T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

KAYIP DEĞER ATAMASINA BULANIK KÜMELEME YAKLAŞIMI

Aleyna KOL

YÜKSEK LİSANS TEZİ

İstatistik Anabilim Dalı

İstatistik Programı

Danışman

Prof. Dr. Gülhayat GÖLBAŞI ŞİMŞEK

Aralık, 2023

T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

KAYIP DEĞER ATAMASINA BULANIK KÜMELEME YAKLAŞIMI

Aleyna KOL tarafından hazırlanan tez çalışması 01.12.2023 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, İstatistik Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Prof. Dr. Gülhayat GÖLBAŞI ŞİMŞEK
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Prof. Dr. Gülhayat GÖLBAŞI ŞİMŞEK, Danışman
Yıldız Teknik Üniversitesi _____

Prof. Dr. Fatma NOYAN TEKELİ, Üye
Yıldız Teknik Üniversitesi _____

Doç. Dr. Seher ARIKAN TEZERGİL, Üye
Marmara Üniversitesi _____

Danışmanım Prof. Dr. Gülhayat GÖLBAŞI ŞİMŞEK sorumluluğunda tarafımca hazırlanan Kayıp Değer Atamasına Bulanık Kümeleme Yaklaşımı başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Aleyna KOL

İmza



Aileme

TEŞEKKÜR

Tez çalışma sürecimde desteğini her zaman yanımda hissettiğim, bilgi ve tecrübeleriyle beni yönlendiren, yoluma ışık olan, sabrı ve güler yüzüyle güven vererek bu yolculukta bana eşlik eden çok değerli danışmanım Prof. Dr. Gülhayat GÖLBAŞI ŞİMŞEK'e teşekkürlerimi sunarım. Tez yazımımda yardımlarını benden esirgemeyen, yoğun sorularıma rağmen yanımda olan ve çözüm sunan Sinan USLU'ya ve hoşgörüleri için İstanbul İstatistik Ofisin'deki çalışma arkadaşlarıma teşekkür ederim. Bu süreçte anlayışını benden esirgemeyen Şefim Akın EMİN'e ve manevi destekleri için Fatih YILMAZ'a teşekkür ederim. Tez sürecimde bilgi ve tecrübeleriyle bana katkıda bulunan sorularımı cevapsız bırakmayan Arş. Gör. Bilge Kağan ÖZBAY'a desteklerinden dolayı teşekkür ederim. Lisans eğitimim sürecinde istatistik alanında gelişmemi sağlayan ve bana bu alanı sevdiren Ankara Üniversitesi hocalarıma teşekkürü borç bilirim. Lisans hayatımda tanıdığım ve her an yanımda olacağına emin olduğum Beyza Nur ÖZKAN'a desteklerinden dolayı teşekkür ederim. Bu süreçte beni motive eden ve destekleyen güzel arkadaşlarım Nagihan'a, Duygu'ya, Ayşe Gül'e, Emine'ye ve Şevval'e teşekkür ederim.

Son olarak, desteklerini her zaman yanımda hissettiğim annem Nezahat'a ve babam Yüksel'e, kardeşlerim Almina ve Alper'e çok teşekkür ederim.

Aleyna KOL

İÇİNDEKİLER

TEŞEKKÜR iv
KISALTMA LİSTESİ vi
ŞEKİL LİSTESİ vii
TABLO LİSTESİ viii
ÖZET ix
ABSTRACT xi
1 GİRİŞ 13
1.1 Literatür Özeti 13
1.2 Tezin Amacı 19
1.3 Tezin İçeriği 20
2 KAYIP VERİ 21
2.1 Kayıp Veri 21
2.2 Kayıp Veri Mekanizmaları 22
2.3 Kayıp Veri Analiz Yöntemleri 24
3 BULANIK YAKLAŞIM 26
3.1 Bulanık Mantık 26
3.2 Bulanık C-Ortalamalar 29
4 BULANIK C-ORTALAMA İLE KAYIP VERİ ATAMASI 33
4.1 Bulanık C-Ortalama ile Kayıp Veri Tahmini 33
4.2 Uygulamada Kullanılan Veri Seti 33
4.3 Uygulamada Kullanılan Performans Kriterleri 34
4.4 Uygulamada Kullanılan Küme Sayıları ve Klasik Yöntemler 36
4.5 Verilere Ait Tablolar 37
5 SONUÇ VE ÖNERİLER 49
KAYNAKÇA 52

KISALTMA LİSTESİ

BCO	Bulanık C-Ortalamalar
RMSE	Kök Ortalama Kare Hata
MAE	Ortalama Mutlak Hata
DASS-21	Depresyon Anksiyete ve Stres Ölçekleri-21
FCM	Fuzzy C-Means
MCAR	Missing Completely At Random
MAR	Missing At Random
MNAR	Missing Not At Random
TRK	Tamamen Rassal Kayıp
RK	Rassal Kayıp
ROK	Rassal Olmayan Kayıp
EÇO	En Çok Olabilirlik
EKK	En Küçük Kareler

ŞEKİL LİSTESİ

Şekil 1.1 Tamamen Rassal Kayıp Mekanizması **22**
Şekil 1.2 Rassal Kayıp Mekanizması **23**
Şekil 1.3 Rassal Olmayan Kayıp Mekanizması **23**
Şekil 1.4 Bulanık Mantık ve Küme Teorisi **28**
Şekil 1.5 Bulanık C-Ortalamalar Algoritmasının Çalışma Prensibi **30**

TABLO LİSTESİ

Tablo 1.1 Bulanık Mantık Avantaj ve Dezavantajları **27**
Tablo 1.2 Kullanılan Küme Sayıları ve Klasik Yöntemler **36**
Tablo 1.3 %1 Kayıp ile RMSE-MAE (738) **37**
Tablo 1.4 %1 Kayıp ile RMSE-MAE (369) **38**
Tablo 1.5 %5 Kayıp ile RMSE-MAE (738) **39**
Tablo 1.6 %5 Kayıp ile RMSE-MAE (369) **40**
Tablo 1.7 %10 Kayıp ile RMSE-MAE (738) **41**
Tablo 1.8 %10 Kayıp ile RMSE-MAE (369) **42**
Tablo 1.9 %20 Kayıp ile RMSE-MAE (738) **43**
Tablo 1.10 %20 Kayıp ile RMSE-MAE (369) **44**
Tablo 1.11 %1 Kayıp ile Cronbach Alfa (738) **45**
Tablo 1.12 %1 Kayıp ile Cronbach Alfa (369) **45**
Tablo 1.13 %5 Kayıp ile Cronbach Alfa (738) **46**
Tablo 1.14 %5 Kayıp ile Cronbach Alfa (369) **46**
Tablo 1.15 %10 Kayıp ile Cronbach Alfa (738) **47**
Tablo 1.16 %10 Kayıp ile Cronbach Alfa (369) **47**
Tablo 1.17 %20 Kayıp ile Cronbach Alfa (738) **48**
Tablo 1.18 %20 Kayıp ile Cronbach Alfa (369) **48**

ÖZET

Kayıp Değer Atamasına Bulanık Kümeleme Yaklaşımı

Aleyna KOL

İstatistik Anabilim Dalı

İstatistik Programı

Yüksek Lisans Tezi

Danışman: Prof. Dr. Gülhayat GÖLBAŞI ŞİMŞEK

Çalışmalarımızda yer alan veri setlerinde hücrelerde bulunan değerlerin gözlenememesi durumu eksik veri, kayıp veri veya tamamlanamamış veri olarak isimlendirilmektedir. Kayıp değerleri bulabilmek ve veri setlerindeki eksikleri tamamlayarak daha sağlıklı sonuçlara ulaşabilmek amaçlanmıştır. Eksik verileri tamamlamak için çeşitli kayıp veri tahmin etme yöntemleri bulunmaktadır. Bu çalışmada kayıp verileri tahmin etme yöntemlerine alternatif bir yöntem olarak bulanık c-ortalamalar yöntemi kullanılmıştır. Yöntemin eksik verilerin tamamlanmasında uygunluğu incelenmiştir.

Bu çalışmada "Depresyon Anksiyete ve Stres Ölçekleri-21" (DASS-21) veri seti kullanılmıştır. DASS-21 veri setinden %1, %5, %10 ve %20 oranlarında rastgele veri kaybedilmiştir. Her veri seti için 100 tekrar yapılarak yöntemlerin performansları karşılaştırılmıştır.

Bu çalışmada belirli küme sayılarına ayrılarak RMSE, MAE ve Cronbach alfa değerleri hesaplanmıştır. Veri setlerindeki kayıpların tahmin edilmesinde örneklem büyüklüğünün bir etkisinin olup olmadığını anlamak amacıyla tam veri seti yarıya indirilerek yapılan işlemler yeniden uygulanarak iki veri setinin sonuçları karşılaştırılmıştır. Bulanık C-Ortalamalar 1'den 11'e kadar küme

sayıları artırılarak sonuçlar incelenmiştir. Çıkan değerlerin, sıklıkla kullanılan ortalama atama, medyan atama ve mod atama yöntemleri ile kıyaslanması yapılmıştır. Araştırmanın uygulaması R Studio programında gerçekleştirilmiştir.
Anahtar Kelimeler: Kayıp veri, bulanık c-ortalamalar, bulanık küme

ABSTRACT

Fuzzy Clustering Approach To Missing Value Imputation

Aleyna KOL

Department of Statistics

Master of Science

Supervisor: Prof. Dr. Gülhayat GÖLBAŞI ŞİMŞEK

In the data sets in our studies, the situation where the values in the cell cannot be observed is called missing data, missing data or incomplete data. It is aimed to find missing values and reach more accurate results by completing the missing data in the data sets. There are various missing data estimation methods to complete missing data. In this study, the fuzzy c-means method was used as an alternative method to estimating missing data. The suitability of the method in completing missing data was examined.

