



**T.C.  
ONDOKUZ MAYIS ÜNİVERSİTESİ  
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ  
ZOOTEKNİ ANA BİLİM DALI**

**BAYES VE BAZI İKİLİ KÜMELEME ALGORİTMALARININ  
ZOOTEKNİ VERİLERİNDE KULLANIMI**

Doktora Tezi

**Lütfi BAYYURT**

**DANIŞMAN  
Prof. Dr. Hasan ÖNDER**

**SAMSUN  
2022**

T.C.  
ONDOKUZ MAYIS ÜNİVERSİTESİ  
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ  
ZOOTEKNİ ANA BİLİM DALI



**BAYES VE BAZI İKİLİ KÜMELEME ALGORİTMALARININ  
ZOOTEKNİ VERİLERİNDE KULLANIMI**

Doktora Tezi

**Lütfi BAYYURT**



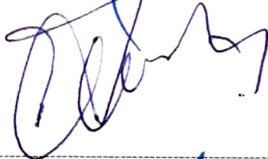


Danışman

**Prof. Dr. Hasan ÖNDER**

SAMSUN  
2022

## TEZ KABUL VE ONAYI

Lütfi BAYYURT tarafından, Prof. Dr. Hasan ÖNDER danışmanlığında hazırlanan “BAYES VE BAZI İKİLİ KÜMELEME ALGORİTMALARININ ZOOTEKNİ VERİLERİNDE KULLANIMI” başlıklı bu çalışma, jürimiz tarafından 21.7.2022 tarihinde yapılan sınav sonucunda oy birliği ile başarılı bulunarak Doktora Tezi olarak kabul edilmiştir.

	Unvanı Adı Soyadı Üniversitesi Ana Bilim/Ana Sanat Dalı	İmza	Sonuç
Başkan	Prof. Dr. Hasan ÖNDER Ondokuz Mayıs Üniversitesi Zootečni Ana Bilim Dalı		<input checked="" type="checkbox"/> Kabul <input type="checkbox"/> Ret
Üye	Prof. Dr. Mustafa ŞAHİN Kahramanmaraş Sütçü İmam Üniversitesi Tarımsal Biyoteknoloji Ana Bilim Dalı		<input checked="" type="checkbox"/> Kabul <input type="checkbox"/> Ret
Üye	Doç. Dr. Taner TUNÇ Ondokuz Mayıs Üniversitesi İstatistik Ana Bilim Dalı		<input checked="" type="checkbox"/> Kabul <input type="checkbox"/> Ret
Üye	Doç. Dr. Yalçın TAHTALI Tokat Gaziosmanpaşa Üniversitesi Zootečni Ana Bilim Dalı		<input checked="" type="checkbox"/> Kabul <input type="checkbox"/> Ret
Üye	Doç. Dr. Samet Hasan ABACI Ondokuz Mayıs Üniversitesi Zootečni Ana Bilim Dalı		<input checked="" type="checkbox"/> Kabul <input type="checkbox"/> Ret

Bu tez, Enstitü Yönetim Kurulunca belirlenen ve yukarıda adları yazılı jüri üyeleri tarafından uygun görülmüştür.

ONAY

... / ... / ...

Prof. Dr. Ali BOLAT  
Enstitü Müdürü

## BİLİMSEL ETİĞE UYGUNLUK BEYANI

Hazırladığım Doktora tezinin bütün aşamalarında bilimsel etiğe ve akademik kurallara riayet ettiğimi, çalışmada doğrudan veya dolaylı olarak kullandığım her alıntıya kaynak gösterdiğimi ve yararlandığım eserlerin Kaynaklar'da gösterilenlerden oluştuğunu, her unsurun enstitü yazım kılavuzuna uygun yazıldığını ve TÜBİTAK Araştırma ve Yayın Etiği Kurulu Yönetmeliği'nin 3. bölüm 9. maddesinde belirtilen durumlara aykırı davranılmadığını taahhüt ve beyan ederim.

Etik Kurul Gerekli mi ?

Evet  (Gerekli ise ekler kısmına ekleyiniz)

Hayır

02 /06 / 2022  
Lütfi BAYYURT

## TEZ ÇALIŞMASI ÖZGÜNLÜK RAPORU BEYANI

**Tez Başlığı :** BAYES VE BAZI İKİLİ KÜMELEME ALGORİTMALARININ  
ZOOOTEKNİ VERİLERİNDE KULLANIMI

Yukarıda başlığı belirtilen tez çalışması için şahsım tarafından 02.06.2022 tarihinde intihal tespit programından alınmış olan özgünlük raporu sonucunda;

Benzerlik oranı : % 15

Tek kaynak oranı : % 4 çıkmıştır.

02 /06/ 2022  
Prof. Dr. Hasan ÖNDER

## ÖZET

### BAYES VE BAZI İKİLİ KÜMELEME ALGORİTMALARININ ZOOTEKNİ VERİLERİNDE KULLANIMI

Lütfi BAYYURT  
Ondokuz Mayıs Üniversitesi  
Lisansüstü Eğitim Enstitüsü  
Zootečni Ana Bilim Dalı  
Doktora, Temmuz/2022  
Danışman: Prof. Dr. Hasan ÖNDER

Bu çalışmada, iki farklı veri seti kullanılarak farklı ikili kümeleme algoritmalarının kümeleme performanslarının karşılaştırılması amaçlanmıştır. Çalışmada ilk olarak Sheepnet (Ağ oluşturma yoluyla Avrupa Birliği (AB) ve Türk koyun verimliliğini artırmaya yönelik uzmanlık ve deneyim paylaşımı) platformuna ait koyun yetiştiriciliğinde gebelik sorunları ve kuzu ölümlerini azaltmak amacıyla geliştirilen çözümler ve faydalarını içeren veri seti kullanılmıştır. Sheepnet veri seti kullanılarak yapılan ikili kümeleme analizi için Bimax, Xmotif ve Bayes algoritmaları kullanılmıştır. Algoritmaların ikili kümeleme performansları CKSB skoru, ortalama küme boyutları kullanılarak karşılaştırılmıştır. Karşılaştırma sonucunda en iyi ikili kümeleme performansını Bayes algoritması vermiştir.

Çalışmada ikinci olarak, 2018 yılında yapılan arıcılık faaliyetlerindeki öneminden dolayı seçilen beş ilden (Artvin, Düzce, Hatay, Kırklareli ve Muğla) dört yaygın bal arısı ırklarının genetik açıdan durumları, filogenetik ilişkileri ve popülasyonlarının genetik yapılarının belirlenmesi amacıyla 150 koloni ve 30 lokus içeren veri seti kullanılmıştır. Bal arısı veri seti kullanılarak yapılan ikili kümeleme analizi için Bimax, Xmotif, Bayes, CC ve Plaid algoritmaları kullanılmıştır. Algoritmaların ikili kümeleme performansları HKO, VAR, CKSB skoru ve ortalama küme boyutları kullanılarak karşılaştırılmıştır. Algoritmaların ikili kümeleme performansları CKSB skoru, ortalama küme boyutu, HKO ve VAR değerleri kullanılarak karşılaştırılmıştır. Karşılaştırma sonucunda en iyi ikili kümeleme performansını Bimax algoritması vermiştir.

**Anahtar Sözcükler:** Kümeleme; İkili kümeleme, Bal arısı, Koyun

## ABSTRACT

### USAGE OF BAYESIAN AND SOME BICLUSTERING ALGORITHMS IN ANIMAL SCIENCE DATA

Lütfi BAYYURT

Ondokuz Mayıs University

Institute of Graduate Studies

Department of Animal Science

Ph.D., July/2022

Supervisor: Prof. Dr. Hasan ONDER

In this study, it was aimed to compare the clustering performances of different binary clustering algorithms using two different data sets. In the study, firstly, the data set of the Sheepnet (Expertise and experience sharing to increase European Union (EU) and Turkish sheep productivity through networking) platform was used, which includes the solutions and benefits developed to reduce gestation problems and lamb deaths in sheep breeding. Bimax, Xmotif and Bayes algorithms were used for biclustering analysis using Sheepnet dataset. The biclustering performances of the algorithms were compared using the CKSB score and average cluster sizes. As a result of the comparison, Bayes algorithm gave the best bicluster performance.

Secondly, in the study, a data set containing 150 colonies and 30 loci was used to determine the genetic status, phylogenetic relationships and genetic structures of the populations of four common honey bee breeds from five provinces (Artvin, Düzce, Hatay, Kırklareli and Muğla) selected due to their importance in beekeeping activities carried out in 2018. Bimax, Xmotif, Bayes, CC and Plaid algorithms were used for biclustering analysis using honey bee dataset. The biclustering performances of the algorithms were compared using MSR, VAR, CKSB score and average cluster sizes. The biclustering performances of the algorithms were compared using the CKSB score, average cluster size, MSR and VAR values. As a result of the comparison, Bimax algorithm gave the best binary cluster performance.

**Keywords:** Clustering; Binary clustering, Honey bee, Sheep

## ÖN SÖZ VE TEŞEKKÜR

Doktora öğrenimimin her aşamasında bana destek veren ve tezimin hazırlanmasında çalışmalarım boyunca yardımını esirgemeyen ve beni teşvik eden, bilgi ve tecrübelerinden yararlandığım değerli danışman hocam Sayın Prof. Dr. Hasan ÖNDER'e teşekkürlerimi sunarım.

Tez İzleme Komitesi üyeleri Prof. Dr. Musa SARICA ve Doç. Dr. Taner TUNÇ hocalarıma ve tez jürisinde yer alan Prof. Dr. Mustafa ŞAHİN, Doç. Dr. Yalçın TAHTALI ve Doç. Dr. Samet Hasan ABACI hocalarıma tezimle alakalı değerli katkılarından dolayı teşekkür ederim.

Sadece bu süreçte değil, hayatım boyunca her zaman her koşulda yanımda olan ve benim için her şeyden kıymetli olan aileme çok teşekkür ederim.

Lütfi BAYYURT

# İÇİNDEKİLER

TEZ KABUL VE ONAYI .....	i
BİLİMSEL ETİĞE UYGUNLUK BEYANI .....	ii
TEZ ÇALIŞMASI ÖZGÜNLÜK RAPORU BEYANI .....	ii
ÖZET .....	iii
ABSTRACT .....	iv
ÖNSÖZ VE TEŞEKKÜR .....	v
İÇİNDEKİLER .....	vi
SİMGELER VE KISALTMALAR .....	vii
ŞEKİLLER DİZİNİ .....	viii
TABLolar DİZİNİ .....	ix
1. GİRİŞ .....	1
2. KAYNAK ÖZETLERİ .....	4
3. MATERYAL VE YÖNTEM .....	8
3.1. Materyal .....	8
3.2. Yöntem .....	9
3.2.1. Çalışmada Kullanılan Yöntemler .....	10
3.2.2. İkili Küme Değerlendirme Ölçüleri .....	19
4. BULGULAR .....	23
4.1. Sheepnet Veri Seti Bulguları .....	23
4.1.1. Bimax Algoritma Sonuçları .....	23
4.1.2. Xmotif Algoritma Sonuçları .....	25
4.1.3. Bayes Algoritma Sonuçları .....	27
4.2. Algoritmaların Karşılaştırılması .....	28
4.3. Bal Arısı Veri Seti Bulguları .....	29
4.3.1. Bimax Algoritması Sonuçları .....	29
4.3.2. Xmotif Algoritması Sonuçları .....	31
4.3.3. Plaid Algoritma Sonuçları .....	33
4.3.4. CC Algoritması Sonuçları .....	35
4.3.5. Bayes Algoritması Sonuçları .....	37
4.4. Algoritmaların Karşılaştırılması .....	40
5. SONUÇ .....	42
KAYNAKLAR .....	45
EKLER .....	49
ÖZ GEÇMİŞ .....	51

## SİMGELER VE KISALTMALAR

BBC	: Bayes ikili küme algoritması
BM	: Birleşmiş Milletler
CC	: Cheng ve Church algoritması
CKSB	: Chia ve Karuturi Ölçüsü
FABIA	: İkili küme edinmek için faktör analizi
ISA	: Yinelemeli işaret algoritması
LSS	: Katman kareler toplamı
HKO	: Hata Kareler Ortalaması
OPSM	: Bayes ikili küme algoritması
QUBIC	: Nitel ikili küme algoritması
VAR	: Varyans Ölçüsü

## ŞEKİLLER DİZİNİ

Şekil 1.1. John Snow kolera haritası .....	1
Şekil 3.1. Klasik Kümeleme gösterimi (solda) ile ikili kümeleme gösterimi (sağda) .....	9
Şekil 4.1. Bimax algoritması ikili küme üyelik grafiği .....	24
Şekil 4.2. Xmotif algoritması ikili küme üyelik grafiği .....	26
Şekil 4.3. Bimax algoritmasına ait çubuk grafiği .....	31
Şekil 4.4. Xmotif algoritması için çubuk grafiği .....	33
Şekil 4.5. Plaid algoritması için çubuk grafiği .....	35
Şekil 4.6. CC algoritması için çubuk grafiği .....	37
Şekil 4.7. Bayes algoritması çubuk grafiği .....	39

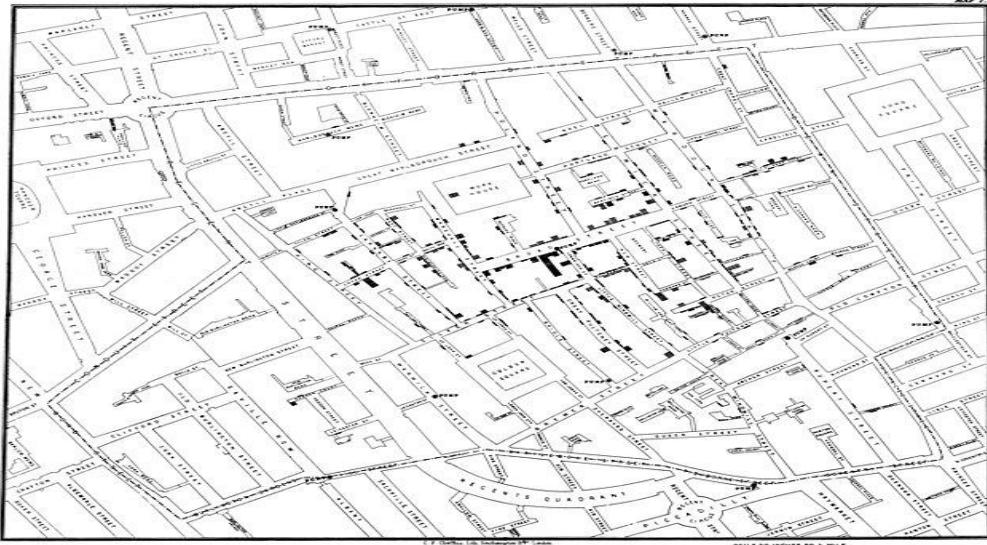
## TABLULAR DİZİNİ

Tablo 3.1. Plaid algoritması .....	16
Tablo 3.2. Bimax algoritması .....	17
Tablo 3.3. CC algoritması .....	18
Tablo 3.4. Xmotif algoritması .....	19
Tablo 4.1. Bimax algoritmasına göre küme içerikleri ve boyutları.....	23
Tablo 4.2. Xmotif algoritmasına göre küme içerikleri .....	25
Tablo 4.3. Bayes algoritmasına göre küme içerikleri .....	27
Tablo 4.4. Algoritmaların karşılaştırma sonucu .....	28
Tablo 4.5. Algoritmaların Jaccard indeks sonuçları .....	28
Tablo 4.6. Bimax algoritmasına göre ikili küme içerikleri .....	30
Tablo 4.7. Xmotif algoritmasına göre ikili küme içerikleri .....	32
Tablo 4.8. Plaid algoritmasına göre ikili küme içerikleri .....	34
Tablo 4.9. CC algoritmasına göre ikili küme içerikleri .....	36
Tablo 4.10. Bayes algoritmasına göre ikili küme içerikleri .....	38
Tablo 4.11. Algoritmaların karşılaştırma sonucu .....	40
Tablo 4.12. Jaccard indeks sonuçları.....	41

## 1. GİRİŞ

İnsanın var olmasından beri nesnelere sınıflandırılması insanlığın ilgilendiği bir konudur. İnsanlar yaşadıkları ortamı veya evreni daha iyi anlamak için sınıflandırma kavramını kullanmıştır. Sınıflandırmaya temel teşkil eden nesnelere sayısının ve özelliklerinin artması nedeni ile nesnelere birbirlerinden ayrılmasının daha da zor olması insanları yeni teknikler keşfetmeye yöneltmiştir. Bu karmaşık ve zor yapının oluşması “kümeleme analizi” kavramını ortaya çıkarmıştır (Öztürk, 2012).

Kümeleme analizi kavramı, Linnaeus’un 1753 yılında bitki ve hayvanları gruplara ayırma üzerine yaptığı çalışmaya dayanan bir araştırma yöntemi olmasının yanında, ilk defa 1854 yılında Londra’da meydana gelen bir kolera salgını sırasında John Snow tarafından uygulanmış olup, bu olay literatürde kümeleme analizi için bilinen ilk uygulama olarak görülmektedir. John Snow, nehirlere ve şehir kuyularının yakınındaki lağım havuzlarına dökülen kanalizasyonunun su kaynaklarını kirletebileceğine ve kolera salgınlarına neden olabileceğini düşünmüştü. John Snow, yerel halkla konuşarak salgının Broad Street’teki kamuya ait su pompası olup olmadığı konusunda şüphelenmeye başlamıştır. Yerel hastane ve kamu kayıtlarından alınan bilgileri kullanarak, halka özellikle pompadan su içip içmediklerini sordu. Bu bilgileri kullanarak, pompanın etrafındaki durum kümelerini göstermek için bir nokta haritası oluşturmuştur (Hofman and Jarvis, 1998). John Snow tarafından oluşturulan nokta haritası Şekil 1’de gösterilmiştir.



Şekil 1.1. John Snow kolera haritası

Özellikle bilgisayar kullanımının yaygınlaşması ile beraber gelişme gösteren kümeleme analizi, farklı disiplinlerde geniş bir uygulama alanına sahip olmuştur. Günümüzde; veri madenciliği, bankacılık, pazarlama, tıp, ziraat, sosyoloji ve ekonomi alanlarında uygulamaları görülmektedir (Alptekin ve Yeşilaydın, 2015).

Kümeleme analizi, bir araştırmada incelenen birimleri aralarındaki benzerliklerine göre belirli gruplar içinde toplayarak sınıflandırma yapmayı, birimlerin ortak özelliklerini ortaya koymayı ve bu sınıflar ile ilgili genel tanımlar yapmayı sağlayan bir yöntemdir. Kümelemenin temel amacı, grup içi benzerliği yüksek, grup dışı benzerliği düşük gruplar oluşturmaktır. Uygulamalı çalışmalarda dünyada ortak bir yapıya sahip verileri gruplama önemli bir sorun olarak ortaya çıkmaktadır (Çelik, 2012).

