MEMNUNİYET_IL_SU	B19.2 İl özel idarenin, şebeke suyu hizmetlerinden memnun musunuz?	1-7	Kategorik	7
68	MEMNUNİYET_IL_ YOL	B19.3 İl özel idarenin, yol yapımı hizmetlerinden memnun musunuz?	1-7	Kategorik	7
69	MEMNUNİYET_IL_ ISKAN	B19.4 İl özel idarenin, imar/iskan/ruhsat işlemlerinden memnun musunuz?	1-7	Kategorik	7
70	MEMNUNİYET_IL_ ENGEL	B19.5 İl özel idarenin, engellilere yönelik düzenlemelerden memnun musunuz?	1-7	Kategorik	7
71	MEMNUNİYET_IL_ HASTA	B19.6 İl özel idarenin, hasta ve yoksullara yardımlarından memnun musunuz?	1-7	Kategorik	7
72	MEMNUNİYET_IL_ SERGI	B19.7 İl özel idarenin, sergi, festival, kermes, konser faaliyetlerinden memnun musunuz?	1-7	Kategorik	7
73	MEMNUNİYET_IL_ KURS	B19.8 İl özel idarenin, meslek edindirme/el becerisi geliştirme kurslarından memnun musunuz?	1-7	Kategorik	7
74	MEMNUNİYET_IL_ ISIK	B19.9 İl özel idarenin, ışıklandırma hizmetlerinden memnun musunuz?	1-7	Kategorik	7
75	MEMNUNİYET_IL_ TEMİZLİK	B19.10 İl özel idarenin, temizlik hizmetlerinden memnun musunuz?	1-7	Kategorik	7
76	MEMNUNİYET_IL_ LEVHA	B19.11 İl özel idarenin, sokak levhaları ve dış kapı numaralandırma hizmetlerinden memnun musunuz?	1-7	Kategorik	7

77	MEMNUNİYET_IL_YESİL	B19.12 İl özel idaresinin, yeşil alan ve ağaçlandırma hizmetlerinden memnun musunuz?	1-7	Kategorik	7
78	SGK_KAYIT	B20 Aşağıdakilerden hangi sosyal güvenlik kurumuna kayıtlısınız?	1-14	Kategorik	14
79	SGK_KIMDEN	B21 Kimin üzerinden sosyal güvenlik kapsamındasınız?	1-4	Kategorik	4
80	SORUN_SAGLIK_HARCAMA	B22.1 Bağlı bulunduğunuz sosyal güvenlik kuruluşu vasıtasıyla aldığınız hizmetlerden, sağlık harcamalarının geri ödenmesi ile ilgili işlemlerde sorun yaşıyor musunuz?	1-3	Kategorik	3
81	SORUN_ILAC	B22.2 Bağlı bulunduğunuz sosyal güvenlik kuruluşu vasıtasıyla aldığınız hizmetlerden, ilaç alımı ile ilgili işlemlerde sorun yaşıyor musunuz?	1-3	Kategorik	3
82	SORUN_SGK_PERSONEL	B22.3 Bağlı bulunduğunuz, Sosyal Güvenlik Kuruluşunda çalışan personelin davranışlarında sorun görüyor musunuz?	1-3	Kategorik	3
83	SORUN_EMEKLİLİK	B22.4 Bağlı bulunduğunuz sosyal güvenlik kuruluşu vasıtasıyla aldığınız hizmetlerden, emeklilik işlemlerinde sorun yaşıyor musunuz?	1-3	Kategorik	3
84	SORUN_MIKTAR_EMAAS	B22.5 Emekli maaşlarının miktarı sizce yeterli mi?	1-3	Kategorik	3
85	SOSGUV_GITME	B23 2014 yılı içinde bağlı olduğunuz Sosyal Güvenlik Kurumuna herhangi bir iş için gittiniz mi?	1-2	Kategorik	2
86	SORUN_SGK	B24 Sosyal Güvenlik Kurumunda sorun yaşadınız mı?	1-2	Kategorik	2
87	HASTANE_BASVURU	B25 Hastalandığınızda genellikle ilk nereye başvurursunuz?	1-7	Kategorik	7
88	SAGLIK_HIZMET	B26 2014 yılı içinde herhangi bir sağlık kuruluşuna başvurarak sağlık hizmeti aldınız mı?	1-2	Kategorik	2

89	SORUN_SAGLIK_KRLS	B27 En son sađlık hizmeti aldığınız sađlık kuruluđu hangisidir?	1-7	Kategorik	7
90	HASTANE_TERCIH_SEBEP	B28 Bu sađlık kuruluđuunu neden tercih ediyorsunuz?	1-7, 98	Kategorik	8
91	SORUN_SAGLIK_HIZALIM	B29 En son sađlık hizmeti alımı sırasında herhangi bir sorun yaşıadınız mı?	1-2	Kategorik	2
92	SORUN_MUAYENE	B30.1 Genel olarak sađlık hizmeti aldığınız kuruluđularla ilgili, muayene ve/veya tahlil için randevu almakta sorun yaşıyor musunuz?	1-3	Kategorik	3
93	SORUN_HIJYEN	B30.2 Genel olarak sađlık hizmeti aldığınız kuruluđularla ilgili, temizlik/hijyen konusunda sorun var mı?	1-3	Kategorik	3
94	MEMNUNİYET_MUAYENE	B30.3 Genel olarak sađlık hizmeti aldığınız kuruluđularla ilgili, yapılan muayeneden memnun musunuz?	1-3	Kategorik	3
95	SORUN_DOKTOR	B30.4 Genel olarak sađlık hizmeti aldığınız kuruluđularla ilgili, doktorların hastalara davranışında sorun var mı?	1-3	Kategorik	3
96	SORUN_HEMSIRE	B30.5 Genel olarak sađlık hizmeti aldığınız kuruluđularla ilgili, hemşirelerin/hastabakıcıların hastalara davranışında sorun var mı?	1-3	Kategorik	3
97	SORUN_YETERLI_SGLKPERSONEL	B30.6 Genel olarak sađlık hizmeti aldığınız kuruluđularla ilgili, doktor ve sađlık personeli sayısını yeterli buluyor musunuz?	1-3	Kategorik	3
98	SORUN_UCRET_MUAYENE	B30.7 Genel olarak sađlık hizmeti aldığınız kuruluđularla ilgili, muayene ve tahlil ücretlerini yüksek buluyor musunuz?	1-3	Kategorik	3
99	SORUN_ILAC_FIYAT	B30.8 Genel olarak sađlık hizmeti aldığınız kuruluđularla ilgili, ilaç fiyatlarında sorun görüyor musunuz?	1-3	Kategorik	3
100	SORUN_SIRABEKE	B30.9 Genel olarak sađlık hizmeti aldığınız kuruluđularla ilgili, muayene ve/veya tahlil için sıra beklemede sorun var mı?	1-3	Kategorik	3