"Depression Anxiety and Stress Scales-21" (DASS-21) data set was used in this study. Random data was lost at rates of 1%, 5%, 10% and 20% from the DASS-21 data set. The performances of the methods were compared by performing 100 repetitions for each data set.

In this study, RMSE, MAE and Cronbach's Alpha values were calculated by dividing into certain cluster numbers. In order to understand whether the sample size has an effect on estimating the losses in the data sets, the results of the two data sets were compared by halving the full data set and applying the operations again. Fuzzy C-Means the results were examined by increasing the number of clusters from 1 to 11.

The resulting values were compared with the frequently used mean assignment, median assignment and mode assignment methods. The application of the research was carried out in the R Studio program.

Keywords: Missing values, fuzzy c-means, fuzzy clustering

1
GİRİŞ

1.1 Literatür Özeti

Eksik veri sorunu, istatistik ve veri analizi alanında geniş bir tarihe sahip olan önemli konulardan biridir. Kayıp veri, veri analizi ve modelleme çalışmaları sırasında önemli bir sorun olduğundan, bu konuda yapılan çalışmaların tarihi oldukça eski bir geçmişe sahiptir. Kayıp veri sorununu gidermek için birçok çalışma yapılmıştır. Bu çalışmaların ilk adımları 1900'lü yılların başlarına dayanmaktadır fakat Rubin'in 1976 yılında yaptığı kayıp veri mekanizmaları ile ilgili çalışma bu problemi çözümlemek için dönüm noktası olmuştur.

1910'da Elderton ile Pearson yaptıkları bir çalışmada ebeveynlerin alkol kullanımının etkilerini incelemek adına eksik değerlere ortalama ile atama yaparak kayıp değerleri tamamlama yoluna gitmiştir. Yaptıkları bu çalışma kayıp olan değerlerin tamamlanması için uygulanan sorun giderici bir çözüm olmuştur.

1911 yılında Elderton ve Pearson'un yaptığı çalışmayı eleştiren ve eksiklikleri ele alan isimler Sturge ve Horsley olmuştur. Bu iki isim kayıp veri sorunu için analitik yöntemler sunarak problemin çözülmesine dair öneriler sunmuştur.

1926 yılında McKendrick özyinelemeli tahmin üzerine ilk sistematik çalışmayı yapmıştır. Bu çalışma beklenti en büyükleme algoritmasının versiyonlarından birisi olarak sayılabilir.

1927, 1932 ve 1933 yıllarına yayılan Kermack ve Mckendrick'in epimedi üzerine yürüttükleri bir çalışma vardır ve bu çalışmada eksik olmayan ve eksik olan değişkenlere ait eşitlikler bulmuşlardır.

1930 yıllında Allan ve Wishart eksik değerler bulunduran deney tasarımı üzerinde bir takım çalışmalar yapmışlardır. Bu çalışmada rassal blok düzeni ve latin kare düzeninde bulunan tasarımlar için eksik bir verinin alabileceği değeri tahmin etmeye çalışarak araştırmalarını sürdürmüşlerdir.

1932 yılında Wilks, parçalı örneklemlerde kitle parametrelerinin dağılımlarının ve momentlerinin bulunması üzerine çalışmalar yürütmüştür. Kitle parametrelerinin belirli koşullarda tahmin edilebilmesi için EÇO yöntemlerini kullanmıştır.

1933 ve 1936 yıllarında Yates, tek bir kayıp verinin tahmin edilebilmesi için hata sapmalarını en küçükleyecek genel bir formül ortaya atmıştır. Bu çalışma eksik veri tahmininin istatistiksel temellerini atmıştır. Eksik olan rassal blok ve eksik olan latin kare düzenini açıklayarak deney tasarımındaki eksik verileri tamamlama konusunda çalışmalar yapmıştır.

1938 yılında eksik veri sorununa farklı bir açıdan yaklaşan Bose ve Mahalanobis, değişkenler arasında sadece doğrusal bir kombinasyon olduğu durumlarda deney tasarımı üzerine çalışmışlardır.

1951 yılında Matthai, Wilks'in 1932 yılında iki değişkenli olarak incelediği durumu üç değişkenli olarak inceleyerek kitle parametrelerini tahmin etme konusunda çalışmalarını sürdürmüştür.

1955 yılında Lord'ta, 1951 yılında Matthai'nin yaptığı gibi Wilks'in iki değişkenli olarak incelediği durumu üç değişkenli olarak inceleyerek, normal dağılıma sahip üç değişkenli bir kitlenin parametrelerini eksik verilerden tahmin etme konusunda çalışmalarını sürdürmüştür.

1956 yılında Edgett, eksik veri sorunu olan bağımsız değişkenler ile çoklu regresyon analizi üzerine çalışmalar yapmıştır. Bu çalışmada, normal dağılan üç değişkenli kitlenin parametrelerini tahmin etmek için eksik veri bulunduran bir değişken için EÇO denklemlerini elde etmek üzere çalışmalarını sürdürmüştür.

1956 yılında Healy ve Westmacott, deney tasarımı konusundaki çalışmalarıyla genelleştirilmiş beklenti en büyükleme algoritmasının temelleri atmışlardır. Bu çalışma beklenti en büyükleme algoritmasının temel yapısını geliştiren ve deney tasarımlarında kullanılan önemli bir yöntemi tanımlayan yapı taşlarından biridir. Bu araştırma, istatistiksel analiz ve deney tasarımı alanındaki gelişmelere katkıda bulunur ve ileride yapılacak olan çalışmalara referans noktası oluşturur.

1957 yılında Anderson, Lord ve Edgett tarafından normal dağılan üç değişkenli bir kitle üzerine yapılan çalışmaları inceleyerek, bu çalışmaların sonuçlarını daha genel bir çerçeveye oturtmuş ve bulguları birçok farklı durumda kolayca uygulanabilir hale getirmiştir. Anderson'un yaptığı bu çalışma istatistiksel analiz yöntemlerinin daha geniş bir havuzda kullanılabilirliğini artırmış ve istatistiksel analiz alanında katkılar sağlamıştır.

1957 yılında Nicholson, Wilks'in 1932 yılında iki değişkenli olarak, Lord'un ise 1955 yılında üç değişkenli olarak incelediği durumu p değişkenli durumlar için inceleyerek çalışmalarını sürdürmüştür. Bu çalışma, istatistiksel analiz yöntemlerini farklı değişken sayıları ve özel durumlar için uygulanabilir hale getirme amacı içermektedir.

1957 yılında Wilkinson, eksik verileri gidermek için kovaryans analizini incelemek üzere çalışmalar yapmıştır. Wilkinson, kayıp gözlemlerin eş zamanlı yapılan ölçümler ile mevcut verilerin analizi arasında ilişkisiz olduğunu ve kayıp verilerin tamamlanan veriler üzerinden kovaryans analizi yapmanın daha iyi bir yaklaşım olduğunu belirterek çalışmalarını ilerletmiştir.

1958 yılında Hartley, eksik veri sorunlarını ele alarak EÇO kestriminin genel bir yaklaşımını ilerletmek adına araştırmalar ve çalışmalar yapmıştır. Bu çalışma kayıp veri durumlarına genel bir çerçeve sunarak, EÇO yönteminin daha geniş bir uygulanabilirlik alanına sahip olmasına katkı sağlamıştır.

1958 yılında Wilkinson, varyans analizi gerçekleştirilirken karşılaşılan eksik veri problemlerini gidermek amacıyla yeni denklemler ve yöntemler ortaya atmıştır.

1959 yılında Dear, eksik veri ile karşılaşıldığı durumlarda çoklu regresyon için temel bileşen analizi gerçekleştirmek üzerine araştırmalar ve çalışmalar sürdürmüştür.

1960 yılında Buck, k değişkenli bir kitlenin içindeki kayıp verilerin regresyon teknikleriyle tahmin edilmesi ve bu tahmin değerlerini kullanarak revize edilmiş bir varyans-kovaryans matrisinin hesaplanması konusunda çalışmalar sürdürmüştür.

1964 yılında Glasser, eksik veri setlerinde regresyon modelinin verimliliğinin ilişkisiz değişkenler arasındaki korelasyona ve hücrelerdeki eksik veri oranına bağlı olduğunu belirtmektedir.

1966 yılında Afifi ve Elashoff kayıp veri hakkında yapılmış önceki çalışmaları incelemişlerdir. Ortalama, korelasyon, doğrusal regresyon ve varyans çalışmalarını inceleyen Afifi ve Elashoff eksik verilerin belli bir düzene göre oluştuğunda uygulanacak yöntemleri basite indirgeyecek yöntemler geliştirmişlerdir. 1967 yılında çalışmalarına devam eden Afifi ve Elashoff iki rassal değişken arasındaki doğrusal regresyon kestiriminde eksik veri bulunduğunda bu eksikliği gidermek

üzere olasılıksal yöntemlere yönelmişlerdir. Hata kare ortalamalarını test etmek için EKK yöntemini kullanarak çalışmalarını sürdürmüşlerdir.