Kümeleme yöntemleri; uzaklık matrisi ya da benzerlik matrisinden yararlanarak birimler ya da değişkenleri kendi içinde homojen ve kendi aralarında heterojen gruplamalar oluşturmayı sağlayan yöntemlerdir (Özdamar, 1999). En çok bilinen ya da kabul gören kümeleme yöntemleri; hiyerarşik ve hiyerarşik olmayan yöntemler biçiminde iki grupta toplanmaktadır (Kaygısız vd., 2005). Hiyerarşik olmayan kümeleme yöntemlerinden en çok kullanılanı k-ortalamlar tekniği, MacQueen tarafından bulunmuş olup, küme sayısının belli olduğu durumlarda birbirine en yakın değerlere sahip elemanları aynı kümede toplamayı amaçlar (Tatlıdil, 1996; Özdamar, 2002; Atalay ve Tortum, 2010). Hiyerarşik kümeleme yöntemleri yığılmacı ve bölücü hiyerarşik kümeleme olarak sınıflandırılabilir. Bu sınıflandırma, hiyerarşik ayrışmanın aşağıdan yukarıya veya yukarıdan aşağıya doğru oluşturulmasına göre ortaya çıkmaktadır. Bu durum, her bir ara seviye, bir sonraki alt seviyeden iki kümeyi birleştirmek veya bir kümeyi bir sonraki daha yüksek seviyeden ayırmak şeklinde ifade edilebilir. Bu ikisi arasından en fazla tercih edilen, okunuşunun ve yorumunun kolay olması nedeniyle, yığılmacı hiyerarşik yöntemdir (Kalaycı, 2009; Tekin, 2020).

Kümeleme tek yönlü veriler için kullanışlı bir yöntem olmasına rağmen, çok yönlü veriler için yetersiz olduğu düşünülmektedir. Kümeleme, her koşulda nesnelere arasındaki ilişkileri bulmaya çalışır. Oysa ilişkiler farklı koşullara bağlı olabilir. Bu nedenle, tek koşula bağlı kalınıp yapılan bir kümeleme yanlış sonuçlar doğurabilmektedir. Bu sebeplerden dolayı ikili küme kavramı tanımlanmış ve bu kavram üzerinde çalışmalar yapılmıştır. Literatür çalışmaları incelendiğinde ikili kümeleme yöntemini matris veya blok kümeleme şeklinde de tanımlayan çalışmalar

bulunmaktadır (Govaert and Nadif, 2008, 2013; Van Mechelen et al., 2004). İkili kümeleme kavramı ilk defa Hartigan (1972)'nin yaptığı çalışmada tanımlanmasına rağmen kullanımı Cheng ve Church (2000)'un çalışmalarıyla başlamıştır. Aslında ikili kümeleme kavramının esas dayanak noktası veri matrisinde bulunan satırlar ve sütunların eş zamanlı kümelenmesidir.

Son yıllarda, birbirleriyle ilişkili bir dizi nesneden oluşan verilerden anlamlı bilgiler elde etmek için çeşitli ikili kümeleme yöntemleri geliştirilmiştir. Bu yöntemlerden bazıları Cheng ve Church (2000)'in geliştirdiği CC, Lazzeroni ve Owen (2000)'nin geliştirdiği Plaid, Murali ve Kasif (2003)'in geliştirdiği Quest ve Xmotif ve Kluger vd. (2003)'nin geliştirdiği Spectral ve Prelic vd. (2006) tarafından geliştirilen Bimax algoritmasıdır.

Bu çalışmanın amacı koyun yetiştiriciliğinde gebelik dönemi sorunlarını azaltmak amacıyla geliştirilen çözümler ve bu çözümlerin getirmiş olduğu faydaların birlikte değerlendirilmesinin koyun yetiştiriciliğindeki uygulamaları daha kullanışlı hale getirmesi ve bal arısı ırklarının genetik açıdan durumları, filogenetik ilişkileri ve popülasyonlarının genetik yapılarının belirlenmesi için Plaid, CC, Xmotif, Bimax ve Bayes ikili kümeleme algoritmalarının karşılaştırılmasıdır.

## 2. KAYNAK ÖZETLERİ

Hartigan (1972), tarafından oluşturulan blok kümeleme, ikili kümeleme yönteminin kullanıldığı ilk çalışmadır. Literatürde ikili kümelemeler yaygın olarak gen ifade verilerinin değerlendirilmesinde kullanılmasına rağmen araştırmacı ikili kümeleme yöntemini Birleşmiş Milletler (BM)'de ki oylama sonuçlarının değerlendirilmesinde kullanmıştır.

Cheng ve Church (2000), çalışmalarında gen ifade verilerinde hata kareler ortalaması (HKO) skorlarına sahip alt matrisleri bulmak için verimli CC algoritması tanımlamışlar, maya ve insan verilerinde ortak düzenleme modellerini bulmada CC algoritmasının iyi performans gösterdiğini belirtmişlerdir. Bu ikili kümeleme ile çift küme kavramını kullanmışlardır. Çift küme kavramını ise hem genlerin hem de koşulların eş zamanlı kümelenişi olarak tanımlamışlardır.

Lazzeroni ve Owen (2002), çalışmalarında gen ifade verilerinde ikili kümeleme algoritmalarından Plaid algoritmasını oluşturup kullanmışlardır. Çalışmaları sonucunda insanda kansere sebep olan gen yapılarının benzerliklerini ortaya koymuşlardır.

Kluger vd. (2003), çalışmalarında RNA ifade seviyelerinin global analizleri, genleri ve genel fenotipleri sınıflandırmak için önemli olduğunu belirtmişlerdir. Belirli koşullar kümelerinde farklı şekilde ifade edilen işaretleyici genleri bulmanın sınıflandırma için çok önemli olduğunu söylemişlerdir. Çalışmalarında genleri ve koşulları eş zamanlı kümeleyen spektral ikili kümeleme algoritmasını geliştirmişlerdir. Genellikle bu sınıflandırma sorunları birbiriyle bağlantılıdır ve belirli "koşullar" kümelerinde farklı şekilde ifade edilen "işaretleyici genler" bulmayı amaçlar. Spektral ikili küme algoritması, gen ifade verilerinin matrislerindeki dama tahtası yapılarının, genler veya koşullar arasındaki karakteristik ifade modellerine karşılık gelen özvektörlerde bulunabileceği fikrine dayanmaktadır.

Murali ve Kasif (2003), çalışmalarında farklı kanser türlerini içeren gen ifade veri setlerinde Xmotif algoritmasını kullanarak gen motifleri oluşturmuşlardır. Çalışmaları sonucunda kanser türlerini birbirinden ayıran gen ifade motiflerini tespit etmişlerdir.

Prelic vd. (2006), yapay gen ifade verileri üzerinde ikili kümeleme algoritmalarından CC, ISA, OPSM, Xmotif, SAMBA ve Bimax algoritmalarının kümeleme performanslarını karşılaştırmışlardır. Bu algoritmaları karşılaştırırken kriter olarak ikili küme algoritmaları sonucunda tespit ettikleri ikili kümelerin benzerliklerini kullanmışlardır. Çalışmaları sonucunda algoritmaların kullanılan veri yapılarına göre birbirinden farklı kümeleme sonucu verdiklerini söylemişlerdir. Çalışmalarını gerçek veri seti kullanarak gerçekleştirdiklerinde ise OPSM, SAMBA ve ISA algoritmalarından diğer yöntemlere göre daha iyi sonuçlar elde etmişlerdir.

Tchagang vd. (2011), çalışmalarında cDNA mikroarray verilerinin analizinde ikili kümeleme algoritmalarını kullanmışlardır. Genler arasındaki benzerliklerin ortaya çıkarılmasında bu yöntemleri tercih etmişlerdir.

Bhattacharya vd. (2012), çalışmalarında farklı gen ifade verileri için farklı küme doğrulama indeksleri bakımından ikili kümeleme algoritmalarını kullanmışlardır. Çalışmalarında CC, bioNMF, ISA, Bimax, SAMBA ve OPSM algoritmalarını kullanmışlardır. Algoritmaların performansları değerlendirildiğinde en iyi sonucu SAMBA, en kötü sonucu ise CC algoritmasından aldıklarını belirtmişlerdir.

Zhao vd. (2012), çalışmalarında Gen ontoloji (GO), metabolik yol haritaları (MPM) ve protein-protein etkileşim skoru olan (PPI) değerlerini kullanarak farklı ikili kümeleme algoritmalarının performanslarını karşılaştırmışlar ve Bayes ikili kümeleme algoritmasının (BBC, Bayesian Biclustering Algorithm) en iyi performans gösterdiğini belirtmişlerdir.

Li vd. (2012), çalışmalarında iki farklı veri seti (GDS1620 ve Pathway) üzerinde ikili kümeleme algoritmalarından FABIA, QUBIC, SAMBA, ISA ve Bimax algoritmalarının performanslarını değerlendirmiştir. İkili kümelerin karşılaştırılmasında Gen ontoloji ağırlıklı güçlendirme değeri (GOWE) ve (PPI) değerleri içeren veri setini kullanılmıştır. Çalışması sonucunda ISA ve Bimax algoritmalarının benzer şekilde iyi sonuçlar verdiğini ifade etmişlerdir.

Padilha ve Campello (2017), çalışmalarında ikili kümeleme algoritmalarını yapay ve gerçek veriler kullanarak karşılaştırmışlardır. Çalışmalarında soft indeksi ve kümeleme skorlarını kullanmışlardır. Çalışmaları sonucunda farklı veri setleri için farklı sonuçlar elde etmişlerdir. Yapay veri setleri için veri boyutu küçük olduğunda

OPSM ve Bimax algoritmalarının, büyük olduğunda ise BiBit algoritmasının daha iyi sonuçlar verdiğini belirtmişlerdir. Gerçek veri seti olarak gen ifade verilerini kullanmış ve ISA ve Plaid algoritmalarının daha iyi sonuçlar verdiklerini ifade etmişlerdir.

Singh vd. (2018), çalışmalarında online alışveriş yöntemlerinde karşılaşılan öneri sistem yapısında ikili kümeleme algoritmalarından yararlanmışlardır. Çalışmalarında müşteri odaklı bilgi kayıtları yaparak müşterilerin karşısına onların ilgilendikleri ürünleri çıkararak daha fazla ürün satışı yapmayı hedeflemişlerdir. Çalışmalarında QUBIC algoritmasını kullanmışlardır.

Kocatürk (2018), çalışmasında ikili kümeleme algoritmalarını gen ifade verilerinde benzer ifade yapıların için gen gruplarını belirlemek amacıyla kullanmıştır. Çalışmada, Bimax, Xmotif, Spectral, Plaid, Quest ve CC algoritmaları kullanılmış ve bu algoritmaların kümeleme performansları karşılaştırılmıştır. Algoritmaların karşılaştırılmasında ise (VAR), (MSR), (CKSB), ve küme boyutu gibi kriterleri kullanılmıştır. Ayrıca çalışmada maya ve fare verileri kullanılmıştır. Sonuç olarak farklı veri yapılarında farklı algoritmaların daha iyi sonuçlar verdiğini ifade etmiştir.

Arslan (2018), çalışmasında kaçakçılıkta kullanılan aletlerin kendi aralarında ilişkisi ve kaçakçılık sırasında yakalanan aletlerin türüne bakılarak suçluların suç profillerinin belirlenmesinde ikili kümeleme algoritmalarını kullanmışlardır. Çalışma kapsamında ikili küme algoritmalarından; CC, Quest, xMotif, Bimax ve Plaid algoritmaları tanıtılıp performansları karşılaştırılmıştır. Çalışma sonucunda işlenen suç türlerine göre şehirler ve kaçakçılıkta yakalanan aletler bakımından suçlular farklı algoritmalar kullanılarak kümelenebilir ve suç işlenen bölgelerle suçlu profil yapıları bulunmuştur.

Wei vd. (2019), çalışmalarında insan sinir kök hücreleri için tema eğilimleri ve bilgi yapılarını bibliyometrik olarak incelemişlerdir. Bu amaçla 2013'ten 2018'e kadar PubMed veri tabanından geri çağırma sözcüğü olarak "Sinirsel Kök Hücreler" kullanarak 2742 makale incelemişlerdir. İnsan sinir kök hücreleri ile ilgili çalışmaların özelliklerini ve popüler temalarını istatistiksel olarak ölçmek için ortak kelime analizini ikili kümeleme yöntemlerini kullanarak yapmışlardır.

Kaban vd. (2019), Bu çalışma, mekansal örüntüyü yakalamayı ve Endonezya'daki ilçelerdeki / şehirlerdeki sosyal kırılmanın nedenlerini çift kümeleme yöntemini kullanarak belirlemeyi amaçlamaktadır. Çalışmalarında CC algoritmasını kullanmışlardır. Çalışmaları sonucunda her bir ilçenin / şehrin kendi sosyal kırılma özelliklerine sahip olduğunu ve her ilçe / şehrin savunmasız yönlerinin farklı olduğunu göstermektedir. Endonezya'daki ulusal afet azaltma politikası için bir referans olarak kullanılabilir olduğunu bildirmişlerdir.

Huang vd. (2019), çalışmalarında kadınlarda daha sık görülen göğüs kanseri tanısı için ikili kümeleme yöntemlerini kullanmışlardır. Çalışmada elde edilen görüntüler matris yapılarına dönüştürülüp ikili kümelemiştir. Çalışma sonucunda ikili kümeleme sonucu, destek vektör makinaları ve bulanık destek vektör makinaları sınıflayıcı sonuçları ile karşılaştırılmıştır. Çalışma sonucunda ikili kümeleme yönteminin göğüs kanseri teşhisinde kullanılabilirliğini göstermişlerdir.

Branders vd. (2019), çalışmalarında benzer ifadeye sahip gen alt kümelerini bulmak için ikili kümeleme algoritmalarından yararlanmışlardır. Bu çalışma ile çeşitli deneysel koşullarda gen ifade modellerini keşfederek hücrel süreçler hakkında fikir edinmeyi amaçlamışlardır.

Nicholls ve Wallace (2021), yaptıkları çalışmada kan hücrelerinde dokular arasındaki farklılıkları belirlemek amacıyla ikili kümeleme algoritmalarından Bayes ve FABIA algoritmalarını kullanmışlardır. Çalışmaları sonucunda Bayes algoritmasının daha iyi sonuçlar verdiğini belirtmişlerdir.

Pang (2022), çalışmasında Çin'de ki makro ekonomik verilere dayanarak ekonomik imar alanında, eyaletlerinin ekonomik bölge bölünme planını ve her bir eyaletin ekonomik büyüme modelini araştırmak için Qubic ikili kümeleme algoritmasından yararlanmışır. İkili kümeleme algoritmasının sonucunda belirli nitelik kümeleri altında çeşitli eyaletlerinin ortak yön avantajlarına sahip olması nedeniyle ekonomik büyüme modelleri için uygulanabilir olduğunu ifade etmiştir.

### 3. MATERYAL VE YÖNTEM

#### 3.1. Materyal

Bu tez kapsamında iki farklı veri seti kullanılmıştır. Kullanılan ilk veri seti Sheepnet (Ağ oluşturma yoluyla AB ve Türk koyun verimliliğini artırmaya yönelik uzmanlık ve deneyim paylaşımı) platformuna ait koyun yetiştiriciliğinde gebelik sorunları ve kuzu ölümlerini azaltmak amacıyla geliştirilen çözümler ve faydalarını içeren veri setidir. Bu veri seti 55 çözüm ve bu çözümlerin sağladığı 10 faydayı içermektedir. Veri seti 0-1 (geliştirilen çözümlerin fayda getirip getirmemesi) ikili değerlerden oluştuğu için çalışmada Bimax, Xmotif ve Bayes algoritmaları kullanılmıştır. Veri seti Ek 1’de gösterilmiştir.

Çalışmada kullanılan ikinci veri seti olan bal arısı verileri ise Karabağ vd. (2020) yaptıkları çalışmadan izin alınarak alınmıştır.

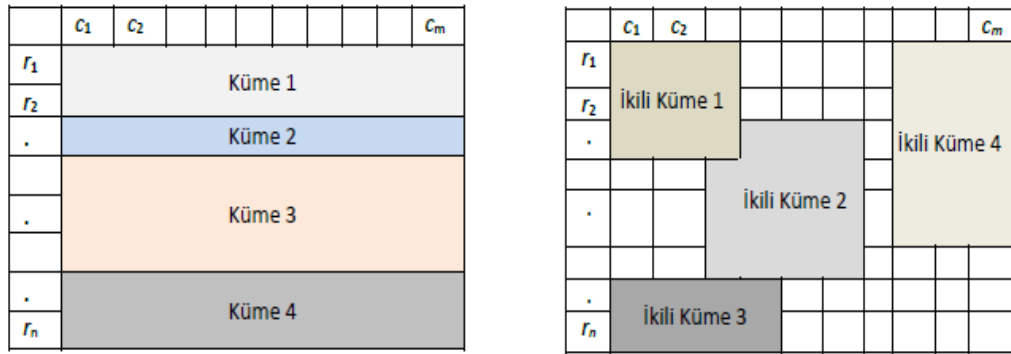
Bu veriler, 2018 yılında yapılan arıcılık faaliyetlerindeki öneminden dolayı seçilen beş ilden (Artvin, Düzce, Hatay, Kırklareli ve Muğla) dört yaygın bal arısı ırkının [*Apis mellifera anatoliaca* Maa, 1953, *Apis mellifera carnica* Pollmann, 1879, *Apis mellifera caucasica* Pollmann, 1889, *Apis mellifera syriaca* Skorikov, 1829 (Hymenoptera: Apidae)] bulunduğu kolonileri içermektedir. Çalışmada kullanılan koloniler, 2018 yılında Türkiye’de arıcılık faaliyetlerindeki önemine göre seçilen beş ilden farklı arıcılardan temin edilmiştir. Her ilden temsili bir örnekleme rastgele yapılmış ve il başına 5 ila 15 arı kovani örneklenmiştir. Böylece, Türkiye’de bilinen dört bal arısı ırkının en yaygın olduğu beş lokasyonun her biri için 30 koloni belirlenmiştir. Çalışmada kullanılan mikrosatellite lokuslar ise Türkiye ve Dünyada yapılan çalışmalarda belirlenmiş ve çalışmaya uygun 30 adet lokuslardan belirlenmiştir. (Bodur et al., 2007; Cánovas et al., 2011; Alburaki et al., 2013; Ilyasov et al., 2016; Ghassemi-Khademi et al., 2018; Hassett et al., 2018). Çalışmada Plaid, CC, Bimax, Xmotif ve Bayes ikili kümeleme algoritmaları kullanılmıştır. Bu çalışmada Bimax algoritması için veri matrisi ikili değer (binary) 0 ve 1 değerlerini içeren yapıya, Xmotif algoritması için ise kesikli değişkenler (discrete) olacak yapıya dönüştürülmüştür. Veri Seti Ek 1’de gösterilmiştir.

Çalışmada kullanılan Plaid, CC, Bimax, Xmotif algoritmalarının analizi R yazılımında “biclust” paketi kullanılarak gerçekleştirilmiştir (R Core Team, 2021).

Bayes algoritmasının kodlamasında <https://jbiclustge.github.io> adresinden erişilebilen java programlama dili ile yapılmıştır. Model karşılaştırma kriterlerinin hesaplanmasında R ücretsiz yazılımında “biclust” paketi kullanılarak yapılmıştır. Çalışmada kullanılan algoritmaların kodları Ek 2’de gösterilmiştir.