101	SORUN_KATKIPAY	B30.10 Genel olarak sağlık hizmeti aldığınız kuruluşlarla ilgili, muayene için katkı payı ücreti ödemeyi sorun olarak görüyor musunuz?	1-3	Kategorik	3
102	TEDAVI_MASRAF_KANAL	B31 Hastalandığınızda, tedavi, ilaç vb. masraflarınızı genellikle hangi kanalla karşılıyorsunuz?	1-6,90,98	Kategorik	8
103	POL_JAN__MUDAHALE	B32.1 Polis veya Jandarma olaylara zamanında müdahale ediyor mu?	1-3	Kategorik	3
104	MEMNUNİYET_POLIS	B32.2 Polis veya jandarmanın vatandaşa davranışından memnun musunuz?	1-3	Kategorik	3
105	MEMNUNİYET_TRAFIK	B32.3 Polis veya jandarmanın verdiği trafik hizmetinden memnun musunuz?	1-3	Kategorik	3
106	KAPKAC	B33.1 2014 yılında kapkaç, yankesicilik vb hırsızlık olayı yaşadınız mı?	1-2	Kategorik	2
107	KAPKAC_POLIS	B33.11 Bu olay için Polise/Jandarmaya başvurduğunuz mu?	1-2	Kategorik	2
108	KAPKAC_BASVURMAMA_NEDEN	B33.12 Başvurmama nedeniniz nedir?	1-5, 98	Kategorik	6
109	GASP	B33.2 2014 yılında gasp olayı yaşadınız mı?	1-2	Kategorik	2
110	GASP_POLIS	B33.21 Bu olay için Polise/Jandarmaya başvurduğunuz mu?	1-2	Kategorik	2
111	GASP_BASVURMAMA_NEDEN	B33.22 Başvurmama nedeniniz nedir?	1-2	Kategorik	2
112	DARP	B33.3 2014 yılında yaralanma, darp olayı yaşadınız mı?	1-2	Kategorik	2
113	DARP_POLIS	B33.31 Bu olay için Polise/Jandarmaya başvurduğunuz mu?	1-2	Kategorik	2
114	DARP_BASVURMAMA_NEDEN	B33.32 Başvurmama nedeniniz nedir?	1-5	Kategorik	5
115	FERT_KOTU	B33.4 2014 yılında aile fertlerinizin herhangi birinden kötü muamele gördünüz mü	1-2	Kategorik	2
116	FERT_KOTU_POLIS	B33.41 Bu olay için Polise/Jandarmaya başvurduğunuz mu?	1-2	Kategorik	2

117	FERT_KOTU_BASVURMAMA_NEDEN	B33.42 Başvurmama nedeniniz nedir?	1-5	Kategorik	5
118	TEHDIT	B33.5 2014 yılında herhangi bir nedenden dolayı şantaj, tehdit olayı yaşadınız mı?	1-2	Kategorik	2
119	TEHDIT_POLIS	B33.51 Bu olay için Polise/Jandarmaya başvurduğunuz mu?	1-2	Kategorik	2
120	TEHDIT_BASVURMAMA_NEDEN	B33.52 Başvurmama nedeniniz nedir?	1-5	Kategorik	5
121	MAGDURIYET	B33.6 2014 yılında cinsel suçlardan dolayı mağduriyet yaşadınız mı?	1-2	Kategorik	2
122	MAGDURIYET_POLIS	B33.61 Bu olay için Polise/Jandarmaya başvurduğunuz mu?	1-2	Kategorik	2
123	MAGDURIYET_BASVURMAMA_NEDEN	B33.62 Başvurmama nedeniniz nedir?	1-5	Kategorik	5
124	DOLANDIR	B33.7 2014 yılında dolandırıcılık sebebiyle herhangi bir mağduriyet yaşadınız mı?	1-2	Kategorik	2
125	DOLANDIR_POLIS	B33.71 Bu olay için Polise/Jandarmaya başvurduğunuz mu?	1-2	Kategorik	2
126	DOLANDIR_BASVURMAMA_NEDEN	B33.72 Başvurmama nedeniniz nedir?	1-5, 98	Kategorik	6
127	BASKASUC	B33.8 2014 yılında bunların dışında herhangi bir sebepten dolayı mağduriyet yaşadınız mı?	1-2	Kategorik	2
128	BASKASUC_POLIS	B33.81 Bu olay için Polise/Jandarmaya başvurduğunuz mu?	1-2	Kategorik	2
129	BASKASUC_BASVURMAMA_NEDEN	B33.82 Başvurmama nedeniniz nedir?	1-5, 98	Kategorik	6
130	SORUN_MAH	B34.1 Genel olarak mahkeme ve adli sistem ile ilgili, mahkemelerdeki işlemlerde sorun var mı?	1-4	Kategorik	4
131	SORUN_DAVA	B34.2 Genel olarak mahkeme ve adli sistem ile ilgili, davaların karara bağlanma süresinde sorun var mı?	1-4	Kategorik	4
132	SORUN_YASA	B34.3 Genel olarak mahkeme ve adli sistem ile ilgili, yasaların herkese adil ve tarafsız uygulanmasında sorun var mı?	1-4	Kategorik	4

133	SORUN_AVUKAT	B34.4 Genel olarak mahkeme ve adli sistem ile ilgili, avukatlık hizmetlerinin kalitesinde sorun var mı?	1-4	Kategorik	4
134	BASVURU_MAH	B35 2014 yılı içinde mahkemeye başvurunuz oldu mu?	1-2	Kategorik	2
135	SORUN_MAHSUREC	B36 Mahkeme sürecinde herhangi bir sorun yaşadınız mı?	1-2	Kategorik	2
136	ULASIM_KARAYOLU	B37.1 Son bir yılda kullanmış olduğunuz, karayolu ulaşımı hizmetleri için memnuniyet derecenizi belirtiniz?	1-7	Kategorik	7
137	ULASIM_DENIZYOLU	B37.2 Son bir yılda kullanmış olduğunuz, denizyolu ulaşımı hizmetleri için memnuniyet derecenizi belirtiniz?	1-7	Kategorik	7
138	ULASIM_HAVAYOLU	B37.3 Son bir yılda kullanmış olduğunuz, havayolu ulaşımı hizmetleri için memnuniyet derecenizi belirtiniz?	1-7	Kategorik	7
139	ULASIM_DEMIRYOLU	B37.4 Son bir yılda kullanmış olduğunuz, demiryolu ulaşımı hizmetleri için memnuniyet derecenizi belirtiniz?	1-7	Kategorik	7
140	GUVEN_EV	B38 Evinizde yalnızken kendinizi ne kadar güvende hissediyorsunuz?	1-5	Kategorik	5
141	GUVEN_CEVRE	B39 Yaşadığınız çevrede, gece yalnız yürürken kendinizi ne kadar güvende hissediyorsunuz?	1-5	Kategorik	5
142	UMUT	B40 Kendi geleceğinizden ne kadar umutlusunuz?	1-4	Kategorik	4
143	UMUT_BASAMAK	B41 Türkiye'de yaşayan insanların refah düzeyini 0 basamağı en düşük 10 basamağı en yüksek düzey olarak düşündüğünüzde, kendinizi hangi düzeyde görüyorsunuz?	1-11	Kategorik	11
144	GECEMIS_KARSILASTIRMA	B42 5 yıl öncesi ile karşılaştırdığınızda bugünkü durumunuzu (maddi veya manevi) nasıl görüyorsunuz?	1-4	Kategorik	4