1968 yılında Haitovsky regresyon veri setlerindeki eksik verileri tamamlamak için iki farklı yöntem üzerinde çalışmıştır. Üzerinde çalıştığı ilk yöntem EKK yöntemidir. Bu yöntemi uygularken veri setindeki eksik verileri veri setinden çıkartmıştır. Üzerinde çalıştığı ikinci yöntem ise kovaryansların normal eşitlikler üzerinden hesaplanmasıdır. Bu yöntemde tüm değişkenlerin birbirleri ile olan ilişkilerine bakılmıştır. Bu iki yöntem arasında eksik verilerin dışarıda bırakılarak tam gözlemlere uygulanan EKK yönteminin daha iyi sonuçlar verdiğini belirtmektedir.

1969(A) yılında Afifi ve Elashoff, x ve y değişkenlerinde kayıp değerler bulunduğunda y değişkeninin x üzerinden oluşturulan regresyon denkleminin parametreleri ile ilgili asimptotik olan dağılımlar bulmuşlardır. Bunların karşılaştırmasını hata kare ortalamalar ile yapmışlardır. 1969(B) yılında çalışmalarına devam eden Afifi ve Elashoff eksik veri olduğu durumlarda doğrusal regresyon ile tahmin edilen parametrelerin örneklem sayısı küçük olduğundaki verimliliklerini değerlendirmişlerdir. Örneklem sayısının çok ve az olduğu durumlarda verimliliğin korelasyon katsayısına ve eksik verilerin nasıl bir veri seti yapısına sahip olduğu ile ilgilidir.

1971 yılında Hartley ve Hocking ile 1972 yılında Orchard ve Woodbury'in araştırmaları ve yaptıkları çalışmalar Beklenti En Büyükleme yöntemi ile ilgilidir. İlgilendikleri bu çalışma gelişmiş Beklenti En Büyükleme yönteminin ilk adımlarını oluşturur.

1976 yılında Rubin, eksik verileri üç kategoriye ayırmıştır. Bu kategorilerden ilki tamamen rassal kayıp, ikincisi rassal kayıp ve üçüncüsü rassal olmayan kayıptır. Bu üç kategori kayıp verilerin hangi yöntem ile tamamlanabileceği konusunda fikir vermektedir. Rubin'in bu çalışması literatürde kayıp veri mekanizmaları olarak bilinmektedir.

1977 yılında Dempster, Laird ve Rubin eksik verilerin tamamlanabilmesi için EÇO'nun bir versiyonu olan Beklenti En Büyükleme algoritmasını geliştirmişlerdir.

1982 yılında Louis eksik verilerin tamamlanmasında EÇO tahminlerini bulmak üzere Beklenen En Büyükleme yönteminin kullanılması durumunda ilgilenilen veri setinin görülmesi ve Beklenti En Büyükleme yönteminin yakınsamasını geliştirmek ve hızlandırmak için bir yöntem geliştirerek çalışmalarını sürdürmüştür.

1987 yılında Little ve Rubin'in yayınladığı kitap eksik veri konusunda yapılan ilk sistematik çalışma olmuştur. Eksik verilerin tamamlanabilmesi için izlenecek adımları örneklendirmişlerdir. Bu çalışma üzerine kayıp verilerin giderilmesi için uygulanacak yöntemler ve adımlar daha da yaygınlaşmıştır.

1992 yılında Little kapsamlı bir çalışma yaparak regresyon analizinde değişkenlerin kayıp olduğu durumları incelemiştir.

1996 yılında Jones, yürüttüğü çalışmalarda doğrusal regresyon analizlerinde eksik veri bulunduran veri setleri olduğu durumlarda dummy değişkeninin kullanılmasını, 1994 yılında ise Vach, Jones'in dummy değişkeni üzerine yaptığı çalışmalar gibi benzer çalışmayı lojistik regresyon üzerinde uygulayarak araştırma ve çalışmalarını sürdürmüştür.

1999 yılında Eberle ve Toutenburg, eksik verileri tamamlama yöntemleri içeren birçok programda değerlendirmeler yapmıştır. Bu programlara Minitap, SAS, SPSS, Stata vb. programlar örnek gösterilebilir. Eberle ve Toutenburg'in değerlendirmeleri sonucunda hiçbir analiz programının eksik verileri gidermek için kullanılacak yöntemlerin tamamını barındırmadığı belirlenmiştir. Eksik verilerin tamamlanması için uygulanacak yöntemler geliştikçe ve arttıkça bu konu üzerinde uygulanacak yöntemlerin olumlu ve olumsuz yönleri ortaya çıkmıştır.

2000 yılında Cool, eksik verileri gidermek için kullanılan veri silme yönteminin örneklem sayısını düşürmesi sebebiyle yapılacak analizleri olumsuz etkileyeceğini, kayıp gözlemleri tamamlamak için veri atama yöntemlerini kullanmanın da testlerin gücünü arttırdığını gösteren araştırma ve çalışmalar yapmıştır.

2001 yılında Enders ve Bandalos, eksik verileri tahmin etme yöntemlerinden olan liste düzeyinde veri silme ve çift yönlü veri silme yöntemlerini incelemişlerdir. Eksik verileri gidermek için hangi yöntemlerin kullanılarak daha iyi sonuçlar verdiğini ve anlamlı sonuçlar çıkardığını belirlemek üzere çeşitli araştırmalar yapılmıştır.

2001 yılında Allison, geleneksel olan kayıp veri tahmin etme yöntemlerine göre maksimum benzerlik ve çoklu veri atama yöntemlerinin eksik verileri gidermek için daha iyi sonuçlar verdiğini söylemektedir.

2005 yılında Molenberg ve Verbeke, Rubin'in kayıp veriyi sınıflandırdığı üç mekanizmadan biri olan rassal olmayan kayıp veri mekanizmasını araştırarak çalışmalar yapmıştır. Eksik verilerin giderilmesi için yöntemler geliştikçe çeşitli analiz programlarında yer alarak daha iyi analizler yapmak ve sonuçlar almak için yaygın olarak kullanılmaya başlanmıştır.

2007 yılında Horton ve Kleinman, eksik verilerin giderilmesi ve sağlıklı sonuçların alınması için istatistiksel yazılım programlarının kullanılması gerektiğini vurgulamıştır.

2010 yılında Enders, Rubin'in kayıp verileri üç sınıfa ayırdığı tamamen rassal kayıp, rassal kayıp ve rassal olmayan kayıp olarak adlandırılan kayıp veri mekanizmaları ile kayıp veri yapılarının farklı anlamlar ifade ettiğini söylemiştir. Kayıp veri mekanizmaları verinin nasıl çözüleceğine ipuçları verirken, kayıp veri yapıları eksik veri içeren veri setlerindeki kayıpların nasıl gruplandığını ve yapısını inceler.

Kayıp değerlerin atamasını gerçekleştirmek için birçok yöntemlerden ve araştırmalardan yukarıda bahsedilmiştir. Kayıp verilerin giderilmesi için yapılacak çalışmaların daha iyi sonuçlar vermesi ve sağlıklı çıktıların yorumlanabilmesi oldukça önemlidir. Eksiklik içerin verilerin tamamlanabilmesi için günümüzde birçok yöntem denenmektedir. Kayıp verilerin giderilmesi için uygulanan yöntemlerden biri de bu araştırma da yer almaktadır. Veri setlerindeki eksikliklerin tamamlanması için bulanık c-ortalamalar kullanılarak, bulanık c-ortalamalar yönteminin nasıl sonuçlar verdiğini inceleyeceğiz.

1.2 Tezin Amacı

Eksik veriler çalışmalarda sıkça karşılaştığımız ve çalışmalarımızı etkileyen büyük bir sorundur. Kayıp verileri gidermenin pek çok yöntemi vardır. Kayıp veriler örneklem sayılarını etkilediği için eksik verileri doğru yöntemle tamamlamak çalışmaların sağlıklı ve güvenilir sonuçlar vermesine yardımcı olacaktır. Kayıp verileri tahmin etmek için kullanılan yöntemler oldukça fazla olup hangi yöntemin hangi veriye uygun olduğu kayıp veri mekanizmalarına göre gruplandırılabilir. Kayıp veri mekanizmaları olarak adlandırılan terimler, istatistik alanında kullanılan terimlerdir. Bu terimler, veri setlerindeki kayıp verilerin nedenlerini ve özelliklerini açıklar. Kayıp veri mekanizmaları üçe ayırılır. Bunlar tamamen rassal kayıp (MCAR-Missing Completely At Random), rassal kayıp (MAR-Missing At Random) ve rassal olmayan kayıptır (MNAR-Missing Not At Random). MCAR, eksik verilerin rastgele ve sistematik bir düzen izlemeden meydana geldiği durumlardır. MAR kayıp verilerin diğer verilere bağlı olarak meydana geldiği fakat nedenlerin gözlemciler tarafından belirlenemediği durumu ifade ederken MNAR, eksik verilerin belirli bir nedenle veya dış faktörlerle meydana geldiği ve nedenlerin gözlemciler tarafından bilindiği durumu ifade eder.

Araştırmanın diğer bir konusu ise bulanık mantıktır. Bulanık mantık kendi içerisinde kesinliğe yer vermez. Gözlemlerin aldığı değerler 0 ile 1 arasındaki derecelerdir. Bulanık mantıkta, bulanık kümeleme algoritmalarından bulanık c-ortalamalar (BCO) araştırmamızdaki anahtar olacaktır. Kayıp verileri bulanık c-ortalamalar ile tahmin etmek için yürüttüğümüz bu çalışmada BCO, klasik kayıp veri tahmin etme yöntemlerine alternatif olarak önerilebilecek bir kayıp veri tahmin etme yöntemi olacak mı sorusuna cevap aratacaktır. Çalışmamızdaki amacın iyi sonuçlar barındırdığını ve BCO'nun alternatif hatta rekabetçi sonuçlar sunduğunu bu çalışmada görmüş olacağız.