### 3.2. Yöntem

İkili kümeleme kavramını ilk kez Hartigan (1972) tarafından kullanılmıştır. Bu yöntemin yaygınlaşması ise Cheng ve Church (2000)’un çalışmalarıyla aktif olarak kullanılmaya başlanmıştır. İkili kümeleme yönteminin temel mantığını veri matrisinin satır ve sütunlarını eş zamanlı olarak kümelmesi oluşturmaktadır. Veri matrisinin satır ve sütunlarının eş zamanlı kümelmesi, klasik kümeleme yöntemlerinden en temel farkını oluşturmaktadır. Fakat bu farklılık ikili kümeleme yönteminin daha zor hesaplanmasına sebep olmaktadır (Zhao et al., 2007; Zhao et al., 2009). Bu sebepten dolayı ikili kümeleme yönteminin gelişip yaygınlaşması bilgisayar sistemlerinin gelişmesine bağlı olarak artmıştır. İkili kümeleme yöntemleri son yıllarda mikro dizi verilerinin analiz edilmesinde yaygın olarak kullanılmaya başlamıştır. Bunun en temel nedeni satırları genlerden sütunları ise koşullardan oluşan mikro dizi verilerinde aynı koşulu ifade eden ortak genlerin tespitinin yapılması gen analizi için önemli bir aşama olmasıdır. Klasik ve ikili kümeleme arasındaki kavramsal fark Şekil 3.1’de verilmiştir. Şekil 3.1 incelendiğinde klasik kümeleme tek yöne bağlı olduğu, ikili kümeleme ise satır ve sütunların aynı anda kümelendiği keyfi alt kümelerden oluşmaktadır.



Şekil 3.1. Klasik Kümeleme gösterimi (solda) ile ikili kümeleme gösterimi (sağda)

İkili kümeleme yapısını oluşturan veri matrisi temelde eşitlik 3.1’de ki gibi ifade edilebilmektedir.

$$A = \dots \begin{bmatrix} a_{11} & \dots & a_{1|J|} \\ \vdots & \ddots & \vdots \\ a_{|I|1} & \dots & a_{|I||J|} \end{bmatrix} \quad (3.1)$$

İkili küme yapısını oluşturan veri matrisinde  $a_{ij}$  i inci satır j inci sütunda bulunan elemanını ifade etmektedir. Eşitlik 3.1’de verilen veri matrisi kullanılarak ikili kümeleme algoritmaları için kullanılan bazı eşitlikler aşağıda verilmiştir.

$$a_{Ij} = \frac{1}{|I|} \sum_{i=1}^{|I|} a_{ij} \quad (3.2)$$

$$a_{iJ} = \frac{1}{|J|} \sum_{j=1}^{|J|} a_{ij} \quad (3.3)$$

$$a_{IJ} = \frac{1}{|I||J|} \sum_{i=1}^{|I|} \sum_{j=1}^{|J|} a_{ij} \quad (3.4)$$

Burada,  $a_{Ij}$  j inci sütunun ortalamasını,  $a_{iJ}$  i inci satırın ortalamasını ve  $a_{IJ}$  genel ortalamayı ifade etmektedir (Kocatürk, 2018).

### 3.2.1. Çalışmada Kullanılan Yöntemler

#### 3.2.1.1. Bayes İkili Kümeleme Algoritması

Bir veri matrisinin  $n$  satır ve  $p$  sütundan oluştuğunu düşünersek,  $i$ . satır ve  $j$ . sütuna gelen değer  $y_{ij}$ ,  $i=1, \dots, n, j=1, \dots, p$  ve eşitlik 3.5’de gösterilmiştir (Gu and Liu, 2008).

$$y_{ij} = \sum_{k=1}^K \left( (\mu_k + \alpha_{ik} + \beta_{jk} + \epsilon_{ijk}) \delta_{ik} \theta_{jk} \right) + e_{ij} \left( 1 - \sum_{k=1}^K \delta_{ik} \theta_{jk} \right) \quad (3.5)$$

Bu eşitlikte  $K$  toplam küme sayısı,  $\mu_k$  ise  $k$  kümesinin ana etkisi,  $\alpha_{ik}$  ve  $\beta_{jk}$  sırasıyla  $k$  kümesinde ki  $i$ . satır ve  $j$ . sütunun etkisi,  $\epsilon_{ijk}$  ise  $k$  kümesinin gürültü terimi

(noise term),  $e_{ij}$  herhangi bir kümeyle ait olmayan veri noktalarını modellemektedir.  $\delta_{ik}$  ve  $\theta_{jk}$  ise ikili değişkenlerdir.  $i$ .sırtarda ki özellik  $k$  kümesine ait ise  $\delta_{ik} = 1$  değil ise  $\delta_{ik} = 0$ ,  $j$ . sütunda ki özellik  $k$  kümesine ait ise  $\theta_{jk} = 1$ , değil ise  $\theta_{jk} = 0$  olmaktadır (Gu and Liu, 2008).

Bayes ikili kümeleme algoritması yalnızca çift kümelerin satır yönünde veya sütun yönünde üst üste binmesine izin verir, bu da bu algoritmanın örtüşmeyen satır çift küme ve örtüşmeyen sütun çift küme şeklinde iki versiyonuyla sonuçlanır. Örtüşmeyen sütun çift kümelemede, bir sütun en fazla bir kümeyle atanabilirken, bir satır birden çok çift kümeyle atanabilir. Bu kısıtlama matematiksel olarak;

$$\sum_{k=1}^K \theta_{jk} \leq 1 \quad (3.6)$$

olarak gösterilmiştir. Örtüşmeyen satır çift kümelemede, bir satır birden çok kümeyle atanabilirken, bir sütun birden fazla kümede olamaz. Bu kısıtlama ise matematiksel olarak;

$$\sum_{k=1}^K \delta_{ik} \leq 1 \quad (3.7)$$

gösterilmektedir. Bu iki versiyonun herhangi birinde, farklı çift kümelerin örtüşmediği bilinmektedir. Böylece,  $\theta$  ve  $\delta$  üyelik göstergelerinin önselleri;

$$\theta_{jk} \sim \text{Bernoulli}(q_k) \quad (3.8)$$

$$P(\delta_{ik} = 1, \delta_{il} = 0, l \neq k) = p_k \quad (3.9)$$

$$P(\delta_{il} = 0, l = 1, 2, \dots, K) = p_0 = 1 - \sum_{k=1}^K p_k \quad (3.10)$$

şeklinde oluşturulmaktadır. Burada  $q_k$  ve  $p_k$  başlangıç değerleri sabit değerler olarak belirlenecektir. Bu değerlerin farklı seçimleri sonucu önemli derecede etkilemeyecektir. Literatürde yapılan çalışmalarda bu değerler  $q_k$  başlangıç değeri için 0.5-0.1 arasında,  $p_k$  değeri için ise;

$$p_k = \frac{1}{K + 1} \quad (3.11)$$

şeklinde alınmaktadır.

Diğer parametreler için önseller ise;

$$\mu_k \sim N(0, \sigma_{\mu_k}^2) \quad (3.12)$$

$$\alpha_{ik} | \delta_{ik} = 1 \sim N(0, \sigma_{\alpha_k}^2) \quad (3.13)$$

$$\beta_{jk} | \theta_{jk} = 1 \sim N(0, \sigma_{\beta_k}^2) \quad (3.14)$$

$$\epsilon_{ijk} \sim N(0, \sigma_{\epsilon_k}^2) \quad (3.15)$$

$$e_{ij} \sim N(0, \sigma_e^2) \quad (3.16)$$

Parametrelerin hiper önselleri (hyperpriors) ters Gamma dağılımı şeklinde kurulmuştur. Tüm hiper önseller bir  $\sigma$  vektörü olarak belirlenmektedir.

Bu şartlar altında  $y_{ij}$  nin olasılık dağılımı;

$$y_{ij} \sim \begin{cases} N(\mu_k + \alpha_{ik} + \beta_{jk}, \sigma_{\epsilon_k}^2) & \text{Eğer } \delta_{ik} \cdot \theta_{jk} = 1 \\ N(0, \sigma_e^2) & \text{Eğer } \delta_{ik} \cdot \theta_{jk} = 0 \end{cases} \quad (3.17)$$

şeklinde olmaktadır.  $y_{ij}$  nin koşullu marjinal dağılımı ise;

$$Y | \delta, \theta \sim N(0, \Sigma) \quad (3.18)$$

şeklinde olmaktadır. Burada  $Y = (Y_0, Y_1, \dots, Y_K)^T$  ile  $Y_k = \{y_{ij} : \delta_{ik} \theta_{jk} = 1\}$ ,  $k \geq 1$ ;  $\Sigma$  ise  $Y$  nin kovaryans matrisidir.  $\delta$  ve  $\theta$  üyelik göstergeleri iteratif bir şekilde güncellendiğinde;

$$\frac{P(Y | \theta_{jk} = 1 | \theta_{[-jk]}, p, \sigma, Y)}{P(Y | \theta_{jk} = 0 | \theta_{[-jk]}, p, \sigma, Y)} \quad (3.19)$$

$$\frac{P(Y|\delta_{ik} = 1|\delta_{[-ik]}, \theta, \sigma, Y)}{P(Y|\delta_{ik} = 0|\delta_{[-ik]}, \theta, \sigma, Y)} \quad (3.20)$$

eşitlikler elde edilmiştir (Gu and Liu, 2008).

### 3.2.1.2. Bayes İkili Kümesi İçin Gibbs Örnekleme

Bayes ikili kümeleme modelinden çıkarım yapmak için bir Gibbs örnekleme yöntemi uygulanır. Bir dizi rastgele atanmış  $\delta$  ve  $\theta$  değerinden başlayarak, aşağıdaki log-olasılık oranını hesaplayarak  $\theta$  sütun (koşul) göstergeleri örneklenir. Bu denklem 3.19 ile  $\theta$  üzerindeki önsel bir dağılımın birleştirilmesiyle;

$$= \log \frac{P(Y|\theta_{jk} = 1|\theta_{[-jk]}, \delta, \sigma)P(\theta_{jk} = 1)}{P(Y|\theta_{jk} = 0|\theta_{[-jk]}, \delta, \sigma)P(\theta_{jk} = 0)} \quad (3.21)$$

elde edilmektedir.

Benzer şekilde denklem 3.20 ile  $\delta$  üzerindeki önsel bir dağılımın birleştirilmesiyle;

$$= \log \frac{P(Y|\delta_{ik} = 1|\delta_{[-ik]}, \theta, \sigma)P(\delta_{ik} = 1)}{P(Y|\delta_{ik} = 0|\delta_{[-ik]}, \theta, \sigma)P(\delta_{ik} = 0)} \quad (3.22)$$

elde edilmektedir (Gu and Liu, 2008).

### 3.2.1.3. Plaid Algoritması

Plaid algoritması ilk olarak Lazzeroni ve Owen (2000)'ın yaptıkları çalışmada önerilmiş fakat geliştirilmesi Turner vd. (2003) tarafından yapılmıştır. Bu algoritmanın temel dayanak noktası veri matrislerinde bulunan katmanların toplamını modellemek ve hataların minimize yapılmasıdır (Turner et al., 2003). Plaid algoritması örtüşen modelleri yani ikili kümelerin üst üste gelmesine imkan tanıyan iki yönlü Anova modellerini içeren bir yapıdadır (Lazzeroni and Owen, 2002). Plaid algoritmasının varsayımları;

Bir satır birden çok kümeye ait olabilir.

Her bir satır kümesi, bazı sütunlara göre tanımlanmıştır.

Plaid algoritmasında veri matrisi içerisinde satır ve sütunların bir araya gelmesiyle dikdörtgensel bölgeler oluşturularak alt matrisler şeklinde bloklama yapılır. Oluşan her blok kendi içinde birbirine yakın değerlerle aynı değere sahip olur.

Eğer bu bloklar grafik şeklinde gösteriliyorsa aynı renklere sahip olurlar. Oluşan bloklar içindeki sütunların cebirsel gösterimi eşitlik 3.23’de verilmiştir.

$$Y_{ij} = \mu_0 + \sum_{k=1}^K \mu_k r_{ik} c_{jk} \quad (3.23)$$

Eşitlik 3.23’de  $\mu_0$  ile arkaplan,  $\mu_k$  ile  $k$  bloğunda var olan renk ifade edilmektedir. Eşitlik 1’de  $k$  bloğunda  $i$ . satır varsa  $r_{ik}$  değeri 1,  $k$  bloğunda  $i$ . satır yoksa bu değer 0 alınır. Ayrıca  $k$  bloğu içerisinde  $j$ . örnek bulunuyorsa  $c_{jk}$  değeri 1, bulunmuyorsa 0 olarak alınır (Lazzeroni and Owen, 2002). Veri matrisi içerisinde her satır bir kümede varsa yani kümeye girmemiş satır bulunmuyorsa veya her sütun bir küme içerisinde bulunuyorsa eşitlik 3.24’de gösterilen kısıtlar oluşturulur.

$$\sum_{k=1}^K r_{ik} = 1 \text{ ve } \sum_{k=1}^K c_{ik} = 1 \quad (3.24)$$

Bazı veri matrislerinde, bir satır veya bir sütunun tek bir kümede değil birden fazla kümede de bulunması mümkündür. Bu durum kümeler arasında örtüşme kavramıyla açıklanmaktadır (Arslan, 2018). Eğer kümeler arasında örtüşme olursa eşitlik 3.24’deki kısıtlar yerine eşitlik 3.25’ deki kısıtlar geçerli olmaktadır.

$$\sum_{k=1}^K r_{ik} \geq 2 \text{ ve } \sum_{k=1}^K c_{ik} \geq 2 \quad (3.25)$$

Eşitlik 3.23’de tanımlanan  $\mu_k$  değeri, veri seti içerisinde bulunan bütün satır ve sütun değerlerinin bulunduğu durumlarda tanımlanan doğrusal model denklemdir. Veri matrisinde bazen sütunlar kümesi içerisinde bazı satırlar kümesi, bazen de satırlar kümesi üzerinde bazı sütunların etkili olduğu gözlenmektedir. Bu etkileri görebilmek için Eşitlik 3.23’deki doğrusal denkleme farklı parametrelerin eklenmesi gerekmektedir. Eklenen parametrelerle oluşturulan yeni denklemler;

$$Y_{ij} = \mu_0 + \sum_{k=1}^K (\mu_k + \alpha_{ik}) r_{ik} c_{jk} \quad (3.26)$$

$$Y_{ij} = \mu_0 + \sum_{k=1}^K (\mu_k + b_{ik}) r_{ik} c_{jk} \quad (3.27)$$

$$Y_{ij} = \mu_0 + \sum_{k=1}^K (\mu_k + \alpha_{ik} + b_{jk}) r_{ik} c_{jk} \quad (3.28)$$

ifade edilmektedir. İkili küme için oluşturulan katmanda satırların etkisi eşitlik 3.26'da verilen  $\alpha_{ik}$  parametresi ile, sütunların etkisi ise eşitlik 3.27'de verilen  $b_{ik}$  parametresi ile gösterilmektedir (Kocatürk, 2018). Eşitlik 3.28'de  $\alpha_{ik}$  ve  $b_{jk}$  parametreleri ile hem satırların hem de sütunların etkisinin belirlendiği Plaid model olarak tanımlanan doğrusal denklem oluşturulmuştur.  $k$  katmanının satır veya sütun değerleri için etkisini bu parametreler hesaplar. Eşitlik 3.26 ve 3.27 incelendiğinde Eğer  $|\mu_k + \alpha_{ik}|$  değeri büyükse satır elemanlarının etkisi,  $|\mu_k + b_{jk}|$  değeri büyükse sütun elemanlarının etkisi daha fazla olduğu söylenir (Busygın et al., 2008). Plaid algoritması için hazırlanan algoritma Tablo 3.1'de verilmiştir (Turner et al., 2003).

Tablo 3.1. Plaid algoritması

---

**Algoritma Girdileri:**

---

$A$  : Kullanılan veri matrisi  
 $\mu$  : Arkaplan ve blok etki değişkeni  
 $\alpha_{ik}$  : Satır etki değişkeni  
 $b_{ik}$  : Sütun etki değişkeni  
 $r_{ik}$  : Satırları ifade eden değişken,  $r_{ik} = 0,1$   
 $c_{jk}$  : Sütunları ifade eden değişken,  $c_{jk} = 0,1$

**Algoritma Adımları:**

1. Kullanılan  $A$  matrisini kullanarak oluşturulan kalıntı matrisi olan ( $Z$ ) bulunur.
  2. Başlangıç değerleri için  $r_{ik}^0$  ve  $c_{jk}^0 = 0$  olacak durumu ayarlanır.  $r_{ik}^{(n-1)}$  ve  $c_{jk}^{(n-1)}$  değerlerine bağlı olarak göre EKK'yı kullanarak  $\mu^{(n)}$ ,  $\alpha_{ik}^{(n)}$  ve  $b_{ik}^{(n)}$  etkilerine ait tahmin edicileri bulunur.
  3.  $\mu^{(n)}$ ,  $\alpha_{ik}^{(n)}$ ,  $b_{ik}^{(n)}$  ve  $c_{jk}^{(n-1)}$  etkilerine ait tahmin edicileri ile EKK'yı beraber kullanarak  $r_{ik}^{(n)}$  ait etki bulunur.
  4.  $\mu^{(n)}$ ,  $\alpha_{ik}^{(n)}$ ,  $b_{ik}^{(n)}$  ve  $r_{jk}^{(n-1)}$  etkilerine ait tahmin edicileri ile EKK'yı kullanarak  $c_{jk}^{(n)}$  ait etki bulunur.
  5. Eğer  $0.5+n/2(N-T)$  sonuç değeri  $0.5$ 'ten büyük olursa  $r_{ik}^{(n)}$  ve  $c_{jk}^{(n)}$  etkilerinden biri eklenir.
  6. Eğer  $0.5-n/2(N-T)$  sonuç değeri  $0.5$ 'ten küçük olursa  $r_{ik}^{(n)}$  ve  $c_{jk}^{(n)}$  etkilerinden biri eklenir aksi durumda adım 3'e dönülür.
  7.  $\mu^{(N+1)}$ ,  $\alpha_{ik}^{(N+1)}$  ve  $b_{ik}^{(N+1)}$  ait etkileri bulunur.
  8. Her bir satır için Aday katman kareler toplamı hesaplanır.  

$$LSS_{aday} = \sum_{i=1}^I \sum_{j=1}^J (\mu + a_i + \hat{b}_j) \hat{r}_i \hat{c}_j$$
  9. Satır elemanlarındaki maksimum  $LSS_{aday}$  sonucunda ikili küme oluşturulur ve adım 2'ye dönülür.
- 

### 3.2.1.4. Bimax Algoritması

Bimax algoritması, iki değer (0-1) değerlerini içeren veri matrisleri içerisinde 1'lerden oluşan alt matrislerin bulunmasını temel alan etkili ve basit bir ikili kümeleme algoritmasıdır (Rodriguez-Baena et al., 2011). İki değer alan veri matrislerinde etkili performans gösteren Bimax algoritmasının temeli Prelic vd. (2006) tarafından oluşturulmuştur. Bu algoritmanın temeli gen ifade verilerinde kullanılmasıyla atılmıştır. Gen ifade verileri için belirlenen bir sınır değerine göre iki değerli bir veri matrisi oluşturulmuştur. Temelde Bimax algoritmasını diğer algoritmalarından ayıran en önemli avantaj veri matrisindeki aykırı değerlerden etkilenmemesidir.