145	GELECEK_KARSILASTIRMA	B43 Gelecek 5 yıllık dönemi düşündüğünüzde, genel olarak durumunuzun nasıl olacağını tahmin ediyorsunuz?	1-4	Kategorik	4
146	GELECEK_HAYAT	B44.1 Önümüzdeki yıl (2015) genel olarak hayatınız sizce nasıl olacak?	1-4	Kategorik	4
147	GELECEK_KISISEL	B44.2 Önümüzdeki yıl (2015), kişisel iş durumunuz sizce nasıl olacak?	1-5	Kategorik	5
148	GELECEK_MALI	B44.3 Önümüzdeki yıl (2015), hanenizin mali durumu sizce nasıl olacak?	1-4	Kategorik	4
149	GELECEK_IS	B44.4 Önümüzdeki yıl (2015), Türkiye'deki iş/çalışma durumu sizce nasıl olacak?	1-4	Kategorik	4
150	GELECEK_EKONOMI	B44.5 Önümüzdeki yıl (2015), Türkiye'nin içinde bulunduğu ekonomik durum sizce nasıl olacak?	1-4	Kategorik	4
151	AB_UYELIK	B45 Türkiye'nin Avrupa Birliği'ne üye olmasının sizin yaşamınızı ne yönde etkileyeceğini düşünüyorsunuz?	1-4	Kategorik	4
152	AB_REFERANDUM	B46 Türkiye'nin Avrupa Birliği'ne üye olması konusunda bir referandum (halk oylaması) yapılırsa, siz ne yönde oy kullanırsınız?	1-3	Kategorik	3
153	EKONOMIK_ACIDAN	B47.1 Sizce Ülkemiz ekonomik açıdan önümüzdeki 5 yılda nasıl değişecek?	1-4	Kategorik	4
154	SOSYAL_HAKLAR	B47.2 Sizce Ülkemiz Sosyal haklar ve özgürlükler açısından önümüzdeki 5 yılda nasıl değişecek?	1-4	Kategorik	4
155	KAMU_HIZMET_SUNUM	B47.3 Sizce Ülkemiz Kamu hizmetlerinin sunumu açısından önümüzdeki 5 yılda nasıl değişecek?	1-4	Kategorik	4
156	DEVLET_SEFFAF	B47.4 Sizce Ülkemiz devletin şeffaflığı açısından önümüzdeki 5 yılda nasıl değişecek?	1-4	Kategorik	4

157	DUNYA_SAYGINLIGI	B47.5 Sizce Ülkemiz Dünyadaki saygınlığı (itibar) açısından önümüzdeki 5 yılda nasıl değişecek?	1-4	Kategorik	4
158	ULKE_DEGISIM_DIGER	B47.6 Sizce Ülkemiz başka hangi konularda önümüzdeki 5 yılda değişecek?(Bu soru boş geçilebilir.)	1-3	Kategorik	3
159	ISE_GIRDI	B48.1 Son bir yılda, işe girdim (iş arayanlar için)	1-3	Kategorik	3
160	ISYERI_ACTI	B48.2 Son bir yılda, Yeni iş yeri açtım	1-3	Kategorik	3
161	ISINI_KAYBETTI	B48.3 Son bir yılda, İşimi kaybettim	1-3	Kategorik	3
162	IFLAS_ETTI	B48.4 Son bir yılda, İflas ettim/dükkan kapattım	1-3	Kategorik	3
163	GELIRI_AZALDI	B48.5 Son bir yılda, Gelirim azaldı	1-3	Kategorik	3
164	GELIRI_ARTTI	B48.6 Son bir yılda, Gelirim arttı	1-3	Kategorik	3
165	TASARRUFLARI_AZALDI	B48.7 Son bir yılda, Tasarruflarımda azalma oldu	1-3	Kategorik	3
166	TASARRUF_YAPTI	B48.8 Son bir yılda, Tasarruf yapmaya başladım	1-2	Kategorik	2
167	UCUZ_URUN	B48.9 Son bir yılda, Daha ucuz ürünlerin tüketimine yöneldim	1-2	Kategorik	2
168	TATIL_KISTI	B48.10 Son bir yılda, Eğlence/tatil harcamalarımı kısıtım	1-3	Kategorik	3
169	GOC_ETTI	B48.11 Son bir yılda, göç ettim.	1-2	Kategorik	2
170	ARABA_ALDI	B48.12 Son bir yılda, Araba aldım	1-2	Kategorik	2
171	EV_ALDI	B48.13 Son bir yılda, Ev, arsa, yazlık vb. aldım	1-2	Kategorik	2
172	ARABA_SATTI	B48.14 Son bir yılda, Arabamı sattım	1-3	Kategorik	3
173	EV_SATTI	B48.15 Son bir yılda, Ev, arsa, yazlık vb. sattım	1-3	Kategorik	3
174	BORCLANDI	B48.16 Son bir yılda, Borçlandım	1-2	Kategorik	2
175	BORC_ODEDI	B48.17 Son bir yılda, borçlarımı ödedim	1-3	Kategorik	3
176	EMEKLİ_OLDU	B48.18 Son bir yılda, emekli oldum	1-2	Kategorik	2
177	EVLENDI	B49.1 Son bir yılda, evlendim	1-2	Kategorik	2
178	BOSANDI	B49.2 Son bir yılda, boşandım	1-3	Kategorik	3
179	COCUK_OLDU	B49.3 Son bir yılda, çocuğum oldu	1-2	Kategorik	2