1.3 Tezin İçeriği

Çalışmamızın adımları şu şekilde planlanmıştır. Bölüm 2'de kayıp verinin tanımı ve kayıp veri çalışmalarının tarihçesine yer verilmiştir. Ardından kayıp veri mekanizmaları anlatılarak tamamen rassal kayıp, rassal kayıp ve rassal olmayan kayıptan söz edilmiştir. Bölüm 2'nin son kısmında kayıp veri analiz yöntemlerinden bahsedilerek, klasik tahmin etme yöntemleri olan ortalama atama, ortanca atama ve mod atama anlatılmıştır. Bölüm 3'de bulanık mantık, bulanık mantık ve küme teorisine yer verilmiştir. Bulanık kümeleme algoritmalarından, bulanık c-ortalamalardan ve BCO algoritmasının çalışma prensibinden de Bölüm 3'te bahsedilerek, BCO ile kayıp veri atamasından söz edilerek Bölüm 3'ten Bölüm 4'e geçilmiştir. Bölüm 4'te BCO ile kayıp değer ataması uygulaması yapılmış ve ilgili tablolara yer verilmiştir.

2
KAYIP VERİ

2.1 Kayıp Veri

Kayıp veriler neredeyse tüm çalışmaları etkileyen bir öneme sahiptir. Longford eksik veriyi en yalın tarifini, gözlenmesi beklenen ve gözlenen veri seti arasındaki farklar olduğunu söylemektedir (Longford, 2004). Çalışmayı yürüten kişi tüm değerlerin eksiksiz tam olarak yanıtlanmasını istese de kayıplar kaçınılmaz bir durumdur. Bu kaçınılmaz durumlar, soruların cevaplanamaması, soruların anlaşılamaması, verilerin gözlenememesi gibi durumlardan kaynaklanabilmektedir (Scheffer, 2002).

Kayıp verilerin olması örneklem sayısını azalttığından ve örneklem sayısının azalması yapılacak testleri ve sonuçları yüksek derecede etkilediğinden kayıp verileri giderme yolları bulunmuştur. Bu sorundan hareketle eksik veri giderme yöntemleri geliştirilmiştir. Eksik veri analizleri ile veri setindeki eksik veriler, mümkün değerler ile tamamlanarak, istatistiksel işlemler yapılır ve sonuçlar ortaya çıkar (Roderick J.A. Little, 1987). Eksik verilerin mümkün oldukça minimize etmek için en iyi yöntem, araştırmanın nasıl yapılacağını iyi tasarlamak ve verileri hassas şekilde toplamaktır (Wisniewski S.R., 2006).

Eksik veriler nedeniyle ortaya önemli sorunlar çıkmaktadır. Bu sorunları Peng ve arkadaşları dört maddede şu şekilde ifade etmektedir. Birincisi yanlılık olacağı, ikincisi istatistiksel güç kaybının olacağı, üçüncüsü bazı model varsayımlarının sağlanamayacağı ve son olarak kaynak kaybı olacağıdır (Peng, 2007) .

Eksik verilerin analizleri ve modellemesiyle ilgili birçok çalışma yapılmıştır. Bu çalışmalar, eksik verilerin etkisini azaltmak için farklı farklı yöntemler sunar. Bu yöntemler, istatistiksel sonuçların doğru olmasını sağlamak için oldukça önemlidir ve çalışmayı yürüten kişilerin veri toplama ve analiz süreçlerinde karşılarına çıkan eksik veri sorunlarına çözüm sağlamalarına yardımcı olacaktır.

2.2 Kayıp Veri Mekanizmaları

Verilerin kayıplarının nasıl ortaya çıktığını ifade eden terime kayıp veri mekanizmaları denir. Kayıp veri mekanizmalarının belirlenmesi kayıp veri sorununu giderecek doğru analizin seçilmesinde, doğru analizin seçilmesi de o verinin değerlendirilmesinde ve doğru sonuca varılmasında etkin rol oynar (Baygül, 2007). Kayıp veriler ilk defa 1976 yılında Rubin tarafından sınıflandırılmıştır. Sınıflandırma konusundaki diğer çalışmaların yine Little (1992) ve Rubin tarafından yapıldığı bilinmektedir. 1976 yılında Rubin tarafından sınıflandırılan kayıp veriler şu şekildedir:

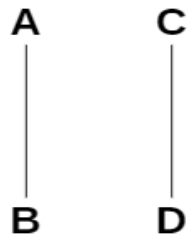
- Tamamen Rassal Kayıp (TRK)
- Rassal Kayıp (RK)
- Rassal Olmayan Kayıp (ROK)

olmak üzere üçe ayrılmaktadır.

2.2.1 Tamamen Rassal Kayıp (Missing Completely At Random-MCAR)

Tamamen rassal kayıpta, kayıp veriler tamamen rassal olarak ortaya çıkmaktadır. Böyle bir durumda, kayıp verilerin meydana gelmesi, diğer verilerden bağımsızdır. Eksik değerlerin meydana gelmesi diğer verilerdeki değerler ile ilgili değildir.

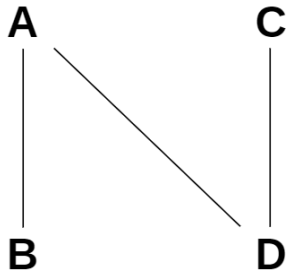
Kayıp veriler hakkında sınıflandırma çalışmaları yapan Little ve Rubin'in çalışmalarından yola çıkarak tamamen rassal kayıp mekanizması şu şekilde ifade edilebilir. A ve B birbirinden farklı iki değişkendir. Değişkenlerden her birine yanıt alınamama olasılığı birbirinden etkilenmeyen bağımsız değişkenlerdir (Oğuzlar, 2001). Diğer bir söylem ile A değişkenindeki eksik veri olasılığının B değişkeni ile bağlantılı olmaması beklenir. Bu durum tamamen rassal kayıp olarak adlandırılır (Roderick J.A. Little, 1987).



Şekil 1.1 Tamamen Rassal Kayıp Mekanizması

2.2.2 Rassal Kayıp (Missing At Random-MAR)

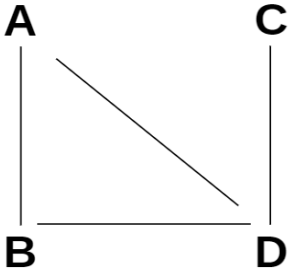
Rassal kayıpta, kayıp veriler diğer gözlemlenen veriler ile ilişkilidir fakat kayıp verilerin kendisi gözlenemez. Kayıp verilerin sınıflandırılması hakkında çalışma yapan Little ve Rubin'in çalışmalarından yola çıkarak rassal kayıp mekanizması şu şekilde ifade edilebilir. A ve B birbirinden farklı iki değişkendir. A değişkenine yanıt alınamama durumunun B değişkeni ile ilişkili olması fakat B'deki eksik veri durumunun A ile ilişkili olmaması durumundaki eksik veri rassal kayıp olarak adlandırılır (Roderick J.A. Little, 1987) (Oğuzlar, 2001).



Şekil 1.2 Rassal Kayıp Mekanizması

2.2.3 Rassal Olmayan Kayıp (Missing Not At Random-MNAR)

Rassal olmayan kayıpta kayıp olan verilerin oluşumu diğer verilerden bağımsız olmaz ve gözlenemeyen faktörlere bağlıdır. Eksik veriler, eksikliği olan değişkenlerin özelliklerine ve durumlarına bağlı olarak değişebilir. Kayıp verilerin sınıflandırılması hakkında çalışma yapan Little ve Rubin'in çalışmalarından yola çıkarak rassal olmayan kayıp mekanizması şu şekilde ifade edilebilir. A ve B birbirinden farklı iki değişkendir. A değişkenine yanıt alınamama durumu hem A hem de B değişkeni ile ilişkili olması durumundaki eksik veri rassal olmayan kayıp olarak adlandırılır (Roderick J.A. Little, 1987) (Oğuzlar, 2001).



Şekil 1.3 Rassal Olmayan Kayıp Mekanizması

2.3 Kayıp Veri Analiz Yöntemleri

Kayıp veri analiz yöntemlerinde temel amaç kayıp olan veriler yerine en uygun ve mümkün değerlerin atanmasıdır. Kayıp veri ile karşılaşan çalışmacıların çoğunluğu genellikle eksik verileri analiz dışı bırakarak bu durumdan kurtulmayı tercih ederler. Araştırmacıların bunu yapabilmesi için öncelikle kayıp veriye neyin neden olduğunu anlaması gerekir (Ergül Demir, 2012). Eksik olan veriler yerine mümkün değerlerin atanabilmesi için kayıp verilerin hangi mekanizmaya ait olduğunun bilinmesi gerekir. Eksik verilerin tamamen rassal kayıp, rassal kayıp ve rassal olmayan kayıp veri mekanizmalarından hangisine ait olduğunun bilinmesi, eksik verilere mümkün değerlerin atanabilmesi için uygun yöntemin seçilmesinde etkin rol oynar (Allison, 2003).