Bimax algoritmasında böl ve yönet stratejisi kullanılarak veri matrisinde bulunan ikili kümeler tespit edilmektedir. Algoritmanın ilk adımında referans olarak rastgele bir satır elemanı ele alınır ve veri matrisi ilk olarak sütun kümesinde 0-1 değerlerini ayırma işlemiyle iki farklı alt kümeye parçalanır. Satır kümesi içinde algoritmanın ikinci adımında aynı yöntem uygulanır. Algoritma en büyük alt matrisi oluşturana kadar her aşamada satır ve sütun elemanlarının sırasını tekrardan düzenler.

Algoritma veri matrisinde parçalama stratejisini uygular, 0 değerini içeren satır ve sütunları silerek yalnız 1 değerini içeren ikili kümeleri oluşturur. Bimax algoritmasının adımları Çizelge 3.2’de verilmiştir (Mina, 2001).

Tablo 3.2. Bimax algoritması

---

**Algoritma Girdileri :**

---

$A_{m \times n}$ : veri matrisi

**Algoritma Adımları:**

1. A veri matrisinden örtüşen şekilde U ve W alt matrislerini bulmak amacıyla tesadüfi bir r satır elemanını belirlenir ve seçilir.
  2. İlk olarak sütunları içeren kümede  $N_u$  ve  $N_w$  şeklinde iki alt küme tanımlanır ve oluşturulur.
  3. Yalnızca  $N_u$  alt kümesi içerisindeki sütunlarla uyum gösteren satırları  $K_u$ ,  $N_u$  ve  $N_w$  alt kümelerinde örtüşen kısımlarda var olan satırları  $K_z$  ve sadece  $N_w$  alt kümesinde var olan sütunlarla uyum gösteren satırları  $K_w$  alt kümesine ayrılır.
  4. Satırlar sonucunda oluşan  $K_u$ ,  $M_z$  ve  $K_w$  alt kümelerine uyum gösteren sütun elemanlarında var olan alt kümelerden  $N_u$  ve  $N_w$  güncellenir.
  5. U, W birbirlerinden tam olarak ayrılana kadar bölme işlemi tekrarlanır.
- 

### 3.2.1.5. CC Algoritması

CC algoritması Cheng ve Church (2000)’un çalışmalarında tanımladıkları ve  $\delta$  - ikili küme şeklinde de ifade edilen, bir ikili kümeleme algoritmasıdır. Cheng ve Church,(2000) çalışmalarında gen ifade verilerinden yararlanarak, benzerlik skoru yüksek olan genlerin ve koşulların her bir alt kümesini ikili küme olarak tanımlamışlardır. Bu algoritmanın temel amacı verilerde düşük hata kareler ortalaması skorları içeren alt matrisleri belirlemeye çalışmaktır.

A bir veri matrisi  $X$  ve  $Y$ ’de sırasıyla satırların ve sütunların bir kümesi olsun. A matrisinde  $\alpha_{ij}$ ,  $i$ . satır ve  $j$ . sütunu gösteren değerdir. Satırların ve sütunların birer alt kümeleri  $I$  ve  $J$  sırasıyla  $I \subset X$  ve  $J \subset Y$  olarak tanımlansın. Eşitlik 7’de MSR skoruna sahip bir  $A_{IJ}$  alt matrisi,  $H$  değeri ile; tanımlanmıştır (Cheng and Church, 2000).

$$H(I,J) = \frac{1}{|I||J|} \sum_{i \in I, j \in J} (\alpha_{ij} - \alpha_{i\cdot} - \alpha_{\cdot j} + \alpha_{\cdot\cdot})^2 \quad (3.29)$$

tanımlanmıştır (Cheng and Church, 2000).

CC algoritmasına göre  $H$  değeri ile küme içerisinde bulunan elemanların birbirine olan benzerlikleri hesaplanmakta ve ikili kümenin anlamlılığı ortaya konulmaktadır. Eğer küme elemanlarının benzerlikleri yüksek ise  $H$  değeri küçük,

küme elemanlarının benzerlikleri düşük ise  $H$  değeri büyük olmaktadır. CC algoritmasında  $H$  değeri ile algoritma için belirlenen bir  $\delta$  sabiti karşılaştırılır. Eğer  $H$  değeri belirlenen  $\delta$  sabitinden daha küçük olursa bulunan ikili kümeye  $\delta$ -ikili küme denilmektedir. Bu algoritmanın kullanılmasındaki esas amaç; maksimum büyüklükte ve  $H$  skoru minimum olan ikili kümeleri tespit etmektir.

Veri matrisi, CC algoritmasında tek bir ikili küme gibi alınarak sürece başlanır. Daha sonra veri matrisinden sırasıyla satır ve sütun silme işlemleriyle devam edilir. Bu işlemlerden sonra tesadüfi olarak herhangi bir satır veya sütunu ekleme işlemi ile  $H$  değeri, belirlenen  $\delta$  değerini geçmeyecek, maksimum  $\delta$  değerine eşit olacak düzeyde artırılır.  $H$  değeri,  $\delta$  değerine ulaşıncaya kadar yeni satır veya sütunlar eklenmeye devam edilir. Algoritmanın ilk aşamasında ilk ikili küme tespit edilir. Sonrasında ise tespit edilen ilk ikili küme veri matrisinden çıkarılarak veri matrisinde güncelleme yapılır ve ikinci iterasyonla bu işleme devam edilir. Çizelge 3.3’de CC algoritmasının adımları gösterilmiştir (Cheng and Church, 2000).

Tablo 3.3. CC algoritması

---

**Algoritma Girdileri:**

$A$  : Başlangıç veri matrisi

$\delta$  : Kabul edilebilir en büyük ortalama kare artık skoru,  $\delta \geq 0$

$\alpha$  : Veri matrisinden çoklu satır silme işlemi için sınır değer,  $\alpha \geq 0$

**Algoritma Adımları:**

1.  $A_{IJ} = A$

2.  $a_{ij}$ ,  $a_{ij}$  ve  $a_{IJ}$  değerleri hesaplanır.

3.  $H(I, J)$  değeri hesaplanır.

4. Eğer  $H(I, J) \leq \delta$  ise  $A_{IJ}$  matrisi bir ikili küme şeklinde tanımlanır. Olumsuz ise sonraki adıma geçilir.

5. Aşağıdaki koşulu sağlayan  $i \in I$  satırları silinir.

$$\frac{1}{|J|} \sum_{j \in J} (a_{ij} - a_{ij} - a_{ij} + a_{IJ})^2 > \alpha H(I, J)$$

6.  $a_{ij}$ ,  $a_{ij}$ ,  $a_{IJ}$  ve  $H(I, J)$  değerleri hesaplanır.

7. Aşağıda verilen koşulu yerine getiren  $j \in J$  sütunları silinir.

$$\frac{1}{|I|} \sum_{j \in J} (a_{ij} - a_{ij} - a_{ij} + a_{IJ})^2 > \alpha H(I, J)$$

8. Eğer tüm satır ve sütunları silme işlevi tamamlandı ise aşama sonlandırılır. Olumsuz ise  $A_{IJ}$  matrisini başlangıç matrisi olan  $A$ ’dan çıkarılır, kalan matris için tekrar adım 1’e dönülür.

---

### 3.2.1.6. Xmotif Algoritması

Xmotif algoritması, veri matrislerinde benzer satır ve sütunları birlikte kümelemek amacıyla Xmotif olarak tanımlanan alt matrisleri belirlemeye çalışan bir algoritmadır. Veri matrisleri incelendiğinde birden fazla Xmotif olarak tanımlanan ikili kümeler bulunabilir. Bu algoritmanın temelinde bulunan Xmotif'lerin maksimum büyüklükte olması yatmaktadır. Ancak bu durum her zaman istenilen şekilde olamayabilir ve veri matrisinde bulunan her satır veya sütun Xmotif'ler içerisinde bulunmayabilir. Bu algoritma maksimum büyüklükte bulunan Xmotif'leri belirlemek amacıyla tekrarlayıcı bir yapıdadır. Algoritma, tüm veri matrisinden maksimum büyüklükteki Xmotif kümeleri çıkarıp geriye kalan veri matrisi içinden yeniden bir Xmotif kümesi tespit ederek bu sürece devam eder. Veri matrisindeki tüm Xmotif alt matrisleri tespit edilinceye kadar iterasyon işlemine devam edilmektedir (Murali and Kasif, 2003). Xmotif algoritmasının yapısı Çizelge 3.4'te ifade edilmiştir (Murali and Kasif 2003).

Tablo 3.4. Xmotif algoritması

---

**Algoritma Girdileri:**

---

$A_{m \times n}$ : Kullanılan veri matrisi

**Algoritma Adımları:**

1. A matrisinde satır elemanları içinden tesadüfi bir örnek (b) belirle ve seç.
  2. A matrisinde sütun elemanları içinden tesadüfi bir K alt kümesi belirle ve seç.
  3. Tüm r satırlarında, eğer r, K alt kümesinde bulunan tüm satırlar ve b'de m durumunda ise (r,m) çiftini  $R_{ij}$  kümesine ekle.
  4.  $B_{ij}$ 'yi  $R_{ij}$  kümesinde bulunan tüm satır durumlarında (b) ile uyum sağlayan bir örnek kümesi şeklinde belirle ve al.
  5.  $B_{ij}$   $m \times n$  boyutlu veri matrisinden daha küçük olursa ( $B_{ij}, R_{ij}$ ) çiftini çıkar.
  6.  $|R_{ij}|$  maksimum boyuta ulaşana kadar (B,R) motifini çalıştır.
- 

## 3.2.2. İkili Küme Değerlendirme Ölçüleri

### 3.2.2.1. Varyans Ölçüsü (VAR)

Bu değerlendirme ölçüsünü, Hartigan (1972) çalışmasında küme içi varyans kavramı şeklinde ikili kümeleri değerlendirme ölçüsü olarak kullanılmıştır. Bu ölçü;

$$VAR(A) = \sum_{i=1}^{|I|} \sum_{j=1}^{|J|} (\alpha_{ij} - \alpha_{IJ})^2 \quad (3.30)$$

şeklinde tanımlanmaktadır. VAR değerinin düşük olması ikili kümenin daha tutarlı bir ikili küme olduğunu göstermektedir. Bir ikili kümenin kusursuz bir şekilde tutarlı olması için VAR değeri 0, bu değer 1 veya 1,5 değerinden fazla bulunmuş ise, bulunan ikili kümenin tutarlı olmadığı söylenebilir (Kaiser, 2011).

### 3.2.2.2. Hata Kareler Ortalama Skoru (HKO)

Bu değerlendirme ölçüsü ikili kümelerin kalitesini ölçmek için yaygın olarak kullanılmaktadır. Bu ölçü;

$$HKO(A) = \frac{1}{|I||J|} \sum_{i=1}^{|I|} \sum_{j=1}^{|J|} (\alpha_{ij} - \alpha_{iJ} - \alpha_{iI} + \alpha_{IJ})^2 \quad (3.31)$$

şeklinde tanımlanmaktadır (Cheng and Church, 2000). İkili kümede bulunan satır ve sütun değerlerinin tutarlılığını incelemek için bu ölçü kullanılabilir. HKO değeri küçüldükçe ikili kümelemenin başarısı artmaktadır. HKO değeri, veri matrisini oluşturan değerlerin ölçeklendirilmesi durumunda aşırı duyarlı olur ve bu duyarlılık ikili kümenin değerlendirilmesinde yetersizlik yaratır (Aguilar-Ruiz, 2005).

Bir ikili küme bazı gereksinimleri aynı anda sağlamalıdır. Bunlardan biri, düşük bir HKO değerinin olmasıdır. Diğer önemli gereksinim ise en büyük boyuta sahip ikili kümeler oluşturulmasıdır (Mitra and Banka, 2006).

### 3.2.2.3. Chia ve Karuturi Ölçüsü (CKSB)

Bir veri matrisinde bulunan sütunların ikili küme içerisinde bulunan ve bulunmayanlar arasındaki benzerliği dikkate alarak değerlendiren bu ölçü ikili kümelerin doğruluğunu belirlemektedir (Chia and Karuturi, 2010). HKO skoru kullanılarak CKSB skoruna ait ikili kümelerin satır etkisi  $T_k(\alpha)$  sütun etkisi  $B_k(\alpha)$  belirlenir. Bu etkiler sırasıyla;

$$T_k(\alpha) = \frac{1}{|I(\alpha)|} \sum_{i=1, i \in \alpha}^{I(\alpha)} \alpha_{ijk}^2 - \frac{HKO_k(\alpha)}{J_k(\alpha)} \quad (3.32)$$

$$B_k(\alpha) = \frac{1}{|J_k(\alpha)|} \sum_{j=1, j \in \alpha}^{J_k(\alpha)} \alpha_{ijk}^2 - \frac{HKO_k(\alpha)}{I_k(\alpha)} \quad (3.33)$$

şeklinde hesaplanmaktadır (Chia and Karuturi, 2010). Eşitlikler incelendiğinde k alt indis değeri ikili kümede bulunan elemanlar için 1 değerini, ikili kümede bulunmayan elemanlar için ise 2 değerini almaktadır. İkili küme içinde bulunan satır sayısı  $I(b)$  değeri ile, İkili küme içinde ve dışında bulunan sütun sayıları ise  $J_k(b)$  değeri ile gösterilmektedir (Chia and Karuturi, 2010; Kocatürk, 2018).

Hesaplanan satır ve sütun etkileri kullanılarak CKSB değeri;

$$CKSB(\alpha) = LOG \left( \frac{\max(T_1(\alpha) + \beta, B_1(\alpha) + \beta)}{\max(T_2(\alpha) + \beta, B_2(\alpha) + \beta)} \right) \quad (3.34)$$

şeklinde belirtilmiştir (Chia and Karuturi, 2010). Eşitlik 3.34'de  $\beta$  değeri küçük boyutlu ikili kümelerin satır ve sütun etkilerini dengelemek için kullanılan ve 0 ile 1 arasında belirlenen bir ölçek faktörüdür. CKSB skoru ikili kümelerin anlamlılığını belirlemek için kullanılan bir değerdir. CKSB skorunun yüksek olması ikili kümenin daha anlamlı olmasını göstermektedir (Chia and Karuturi, 2010).

#### 3.2.2.4. Jaccard İndeksi

Jaccard indeksi, algoritmalar için tümüyle bir biri ile benzer olmayan alt matrisleri araştırmak için kullanılmaktadır. Jaccard indeksi kullanılarak belirlenen bir ikili kümenin geçerli olabilmesi için en önemli şart, diğer ikili kümeler ile veya en doğru olduğu düşünülen ikili küme ile karşılaştırılmasına dayanmaktadır. Aşağıdaki eşitlikte Jaccard indeksinin standart kümeleme yöntemine göre düzenlenmiş hali;

$$jac(BC_i, BC_j) = jac_{ij} = \frac{|BC_i \cap BC_j|}{|BC_i \cup BC_j|} \quad (3.35)$$

gösterilmiştir (Jaccard, 1901). Eşitlik 3.35’de  $BC_i$ , i. ikili kümeyi,  $BC_j$  ise j. ikili kümeyi belirtmektedir. Jaccard indeksi, ikili kümeleri karşılaştırmak amacıyla belirlenen her iki alt kümelerde satır ile sütun kesişimlerinin, en az bir alt kümede bulunan tüm satır ile sütun birleşimlerine bölünmesi ile belirlenir.

## 4. BULGULAR

### 4.1. Sheepnet Veri Seti Bulguları

#### 4.1.1. Bimax Algoritma Sonuçları

Bimax algoritması kullanılarak oluşturulan ikili kümelerin içerikleri ve küme boyutları Tablo 4.1’de verilmiştir. Bimax algoritması sonucunda 5 farklı ikili küme oluşmuştur. İkili kümeler büyüklüklerine göre incelendiklerinde, küme 1’in 55 çözümden 17 çözümü içerdiği (%30.9), küme 2’nin 13 çözümü içerdiği (%23.6), küme 3’ün 10 çözümü içerdiği (%18.1) ve küme 4’ün 8 çözümü içerdiği (%14.54), küme 5’in 20 çözümü içerdiği (%36.36) Tablo 4.1’de görülmektedir.

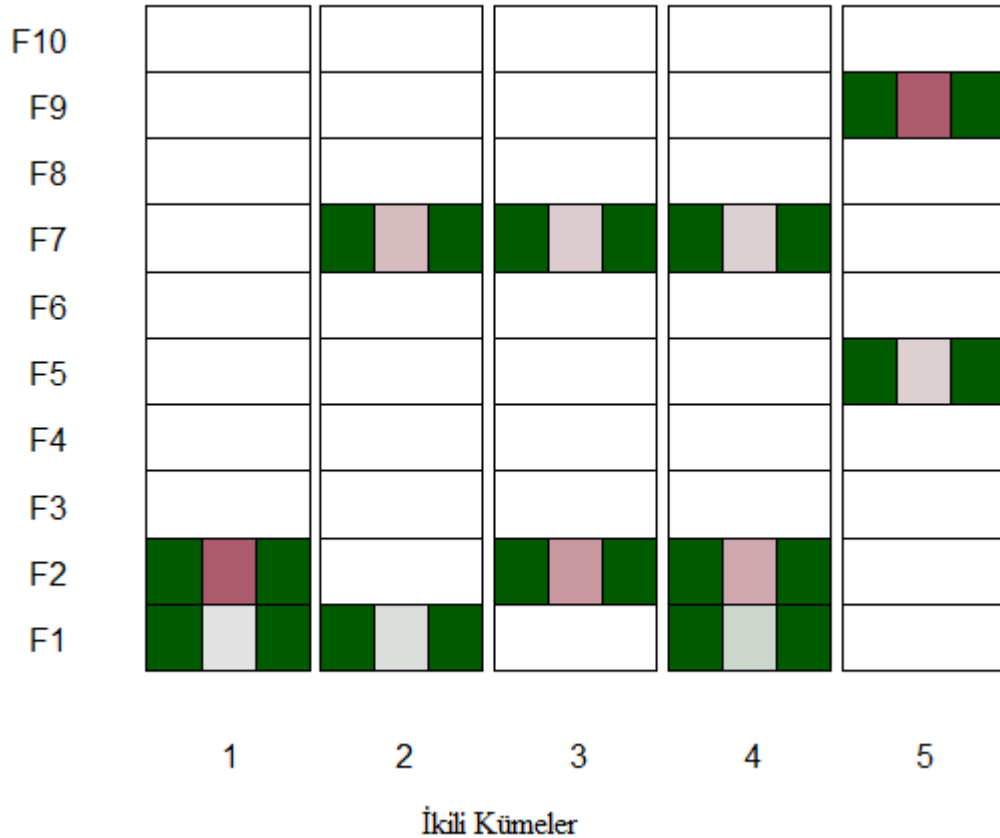
Tablo 4.1. Bimax algoritmasına göre küme içerikleri ve boyutları

İkili küme	Çözümler(Ç)	Faydalar(F)	Büyüklik	Çözümlere göre oranları	ÇözümxFayda oranları
1	Ç1 Ç2 Ç3 Ç5 Ç16 Ç17 Ç18 Ç20 Ç21 Ç24 Ç25 Ç32 Ç33 Ç40 Ç48 Ç49 Ç55	F1 F2	17x2	%30.9	%6.18
2	Ç1 Ç2 Ç5 Ç6 Ç16 Ç18 Ç20 Ç21 Ç33 Ç39 Ç42 Ç44 Ç46	F1 F7	13x2	%23.6	%4.73
3	Ç1 Ç2 Ç5 Ç9 Ç16 Ç18 Ç20 Ç21 Ç22 Ç33	F2 F7	10x2	%18.1	%3.6
4	Ç1 Ç2 Ç5 Ç16 Ç18 Ç20 Ç21 Ç33	F1 F2 F7	8x3	%14.54	%4.4
5	Ç2 Ç5 Ç6 Ç8 Ç10 Ç11 Ç12 Ç13 Ç14 Ç15 Ç35 Ç36 Ç37 Ç38 Ç39 Ç43 Ç46 Ç53 Ç54 Ç55	F5 F9	20x2	%36.36	%7.3

İkili kümelere ait üyelik grafiği Şekil 4.1’de gösterilmiştir. Şekil 4.1 incelendiğinde; satırlar farklı faydaları, sütunlar ise farklı ikili kümeleri belirtmektedir. Yeşil dikdörtgen bölümler çözümlerden elde edilen ortak faydaları, yeşil alanların iç kısımlarında bulunan orta dikdörtgenler ise 0 (beyaz) ile 1 (kırmızı) arasında değişen, çözümlerden elde edilen faydaların ortalama değerini belirtmektedir. Kırmızı alan ne kadar koyu renk verirse, ikili kümelerdeki faydalar ile elde edilen tüm faydaların uyum düzeyi de o kadar artmaktadır.