180	VEFAT	B49.4 Son bir yılda, Ailemden vefat eden oldu	1-2	Kategorik	2
181	TASINMA	B49.5 Son bir yılda, Başka bir eve taşındım	1-2	Kategorik	2
182	UNIVERSITE_KAZANMA	B49.6 Son bir yılda, Üniversiteyi kazandım	1-2	Kategorik	2
183	OKUL_MEZUN	B49.7 Son bir yılda, Okuldan mezun oldum	1-2	Kategorik	2
184	SAGLIK_PROBLEM	B49.8 Son bir yılda, Ciddi bir sağlık problemi yaşadım.	1-2	Kategorik	2
185	ASKERLIK	B49.9 Son bir yılda, Askerliğimi bitirdim (terhis oldum).	1-3	Kategorik	3
186	TAHLIYE	B49.10 Son bir yılda, Tahliye oldum	1-3	Kategorik	3
187	TERFI_ETTI	B49.11 Son bir yılda, Terfi ettim	1-2	Kategorik	2
188	ITIBAR_SIRA1	B50.1 Toplumda itibarlı olmak için sizce aşağıdakilerden hangisi 1. derecede önemlidir?	1-6, 98	Kategorik	7
189	ITIBAR_SIRA2	B50.2 Toplumda itibarlı olmak için sizce aşağıdakilerden hangisi 2. derecede önemlidir?	1-6, 98	Kategorik	7
190	ITIBAR_SIRA3	B50.3 Toplumda itibarlı olmak için sizce aşağıdakilerden hangisi 3. derecede önemlidir?	1-6, 98	Kategorik	7
191	KIYAFET_ONEM	B51.1 Çevrenizdeki diğer kişilerin, kılık kıyafetleri sizin için ne kadar önemlidir?	1-3	Kategorik	3
192	AILE_YASAM_ONEM	B51.2 Çevrenizdeki diğer kişilerin, aile yaşam biçimleri sizin için ne kadar önemlidir?	1-3	Kategorik	3
193	KISISEL_ESYA_ONEM	B51.3 Çevrenizdeki diğer kişilerin, ev ve kişisel eşyaları (cep tel, araba vb.) sizin için ne kadar önemlidir?	1-3	Kategorik	3
194	ARKADAS_CEVRE_ONEM	B51.4 Çevrenizdeki diğer kişilerin, arkadaş çevreleri sizin için ne kadar önemlidir?	1-3	Kategorik	3
195	COCUK_BASARI_ONEM	B51.5 Çevrenizdeki diğer kişilerin, çocuklarının başarıları sizin için ne kadar önemlidir?	1-3	Kategorik	3
196	EVE_GIRIS_SAAT_ONEM	B51.6 Çevrenizdeki insanların, eve giriş çıkış saatleri sizin için ne kadar önemlidir?	1-3	Kategorik	3

197	YAPTIKLARI_IS_ONEM	B51.7 Çevrenizdeki diğer kişilerin, yaptıkları işleri sizin için ne kadar önemlidir?	1-3	Kategorik	3
198	GELIR_DUZEYLERI_ONEM	B51.8 Çevrenizdeki diğer kişilerin, gelir düzeyleri sizin için ne kadar önemlidir?	1-3	Kategorik	3
199	DINI_INANC_ONEM	B51.9 Çevrenizdeki diğer kişilerin, dini inançları sizin için ne kadar önemlidir?	1-3	Kategorik	3
200	SIYASI_DUSUNCE_ONEM	B51.10 Çevrenizdeki diğer kişilerin, siyasi düşünceleri sizin için ne kadar önemlidir?	1-3	Kategorik	3
201	EGITIM_DUZEYI_ONEM	B51.11 Çevrenizdeki diğer kişilerin, eğitim düzeyleri sizin için ne kadar önemlidir?	1-3	Kategorik	3
202	KIYAFETINIZ_ONEM	B52.1 Çevrenizdeki kişilerin (akraba, arkadaş, komşu vb.) sizinle ilgili kıyafetiniz hakkındaki düşünceleri sizin için ne kadar önemlidir?	1-3	Kategorik	3
203	AILE_YASAM_BICIMI_ONEM	B52.2 Çevrenizdeki kişilerin (akraba, arkadaş, komşu vb.) sizinle ilgili aile yaşam biçiminiz hakkındaki düşünceleri sizin için ne kadar önemlidir?	1-3	Kategorik	3
204	KISISEL_ESYANIZ_ONEM	B52.3 Çevrenizdeki kişilerin (akraba, arkadaş, komşu vb.) sizinle ilgili ev ve kişisel eşyalarınız hakkındaki düşünceleri (cep tel, araba vb.) sizin için ne kadar önemlidir?	1-3	Kategorik	3
205	ARKADAS_CEVRENIZ_ONEM	B52.4 Çevrenizdeki kişilerin (akraba, arkadaş, komşu vb.) sizinle ilgili arkadaş çevreniz hakkındaki düşünceleri sizin için ne kadar önemlidir?	1-3	Kategorik	3
206	COCUGUN_BASARISI_ONEM	B52.5 Çevrenizdeki kişilerin (akraba, arkadaş, komşu vb.) sizinle ilgili çocuklarınızın başarıları hakkındaki düşünceler sizin için ne kadar önemlidir?	1-4	Kategorik	4

207	EVE_GIRISCIKIC_SAA T_ONEM	B52.6 Çevrenizdeki kişilerin (akraba, arkadaş, komşu vb.) sizinle ilgili eve giriş çıkış saatiniz hakkındaki düşünceleri sizin için ne kadar önemlidir?	1-3	Kategorik	3
208	YAPTIGINIZ_IS_ONEM	B52.7 Çevrenizdeki kişilerin (akraba, arkadaş, komşu vb.) sizinle ilgili yaptığınız işiniz hakkındaki düşünceleri sizin için ne kadar önemlidir?	1-3	Kategorik	3
209	GELIR_DUZEY_ONEM	B52.8 Çevrenizdeki kişilerin (akraba, arkadaş, komşu vb.) sizinle ilgili geliriniz hakkındaki düşünceleri sizin için ne kadar önemlidir?	1-4	Kategorik	4
210	DINI_INANCI_ONEM	B52.9 Çevrenizdeki kişilerin (akraba, arkadaş, komşu vb.) sizinle ilgili dini inançlarınız hakkındaki düşünceleri sizin için ne kadar önemlidir?	1-3	Kategorik	3
211	SIYASI_DUSUNCENIZ_ONEM	B52.10 Çevrenizdeki kişilerin (akraba, arkadaş, komşu vb.) sizinle ilgili siyasi düşünceleriniz hakkındaki düşünceleri sizin için ne kadar önemlidir?	1-3	Kategorik	3
212	EGITIM_DURUMU_ONEM	B52.11 Çevrenizdeki kişilerin (akraba, arkadaş, komşu vb.) sizinle ilgili eğitim düzeyiniz hakkındaki düşünceleri sizin için ne kadar önemlidir?	1-3	Kategorik	3
213	CINSIYET_BASKI	B53.1 Cinsiyetinizden dolayı üzerinizde herhangi bir toplumsal baskı hissediyor musunuz?	1-4	Kategorik	4
214	MEDENI_DURUM_BASKI	B53.2 Medeni durumunuzdan dolayı üzerinizde herhangi bir toplumsal baskı hissediyor musunuz?	1-4	Kategorik	4
215	YAS_BASKI	B53.3 Yaşınızdan dolayı üzerinizde herhangi bir toplumsal baskı hissediyor musunuz?	1-4	Kategorik	4