Eksik verilerin giderilmesi için birçok yöntem bulunmaktadır. Bu yöntemler, veri setinin özelliklerine, eksik verilerin ait olduğu mekanizmalara ve diğer faktörlere bağlı olarak değişkenlik gösterebilir. Kayıp verilerin tahmin edilebilmesi için kullanılan basit yöntemlere; liste bazında veri silme, çiftler bazında veri silme, ağırlıklandırma yöntemleri, ortalama atama, ortanca atama, mod atama ve regresyon atama yöntemleri örnek gösterilebilir. Kayıp verilere mümkün değerler atamak için kullanılan bir başka yöntem ise en çok olabilirlik yöntemidir. En çok olabilirlik yöntemine; beklenti en büyükleme, çoklu veri yükleme ve bayes ile veri atama örnek verilebilir. Eksik verileri tahmin etmek ve bu sorundan kurtulmak için kullanılan diğer bir yöntem ise makine öğrenmesidir. Makine öğrenmesi yöntemlerine; en yakın k-komşu, ağırlıklandırılmış en yakın k-komşu, k-ortalama kümeleme ile veri atama, bulanık k-ortalama kümeleme ile veri atama, destek vektör makine ile kayıp veri giderme ve kısmi EKK örnek gösterilebilir.

Veri setlerini değerlendirirken, veri setlerinin eksiksiz olması istenilen, veri setlerinin kayıplardan oluşması ise istenilmeyen ve sıklıkla karşılaşılan durumdur. Sıklıkla karşılaşılan bu durumu aydınlatmak için titiz bir çalışma gerekmektedir. Veri setlerinden doğru sonuçlar alabilmek için bu titizlik gerektiren duruma dikkat etmek gerekir. Bu sebeple doğru analiz, doğru sonuç ve doğru bilgilendirme için yöntemlerin yapısına, veri setlerinin uygunluğuna ve analizlerde tercih edilecek yöntemlerin belirlenmesine dikkat edilmesi gerekir.

2.3.1 Ortalama Atama

Kayıp verilerin giderilmesi için sıklıkla başvurulan yöntemlerden birisi ortalama atama yöntemidir. Eksik verileri ortalama atama yöntemi ile tamamlama fikri Wilks'in (1932) çalışmalarına dayanmaktadır. Gözlemlenen değişkene ait kayıp değerlerin, o değişkene ait tam değerlerin ortalaması ile yer değiştirmesidir. Bu atama yöntemi sonucunda, tam ve atanmış veriler için çıkan ortalama eksiksiz verilerle elde edilen ortalamaya denk çıkacaktır. Bu yöntemin olumlu ve olumsuz yönleri vardır. Kolayca uygulanabiliyor olması ve eksiksiz verileri dikkate alarak işlem yapması olumlu yönleridir. Tahmin edilmesi istenen eksik değişkenlerin tek bir değer ile doldurulması, değişkenlerin asıl dağılımından sapmasına sebep olmakta ve verideki gerçek varyans değeri olması gereken değerden küçük çıkmaktadır. Bu iki sorun ortalama ataması yönteminin olumsuz yönleri olmaktadır (Alpar, Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş 1, 2003).

2.3.2 Ortanca Atama

Ortanca atama yöntemi değişkenlerin sahip olduğu değerlerin küçükten büyüğe doğru sıralanarak ortalarında bulunan değerin ele alınmasıdır. Sütunca ya da satırca bulunan verilerin hücreler içerisinde barındırdığı değerler işleme alınır. Kayıp hücreler dışarıda bırakılarak mevcutta eksik veri içermeyen hücrelerin değerleri küçükten büyüğe sıralanır ve ortanca değer kayıp olan veriler yerine atanır. Böylelikle kayıp veriler ortanca atama yöntemi ile tamamlanmış olur.

2.3.3 Mod Atama

Mod atama yöntemi değişkenlerin sahip olduğu değerlerden en çok tekrar eden değerin seçilmesidir. Diğer bir adıyla tepe değer olan bu atama yöntemi, sütunca ya da satırca bulunan verilerin hücreler içerisinde barındırdığı değerler işleme alınarak yapılır. Veri setindeki kayıp hücreler dışarıda bırakılarak eksik veri olmayan veri seti kendi içerisinde incelenir ve en çok tekrar eden değer kayıp olan hücreler yerine atanır. Bu yöntemin her yöntem gibi avantaj ve dezavantajları mevcuttur. Hesaplamasının ve anlaşılmasının kolay olması, dağılımdaki aşırı ve aykırı değerlerden etkilenmemesi avantajları olurken incelenen veri setlerinde tepe değer bulunmaması yahut bazı veri setlerinde birden fazla tepe değer olması ve tepe değerin aritmetik işlemler için elverişsiz olması dezavantajları arasındadır.

3
BULANIK YAKLAŞIM

3.1 Bulanık Mantık

Bulanık mantık ilk kez 1965 yılında Zadeh tarafından ortaya atılmıştır. Bulanık mantığın temel amacı belirsiz ve karmaşık durumda olan ifadelerin ele alınması için tasarlanmış bir matematiksel ve bilimsel bir yaklaşım olmasıdır. Klasik mantık ve bulanık mantık arasındaki esas farklılık şu şekilde ifade edilebilir, bulanık mantığın değerleri [0,1] aralığında olurken klasik mantığın değerleri doğru veya yanlış olmak zorundadır (Türkşen, 2015).

Bulanık mantık, net olmayan, bir belirginliğe sahip olmayan yani belirsizlik barındıran ve kesin sınırları bulunmayan durumların ifade edilmesi için kullanılır. Böyle durumlarda herhangi bir durumun kesinlikle doğru olduğunu ya da kesinlikle yanlış olmadığını fakat bir değere ve dereceye sahip olduğunu belirtir. Bulanık mantığın ifade edilişi bu iki kesinlik arasında değer alabilecek geniş bir havuza sahiptir. Bulanık mantık, belirgin olmayan belirsizlikle başa çıkan ve karmaşık halde bulunan sistemlerin modellenmesi için kullanılan güçlü bir yaklaşımdır.

Bulanık mantığın genel özellikleri Zadeh tarafından şu şekilde ifade edilmiştir;

- Bulanık mantıkta, net değerlere dayanan ifadeler yerine yaklaşık ifadeler kullanılır.
- Bulanık mantıkta ifadeler [0,1] aralığında değerler ile gösterilir.
- Bulanık mantıkta ifadeler kesin, tamamen gibi ifadeler yerine büyük-küçük, çok-az gibi dilsel ifadeler ile belirtilir.
- Her mantıksal sistem bulanık olarak ifade edilebilir.
- Bulanık mantık, belirsizlikle başa çıkan ve karmaşık halde bulunan sistemlerin modellenmesi için oldukça uygun bir yöntemdir.
- Bulanık mantık, net olarak bilinmeyen veya eksiklik içeren bilgilere göre işlem yapabilme yeteneğine sahiptir. (Alavala, 2008)

3.1.1 Bulanık Mantık Avantaj ve Dezavantajları

Bulanık mantıkla ilgili çeşitli araştırma ve geliştirmeler yapılmaktadır. Bu araştırmalar ve gelişmeler için Japonya oldukça fazla harcama yapmaktadır. Bulanık mantığın avantaj ve dezavantajları aşağıdaki gibidir (F. Martin McNeill, 1994).

Tablo 1.1 Bulanık Mantık Avantaj ve Dezavantajları

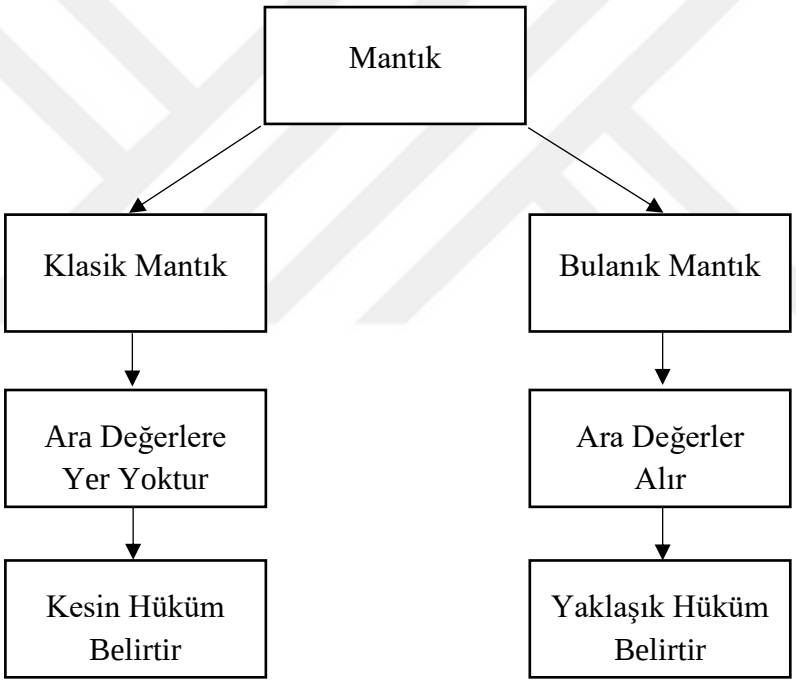
<table>
<tr><th>Avantaj</th><th>Dezavantaj</th></tr>
<tr><td>Az miktarda değerler, kurallar ve kararlar gereklidir.</td><td>Bir bulanık sistemden model geliştirmek zordur.</td></tr>
<tr><td>Gözlemlenen çok miktarda değişken değerlendirilebilir.</td><td rowspan="4">Geleneksel sistemlerden daha kolay ve hızlı prototip oluşturulur ancak bulanık sistemler kullanılmadan önce ince ayarlar ve benzetimler yapılması gerekir.</td></tr>
<tr><td>Dilsel, nümerik olmayan değişkenler kullanılır, bu insan düşünce yapısına benzer.</td></tr>
<tr><td>Çıktı ile girdi arasında, bütün değişkenlerin anlaşılmasına gerek kalmadan bağlantı kurar, geleneksel kontrol sistemlerine göre daha hassas ve stabil sistemler tasarımlanmasına olanak verir.</td></tr>
<tr><td>Basitliği daha önceden çözülememiş problemlerin çözümüne olanak sağlar.</td></tr>
<tr><td>Sistem tasarımcısının, çalışmaya başlamadan önce sistem hakkında her şeyi öğrenmesine gerek olmadığı için prototip oluşturmayı çabuklaştırır.</td><td rowspan="5">En büyük zorluğu, Amerika’da kültürel temayüllerin kontrol sistemleri için matematiksel hassasiyet veya kesin sistemler ve lineer sistemler olmasıdır.</td></tr>
<tr><td>Tasarımlanması kolay olduğu için geleneksel kontrol sistemlerine göre daha ucuzdur.</td></tr>
<tr><td>Daha güçlüdür.</td></tr>
<tr><td>Bilgi kazanımını ve gösterimini kolaylaştırır.</td></tr>
<tr><td>Karmaşık durumlar için az sayıda kurallar vardır.</td></tr>
</table>