Küme 1’de yer alan çözümlerin (17 çözüm, tüm çözümlerin %30.9’u) F1 ve F2 faydalarını sağladığı görülmektedir. Şekil 4.1 incelendiğinde orta bölgede bulunan dikdörtgenler değerlendirildiğinde bu küme için F2 faydasını sağlama oranının F1 faydasını sağlama oranından daha fazla olduğu söylenebilir. Küme 2’de yer alan

çözümlerin (13 çözüm, tüm çözümlerin %23.6'sı) F1 ve F7 faydalarını sağladığı görülmektedir. Orta bölgede bulunan dikdörtgenler değerlendirildiğinde bu küme için F7 faydasını sağlama oranının F1 faydasını sağlama oranından az da olsa daha fazla olduğu söylenebilir. Küme 3'de yer alan çözümlerin (10 çözüm, tüm çözümlerin %18.1'i) F2 ve F7 faydalarını sağladığı görülmektedir. Orta bölgede bulunan dikdörtgenler değerlendirildiğinde bu küme için F2 faydasını sağlama oranının F7 faydasını sağlama oranından daha fazla olduğu söylenebilir. Küme 4'de yer alan çözümlerin (8 çözüm, tüm çözümlerin %14.54'ü) F1, F2 ve F7 faydalarını sağladığı görülmektedir. Orta bölgede bulunan dikdörtgenler değerlendirildiğinde bu küme için F2 faydasını sağlama oranının F7 faydasından, F7 faydasının ise F1 faydasını sağlama oranından daha fazla olduğu söylenebilir. Küme 5'de yer alan çözümlerin (20 çözüm, tüm çözümlerin %36.36'sı) F5 ve F9 faydalarını sağladığı görülmektedir. Orta bölgede bulunan dikdörtgenler değerlendirildiğinde bu küme için F9 faydasını sağlama oranının F5 faydasını sağlama oranından daha fazla olduğu söylenebilir.



Şekil 4.1. Bimax algoritması ikili küme üyelik grafiği

#### 4.1.2. Xmotif Algoritma Sonuçları

Xmotif algoritması kullanılarak oluşturulan ikili kümelerin içerikleri Tablo 4.2’de verilmiştir. Xmotif algoritması sonucunda 6 farklı ikili küme oluşmuştur. İkili kümeler büyüklüklerine göre incelendiklerinde, küme 1’in 55 çözümden 22 çözümü içerdiği (%40.00), küme 2’nin 13 çözümü içerdiği (%23.64), küme 3’ün 8 çözümü içerdiği (%14.55) ve küme 4’ün 5 çözümü içerdiği (%9.10), küme 5’in 3 çözümü içerdiği (%5.45) ve küme 6’nın 2 çözümü içerdiği (%3.64) Tablo 4.2’de görülmektedir.

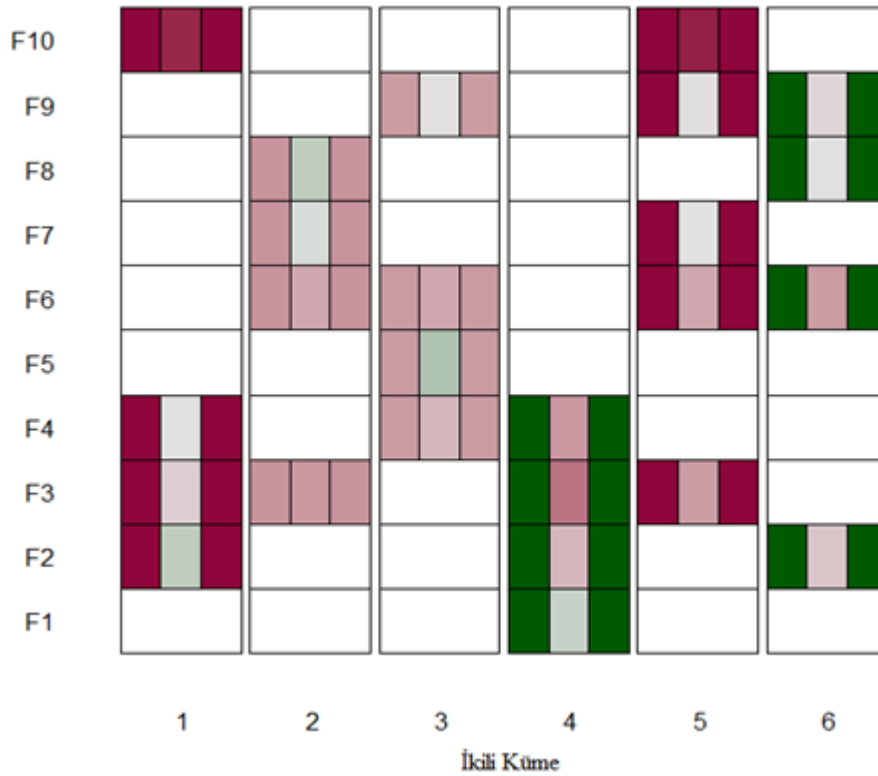
Tablo 4.2. Xmotif algoritmasına göre küme içerikleri

İkili küme	Çözümler(Ç)	Faydalar(F)	Büyüklik	Çözümlere göre oranları	ÇözümxFayda oranları
1	Ç7 Ç12 Ç13 Ç14 Ç15 Ç23 Ç28 Ç29 Ç30 Ç31 Ç35 Ç36 Ç37 Ç38 Ç39 Ç41 Ç44 Ç45 Ç51 Ç52 Ç53 Ç54	F2 F3 F4 F10	22x4	%40.00	%16
2	Ç3 Ç5 Ç10 Ç17 Ç19 Ç24 Ç25 Ç34 Ç40 Ç42 Ç43 Ç47 Ç50	F3 F6 F7 F8	13x4	%23.64	%9.5
3	Ç1 Ç4 Ç6 Ç11 Ç16 Ç18 Ç22 Ç33	F4 F5 F6 F9	8x4	%14.55	%5.8
4	Ç2 Ç20 Ç21 Ç48 Ç55	F1 F2 F3 F4	5x4	%9.10	%3.6
5	Ç26 Ç27 Ç49	F3 F6 F7 F9 F10	3x5	%5.45	%2.7
6	Ç8 Ç9	F2 F6 F8 F9	2x4	%3.64	%1.5

Şekil 4.2’ye göre; her bir satır faydaları, her bir sütun ise bir ikili kümeyi belirtmektedir. İkili küme 1, 2 ve 5 ikili kümeleri için kırmızı bölümler çözümlerden elde edilen ortak faydaları, bu kırmızı alanların iç kısmında bulunan orta dikdörtgenler ise (beyaz) ile (kırmızı) arasında değişen, çözümlerden elde edilen faydaların ortalama değerini belirtmektedir. Kırmızı alan ne kadar açık renk verirse, ikili kümelerdeki faydalar ile elde edilen tüm faydaların uyum düzeyi de o kadar artmaktadır.

Küme 1’de yer alan çözümlerin (22 çözüm, tüm çözümlerin %40’ı) F2,F3,F4 ve F10 faydalarını sağladığı görülmektedir. Şekil 4.2 incelendiğinde orta bölgede bulunan dikdörtgenler değerlendirildiğinde bu küme için F10 faydasını sağlama oranının F2, F3 ve F4 faydasını sağlama oranından daha fazla olduğu söylenebilir. Küme 2’de yer alan çözümlerin (13 çözüm, tüm çözümlerin %23.64’ü) F3, F6, F7 ve

F8 faydalarını sağladığı görülmektedir. Orta bölgede bulunan dikdörtgenler değerlendirildiğinde bu küme için F6 ve F8 faydasını sağlama oranının F3 ve F7 faydasını sağlama oranından daha fazla olduğu söylenebilir. Küme 3’de yer alan çözümlerin (8 çözüm, tüm çözümlerin %14.55’i) F4, F5, F6 ve F9 faydalarını sağladığı görülmektedir. Orta bölgedeki dikdörtgenler değerlendirildiğinde bu küme için F6 faydasını sağlama oranının F4, F5 ve F9 faydasını sağlama oranından daha fazla olduğu söylenebilir. Küme 4’de yer alan çözümlerin (5 çözüm, tüm çözümlerin %9.10’u) F1, F2, F3 ve F4 faydalarını sağladığı görülmektedir. Orta bölgede bulunan dikdörtgenler dikkate alındığında bu küme için F3 faydasını sağlama oranının F1, F2 ve F4 faydasını sağlama oranından daha fazla olduğu söylenebilir. Küme 5’de yer alan çözümlerin (3 çözüm, tüm çözümlerin %5.45’i) F3, F6, F7, F9 ve F10 faydalarını sağladığı görülmektedir. Orta bölgedeki dikdörtgenler değerlendirildiğinde bu küme için F10 faydasını sağlama oranının F3, F6, F7 ve F9 faydalarını sağlama oranından daha fazla olduğu söylenebilir. Küme 6’da yer alan çözümlerin (2 çözüm, tüm çözümlerin %3.64’ü) F2, F6, F8 ve F9 faydalarını sağladığı görülmektedir. Orta bölgedeki dikdörtgenler dikkate alındığında F6 faydasını sağlama oranının F2, F8 ve F9 faydalarını sağlama oranından daha fazla olduğu söylenebilir.



Şekil 4.2. Xmotif algoritması ikili küme üyelik grafiği

### 4.1.3. Bayes Algoritma Sonuçları

Bayes algoritması kullanılarak oluşturulan ikili kümelerin içerikleri Tablo 4.3’de verilmiştir. Bayes algoritması sonucunda 4 farklı ikili küme oluşmuştur. Büyüklüklerine göre ikili kümeler incelendiğinde, küme 1’in 55 çözümden 17 çözümü içerdiği (%30.90), küme 2’nin 19 çözümü içerdiği (%34.55), küme 3’ün 4 çözümü içerdiği (%7.27) ve küme 4’ün 7 çözümü içerdiği (%12.72) Tablo 4.3’de görülmektedir.

Tablo 4.3. Bayes algoritmasına göre küme içerikleri

İkili küme	Çözümler(Ç)	Faydalar(F)	Büyüklik	Çözümlere göre oranları	ÇözümxFayda oranları
1	Ç12 Ç13 Ç14 Ç15 Ç23 Ç28 Ç29 Ç30 Ç31 Ç35 Ç36 Ç37 Ç38 Ç39 Ç51 Ç53 Ç54	F2 F3 F4 F5 F10	17x5	%30.90	%15.5
2	Ç1 Ç3 Ç7 Ç16 Ç17 Ç18 Ç20 Ç21 Ç24 Ç25 Ç26 Ç27 Ç32 Ç33 Ç40 Ç41 Ç45 Ç48 Ç49	F1 F6 F9 F10	19x4	%34.55	%13.8
3	Ç4 Ç9 Ç22 Ç52	F1 F3 F4 F5 F7 F8	4x6	%7.27	%4.4
4	Ç2 Ç6 Ç34 Ç43 Ç46 Ç50 Ç55	F1 F4 F8 F9 F10	7x5	%12.72	%6.4

Bayes algoritması sonucunda ikili küme üyelik grafiği kullanılan veri seti için oluşturulamamıştır. Tablo 4.3’de verilen İkili kümeler içerdiği faydalar bakımından incelendiğinde küme 1’in F2, F3, F4, F5, F10; küme 2’nin F1, F6, F9, F10; küme 3’ün F1, F3, F4, F5, F7, F8; küme 4’ün ise F1, F4, F8, F9, F10 olduğu görülmüştür. F1, F4 ve F10 faydaları üç farklı ikili küme içerisinde bulunmuştur. Bu durumun geliştirilen farklı çözümlerin bu üç faydayı ortak sağladığını gösterdiği düşünülmektedir.

## 4.2. Algoritmaların Karşılaştırılması

Sheepnet verisi kullanılarak elde edilen ikili kümeler Bimax, Xmotif ve Bayes algoritmaları sonucunda ortaya çıkmıştır. Bimax, Xmotif algoritmaları kullanılarak oluşturulan ikili kümelerin eleman değerleri (0-1) yapısında olduğundan dolayı VAR ve MSR bulunamamıştır. Bu algoritmaların karşılaştırması sadece Tablo 4.4'de gösterilen Chia Karuturi ikili küme uygunluk skoru ve Ortalama satır ve sütun büyüklüklerine göre değerlendirilecektir. Tablo 4.4 incelendiğinde en yüksek CKSB skoru Bayes algoritmasının sonucunda elde edilmiştir. Ortalama küme boyutları da incelendiğinde ise Bayes algoritması sonucunda daha büyük boyutlarda kümeler ortaya çıkmıştır.

Tablo 4.4. Algoritmaların karşılaştırma sonucu

	CKSB	Ort.Satır Boyutu	Ort.Sütun Boyutu	Ort. Küme Boyutu
Bimax	0.538	13.6	2.2	29.92
Xmotif	0.617	8.83	4.17	36.82
Bayes	0.742	11.75	5.00	58.75

Algoritmaların jaccard indeksine göre karşılaştırılması Tablo 4.5'de gösterilmiştir. Jaccard indeksi incelendiğinde Bimax algoritması ile Bayes algoritmasının daha düşük değere sahip olduğu görülmektedir. Bimax algoritması ile elde edilen ikili kümelerin Xmotif algoritması ile sonucu belirlenen kümelere göre, Bayes algoritması sonucu belirlenen ikili kümelere daha farklı olduğu söylenebilir.

Bimax, Xmotif algoritmaları kullanılarak tespit edilen ikili kümelerin eleman değerleri (0-1) yapısında olduğundan dolayı VAR ve MSR sonucu sıfır çıkmıştır. Bu yüzden bu kriterlere göre karşılaştırma yapılamamıştır.

Tablo 4.5. Algoritmaların Jaccard indeks sonuçları

Algoritma	Bimax	Xmotif	Bayes
Bimax	1	0.0775	0.0564
Xmotif		1	0.1380
Bayes			1

### **4.3. Bal Arısı Veri Seti Bulguları**

#### **4.3.1. Bimax Algoritması Sonuçları**

Bimax algoritması kullanılarak oluşturulan ikili kümelerin içerikleri Tablo 4.6'da verilmiştir. Bimax algoritması sonucunda 3 farklı ikili küme oluşmuştur. Büyüklüklerine göre ikili kümeler incelendiğinde, küme 1'in 150 koloniden tamamını içerdiği (%100), küme 2'nin 20 koloni (%13.33), küme 3'ün 4 koloni (%20.00) içerdiği Tablo 4.6'da görülmektedir.

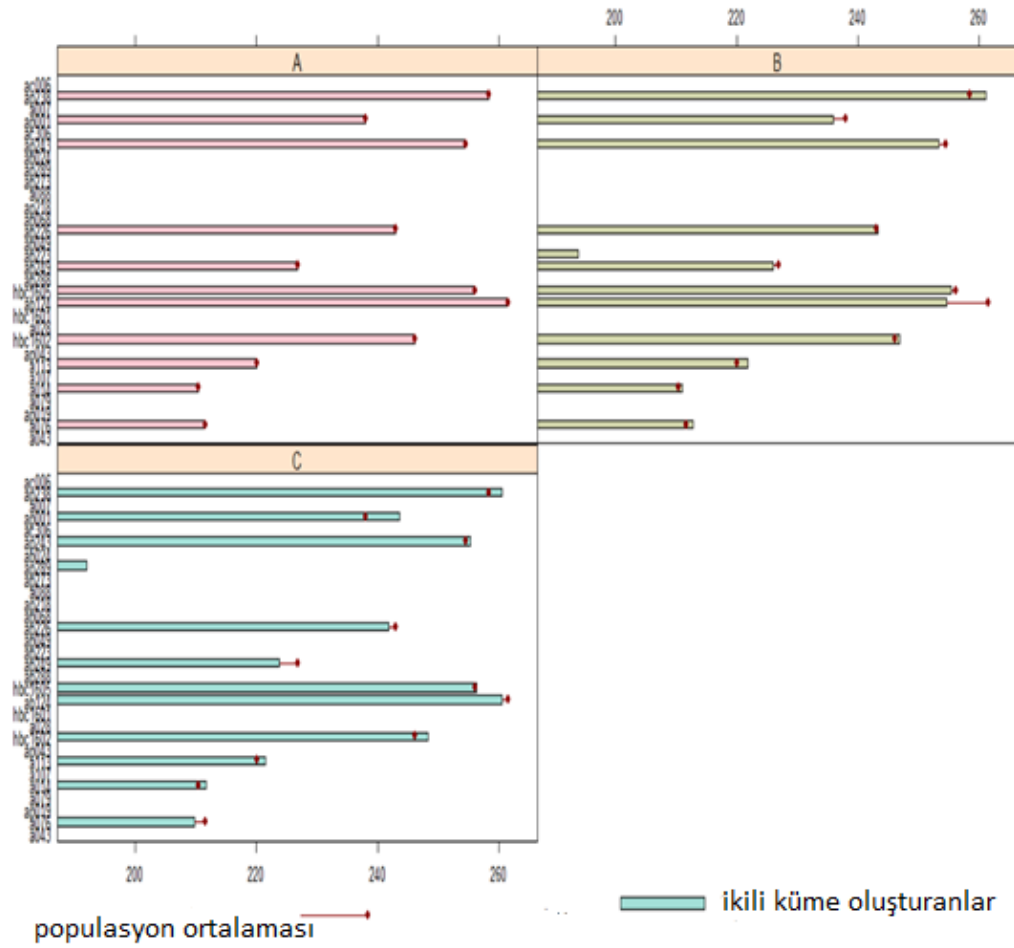
Tablo 4.6. Bimax algoritmasına göre ikili küme içerikleri

İkili küme	Koloniler	Lokuslar	Büyüklik	Kolonilere göre oranları	KoloniXLokus Büyüklük oranları
A	M1 M2 M3 M4 M5 M6 M7 M8 M9 M10 M11 M12 M13 M14 M15 M16 M17 M18 M19 M20 M21 M22 M23 M24 M25 M26 M27 M28 M29 M30 H1 H2 H3 H4 H5 H6 H7 H8 H9 H10 H11 H12 H13 H14 H15 H16 H17 H18 H19 H20 H21 H22 H23 H24 H25 H26 H27 H28 H29 H30 K1 K2 K3 K4 K5 K6 K7 K8 K9 K10 K11 K12 K13 K14 K15 K16 K17 K18 K19 K20 K21 K22 K23 K24 K25 K26 K27 K28 K29 K30 A1 A2 A3 A4 A5 A6 A7 A8 A9 A10 A11 A12 A13 A14 A15 A16 A17 A18 A19 A20 A21 A22 A23 A24 A25 A26 A27 A28 A29 A30 D1 D2 D3 D4 D5 D6 D7 D8 D9 D10 D11 D12 D13 D14 D15 D16 D17 D18 D19 D20 D21 D22 D23 D24 D25 D26 D27 D28 D29 D30	ap238 ap001 ap243 ap226 ap249 hbc1605 ab124 hbc1602 a113 a014 a076	150x11	%100	%36
B	M5 M6 M8 M16 M22 M24 M30 K9 K10 K11 A7 A8 A9 A10 A11 A15 A16 D8 D9 D27	ap238 ap001 ap243 ap226 ap223 ap249 hbc1605 ab124 hbc1602 a113 a014 a076	20x12	%13.33	%5.3
C	H27 K8 K11 K29	ap238 ap001 ap243 ap289 ap226 ap249 hbc1605 ab124 hbc1602 a113 a014 a076	4x12	%2.67	%1.1

Bimax algoritması sonucu belirlenen ikili kümelere ait çubuk grafiği Şekil 4.3'de gösterilmiştir. Çubuk grafiğinde her ikili küme ayrı bir harf ile ifade edilmiş (A, B, C) ve küme içerisinde yer alan Lokuslar gösterilmiştir. Tablo 4.6 ve Şekil 4.3 incelendiğinde ap238, ap001, ap243, ap289 ap226, ap249, hbc1605, ab124, hbc1602,

a113, a014, a076 lokusları her 3 ikili kümede bulunmuşlardır. Ayrıca ikili küme B içerisinde ap223 ve ikili küme C içerisinde ap289 lokusu bulunmaktadır. İkili küme A tüm kolonileri içerisine alan büyük boyutlu bir küme olmuştur. İkili küme B içerisinde bulunan M5, M6, M8, M16, M22, M24, M30, K9, K10, K11, A7, A8 A9, A10, A11, A15, A16, D8, D9, D27 kolonileri ise ikili küme 1 içerisinde de mevcuttur.