216	GELENEK_GORENEK	B53.4 Gelenek göreneklerinizden dolayı üzerinizde herhangi bir toplumsal baskı hissediyor musunuz?	1-4	Kategorik	4
217	DINI_INANCDAN	B53.5 Dini inanç ve davranışlarınızdan dolayı üzerinizde herhangi bir toplumsal baskı hissediyor musunuz?	1-4	Kategorik	4
218	SIYASI_GORUSTEN	B53.6 Siyasi görüşünüzden dolayı üzerinizde herhangi bir toplumsal baskı hissediyor musunuz?	1-4	Kategorik	4
219	MEMLEKETINDEN	B53.7 Memleketinizden dolayı üzerinizde herhangi bir toplumsal baskı hissediyor musunuz?	1-4	Kategorik	4
220	ISINDEN_DOLAYI	B53.8 İşinizden dolayı üzerinizde herhangi bir toplumsal baskı hissediyor musunuz?	1-4	Kategorik	4
221	KILIK_KIYAFET	B53.9 Kılık kıyafetinizden dolayı üzerinizde herhangi bir toplumsal baskı hissediyor musunuz?	1-4	Kategorik	4
222	ISSIZ_OLMA	B53.10 İşsiz, çalışmıyor olmaktan dolayı üzerinizde herhangi bir toplumsal baskı hissediyor musunuz?	1-4	Kategorik	4
223	GELIR_DUZEY_DURUM	B53.11 Gelir düzeyinizden dolayı üzerinizde herhangi bir toplumsal baskı hissediyor musunuz?	1-4	Kategorik	4
224	CEVRE_DOGA	B54.1 Çevre, doğa sorunları(hava, su kirliliği vb.) ile ne derece ilgilisiniz?	1-4	Kategorik	4
225	BILIM	B54.2 Bilim ve teknoloji ile ne derece ilgilisiniz?	1-4	Kategorik	4
226	KULTUR	B54.3 Kültür, sanat, edebiyat ile ne derece ilgilisiniz?	1-4	Kategorik	4
227	SENDIKA	B54.4 Sendika/Dernek faaliyetleri ile ne derece ilgilisiniz?	1-4	Kategorik	4
228	EKONOMI	B54.5 Ekonomi ile ne derece ilgilisiniz?	1-4	Kategorik	4
229	SIYASET	B54.6 Siyaset ile ne derece ilgilisiniz?	1-4	Kategorik	4
230	SPOR	B54.7 Spor ile ne derece ilgilisiniz?	1-4	Kategorik	4

231	MODA	B54.8 Moda ile ne derece ilgilisiniz?	1-4	Kategorik	4
232	DİN	B54.9 Din ile ne derece ilgilisiniz?	1-4	Kategorik	4
233	SAGLIK_KONU	B54.10 Sağlık ile ilgili konularda ne derece ilgilisiniz?	1-4	Kategorik	4
234	MUZİK	B54.11 Müzik ile ne derece ilgilisiniz?	1-4	Kategorik	4
235	ULKE_SORUN_SIRA1	B55-1 Sizce Ülkemizin en önemli 1. sorunu nedir?	1-12, 98	Kategorik	13
236	ULKE_SORUN_SIRA2	B55-2 Sizce Ülkemizin en önemli 2. sorunu nedir?	1-12, 98	Kategorik	13
237	ULKE_SORUN_SIRA3	B55-3 Sizce Ülkemizin en önemli 3. sorunu nedir?	1-12, 98	Kategorik	13
238	FAKTOR_FERT	Anket yapılan fert hesaplanan ağırlık değeri	1851,872-26236,192	Sürekli	-

KAYNAKLAR

1. Bonikowska, A., Helliwell, J.F., Hou, F., Schellenberg, G., 2014. An assesment of life satisfaction responses on recent statics Canada surveys. *Social Indicators Research*, 118(2): 617-643.
2. Chou, K.L., Chi, I., 2002. Financial strain and life satisfaction in Hong kong elderly Chinese: Moderation effect of life management strategies including selection, optimization, and compensation. *Aging-Mental Health*, 6(2): 172-177.
3. Boyce, C.J.,Wood, A.M., Ferguson, E., 2016. For better or for worse: the moderating effects of personality on the marriage-life satisfaction link. *Personality and Individual Differences*, 97: 61-66.
4. Fleming, C.M., Manning, M., Ambrey, C.L., 2016. Crime, green space and life satisfaction: An evaluation of the New Zealand experience. *Landscape and Urban Planning*, 149: 1-10.
5. York, R., Bell, S.E., 2014. Life satisfaction across nations: the effects of women's political status and public priorities. *Social Science Research*, 48: 48-61.
6. Chen, J., Davis, D. S., Wu, K., Dai, H., 2015. Life satisfaction in urbanizing China: The effect of city size and path ways to urban residency. *Cities*, 49: 88-97.
7. Du, H., Bernardo, A.B.I., Yeung, S. S., 2015. Locus-of-hope and life satisfaction: the mediating roles of personal self-esteem and relational self-esteem. *Personality and Individual Differences*, 83: 228-233.
8. Haar, J.M., Russo, M., Sune, A., Ollier-Malaterre, A., 2014. Out comes of work–life balance on job satisfaction, life satisfaction and mental health: A study across seven cultures. *Journal of Vocational Behavior*, 85(3): 361-373.
9. Samaha, M., Hawi, N.S., 2016. Relationships among smart phone addiction, stress, academic performance, and satisfaction with life. *Computers in Human Behavior*, 57: 321-325.
10. Demircan, S., 2015. TÜİK Yaşam Memnuniyeti Anketleri Üzerine Veri Madenciliği Uygulaması. Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 142s.
11. Özkan, Y., 2008. Veri madenciliği Yöntemleri. Papatya Yayıncılık, İstanbul, 214s.
12. Frawley, W. J., Piatetsky-Shapiro, G., Matheus, C. J., 1991. Knowledge discovery databases: An overview, pp. 1-27. In *Knowledge Discovery In Databases* (G. Piatetsky-Shapiroand W. J.,Frawleyeds.), MA:AAAI/MIT, Cambridge.

13. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., 1996. From data mining to knowledge discovery: An overview. *AI Magazine*, 17(3): 37-54.
14. Piatetsky-Shapiro, G., 1991. Knowledge discovery in real databases: A workshop report. *AI Magazine*, 11(5): 68-70.
15. Albayrak, A.S., Koltan Yılmaz, Ş., 2009. Veri madenciliği: Karar ağacı algoritmaları ve İMKB verileri üzerine bir uygulama. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 14(1): 31-52.
16. Akpınar, H., 2000. Veri tabanlarında bilgi keşfi ve veri madenciliği. *İ.Ü. İşletme Fakültesi Dergisi*, 29(1): 1-22.
17. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., 1996. The KDD rocess for extracting useful knowledge from volumes of data. *Communications of ACM*, 39(11): 27-34.
18. Kulluk, S., 2009. Karınca Koloni Optimizasyonu ile Yapay Sinir Ağlarından Kural Çıkarımı. *Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi, Kayseri*, 305s.
19. Tüzüntürk, S., 2010. Veri madenciliği ve istatistik. *Uludağ Üniversitesi İktisadi ve İdari Bilimler Dergisi*, XXIX(1): 65-90.
20. Savaş, S., Topaloğlu, N., Yılmaz, M., 2012. Veri madenciliği ve Türkiye'deki uygulama örnekleri. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 21: 1-23.
21. Kaçmaz Dağ, B., 2015. Veri Madenciliği ve Hizmet Sektöründe Bir Uygulama. *Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi*, 71s.
22. Aggarwal, C.C., 2014. Introduction to data classification, pp.1-36. In: *Data Classification Algorithms and Applications*(Eds. V. Kumar). Taylor & Francis Group, Boca Raton.
23. Hall, M. A., 1999. Correlataion-based Feature Selection for Machine Learning. *The University of Waikoto, Department of Computer Science, Ph. D. Theses*, 178pp.
24. Gökgöz, K.B., 2015. Veri Madenciliği ile Tıbbi Cihaz Bakım Karar Modeli. *Başkent Üniversitesi, Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, Ankara*, 114s.
25. Quinlan, J.R., 1986. Introduction of decision trees. *Machine Learning*, 1: 81-106.