(Turan Paksoy, 2013)

3.1.2 Bulanık Mantık ve Küme Teorisi

Zadeh'e göre bulanıklık derecelendirme ile bağlantılıdır. Bulanıklığın derecelendirme ile ilgili olduğunu savunan Zadeh bu durumu bulanık kümelerle açıklığa kavuşturmaya çalışır. Klasik küme teorisinde bir nesne bir kümenin elemanı olabilir ya da olmayabilir. Klasik küme teorisi bu iki seçenek üzerine kurulur. Bulanık küme teorisi ise iki değerli üyeliği çok değerliliğe taşır. Bulanık küme (0,1) arasında geniş bir değer havuzuna sahiptir. Klasik kümede "0" değeri ait olmamayı "1" değeri ise ait olmayı gösterir. (0,1) arasındaki değerler ise bulanık kümenin değerlerini göstermektedir (Karataş, 29 Ekim 2018).

Mantık biliminin genel çerçevesi aşağıdaki gibi sunulabilir (Prof. Dr. Necdet Tekin, 1995 Ocak).



Şekil 1.4 Bulanık Mantık ve Küme Teorisi

3.2 Bulanık C-Ortalamalar

Kümeleme işlemi, değerleri veya gözlemleri benzerliklerine göre alt kümelere ayırmak ve kendi aralarında gruplamak için kullanılan yaygın bir yöntemdir. Kümeleme yöntemi uygun olduğunda atama yöntemi, bir benzerlik ölçüsü ele alarak faydalı değişken alt gruplarını bulma yöntemi bütün değişkenlerin gerçek aralığı ile değiştirilerek yapılır. Kümeleme işlemi yapılırken kullanılan en yaygın yöntemlerden birisi bulanık c-ortalamalar (BCO)'dur. Bulanık c-ortalamalar yöntemi, bir değişkenin üyelik değerlerine göre iki veya ikiden fazla gruba ait olmasına izin veren bir kümeleme yapma yöntemidir. BCO yöntemi ilk olarak 1973 yılında Dunn tarafından öne sürülmüş ve 1981 yılında Bezdek tarafından geliştirilmiştir. Bulanık c-ortalamalar algoritması genellikle değerleri 0 veya 1 üyelik değerine sahip bir grubun özelliğine böler. Bulanık kümeleme algortiması kısmi gözlemleri birkaç kümeye yerleştirir. Bulanık kümedeki üyelik düzeyi değişkenin kümenin merkezine olan yakınlığına bağlıdır. Gözlem noktası küme merkezine yaklaştıkça üyelik değeri artar, bunun tersi de doğrudur. Üyelik değeri büyükse gözlem noktası küme merkezine yaklaşır. Bulanık c-ortalamalar yönteminin amacı, bir kümedeki gözlemler arasındaki benzerliği artırmak ve diğer kümelerle karşılaştırıldığında gözlemler arasındaki benzerliği azaltmak için amaç fonksiyonunu en aza indirme amacındadır (Kiki Aristiawati, 2019).

$$J = \sum_{i=1}^{n}\sum_{j=1}^{c} \mu_{ij}^{m} \left\| x_i - r_j \right\|^2 \quad , \quad \sum_{i=1}^{n} \mu_{ij} = 1 \qquad (1.1)$$

J = Amaç fonksiyonu

n = Gözlem sayısı

c = Küme sayısı

x_i = i. gözlemler

r_i = j. merkez

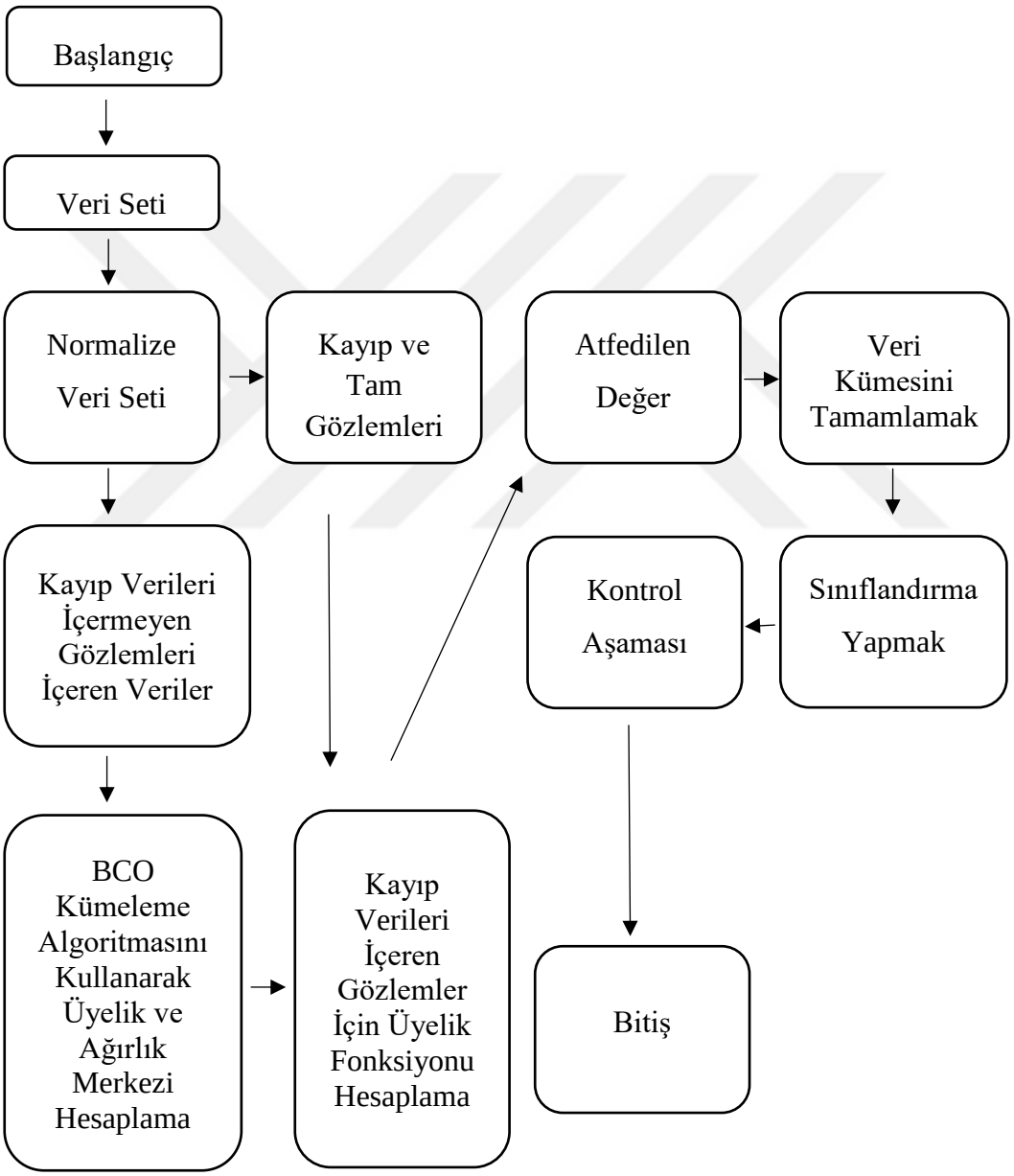
μ_{ij} = j. ağırlık merkezinin i. gözlemi için üyelik fonksiyonu

m = Bulanık derece, yaygın olarak kullanılan m=2 değeri

$\left\| x_i - r_j \right\|$ = j. ağırlık merkezinin i. gözlemi arasındaki Öklid mesafesi

3.2.1 Bulanık C-Ortalamalar Algoritmasının Çalışma Prensibi

BCO ataması, bulanık c-ortalamalar kullanan kümeleme algoritmalarına dayanan bir atama yöntemidir. Bu algoritmadaki asıl mesele her değişkenin ağırlık merkezi değerini bulmak ve her bir küme ile eksik değerler arasındaki mesafeyi hesaplamak için ilgili hücreleri bulanık c-ortalamalar ile aynı değişkende gruplamaktır. Kayıp değerlerin atamasını gerçekleştirmek için izlenen adımların aşamaları aşağıda belirtilmektedir (Kiki Aristiawati, 2019).