İkili kümeler içerdikleri lokuslar bakımından incelendiğinde; ab124 lokusunun A kümesi, ap238 lokusunu B kümesi, ap238 ve ab124 lokuslarının C kümesi içerisinde en fazla etkiye sahip olan lokuslar olduğu görülmektedir.



Şekil 4.3. Bimax algoritmasına ait çubuk grafiği

#### 4.3.2. Xmotif Algoritması Sonuçları

Xmotif algoritması kullanılarak oluşturulan ikili kümelerin içerikleri Tablo 4.7’de verilmiştir. Xmotif algoritması sonucunda 3 farklı ikili küme oluşmuştur. Büyüklüklerine göre ikili kümeler incelendiğinde, küme 1’in 150 koloniden 86

koloni içerdiği (%57.33), küme 2'nin 33 koloni (%22.00), küme 3'ün 30 koloni (%20.00) içerdiği görülmektedir.

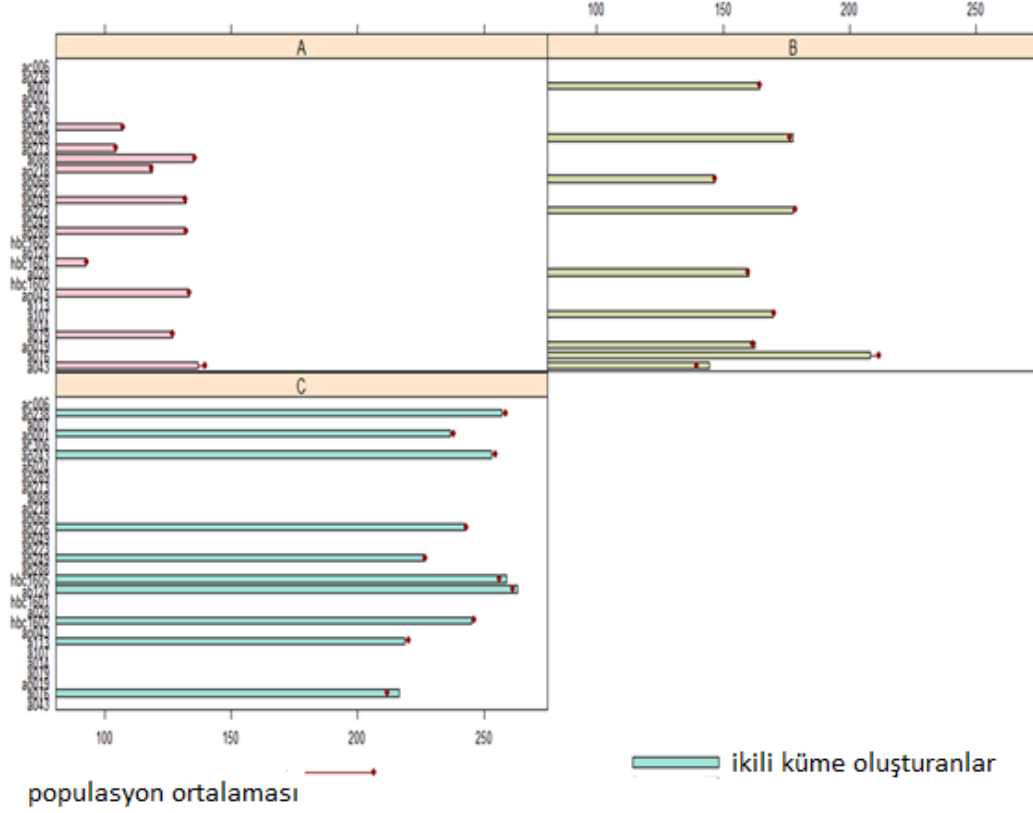
Tablo 4.7. Xmotif algoritmasına göre ikili küme içerikleri

İkili küme	Koloniler(K)	Lokuslar(L)	Büyüklik	Kolonilere göre oranları	KoloniXLokus Büyüklük oranları
A	M1 M2 M7 M13 M14 M15 M16 M17 M18 M19 M24 M25 M26 M27 M28 M29 H2 H3 H7 H8 H9 H10 H12 H16 H20 H21 H22 H24 H25 H29 K3 K4 K6 K7 K8 K9 K10 K12 K13 K14 K16 K17 K18 K19 K22 K23 K24 K28 K30 A3 A7 A9 A11 A12 A13 A14 A15 A16 A18 A19 A20 A21 A24 A25 A26 A27 A28 D1 D2 D3 D4 D6 D7 D8 D9 D16 D20 D21 D22 D23 D24 D25 D26 D27 D28 D29	ab024 ap273 a088 ap218 ap049 ap288 hbc1601 ap043 a079 a043	86x10	%57.33	%19.1
B	M3 M4 M9 M10 M11 M20 M22 M23 H4 H17 H18 H19 H26 H27 H30 K11 K20 K21 K25 K27 A2 A4 A5 A10 A17 A22 A30 D5 D13 D14 D15 D19 D30	a007 ap289 ap068 ap223 a028 a107 ap019 a076 a043	33x9	%22	%6.6
C	M5 M6 M8 M12 M21 M30 H1 H5 H6 H11 H13 H14 H15 H28 K1 K2 K5 K15 K26 K29 A1 A6 A8 A23 A29 D10 D11 D12 D17 D18	ap238 ap001 ap243 ap226 ap249 hbc1605 ab124 hbc1602 a113 a076	30x10	%20	%6.7

Xmotif algoritması sonucu belirlenen ikili kümelere ait çubuk grafiği Şekil 4.4'de gösterilmiştir. Çubuk grafiğinde her ikili küme ayrı bir harf ile ifade edilmiş (A, B, C) ve küme içerisinde yer alan Lokuslar gösterilmiştir. Tablo 4.7 ve Şekil 4.4 incelendiğinde Xmotif algoritmasının kolonilere göre kümeleri ayırmada örtüşme sağlamadığı görülmektedir. Yani her 3 ikili kümede de farklı koloniler bulunmaktadır.

Oluşan ikili kümelerin içerdiği lokuslar incelendiğinde ise a043 lokusu hem küme A'da hem de küme B'de, a076 lokusu ise ikili küme B ve C içerisinde bulunmaktadır.

İkili kümeler içerdikleri lokuslar bakımından incelendiğinde; a088 ve a043 lokuslarının A kümesi, a016 lokusunun B kümesi, ab124 lokusunun ise C kümesi içerisinde en fazla etkiye sahip olan lokuslar olduğu görülmektedir.

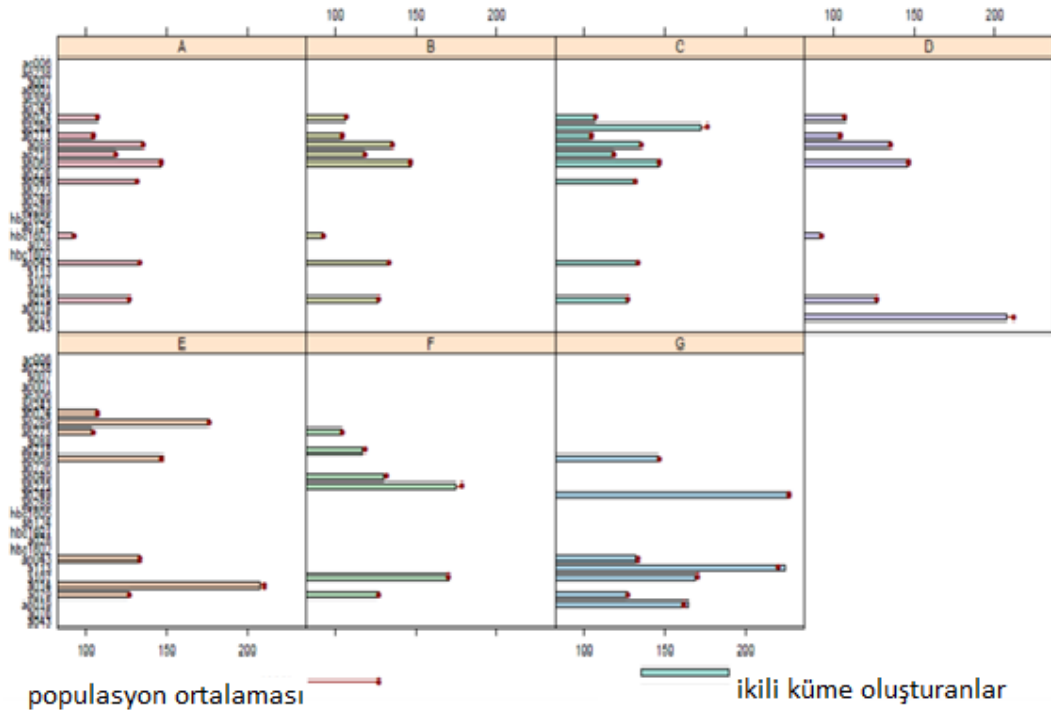


Tablo 4.8. Plaid algoritmasına göre ikili küme içerikleri

İkili küme	Koloniler(K)	Lokuslar(L)	Büyükük	Kolonilere göre oranları	KoloniXLokus Büyükükük oranları
A	M14 M15 M16 M17 M18 M25 M26 H1 H8 H11 H18 H25 H26 H27 H28 H29 K2 K3 K4 K8 K15 K22 K26 A2 A6 A7 A12 A13 A14 A15 A16 A17 A19 A20 A22 A24 A25 A26 A29 A30 D1 D2 D3 D7 D8 D11 D13 D15 D16 D17 D19 D20 D22 D24 D25 D27 D28 D29 D30	ab024 ap273 a088 ap218 ap068 ap049 hbc1601 ap043 a079	59x9	%39.33	%11.8
B	M3 M4 M5 M6 M7 M8 M10 M11 M19 M20 M23 M27 M28 M29 M30 H3 H4 H5 H12 H13 H15 H17 H19 H20 H21 H22 H24 K1 K6 K7 K9 K16 K17 K27 A1 A9 A11 A21 D14 D18 D26	ab024 ap273 a088 ap218 ap068 hbc1601 ap043 a079	41x8	%27.33	%7.3
C	M2 M13 M22 K14 K23 K30 A3 A4 A23 A27 A28 D5 D6 D10 D21	ab024 ap289 ap273 a088 ap218 ap068 ap049 ap043 a079	15x9	%10	%3.0
D	M9 M12 M24 H7 H10 H16 H30 K11 K12 K13 K24 K25 A10 A18 D4 D23	ab024 ap273 a088 ap068 hbc1601 a079 a076	16x7	%10.67	%2.3
E	M21 H2 K18 K20 K28 A5 A8 D9	ab024 ap289 ap273 ap068 ap043 a014 a079	8x7	%5.33	%1.2
F	M1 H9 H14 K29 D12	ap273 ap218 ap049 ap223 a107 a079	5x6	%3.33	%0.6
G	H6 H23 K5 K10 K21	ap068 ap249 ap043 a113 a107 a079 ap019	5x7	%3.33	%0.8

Plaid algoritması sonucu belirlenen ikili kümelere ait çubuk grafiği Şekil 4.5’de gösterilmiştir. Çubuk grafiğinde her ikili küme ayrı bir harf ile ifade edilmiş (A, B, C, D, E, F, G) ve küme içerisinde yer alan Lokuslar gösterilmiştir. Çizelge 4.8 ve Şekil 4.5 incelendiğinde Plaid algoritmasının ikili kümeleri ayırmada örtüşmediği görülmektedir. Algoritma sonucunda oluşan ikili kümeler lokuslar bakımından incelendiğinde a079 lokusu bütün ikili kümeler içerisinde bulunmaktadır. Ayrıca ap273 ve ap068 lokusu 6 farklı, ab024 ve ap043 5 farklı, a088 4 farklı, ap218 ve hbc1601 3 farklı ikili küme içerisinde bulunmaktadır.

İkili kümeler içerdikleri lokuslar bakımından incelendiğinde; ap068 lokusunun A ve B kümesi, ap289 lokusunun C kümesi, a079 lokusunun D kümesi, a014 lokusunun E kümesi, ap223 lokusunun F ve ap249 lokusunun G kümesi içerisinde en fazla etkiye sahip olan lokuslar olduğu görülmektedir.



Şekil 4.5. Plaid algoritması için çubuk grafiği

#### 4.3.4. CC Algoritması Sonuçları

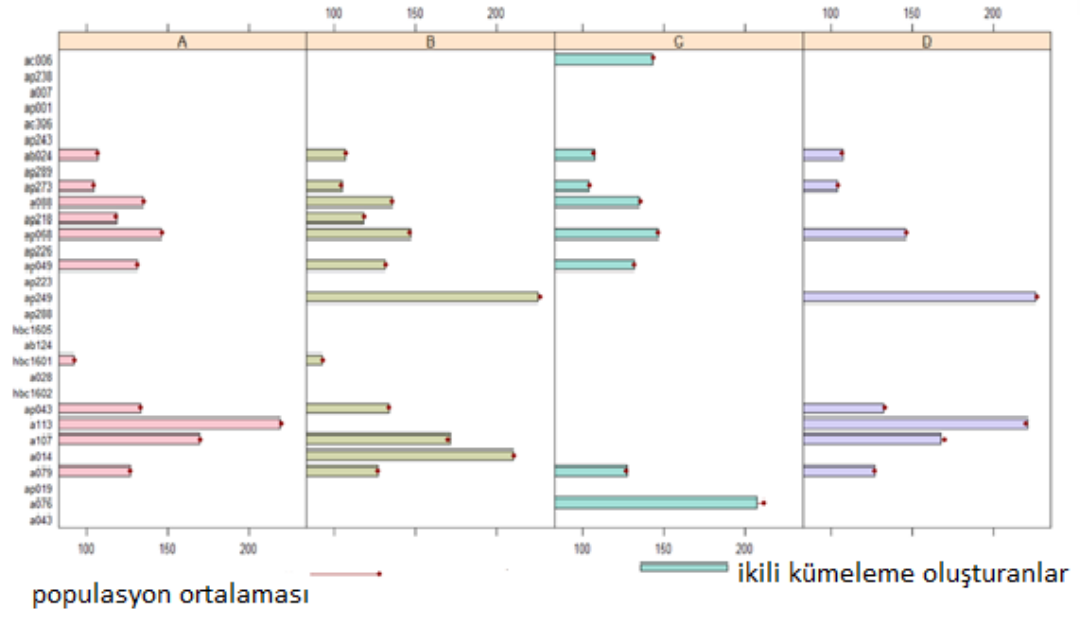
CC algoritması kullanılarak oluşturulan ikili kümelerin içerikleri Çizelge 4.9’da verilmiştir. CC algoritması sonucunda 4 farklı ikili küme oluşmuştur.

Tablo 4.9. CC algoritmasına göre ikili küme içerikleri

İkili küme	Koloniler(K)	Lokuslar(L)	Büyüklik	Kolonilere göre oranları	KoloniXLokus Büyüklük oranları
A	M3 M14 M15 M16 M17 M18 M25 M26 H1 H8 H11 H18 H19 H25 H26 H27 H28 H29 K2 K3 K4 K6 K7 K8 K15 K18 K22 K26 A2 A3 A6 A7 A12 A13 A14 A15 A16 A17 A18 A19 A20 A22 A23 A24 A25 A26 A27 A28 A29 A30 D1 D2 D3 D6 D7 D8 D9 D10 D11 D13 D15 D16 D17 D19 D20 D21 D22 D24 D25 D27 D28 D29 D30	ab024 ap273 a088 ap218 ap068 ap049 hbc1601 ap043 a113 a107 a079	73x11	%48.67	%17.8
B	M1 M4 M5 M6 M7 M8 M9 M10 M11 M12 M13 M19 M20 M21 M23 M27 M28 M29 M30 H2 H3 H4 H5 H12 H13 H14 H15 H17 H20 H21 H22 H24 H30 K1 K9 K14 K16 K17 K24 K27 K30 A1 A4 A9 A11 A21 D5 D14 D18 D26	ab024 ap273 a088 ap218 ap068 ap049 ap249 hbc1601 ap043 a107 a014 a079	50x12	%33.33	%13.3
C	M2 M22 M24 H7 H9 H10 H16 K11 K13 K19 K20 K23 K25 A5 A10 D4 D23	ac006 ab024 ap273 a088 ap068 ap049 a079 a076	17x8	%11.33	%3.0
D	H6 K5 K10 K12 K21 K28 K29 A8 D12	ab024 ap273 ap068 ap249 ap043 a113 a107 a079	9x8	%6.00	%1.6

CC algoritması sonucu belirlenen ikili kümelere ait çubuk grafiği Şekil 4.6'da gösterilmiştir. Çubuk grafiğinde her ikili küme ayrı bir harf ile ifade edilmiş (A, B, C, D) ve küme içerisinde yer alan Lokuslar gösterilmiştir. Tablo 4.9 ve Şekil 4.6'da incelendiğinde CC algoritmasının ikili kümeleri ayırmada örtüşmediği görülmektedir. Algoritma sonucunda oluşan ikili kümeler lokuslar bakımından incelendiğinde ab024, ap273, ap068 ve a079 lokusları tüm ikili kümelerde bulunmaktadır. Ayrıca a088, ap049, ap043 ve a107 lokusları üç farklı, hbc1601, a113 ve ap249 iki farklı kümede bulunmaktadır.