26. Nenonen, N., 2013. Analysing factors related to slipping, stumbling, and falling accidents at work: Application of data mining methods to Finnish occupational accidents and diseases statistics database. *Applied Ergonomics*, 44: 215-224.
27. McCarthy, J., Jacob, T., Atkinson, D., 2009. Innovative uses of data mining techniques in the production of official statistics. (Web page: https://fcsmls.usa.gov/files/2014/05/2009FCSM_McCarthy_X-A.pdf), (Date accessed: April 2015).
28. McCharthy, J., Jacob, T., Atkinson, D., 2009. Innovative uses of data mining techniques in the production of official statistics. United Nations 2009 Statistical Commission Session on Innovations in Official Statistics. (Web page: http://www.fcsmls.gov/09papers/McCarthy_X-A.pdf), (Date accessed: April 2015).
29. Miller, D., McCarthy, J., Zakzeski A., 2009. A fresh approach to agricultural statistic: Data mining and remote sensing, pp. 3144-3155. American Statistical Association Proceedings of the Joint Statistical Meetings 2009, Washington DC.
30. Hassani, H., Gheitanchi, S., Yeganegi, M. R., 2008. On the application of data mining to official data. *Journal of Data Science*, 8: 75-89.
31. Schouten, B., Nooji, G.D., 2005. Nonresponse Adjustment Using Classification Trees. Statistics Netherlands Discussion Paper 05001, Statistics Netherlands, Vooburg/Heerlen, 26pp.
32. Değirmenci, T., 2014. Resmi İstatistiklerde Veri Madenciliği Yaklaşımı. Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 120s.
33. Cura, T., 2008. Kümeleme için bir benzetilmiş tavlama algoritması yaklaşımı. *Yönetim*, 59: 43-51.
34. Xu Z., Wang, L., Lou, J., Zhang, J., 2005. A modified clustering algorithm for data mining, pp. 741-744. IEEE Geoscience and Remote Sensing Symposium, Seoul.
35. Jose-Garcia, A., Gomez-Flores, W., 2016. Automatic clustering using nature-inspired metaheuristics: A survey. *Applied Soft Computing*, 41: 192-213.
36. Madhulatha, T.S., 2012. An overview on clustering methods. *IOSR Journal of Engineering*, 2(4): 719-725.
37. Al-Jabbouli, H., 2009. Data Clustering Using the Bees Algorithm and the Kd-tree Structure. Cardiff University, Ph. D. Theses, 181 pp.

38. Durmuş, M. S., 2005. Veri Kümeleme Algoritmalarının Performansları Üzerine Karşılaştırmalı Bir Çalışma. Pamukkale Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Denizli, 119s.
39. Ada, M., Altunay, F., Civelek, M., Kaplan, S., Koç, P., 2011. Kümeleme analizi.(Websayfası:<http://tip.baskent.edu.tr/kw/upload/600/dosyalar/cg/sempozyum/ogrsmpzsnm13/13.P9.pdf>),(Erişim Tarihi: Nisan 2015).
40. Hançer, E., 2016. Yapay Arı Koloni Algoritması Tabanlı Görüntü Kümeleme Yöntemlerinin Geliştirilmesi. Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi, Kayseri, 179s.
41. Jain, A. K., Murty, M. N., Flynn, P. J., 1999. Data clustering: A review. ACM Computing Surveys, 31(3): 264-323.
42. Sarstedt, M., Mooi, E., 2011. A Concise Guide to Market Research: The Process, Data, and Methods Using IBM SPSS Statistics. Springer-Verlag Berlin Heidelberg, 308pp.
43. Zhang, C., Ouyang, D., Ning, J., 2010. An artificial bee colony approach for clustering. Expert Systems with Applications, 37(7): 4761-4767.
44. Shelokar, P. S. Jayaraman, V. K., Kulkarni, B. D., 2004. An ant colony approach for clustering. Analytica Chimica Acta, 509(2): 187-195.
45. Cura, T., 2012. A particle swarm optimization approach to clustering. Expert Systems with Applications, 39: 1582-1588.
46. Durak, B., 2011. A Classification Algorithm Using Mahalanobis Distance Clustering Of Data With Applications On Biomedical Data Sets. Middle East Technical University, The Graduate School Of Natural And Applied Sciences, Master of Science Thesis, 91pp.
47. Khan, S. S., Kant, S., 2007. Computation of Initial Modesfor K-modes Clustering Algorithm Using Evidence Accumulation, pp. 2784-2789. International Joint Conference on Artificial Intelligence, Hyderabad, India.
48. Sharma, N., Gaud, N., 2015. K-modes clustering algorithm for categorical data. International Journal of ComputerApplications, 127(17): 1-6.
49. Wilson, D. R., Martinez, T. R., 1997. Improved heterogeneous distance functions. Journal of Artificial Intelligence Research, 6: 1-34.
50. Stanfill, C., Waltz, D., 1986. Toward memory-based reasoning. Communications of the ACM, 29(12): 1213-1228.

51. Malhotra, N., Birks, D., 2007. Marketing Research-An Applied Approach. Prentice Hall/Financial Times, 835pp.
52. Miguel, P, Gonçalves, J., Neves, L., Martins, A. G., 2016. Using clustering techniques to provide simulation scenarios for the smart grid. Sustainable Cities and Society(In press).
53. Manning, C.D., Schütze, H., 1999. Foundations of Statistical Natural Language Processing. MIT Press, Cambridge, 680pp.
54. Karaboğa, D., Ozturk, C., 2011. A novel clustering approach: Artificial bee colony(ABC) algorithm. Applied Soft Computing, 11(1): 652-657.
55. MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations, pp. 281–297. 5th Berkeley Symposium Mathematical Statistics and Probability.
56. Ozturk, C., Hançer, E., Karaboğa, D. 2014. Küresel en iyi yapay arı koloni algoritması ile otomatik kümeleme. Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi 29(4): 677-687.
57. Tan, P. N., Steinbach, M., Kumar, V., 2005. Introduction to Data Mining. Pearson, 769pp.
58. Kaufman, L., Rousseeuw, P.J., 1987. Clustering by means medoids, pp. 405-416. Y. Dodge(Ed.), Statistical Data Analysis Based on the L1 Norm, North-Holland.
59. Arora, P., Deepali, Dr., Varshney, S., 2016. Analysis of k-means and k-medoids algorithm for big data. Procedia Computer Science,78: 507-512.
60. Park, H. S., Jun, C. H., 2009. A simple and fast algorithm for k-medoids clustering. Expert Systems with Applications, 36(2): 3336-3341.
61. Kaufman, L., Rousseeuw, P. J., 1990. Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley&Sons, 342pp.
62. Raymond, T. N., Han, J., 2002. CLARANS: A method for clustering objects for spatial data mining, IEEE Transactions On Knowledge And Data Engineering, 14(5): 1003-1016.
63. Revathi, T., Sumathi, P., 2013. A survey on data mining using clustering techniques. International Journal of Scientific&Engineering Research, 4(1): 1-5.
64. Han, J., Kamber, M., 2001. Data Mining: Concepts and Techniques. Morgan Kaufman, 703pp.