Şekil 1.5 Bulanık C-Ortalamalar Algoritmasının Çalışma Prensibi

3.2.2 Bulanık C-Ortalamalar ile Kayıp Veri Ataması

Bulanık c-ortalamalar belli bir algoritmaya sahiptir. Veri setlerindeki eksikliklerin giderilmesi ve kayıpsız veri setlerine ulaşabilmek için bir çok yöntem izlenmektedir. Alışagelmiş kayıp veri tahmin etme yöntemlerinin yanı sıra bulanık c-ortalamalar algoritması ile eksik verilerin giderilmesi şu adımlar izlenecektir.

1. Veri setindeki kayıp değerleri (x1, x2, ……,xn) girilir.
2. Veri seti aşağıdaki formül kullanılarak normalleştirilir.

$$x_{normalize} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{1.2}$$

formülde kullanılan

x_i = i. gözlem verileri

x_{min} = minimum değeri

x_{max} = maximum değeri

$x_{normalize}$ = normalleştirilmiş değeri ifade eder.

Normalleştirilmiş bir veri asıl değerine döndürülmek istendiğinde;

$$x_i = x_{min} + x_{normalize}(x_{max} - x_{min}) \tag{1.3}$$

3. Veri seti iki alt veriye dönüştürülür. Bir veri, eksik değerleri içerirken diğer veri, eksik değerleri içermeyen veri setidir.
4. Eksik değerleri içermeyen veri seti üzerinden üyelik değerleri ve küme merkezleri bulanıkc-ortalamalar algoritması kullanılarak bulunur.
5. Eksik değerleri içeren veri seti için, eksik değerleri içermeyen veri setinden hesaplanan üyelik değerleri ve küme merkezleri kullanılarak eksik değerleri içeren her hücre için üyelik değeri hesaplanır. Hesaplama için kullanılan değerler her kümenin ağırlık merkezi değeridir. Üyelik değeri için;

$$\mu_{ijM} = \frac{1}{\sum_{k=1}^{c} \frac{\left\|x_{iL} - r_{jL}\right\|^{\frac{2}{m-1}}}{\left\|x_{iL} - r_{kL}\right\|}} \tag{1.4}$$

hesaplaması yapılır.

Formülde kullanılan;

μ_{ijM} = Eksik değerleri içeren j.kümeye i.gözlem için üyelik fonksiyonu

x_{iL} = Eksik değerleri içeren alt verilerdeki i. gözlem için mevcut değeri

r_{jL}= Eksik değerleri içermeyen değişken için j. ağırlık merkezi

r_{kL}= Eksik değerleri içermeyen değişken için k. ağırlık merkezi değerini ifade eder.

6. Bulunan üyelik fonksiyonu ve 4. Adımda bulunan ağırlık merkezleri kullanılarak atama değeri hesaplanır. Atama değeri;

$$x_{ilM} = \sum_{j=1}^{c} \mu_{ijM} r_{jM} \tag{1.5}$$

formülü kullanılarak hesaplanır.

Formülde kullanılan;

x_{ilM} = Eksik değerleri bulunduran alt verilerdeki i. gözlem için mevcut değeri

μ_{ijM} = Eksik değerleri içeren j. kümeye i. gözlem için üyelik değerini

r_{jM} = Eksik değerleri içermeyen değişken için j. ağırlık merkezini ifade eder (Kiki Aristiawati, 2019).

4
BULANIK C-ORTALAMA İLE KAYIP VERİ ATAMASI

4.1 Bulanık C-Ortalama ile Kayıp Veri Tahmini

BCO ile kayıp veri tahmini yapmak üzere "Depresyon Anksiyete ve Stres Ölçekleri-21" (DASS-21) veri seti kullanılmıştır. DASS-21 veri setinden %1, %5, %10 ve %20 oranlarında rastgele veri kaybedilmiştir. Her veri seti için 100 tekrar yapılarak yöntemlerin performansları karşılaştırılmıştır.

Bu çalışmada belirli küme sayılarına ayrılarak RMSE, MAE ve Cronbach Alfa değerleri hesaplanmıştır. Veri setlerindeki kayıpların tahmin edilmesinde örneklem büyüklüğünün bir etkisinin olup olmadığını anlamak amacıyla tam veri seti yarıya indirilerek yapılan işlemler yeniden uygulanarak iki veri setinin sonuçları karşılaştırılmıştır. Bulanık C-Ortalamalar 1'den 11'e kadar küme sayıları artırılarak sonuçlar incelenmiştir. Çıkan değerlerin, sıklıkla kullanılan ortalama atama, medyan atama ve mod atama yöntemleri ile kıyaslanması yapılmıştır. Araştırmanın uygulaması R Studio programında gerçekleştirilmiştir.

4.2 Uygulamada Kullanılan Veri Seti

Bu çalışmada "Depresyon Anksiyete ve Stres Ölçekleri-21" (DASS-21) veri seti kullanılmıştır. Lovibond P.F ve Lovibond S.H. (1995) tarafından geliştirilen DASS ölçeği 42 maddeden oluşmaktadır. DASS-42 ölçeğinin kısa formu DASS-21 ölçeğidir. DASS-21 ölçeği Henry ve Crawford, (2005) ve Mahmoud vd. (2010) çalışmalarından alınarak hazırlanmıştır. DASS ölçeği kısa formunun Türkçeye uyarlaması Yılmaz ve arkadaşları (2017) tarafından yapılmıştır. Uygulamamızda kullandığımız veri seti Gomez, Stavropoulos, Griffiths'in (2020) makalesinde paylaşılan gerçek bir veri setidir. DASS-21 veri setinden %1, %5, %10 ve %20 oranlarında rastgele veri kaybedilmiştir. Bu çalışmada belirli küme sayılarına ayrılarak RMSE, MEA ve Cronbach alfa değerleri hesaplanmıştır. Çıkan değerlerin, sıklıkla kullanılan ortalama atama, medyan atama ve mod atama yöntemleri ile kıyaslanması yapılmıştır.

4.3 Uygulamada Kullanılan Performans Kriterleri

BCO ile kayıp değer ataması yapılırken bulunan değerlerin iyi sonuçlar verip vermediğini anlamak için kök ortalama kare hatalara (RMSE), ortalama mutlak hatalara (MAE) ve Cronbach Alfa katsayılarına yer verilmiştir.

4.3.1 Kök Ortalama Kare Hata (RMSE)

Kök Ortalama Kare Hata (RMSE), tahmin edici tarafından tahmin edilen değer ile gerçek değer arasındaki farkları ölçmek için yaygınca kullanılan hata ölçüm tekniklerinden biridir. Kök Ortalama Kare Hata herhangi bir değişken için farklı tahmin edicilerden tahmin edilen hataların farklarını bulmak için kullanılan doğruluk ölçülerinden biridir. (Fırat Artuğer, 2021)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{1.6}$$

4.3.2 Ortalama Mutlak Hata (MAE)

Ortalama Mutlak Hata, tahmin edilen değer ile gerçek değer arasındaki mutlak farktır. Ortalama Mutlak Hata değerinin düşük olması iyi senaryodur. Düşük olan değerin performansı daha iyidir şeklinde söylenmektedir. (Kırğıl, 2022)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{1.7}$$

4.3.3 Cronbach Alfa Katsayısı

Cronbach alfa katsayısı ölçekte bulunan maddelerin iç tutarlılığının ölçüsüdür. Cronbach alfa ölçekteki maddelerin homojen yapısını açıklamak üzere kullanılan bir katsayıdır. Cronbach alfa katsayısının yüksek çıkması istenilen senaryodur. Katsayının yüksek olması ölçekteki maddelerin birbirleriyle tutarlı ve aynı özelliği ölçen maddelerden meydana geldiğini gösterir (Alpar, 2013). Cronbach alfa formülü;

$$Cronbach\ \alpha = \frac{n}{n-1}\left(1 - \frac{\sum S_i^2}{S_x^2}\right) \tag{1.8}$$

şeklindedir. Formülde kullanılan;

n = Testteki madde sayısı

$\sum S_i^2$ = Madde varyanslarının toplamı

S_x^2 = Test varyansını ifade eder.

Cronbach alfa aşağıdaki gibi ifade edilir:

- $0 < \alpha < 0.40$ ise güvenilir değil
- $0.40 < \alpha < 0.60$ ise düşük güvenilirlikte
- $0.60 < \alpha < 0.80$ ise oldukça güvenilir
- $0.80 < \alpha < 1.00$ ise yüksek güvenilirliktedir.

4.4 Uygulamada Kullanılan Küme Sayıları ve Klasik Yöntemler

BCO ile kayıp değer ataması gerçekleştirmek üzere BCO için 1'den 11'e kadar küme sayısı belirlenmiş ve RMSE, MAE ve Cronbach Alfa katsayıları incelenmiştir. BCO değerlerinin karşılaştırılması için sıklıkla kullanılan klasik kayıp veri tahmin etme yöntemlerinden ortalama atama, ortanca atama ve mod atama değerleri içinde RMSE, MAE ve Cronbach Alfa değerleri bulunarak BCO ve klasik yöntemler karşılaştırılmıştır.

Yöntemler karşılaştırılırken her veri setine 100 tekrar yapılmıştır. 100 tekrar sonucu çıkan değerlerin ortalaması alınarak tablolara işlenmiş ve karşılaştırmalar gerçekleştirilmiş ardından sonuçlar yorumlanmıştır.