İkili kümeler içerdikleri lokuslar bakımından incelendiğinde; a113 lokusunun A kümesi, ap249 lokusunu B kümesi, a076 lokusunun C kümesi, ap249 lokusunun D kümesi içerisinde en fazla etkiye sahip olan lokuslar olduğu görülmektedir.



Şekil 4.6. CC algoritması için çubuk grafiği

#### 4.3.5. Bayes Algoritması Sonuçları

Bayes algoritması kullanılarak oluşturulan ikili kümelerin içerikleri Çizelge 4.10'da verilmiştir. Bayes algoritması sonucunda 4 farklı ikili küme oluşmuştur.

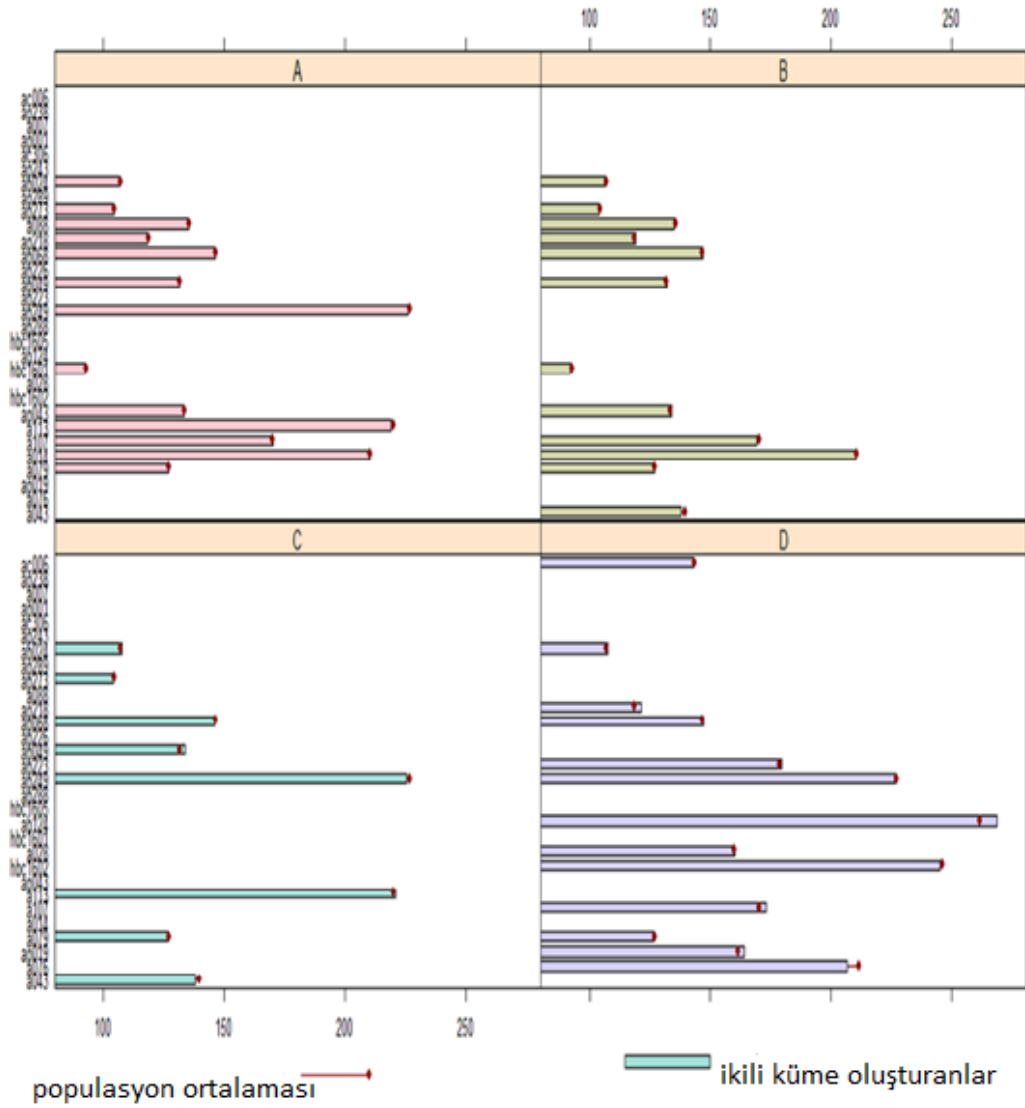
Tablo 4.10. Bayes algoritmasına göre ikili küme içerikleri

İkili küme	Koloniler(K)	Lokuslar(L)	Büyüklik	Kolonilere göre oranları	KoloniXLokus Büyüklük oranları
A	M3 M4 M5 M6 M7 M8 M9 M10 M11 M12 M13 M14 M15 M16 M17 M18 M19 M22 M23 M24 M25 M26 M27 M29 M30 H1 H3 H4 H5 H11 H12 H13 H14 H15 H17 H18 H19 H20 H25 H26 H27 H28 H29 H30 K1 K2 K3 K4 K6 K7 K8 K15 K16 K17 K18 K22 K26 K27 K28 K30 A1 A2 A3 A4 A5 A6 A9 A11 A12 A13 A14 A15 A16 A17 A18 A19 A20 A21 A22 A23 A26 A30 D1 D2 D3 D5 D6 D7 D8 D9 D10 D11 D13 D14 D15 D16 D17 D18 D19 D20 D21 D22 D25 D26 D27 D28 D29 D30	ab024 ap273 a088 ap218 ap068 ap049 ap249 hbc1601 ap043 a113 a107 a014 a079	108x13	%72	%31.2
B	M1 M20 M21 M28 H2 H7 H8 H9 H10 H16 H21 H22 H24 K9 K11 K12 K14 K24 K25 A7 A10 A24 A25 A27 A28 A29 D23 D24	ab024 ap273 a088 ap218 ap068 ap049 hbc1601 ap043 a107 a014 a079 a043	28x12	%18.67	%7.5
C	M2 H6 K5 K10 K13 K21 K23 K29 A8 D4 D12	ab024 ap273 ap068 ap049 ap249 a113 a079 a043	11x8	%7.33	%2.0
D	H23 K19 K20	ac006 ab024 ap218 ap068 ap223 ap249 ab124 a028 hbc1602 a107 a079 ap019 a076	3x13	%2	%0.9

Bayes algoritması ile elde edilen ikili kümelere ait çubuk grafiği Şekil 4.7’de gösterilmiştir. Çubuk grafiğinde her ikili küme ayrı bir harf ile ifade edilmiş ve küme içerisinde yer alan Lokuslar gösterilmiştir. Tablo 4.10 ve Şekil 4.7 incelendiğinde Bayes algoritmasının ikili kümeleri ayırmada örtüşmediği görülmektedir. Algoritma

sonucunda oluşan ikili kümeler lokuslar bakımından incelendiğinde ap068, a079 ve ab024 lokusları tüm ikili kümelerde bulunmaktadır. Ayrıca ap273, ap218, ap049, ap249 ve a107 lokusları üç farklı, a088, ap043, a113 ve a014 lokusları ise iki farklı ikili kümede bulunmuştur.

İkili kümeler içerdikleri lokuslar bakımından incelendiğinde; ap249 lokusunun A kümesi, a014 lokusunun B kümesi, ap249 lokusunun C kümesi ve ab124 lokusunun D kümesi içerisinde en fazla etkiye sahip olan lokuslar olduğu görülmektedir.



Şekil 4.7. Bayes algoritması çubuk grafiği

#### 4.4. Algoritmaların Karşılaştırılması

Arı verisi kullanılarak elde edilen ikili kümeler CC, Bimax, Plaid ve Xmotif algoritmaları ile elde edilmiştir. Bimax ve Xmotif algoritması sonucunda ortaya çıkan ikili kümeler tek değerli elemanlardan oluştuğu için değerlendirme ölçüleri olan VAR ve HKO bulunamamıştır. Bu algoritmalar sadece CKSB skoruna göre karşılaştırılması yapılmıştır.

Tablo 4.11 incelendiğinde, Bimax ve Xmotif algoritması sonucunda ortaya çıkan ikili kümelerden Bimax algoritması CKSB skoruna göre daha iyi sonuçlar ortaya çıkarmıştır. Plaid, CC ve Bayes algoritmalarıyla elde edilen ikili kümelerin karşılaştırılmasında VAR, HKO ve CKSB ölçüleri kullanılmıştır. CKSB skoru ve küme boyutları incelendiğinde Bimax algoritması en iyi sonucu vermiştir. Algoritmaların VAR ve HKO ölçüleri sonuçları değerlendirildiğinde ise Bayes algoritması daha düşük değere sahiptir ve Plaid ve CC algoritmalarına göre daha başarılı ikili kümeleme yaptığı söylenebilir. Karşılaştırma tablosu incelendiğinde Plaid algoritmasının ikili küme oluşturmada en başarısız algoritma olduğu görülmektedir.

Tablo 4.11. Algoritmaların karşılaştırma sonucu

	MSR	VAR	CKSB	Ort.Satır Boyutu	Ort.Sütun Boyutu	Ort. Küme Boyutu
Plaid	0.1123	15.121	0.512	21.29	7.57	161.17
Xmotif	-	-	0.946	39.66	9.66	383.12
Bimax	-	-	1.261	58.00	11.66	676.28
CC	0.0724	10.084	0.897	37.25	9.75	363.19
Bayes	0.0641	9.211	1.072	37.50	11.5	431.25

Tablo 4.12'de verilen Jaccard indeksi incelendiğinde Plaid algoritması sonucunda ortaya çıkan ikili kümeler Bimax algoritması sonucunda ortaya çıkan ikili kümelerden, diğer algoritmalara göre daha farklı bulunmuştur.

Xmotif algoritması ile CC algoritması sonucunda elde edilen ikili kümeler karşılaştırıldığında, iki algoritma sonucunda elde edilen ikili kümelerin birbirine en benzer ikili kümeler olduğu görülmüştür.

Tablo 4.12. Jaccard indeks sonuçları

Algoritma	Plaid	Xmotif	Bimax	CC	Bayes
Plaid	1	0.0882	0.0515	0.0814	0.0762
Xmotif		1	0.0624	0.1412	0.0817
Bimax			1	0.0558	0.0769
CC				1	0.0834
Bayes					1

## 5. SONUÇ

Çalışmada ilk veri seti olarak koyun yetiştiriciliğinde gebelik dönemi sorunları azaltmak amacıyla geliştirilen çözümler ve bu çözümlerin getirmiş olduğu faydaların ikili kümelenmesi ile çözümlerin ve faydaların birlikte değerlendirilerek, koyun yetiştiriciliğindeki uygulamaların daha kullanışlı hale getirilmesi amaçlanmıştır. İkinci veri setinde ise ikili kümeleme algoritmaları kullanılarak çeşitli bal arılarının genetik açıdan durumlarını, filogenetik ilişkilerini ve popülasyonlarının genetik yapılarını belirlenmesi amaçlanmıştır.

Sheepnet veri seti kullanılarak yapılan ikili kümeleme analizi için Bimax, Xmotif ve Bayes algoritmaları kullanılmıştır. Algoritmaların ikili kümeleme performansları CKSB skoru, ortalama küme boyutları kullanılarak karşılaştırılmıştır. Bimax algoritması kullanıldığında elde edilen CKSB skoru 0.538 ve küme boyutu 29.92 olarak bulunmuştur. Xmotif algoritması kullanılarak elde edilen CKSB skoru 0.617 ve küme boyutu 36.82 ve Bayes algoritması sonucunda CKSB skoru 0.742 ve küme boyutu 58.75 olarak bulunmuştur. Karşılaştırma kriterleri dikkate alındığında en yüksek CKSB skoru ve en büyük ortalama küme boyutu Bayes algoritması sonucunda bulunmuştur. Bu sonuçlar ışığında Bayes algoritması en iyi ikili kümeleme performansını sunmuştur. Cheng ve Church (2000), çalışmalarında gen ifade verilerinde HKO skorlarına sahip alt matrisleri bulmak için CC algoritması tanımlamışlar ve maya ve insan verilerinde ortak düzenleme modellerini bulmada CC algoritmasının iyi performans gösterdiğini belirtmişlerdir. Ayrıca Kocatürk (2018), çalışmasında gen ifade verilerinde benzer ifade yapılarında benzer gen gruplarını belirlemek için ikili kümeleme algoritmalarından CC, Bimax, Plaid ve Xmotif algoritmalarını kullanmış olup algoritmaların kümeleme performanslarını karşılaştırmada VAR, MSR ve CKSB skorunu kullanmıştır. Arslan (2018), ise ikili kümelerin performanslarını karşılaştırmada ise ortalama küme boyutundan yararlanmıştır. Ayrıca oluşturulan ikili kümelerin benzerlik yapılarını ortaya koymak için Jaccard indeksinden yararlanılmıştır. Jaccard indeksine göre algoritmalar değerlendirildiğinde ise Bayes algoritması ile Bimax algoritma sonucunda oluşan ikili kümelerin birbirinden en farklı kümeler olduğu belirlenmiştir. Prelic vd. (2006), yapay gen ifade verileri üzerinde ikili kümeleme algoritmalarından CC, Bimax, OPSM, SAMBA, Xmotif ve ISA performanslarını incelemişlerdir. Karşılaştırma kriteri olarak ikili küme algoritmalarından elde edilen ikili kümelerin benzerliklerini belirleyen

Jaccard indeksini kullanmışlar ve en iyi sonucu Bimax algoritması ile elde etmişlerdir.

Bayes algoritması sonucunda dört farklı ikili küme oluşmuştur. Bu kümelerden küme 1'in 55 çözümden 17 çözümü içerdiği (%30.90), küme 2'nin 19 çözümü içerdiği (%34.55), küme 3'ün 4 çözümü içerdiği (%7.27) ve küme 4'ün 7 çözümü içerdiği (%12.72) belirlenmiştir. İkili kümeler içerdiği faydalar bakımından incelendiğinde küme 1'in F2, F3, F4, F5, F10; küme 2'nin F1, F6, F9, F10; küme 3'ün F1, F3, F4, F5, F7, F8; küme 4'ün ise F1, F4, F8, F9, F10 olduğu görülmüştür. F1, F4 ve F10 faydaları üç farklı ikili küme içerisinde bulunmuştur. Bu durumun geliştirilen farklı çözümlerin bu üç faydayı ortak sağladığını gösterdiği düşünülmektedir. Ayrıca Bayes algoritması sonucunda Ç5, Ç8, Ç10 Ç11, Ç19, Ç42, Ç44 ve Ç47 hiçbir ikili kümede bulunmayıp herhangi bir fayda getirmediği belirlenmiştir.

Bal arısı veri seti kullanılarak yapılan ikili kümeleme analizi için Bimax, Xmotif, CC, Plaid ve Bayes algoritmaları kullanılmıştır. Algoritmaların ikili kümeleme performansları HKO, VAR, CKSB skoru ve ortalama küme boyutları kullanılarak kullanılarak karşılaştırılmıştır. Bimax, Xmotif algoritmaları kullanılarak tespit edilen ikili kümelerin eleman değerleri (0-1) yapısında olduğundan dolayı VAR ve MSR sonucu hesaplanamamıştır. Bu yüzden bu kriterlere göre Bimax ve Xmotif algoritmalarını karşılaştırma yapılamamıştır. Plaid algoritması kullanıldığında elde edilen HKO değeri 0.1123, VAR değeri 15.121, CKSB değeri 0.512 ve ortalama küme boyutu 161.17 olarak bulunmuştur. Xmotif algoritmasında CKSB değeri 0.946 ve ortalama küme boyutu 383.12 olarak tespit edilmiştir. Bimax algoritmasında ise CKSB değeri 1.261 ve ortalama küme boyutu 676.28 olarak bulunmuştur. Ayrıca CC algoritması sonucunda MSR değeri 0.0724, VAR değeri 10.084, CKSB değeri 0.897 ve ortalama küme boyutu 363.19; Bayes algoritması sonucunda ise HKO değeri 0.0641, VAR değeri 10.084, CKSB değeri 1.072 ve ortalama küme boyutu 431.25 olarak tespit edilmiştir. Plaid, CC ve Bayes algoritmaları sonucunda elde edilen HKO, VAR, CKSB ve ortalama küme boyutu kriterleri incelendiğinde en küçük HKO ve VAR ve en büyük CKSB değerine Bayes algoritması sonucunda ulaşılmıştır. Ancak Bimax ve Xmotif algoritmalarını da karşılaştırmaya dahil ettiğimizde CKSB skoru en yüksek olan ve en büyük ortalama küme boyutuna sahip olan algoritmanın Bimax algoritması olduğu tespit edilmiştir. Ayrıca karşılaştırma kriterleri incelendiğinde en kötü ikili kümeleme performansını sırası ile Plaid ve CC algoritmaları göstermiştir. Bhattacharya vd. (2012), çalışmalarında maya verilerine çeşitli ikili kümeleme

algoritmaları uygulayarak karşılaştırmış ve en düşük performansı CC algoritmasında belirlemişlerdir. Fare verisinde ise diğer algoritmalara göre Bimax algoritmasıyla elde edilen ikili kümeler daha anlamlı sonuç vermiştir. Prelic vd. (2006), çalışmalarında Bimax algoritmasının, gen eşleşme skorunda diğer algoritmalara göre daha iyi performans sergilediğini ve Bozdağ vd. (2010), çalışmalarında, CC algoritmasının daha kötü performans sergilediği de bu durumu desteklemektedir. Algoritmaların sonucunda oluşan ikili kümelerin benzerliklerini karşılaştırmada kullanılan Jaccard indeks değerleri incelendiğinde ise birbirinden en farklı ikili kümeler Bimax ve CC algoritması sonucunda oluşmuştur. Bu sonuç diğer karşılaştırma kriterlerinin tutarlılığını desteklemektedir.

Bimax algoritması sonucunda üç farklı küme oluşmuştur. A kümesi 150 koloninin tamamını içerisinde bulundurmakta ve ap238, ap001, ap243, ap226, ap249, hbc1605, ab124 hbc1602, a113, a014, a076 lokusları ile bir ikili küme oluşturmaktadır. B kümesi 20 koloni içermekte ve ap238 ap001 ap243 ap226 ap223 ap249 hbc1605 ab124 hbc1602 a113 a014 a076 lokusları ile bir ikili küme oluşturmaktadır. Son olarak C kümesi ise 4 koloni içermekte ve ap238 ap001 ap243 ap289 ap226 ap249 hbc1605 ab124 hbc1602 a113 a014 a076 lokusları ile ikili küme oluşturmaktadır. Ayrıca ap238, ap001, ap243, ap289 ap226, ap249, hbc1605, ab124, hbc1602, a113, a014, a076 lokusları her üç ikili kümede bulunmuşlardır. Ayrıca ikili küme B içerisinde ap223 ve ikili küme C içerisinde ap289 lokusu bulunmaktadır. İkili küme A tüm kolonileri içerisine alan büyük boyutlu bir küme olmuştur. İkili küme B içerisinde bulunan M5, M6, M8, M16, M22, M24, M30, K9, K10, K11, A7, A8 A9, A10, A11, A15, A16, D8, D9, D27 kolonileri ise ikili küme A içerisinde de mevcuttur.