65. Ester, M., Kriegel, H. P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD*, 96(34): 223-231.
66. Lv, Y., Ma, T., Tang, M., Cao, J., Tian, Y., Al-Dhelaan, A., Al-Rodhaan, M., 2016. An efficient and scalable density-based clustering algorithm for datasets with complex structures. *Neurocomputing*, 171: 9-22.
67. Abdullah, M., Eldin, H. N., Al-Moshadak, T., Alshaik, R., Al-Anesi, I., 2015. Density grid-based clustering for wireless sensors networks. *Procedia Computer Science*, 65: 35-47.
68. Kuo, R.J., Ho, L.M., Hu, C.M., 2002. Cluster analysis in industrial market segmentation through artificial neural network. *Computers&Industrial Engineering*, 42(2-4): 391-399.
69. Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43: 56-69.
70. Michalewicz, Z., 1996. Genetic Algorithms + Data Structure = Evolutionary Programs. 2 nd Edition, Springer-Verlag, Berlin, 387pp.
71. Engelbrecht, A.P., 2005. Fundamentals of Computational Swarm Intelligence. John Wiley and Sons Ltd, England, 672pp.
72. Kennedy, J., Eberhart, R., 1995. Particle Swarm Optimization, pp. 1942-1948. In *Proceedings of The IEEE International Conference on Neural Networks*, Piscataway, NJ.
73. Dorigo, M., 1992. Optimization, Learning and Natural Algorithms. Dipartimento di Elettronica, Politecnico di Milano, Italy, Ph.D Thesis, 140pp.
74. Simon, D., 2008. Biogeography-based Optimization, *IEEE Transactions on Evolutionary Computing*, 12(6): 702–713.
75. Rashedi, E., Nezamabadi-Pour, H., Saryazdi, S., 2009. GSA: A gravitational search algorithm. *Information Sciences*, 179(13): 2232–2248.
76. Kaveh, A., Khayatazad, M., 2012. A new meta-heuristic method: Ray optimization, *Computers& Structures*, 112-113: 283–294.
77. Erol, O. K., Eksin, I., 2006. A new optimization method: Big bang–big crunch, *Advances in Engineering Software*, 37(2): 106–111.
78. Alatas, B., 2011. ACROA: Artificial chemical reaction optimization algorithm for global optimization, *Expert Systems with Applications*, 38(10): 13170–13180.

79. Hatamlou, A., 2013. Black hole: A new heuristic optimization approach for data clustering. *Information Sciences*, 222: 175–184.
80. Cuevas, E., Echavarría, A., Ramírez-Ortegón, M.A., 2014. An optimization algorithm inspired by the states of matter that improves the balance between exploration and exploitation. *Applied Intelligence*, 40(2): 256–272.
81. Moein, S., Logeswaran, R., 2014. KGM0. A swarm optimization algorithm based on the kinetic energy of gas molecules. *Information Sciences*, 275: 127–144.
82. Nanda, S.J., Panda, G., 2014. A survey on nature inspired metaheuristic algorithms for partitional clustering. *Swarm and Evolutionary Computation*, 16: 1-18.
83. Güngör, Z., Ünler, A., 2007. K-harmonik means data clustering with simulated annealing heuristic. *Applied Mathematics and Computation*, 184(2): 199-209.
84. Kao, Y.T, Zahara, E., Kao, I.W., 2008. A hybridized approach to data clustering. *Expert Systems with Applications*, 34(3): 1754-1762.
85. Carpenter, G., Grossberg, S., 1991. *Pattern Recognition by Self-Organizing Neural Networks*, Cambridge, Massachusetts, USA, 710pp.
86. Krishna, K., Murty, M., 1999. Genetic k-means algorithm. *IEEE Transactions on Systems, Man and Cybernetics*, 29(3): 433-439.
87. Ruspini, E.H., 1969. A new approach to clustering. *Information and Control*, 15(1): 22-32.
88. Jensi, R., Jiji, W., 2015. Hybrid data clustering approach using k-means and flower pollination algorithm. *Advanced Computational Intelligence: An International Journal (ACII)*, 2(2): 15-25.
89. Huang, C.L., Huang, W.C., Chang, H.Y., Yeh, Y.C., Tsai, C.Y., 2013. Hybridization strategies for continuous ant colony optimization and particle swarm optimization applied to data clustering. *Applied Soft Computing*, 13(9): 3864-3872.
90. İnkaya, T., Kayalığil, S., Özdemirel, N. E., 2015. Ant colony optimization based clustering methodology. *Applied Soft Computing*, 28: 301-311.
91. Akarsu, E., Karahoca, A., 2011. Simultaneous feature selection and ant colony clustering. *Procedia Computer Science*, 3: 1432-1438.
92. Shabanzadeh, P., Yusof, R., 2015. An efficient optimization method for solving unsupervised data classification problems. *Computational and Mathematical Methods in Medicine*, Hindawi Publishing Corporation, 9pp.

93. Gao, Y., Ren, Z., Xu, C., 2007. A branch and bound-pso hybrid algorithm for solving integer seperable concave programming problems. *Applied Mathematical Sciences*, 1(11): 517-525.
94. Baş, F., 2013. Kategorik Veri Analizinde Kullanılan Algoritmaların Performanslarının Karşılaştırılması Üzerine Bir Çalışma. Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 51s.
95. Izakian, H., Abraham, A., Snasel, V., 2009. Clustering categorical data using a swarm-based method, pp. 1720-1724. *World Congress on Nature &Biologically Inspired Computing(NaBIC 2009)*.
96. Cao, F., Liang, J., Li, D., Bai, L., Dang, C., 2012. A dissimilarity measure for the k-modes clustering algorithm. *Knowledge Based Systems*, 26: 120-127.
97. Qin, H., Ma, X., Herawan, T., Zain, J.M., 2014. MGR: An information theory based hierarchical divisive clustering algorithm for categorical data. *Knowledge Based Systems*, 67: 401-411.
98. Agarwal, P., Alam, M. A., Biswas, R., 2011. Issues, challenges and tools of clustering algorithms. *IJCSI International Journal of Computer Science Issues*, 8(3): 523-528.
99. Javidy, B., Hatamlou, A., Mirjalili, S., 2015. Ion motions algorithm for solving optimization problems. *Applied Soft Computing*, 32: 72-79.
100. Baykasoğlu, A., Akpınar, Ş., 2015. Weighted superposition attraction(WSA) : A swarm intelligence algorithm for optimization problems-Part 1: Unconstrained optimization. *AppliedSoft Computing*, (In press).
101. Baykasoğlu, A., Akpınar, Ş., 2015. Weighted superposition attraction(WSA) : A swarm intelligence algorithm for optimization problems-Part 2: Constrained optimization. 37: 396-415.
102. Taguchi, G., 1986. *Introduction to Quality Engineering: Designing Quality into Products and Processes*. Quality Resources, 191pp.
103. Phadke, S.M., 1989. *Quality Engineering Using Robust Design*. Prentice Hall, Englewood Cliffs, N.J., 333pp.
104. Rini, D.P., Shamsuddin, S.M., Yuhani, S.S., 2011. Particle swarm optimization: Technique, system and challenges. *International Journal of Computer Applications*, 14(1): 19-27.