Elde edilen sonuçlar değerlendirildiğinde her iki veri setinin farklı yapılarda olması kullanılan algoritmaların farklı hassasiyetle çalışmasına neden olmuştur. İlk veri seti için Bayes algoritması sonucunda daha iyi sonuçlar elde edilirken, diğer veri seti için Bimax algoritması daha iyi sonuçlar vermiştir. İkili kümeleme algoritmalarının zootekni alanında farklı veri setlerinde kullanılmasıyla daha etkili sonuçlar elde edileceği ve kümeleme analizine farklı bir bakış açısı getireceği düşünülmektedir.

## KAYNAKLAR

- Aguilar-Ruiz, J. S. (2005). Shifting and scaling patterns from gene expression data. *Bioinformatics*. 21(20). 3840-3845.
- Alburaki, M., Bertrand, B., Legout, H., Moulin, S., Alburaki, A., Sheppard, W. S., Garnery, L. (2013). A fifth major genetic group among honeybees revealed in Syria. *BMC Genetics*. 14(1). 117-127.
- Alptekin, N., Yeşilaydın, G. (2015). OECD ülkelerinin sağlık göstergelerine göre bulanık kümeleme analizi ile sınıflandırılması. *İşletme Araştırmaları Dergisi*. 7(4). 137-155.
- Arslan R. (2018). Suç Verilerinin Analizinde İkili Kümeleme Yöntemi: Kaçakçılık Verileri Üzerine Bir Uygulama. Doktora Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Atalay, A., Tortum, A. (2010). Türkiye'deki illerin 1997-2006 yılları arası trafik kazalarına göre kümeleme analizi. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*. 16(3). 1997-2006.
- Bhattacharya, A., Chowdhury, N., K. De, R. (2012). Comparative analysis of clustering and biclustering algorithms for grouping of genes: co-function and co-regulation. *Current Bioinformatics*. 7(1). 63-76.
- Bodur, C., M. Kence., A. Kence, 2007. Genetic structure of honey bee, *Apis mellifera* L. (Hymenoptera: Apidae) populations of Turkey inferred from microsatellite analysis. *Journal of Apicultural Research*. 46 (1). 50-56.
- Bozdağ, D., Kumar, A. S., Çatalyürek, U. V. (2010). Comparative analysis of biclustering algorithms. *In Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology*. 265-274.
- Branders, V., Schaus, P., Dupont, P. (2019). Identifying gene-specific subgroups: an alternative to biclustering. *BMC bioinformatics*. 20(1). 1-13.
- Busygın, S., Prokopyev, O. A., Pardalos, M. P. (2008). Biclustering in data mining. *Computers and Operations Research*. 35(9). 2964-2987.
- Canovas, F., De la Rua, P., Serrano, J., Galian, J., (2011). Microsatellite variability reveals beekeeping influences on Iberian honeybee populations. *Apidologie*. 42(3). 235-251.
- Cengiz, D., Öztürk, F. (2012). Türkiye'de illerin eğitim düzeylerine göre kümeleme analizi ile incelenmesi. *Trakya Üniversitesi Sosyal Bilimler Dergisi*. 14(1). 69-84.
- Cheng, Y., Church G. M. (2000). Biclustering of expression data. *International Conference on Intelligent Systems for Molecular Biology*. (8). 93-103.
- Chia, B. K. H., Karuturi, R. K. M. (2010). Differential co-expression framework to quantify goodness of biclusters and compare biclustering algorithms. *Algorithms for Molecular Biology*. 5(1). 23.
- Çelik, S. (2012). *Biclustering using nonparametric Bayesian methods*. Doktora Tezi, Boğaziçi Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- Ghassemi-Khademi, T., Rajabi-Maham, H., Pashaei-Rad, S. (2018). Genetic diversity evaluation of Persian honeybees (*Apis mellifera meda*) in North West of Iran, using microsatellite markers. *Journal of Wildlife and Biodiversity*. 2(2). 37-46.

- Govaert, G., Nadif, M. (2008). Block clustering with Bernoulli mixture models: Comparison of different approaches. *Computational Statistics & Data Analysis*. 52(6). 3233-3245.
- Gu, J., Liu, J. S. (2008). Bayesian biclustering of gene expression data. *BMC genomics*. 9(1). 1-10.
- Hartigan, John A. (1972). Direct clustering of a data matrix. *Journal of the American Statistical Association*. 67(337): 123-129.
- Hassett, J., Browne, K. A., McCormack, G. P., Moore, E., Society, N. I. H. B., Soland, G., Geary, M. 2018. A significant pure population of the dark European honey bee (*Apis mellifera mellifera*) remains in Ireland. *Journal of Apicultural Research*. 57(3). 337-350.
- Hofman, I., Jarvis, R. (1998). Robust and efficient cluster analysis using a shared near neighbours approach. In Proceedings. *Fourteenth International Conference on Pattern Recognition*. (1). 243-247.
- Huang, Q., Chen, Y., Liu, L., Tao, D., Li, X. (2019). On combining biclustering mining and AdaBoost for breast tumor classification. *IEEE Transactions on Knowledge and Data Engineering*. 32(4). 728-738.
- Ilyasov, R. A., Poskryakov, A. V., Petukhov A. V., Nikolenko, A. G. (2016) . Molecular genetic analysis of five extant reserves of black honeybee *Apis melifera melifera* in the Urals and the Volga region. *Russian Journal of Genetics*. 52(8). 828-839.
- Jaccard, P. (1901). Distribution de la flore alpine dans le bassin des dranses et dans quelques regions voisines. *Bulletin de la Societe Vaudoise des Sciences Naturelles*. 37(140). 241-272.
- Kaban, P. A., Kurniawan, R., Caraka, R. E., Pardamean, B., Yuniarto, B. (2019). Biclustering method to capture the spatial pattern and to identify the causes of social vulnerability in Indonesia: A new recommendation for disaster mitigation policy. *Procedia Computer Science*. 157. 31-37.
- Kaiser, S. (2011). *Biclustering: Methods, Software and Application*. Doctoral dissertation, Munich Ludwig-Maximilians-University Department of Statistics, Munich.
- Kalaycı, Ş. (2009). *SPSS uygulamalı çok değişkenli istatistik teknikleri*. Asil Yayın Dağıtım.
- Karabağ, K., Tunca, R. İ., Sevim, E. T., Doğaroğlu, T. (2020). Current genetic status of honey bees in Anatolia in terms of thirty polymorphic microsatellite markers. *Turkish Journal of Entomology*. 44(3). 333-346.
- Kaygısız, Z., Saraçlı, S., Dokuzlar, K. U. (2005). İllerin gelişmişlik düzeyini etkileyen faktörlerin Path analizi ve Kümeleme analizi ile incelenmesi, *VII. Ulusal Ekonometri ve İstatistik Sempozyumu*, İstanbul.
- Kluger, Y., Basri, R., Chang, J. T., Gerstein, M. (2003). Spectral biclustering of microarray data: coclustering genes and conditions. *Genome Research*. 13(4). 703-716.
- Kocatürk A. (2018). İkili Kümeleme Algoritmalarının Görsel Ve Sayısal Açından Karşılaştırılması. Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Lazzeroni, L., Owen, A. (2002). Plaid models for gene expression data. *Statistica Sinica*, 61-86.

- Li, L., Guo, Y., Wu, W., Shi, Y., Cheng, J., Tao, S. (2012). A comparison and evaluation of five biclustering algorithms by quantifying goodness of biclusters for gene expression data. *BioData mining*. 5(1). 1-10.
- Mina, E. (2001). *Applying Biclustering to understand the molecular basis of phenotypic diversity*, Thesis, Utrecht University Department of Information and Computing Sciences, Utrecht.
- Mitra, S., Banka, H. (2006). Multi-objective evolutionary biclustering of gene expression data. *Pattern Recognition*. (39). 2464-2477.
- Murali, T. M., Kasif, S. (2003). Extracting conserved gene expression motifs from gene expression data. *Pacific Symposium on Biocomputing*. (8). 77-88.
- Nicholls, K., Wallace, C. (2021). Comparison of sparse biclustering algorithms for gene expression datasets. *Briefings in bioinformatics*. 22(6). 140.
- Özdamar, K. (1999). *Paket Programlar ile İstatistiksel Veri Analizi-2*, II. Baskı, Kaan Kitabevi, Eskişehir.
- Özdamar, K. (2002). *Paket Programlar ile İstatistiksel Veri Analizi (Çok değişkenli Analizler)*, IV. Baskı, Kaan Kitabevi, Eskişehir.
- Padilha, V. A., Campello, R. J. (2017). A systematic comparative evaluation of biclustering techniques. *BMC bioinformatics*. 18(1). 1-25.
- Pang, C. (2022). Construction and Analysis of Macroeconomic Forecasting Model Based on Biclustering Algorithm. *Journal of Mathematics*.
- Prelic, A., Bleuler, S., Zimmermann, P. Wil, A., Buhlmann, P., Gruissem, W., Hennig, L., Thiele, L., Zitzler, E. (2006). A Systematic Comparison and Evaluation of Biclustering Methods for Gene Expression Data Bioinformatics. *Oxford Univ Press*. (22). 1122-1129.
- R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rocha, O., Mendes, R. (2018). JBiclustGE: Java API with unified biclustering algorithms for gene expression data analysis. *Knowledge-Based Systems*. 155. 83-87.
- Rodriguez-Baena, D. S., Perez-Pulido, A. J., Aguilar– Ruiz, J. S. (2011). A Biclustering algorithm for extracting bit-patterns from binary datasets. *Bioinformatics*. 27(19). 2738-2745.
- Singh, M., Mehrotra, M. (2018). Impact of biclustering on the performance of biclustering based collaborative filtering. *Expert Systems With Applications*. (113). 443-456.
- Tatlıdil, H. (1996). *Uygulamalı Çok Değişkenli İstatistiksel Analiz*, Ankara.
- Tchagang, A. B., Pan, Y., Famili, F., Tewfik, A. H., Benos, P. V. (2013). Biclustering of dna microarray data: Theory, evaluation, and applications. In *Bioinformatics: Concepts, Methodologies, Tools, and Applications*. 513-551.
- Tekin, B. (2020). Covid-19 Pandemisi Döneminde Ülkelerin Covid-19, Sağlık Ve Finansal Göstergeler Bağlamında Sınıflandırılması: Hiyerarşik Kümeleme Analizi. *Finans Ekonomi Ve Sosyal Araştırmalar Dergisi*. 5(2). 336-349.
- Turner, H., Bailey, T., Krzanowski, W. (2003). Improved biclustering of microarray data demonstrated through systematic performance tests. *Computational Statistics & Data Analysis*. 48(2). 235-254.

- Van Mechelen, I., Bock, H. H., De Boeck, P. (2004). Two-mode clustering methods: a structured overview. *Statistical methods in medical research*. 13(5). 363-394.
- Wei, W. J., Shi, B., Guan, X., Ma, J. Y., Wang, Y. C., Liu, J. (2019). Mapping theme trends and knowledge structures for human neural stem cells: a quantitative and co-word biclustering analysis for the 2013–2018 period. *Neural regeneration research*. 14(10). 1823.
- Zhao, H., Chan, K. L., Cheng, L. M., Hong, Y. (2009). A probabilistic relaxation labeling framework for reducing the noise effect in geometric biclustering of gene expression data. *Pattern Recognit*. 42(11). 2578–2588.
- Zhao, H., Liew, A. W. C., Xie, X. Yan, H. (2007). A new geometric biclustering algorithm based on the Hough transform for analysis of large-scale microarray data. *Journal of Theoretical Biology*. 251(3). 264–74.
- Zhao, H., Wee-Chung Liew, A., Z Wang, D., Yan, H. (2012). Biclustering analysis for pattern discovery: current techniques, comparative studies and applications. *Current Bioinformatics*. 7(1). 43-55.

## EKLER

(Ek 1)

Bal Arısı Veri Seti

Populasyonlar	Lokuslar								
	ac006	ap238	a007	...	a043				
Muğla M1	140	140	260	278	164	164	...	132	138
Muğla M2	138	138	256	256	162	162	...	140	140
Muğla M3	141	145	260	260	153	153	...	144	144
Muğla M4	138	138	256	256	164	164	...	146	146
...	...	...	...	...	...	...	...	...	...
Hatay H1	136	136	264	266	160	162	...	216	142
Hatay H2	150	162	246	246	156	156	...	218	138
Hatay H3	148	148	246	246	162	162	...	212	140
Hatay H4	145	145	245	245	163	174	...	212	140
...	...	...	...	...	...	...	...	...	...
Kırklareli K1	145	145	246	246	162	162	...	212	216
Kırklareli K2	145	145	246	246	156	159	...	212	218
Kırklareli K3	145	145	244	244	156	156	...	216	218
Kırklareli K4	149	149	246	246	163	163	...	216	220
...	...	...	...	...	...	...	...	...	...
Artvin A1	140	141	246	246	156	156	...	220	220
Artvin A2	150	150	278	278	162	162	...	206	206
Artvin A3	141	141	246	246	162	178	...	204	204
Artvin A4	138	138	262	262	173	173	...	206	206
...	...	...	...	...	...	...	...	...	...
Düzce D1	140	140	262	264	160	160	...	214	216
Düzce D2	143	143	258	258	160	160	...	208	212
Düzce D3	140	140	278	278	160	160	...	206	212
Düzce D4	140	140	258	258	156	160	...	206	208
...	...	...	...	...	...	...	...	...	...
Düzce D30	140	140	264	276	162	162	...	140	144

## Sheepnet Veri Seti

Çözümler	Faydalar				
	F1: Üretimi Artırma	F2: Döl Verimi	F3: Doğurganlık	...	F10: Sürünün daha iyi yönetimi
Ç1: Elektronik alfa dedektör( Altlık aktivitesinin tespiti)	1	1	0	...	0
Ç2: Parazit yükünün değerlendirilmesi ve toplanmış dışkı örneklerinin kullanımı	1	0	0	...	
Ç3: Düşüklerin standartlaştırılmış ayırıcı tanısı: abortif bulaşıcı hastalıkların teşhisine küresel bir yaklaşım	0	0	0	...	1
Ç4: Gebe koyunlarda gebeliğin tespiti için transabdominal ultrason taraması	1	1	1	...	1
Ç5: İyi bir koyun verimliliği için anahtar uygulamaların planlanması	1	1	0	...	0
Ç6: Sağlıklı kuzular: iyi yetiştirme uygulamaları ve uyarlanmış ahır	1	1	1	...	0
...	...	...	...	...	...
Ç55: EBV koçlarını kullanan teknik not	1	0	0	...	1

(Ek 2)

### Xmotif algoritma kodları

```
## S4 method for signature 'matrix,BCXmotifs'
biclust(x, method=BCXmotifs(), ns=10, nd=10, sd=5, alpha=0.05, number=100)
```

### Bimax algoritma kodları

```
## S4 method for signature 'matrix,BCBimax'
biclust(x, method=BCBimax(), minr=2, minc=2, number=100)
## S4 method for signature 'matrix,BCrepBimax'
biclust(x, method=BCrepBimax(), minr=2, minc=2, number=100, maxc=12)
```

### CC algoritma kodları

```
## S4 method for signature 'matrix,BCCC'
biclust(x, method=BCCC(), delta = 1.0, alpha=1.5, number=100)
```

### Plaid algoritma kodları

```
## S4 method for signature 'matrix,BCPlaid'
biclust(x, method=BCPlaid(), cluster="b", fit.model = y ~ m + a + b,
background = TRUE, background.layer = NA, background.df = 1, row.release = 0.7,
col.release = 0.7, shuffle = 3, back.fit = 0, max.layers = 20, iter.startup = 5,
iter.layer = 10, verbose = TRUE)
```

### Bayes algoritma kodları

<https://jbiclustge.github.io>

## ÖZ GEÇMİŞ

Lütfi Bayyurt, Erzurum Anadolu Öğretmen Lisesi'ni bitirdikten sonra Atatürk Üniversitesi Kazım Karabekir Eğitim Fakültesi, Matematik bölümünden 28.06.2010 tarihinde mezun oldu. 2014-2017 yılları arasında Tokat Gaziosmanpaşa Üniversitesi Zootekni Bölümünde araştırma görevlisi olarak çalıştı. 2017 yılında Ondokuz Mayıs Üniversitesi Lisansüstü Eğitim Enstitüsü Zootekni Ana Bilim dalında doktora programına başladı. Lisans mezuniyetinden bu yana öğretmen ve sonrasında da araştırma görevlisi olarak görev yapan Lütfi Bayyurt, iyi/orta derecede İngilizce bilmektedir (YDS:72,50). Temel ilgi alanları, Kitap okumak, film ve dizi izlemek, futbol oynamaktır 26.05.2022.

### İletişim Bilgileri

ORCID ID : [0000-0003-2613-9302](https://orcid.org/0000-0003-2613-9302)

### Yayınlar:

1. Yılmaz, F., Bayyurt, L., Abacı, S. H., Tahtalı, Y. (2020). Comparison of Least Squares and Some Bias Estimators in Multicollinearity. *Turkish Journal of Agriculture-Food Science and Technology*. 8(3). 793-799.
2. Bayyurt, L., Bayyurt, B. (2020). Forecasting of COVID-19 cases and deaths using ARIMA models. *medrxiv*.
3. Abacı, S. H., Bayyurt, L., Tahtalı, Y., Şekeroğlu, A., Duman, M. (2020). Determination of Relationship between Cage Position and Laying Time by Correspondence Analysis. *Turkish Journal of Agriculture-Food Science and Technology*. (5). 1211-1215.
4. Tahtalı, Y., Sahin, M., Bayyurt, L. (2020). Comparison of different growth curve models in Romanov lambs. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi*. 26(5).
5. Bayyurt, L., Akdag, A., Tirink, C. (2018). Canonical correlation analysis for estimation of relationships between some characteristics and color of broiler meat. *IOSR J Agric Vet Sci*. 11. 61-5.
6. Tirink, C., Bayyurt, L., Abacı, S. H., Önder, H. (2019). Evaluation of dairy cattle milk production in Black Sea region by cluster analysis. *Black Sea Journal of Engineering and Science*. 2(2). 39-41.
7. Okuyucu, I. C., Bayyurt, L., Akdag, A., Bulu, A. (2018). Evaluation of Enterprises Raising Water Buffalo in Samsun Province of Turkey. *Gene*. 4(8).

**8.** Tahtalı, Y., Bayyurt, L., Abacı, S. H. (2019). A Path Analysis for Determination of Factors Affecting 305-Day Milk Yield in Jersey Cows. *Gaziosmanpaşa Üniversitesi Ziraat Fakültesi Dergisi*. 36(3). 172-177.