105. Bratton, D., Kennedy, J., 2007. Defining a standard for particle swarm optimization, pp. 120-127. IEEE Swarm Intelligence Symposium, April 2007, Honolulu.
106. Küçükdeniz, T., 2009. Sürü Zekası Optimizasyon Tekniği ve Tedarik Zinciri Yönetiminde Bir Uygulama. İstanbul Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi, 93s.
107. Özyön, S., Yaşar, C., Temurtaş, H., 2011. Particle swarm optimization algorithm for the solution of nonconvex economic dispatch problem with valve point effect. pp. 101-105. ELECO 2011 7th International Conference on Electrical and Electronics Engineering, (December 2011, Bursa, Turkey).
108. Trelea, I.C., 2003. The particle swarm optimization algorithm: Convergence analysis and parameter selection. Information Processing Letters, 85: 317-325.
109. Vaneshani, S., Jazayeri-Rad, H., 2011. Optimized Fuzzy Control by Particle Swarm Optimization Technique for Control of CSTR. International Journal of Computer, Electrical, Automation, Control and Information Engineering, 5(11): 1243-1248.
110. Soliman, O.S., AboElhamd, E., 2014. Classification of Diabetes Mellitus using Modified Particle Swarm Optimization and Least Squares Support Vector Machine. International Journal of Computer Trends and Technology (IJCTT), 8(1): 38-44.
111. Rahmani, M., 2008. Particle swarm optimization of artificial neural networks for autonomous robots. Chalmers University of Technology, Department of Applied Physics, Ph.D. Thesis, 40pp.
112. Karaboga, D., 2005. An idea based on honey bee swarm for numerical optimization. (Web page: http://mf.erciyes.edu.tr/abc/pub/tr06_2005.pdf), (Date accessed: June 2015).
113. Gupta, T., Kumar, D., 2014. Optimization of clustering problem using population based artificial bee colony algorithm: A review. International Journal of Advanced Research in Computer Science and Software Engineering, 4(4): 491502.
114. Karaboga, D., Basturk, B., 2007. Artificial Bee Colony (ABC) Optimization Algorithm for Solving Constrained Optimization Problems, pp. 789-798. 12th

- International Fuzzy Systems Association World Congress, IFSA 2007, (June 2007, Cancun, Mexico).
115. Karaboğa, D., Akay, B., 2009. A comparative study of artificial bee colony algorithm. *Applied Mathematics and Computation*, 214(1): 108-132.
 116. Gong, A., Gao, Y., Ma, X., Gong, W., Li, H., Gao, Z., 2016. An optimized artificial bee colony algorithm for clustering. *International Journal of Control and Automation*, 9(4): 107-116.
 117. Xiang, X., 2013. Improved Artificial Bee Colony Algorithm Based on Neighborhood Search. *International Journal of Advancements in Computing Technology(IJACT)*, 5(4): 125-132.
 118. Kuo, R.J., Huang, Y.D., Lin, C.C., Wu, Y.H., Zulvia, F.E., 2014. Automatic kernel clustering with bee colony optimization algorithm. *Information Sciences*, 283: 107-122.
 119. Zou, W., Zhu, Y., Chen, H., Sui, X., 2010. A clustering approach using cooperative artificial bee colony algorithm. *Discrete Dynamics Nature and Society*, 16pp.
 120. Liu, R., Chen, Y., Jiao, L., Li, Y., 2014. A particle swarm optimization based simultaneous learning framework for clustering and classification. *Pattern Recognition*, 47(6): 2143-2152.
 121. Tiejun, Z., Jing, Y., Jianpei, Z., 2011. Clustering algorithm for mixed attributes data based on restricted particle swarm optimization, pp. 350-354. 2011 International Conference on Computer Science and Information Technology, Singapore.
 122. Matsui, T., Kato, K., Sakawa, M., Uno, T., Matsumoto, K., 2008. Particle Swarm Optimization for Nonlinear Integer Programming Problems, pp. 978-988. International MultiConference of Engineers and Computer Scientists, March 2008, Hong Kong.
 123. Ji, J., Pang, W., Zheng, Y., Wang, Z., Ma, Z., 2015. A novel artificial bee colony based clustering algorithm for categorical data. *PLoS ONE*, 10(5): 1-17.
 124. Marinakis, Y., Marinaki, M., Matsatsinis, N., 2009. A hybrid discrete artificial bee colony-GRASP Algorithm for clustering, pp. 548-553. CIE 2009, International Conference on Computers & Industrial Engineering, Troyes, Paris.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı, Soyadı: Fatma TURNA

Uyruğu: Türkiye (TC)

Doğum Tarihi ve Yeri: 17 Nisan 1982, Zonguldak

Medeni Durumu: Evli

Tel: +90 551 150 70 20

email: fatmaturna2007@gmail.com

Yazışma Adresi: Konaklar Mah. Avanos Cad. Kınaş Koru Sitesi No:174/17 Melikgazi
KAYSERİ

EĞİTİM

Derece	Kurum	Mezuniyet Tarihi
Yüksek Lisans	EÜ Fen Bilimleri Enstitüsü	2011
Lisans	Kırıkkale Üniversitesi Mühendislik Fakültesi Endüstri Mühendisliği	2003
Lise	Sultanbeyli İmam-Hatip Lisesi	1999

İŞ DENEYİMLERİ

Yıl	Kurum	Görev
2010- Halen	Kayseri Büyükşehir Belediyesi	Mühendis
2009-2010	İstanbul Büyükşehir Belediyesi	Mühendis
2004-2006	Bolero Kırtasiye A.Ş.	Mühendis

YABANCI DİL

İngilizce

YAYINLAR

1. Turna F., Gürbüz, F., “Veri Madenciliği Teknikleriyle Tramvay Arıza Kayıtlarından Kural Çıkarımı”, Yöneylem Araştırması ve Endüstri Mühendisliği 32. Ulusal Kongresi, İstanbul, Türkiye, 20-22 Haziran 2012, ss.1-10.