Tablo 1.2 Kullanılan Küme Sayıları ve Klasik Yöntemler

No	**Kümeler ve Klasik Yöntemler**	**No**	**Kümeler ve Klasik Yöntemler**
1	Fuzzy 1	8	Fuzzy 8
2	Fuzzy 2	9	Fuzzy 9
3	Fuzzy 3	10	Fuzzy 10
4	Fuzzy 4	11	Fuzzy 11
5	Fuzzy 5	12	Ortalama Atama
6	Fuzzy 6	13	Ortanca Atama
7	Fuzzy 7	14	Mod Atama

4.5 Verilere Ait Tablolar

4.5.1 %1 Kayıp ile RMSE-MAE(738)

Aşağıdaki tabloda bulanık c-ortalamalar 1'den 11'e kadar artan küme sayısı ve klasik kayıp veri bulma yöntemlerinden ortalama atama, ortanca atama ve mod atama yöntemleri 21 sütun 738 satırdan oluşan veri için incelenecektir.

%1 kayıp ile oluşturulan veri için ortalama ve standart sapma RMSE ve MAE sonuçları tablodaki gibidir. RMSE ve MAE değerleri incelendiğinde Fuzzy 3 ve devamında düşük değerlere ulaştığımız görülmektedir.

RMSE değerinde en iyi performansı BCO'da Fuzzy 5, klasik yöntemlerde ortalama ataması yöntemi vermektedir. MAE değerinde en iyi performansı BCO'da Fuzzy 10, klasik yöntemlerde ortalama ataması yöntemi vermektedir.

Tablo 1.3 %1 Kayıp ile RMSE-MAE (738)

Kümeler ve Atama Yöntemleri	**RMSE**		**MAE**	
	Ortalama	**Standart Sapma**	**Ortalama**	**Standart Sapma**
Fuzzy 1	1.393348	0.200579	1.21375	0.233208
Fuzzy 2	1.377286	0.202465	1.1825	0.240619
Fuzzy 3	0.915002	0.195326	0.74	0.218089
Fuzzy 4	0.935260	0.185423	0.74875	0.209085
Fuzzy 5	**0.882653**	0.176734	0.71	0.199336
Fuzzy 6	0.911315	0.183275	0.71375	0.201239
Fuzzy 7	0.899471	0.176815	0.715	0.202634
Fuzzy 8	0.897944	0.191211	0.715	0.219819
Fuzzy 9	0.923824	0.178512	0.75	0.204895
Fuzzy 10	0.882854	0.192845	**0.69875**	0.221973
Fuzzy 11	0.890653	0.182537	0.70875	0.203392
Ortalama Atama	0.939926	0.188797	0.73625	0.205584
Ortanca Atama	0.951502	0.226731	0.73625	0.238296
Mod Atama	1.166776	0.319616	0.8425	0.287370

4.5.2 %1 Kayıp ile RMSE-MAE(369)

Aşağıdaki tabloda bulanık c-ortalamalar 1'den 11'e kadar artan küme sayısı ve klasik kayıp veri bulma yöntemlerinden ortalama atama, ortanca atama ve mod atama yöntemleri 21 sütun 369 satırdan oluşan veri için incelenecektir.

%1 kayıp ile oluşturulan veri için ortalama ve standart sapma RMSE ve MAE sonuçları tablodaki gibidir. RMSE ve MAE değerleri incelendiğinde Fuzzy 3 ve devamında düşük değerlere ulaştığımız görülmektedir.

RMSE değerinde en iyi performansı BCO'da Fuzzy 8, klasik yöntemlerde ortalama ataması yöntemi vermektedir. MAE değerinde en iyi performansı BCO'da Fuzzy 8, klasik yöntemlerde ortanca ataması yöntemi vermektedir.

Tablo 1.4 %1 Kayıp ile RMSE-MAE (369)

Kümeler ve Atama Yöntemleri	**RMSE**		**MAE**	
	Ortalama	**Standart Sapma**	**Ortalama**	**Standart Sapma**
Fuzzy 1	1.361183	0.317309	1.1925	0.369026
Fuzzy 2	1.204124	0.358990	1.0025	0.385165
Fuzzy 3	0.883051	0.251702	0.7175	0.276739
Fuzzy 4	0.921685	0.257209	0.755	0.286435
Fuzzy 5	0.938788	0.248583	0.7625	0.278376
Fuzzy 6	0.936756	0.279767	0.76	0.303431
Fuzzy 7	0.892113	0.249511	0.7575	0.287480
Fuzzy 8	**0.851105**	0.237026	**0.69**	0.266192
Fuzzy 9	0.871515	0.262005	0.7025	0.274540
Fuzzy 10	0.893395	0.259847	0.74	0.299241
Fuzzy 11	0.912401	0.265953	0.7575	0.302462
Ortalama Atama	0.938306	0.296580	0.7675	0.297835
Ortanca Atama	0.943139	0.310567	0.75	0.315828
Mod Atama	1.090454	0.485557	0.8275	0.430373

4.5.3 %5 Kayıp ile RMSE-MAE(738)

Aşağıdaki tabloda bulanık c-ortalamalar 1'den 11'e kadar artan küme sayısı ve klasik kayıp veri bulma yöntemlerinden ortalama atama, ortanca atama ve mod atama yöntemleri 21 sütun 738 satırdan oluşan veri için incelenecektir.

%5 kayıp ile oluşturulan veri için ortalama ve standart sapma RMSE ve MAE sonuçları tablodaki gibidir. RMSE ve MAE değerleri incelendiğinde Fuzzy 3 ve devamında düşük değerlere ulaştığımız görülmektedir.

RMSE değerinde en iyi performansı BCO'da Fuzzy 5, klasik yöntemlerde ortalama ataması yöntemi vermektedir. MAE değerinde en iyi performansı BCO'da Fuzzy 5, klasik yöntemlerde ortanca ataması yöntemi vermektedir.

Tablo 1.5 %5 Kayıp ile RMSE-MAE (738)

Kümeler ve Atama Yöntemleri	**RMSE**		**MAE**	
	Ortalama	**Standart Sapma**	**Ortalama**	**Standart Sapma**
Fuzzy 1	1.400494	0.090317	1.200270	0.099217
Fuzzy 2	1.371014	0.112283	1.154324	0.135123
Fuzzy 3	0.923754	0.086974	0.731621	0.105640
Fuzzy 4	0.917211	0.081406	0.719189	0.089753
Fuzzy 5	**0.909440**	0.077120	**0.710810**	0.086539
Fuzzy 6	0.919001	0.093634	0.717027	0.104812
Fuzzy 7	0.922047	0.078135	0.722162	0.081587
Fuzzy 8	0.932979	0.083162	0.734594	0.093338
Fuzzy 9	0.924433	0.093560	0.729189	0.102068
Fuzzy 10	0.930232	0.083058	0.733243	0.094068
Fuzzy 11	0.915653	0.083986	0.711891	0.096742
Ortalama Atama	0.969968	0.087265	0.758648	0.098258
Ortanca Atama	0.983855	0.094204	0.744324	0.099018
Mod Atama	1.203222	0.144368	0.837567	0.138585

4.5.4 %5 Kayıp ile RMSE-MAE(369)

Aşağıdaki tabloda bulanık c-ortalamalar 1'den 11'e kadar artan küme sayısı ve klasik kayıp veri bulma yöntemlerinden ortalama atama, ortanca atama ve mod atama yöntemleri 21 sütun 369 satırdan oluşan veri için incelenecektir.

%5 kayıp ile oluşturulan veri için ortalama ve standart sapma RMSE ve MAE sonuçları tablodaki gibidir. RMSE ve MAE değerleri incelendiğinde Fuzzy 3 ve devamında düşük değerlere ulaştığımız görülmektedir.

RMSE değerinde en iyi performansı BCO'da Fuzzy 6, klasik yöntemlerde ortalama ataması yöntemi vermektedir. MAE değerinde en iyi performansı BCO'da Fuzzy 6, klasik yöntemlerde ortanca ataması yöntemi vermektedir.

Tablo 1.6 %5 Kayıp ile RMSE-MAE (369)

Kümeler ve Atama Yöntemleri	**RMSE**		**MAE**	
	Ortalama	**Standart Sapma**	**Ortalama**	**Standart Sapma**
Fuzzy 1	1.385499	0.131258	1.186666	0.155843
Fuzzy 2	1.159746	0.178816	0.931111	0.187547
Fuzzy 3	0.966101	0.133936	0.761111	0.143986
Fuzzy 4	0.954685	0.121800	0.760555	0.137608
Fuzzy 5	0.922488	0.114170	0.717222	0.132624
Fuzzy 6	**0.902728**	0.125675	**0.708333**	0.138664
Fuzzy 7	0.931969	0.119804	0.736111	0.131509
Fuzzy 8	0.926405	0.118900	0.735555	0.130258
Fuzzy 9	0.911317	0.138331	0.721666	0.152400
Fuzzy 10	0.936058	0.129381	0.745	0.138150
Fuzzy 11	0.914155	0.122582	0.711666	0.138060
Ortalama Atama	0.967948	0.133029	0.758888	0.143056
Ortanca Atama	0.982170	0.139371	0.755	0.142154
Mod Atama	1.190544	0.212785	0.836666	0.196746

4.5.5 %10 Kayıp ile RMSE-MAE(738)

Aşağıdaki tabloda bulanık c-ortalamalar 1'den 11'e kadar artan küme sayısı ve klasik kayıp veri bulma yöntemlerinden ortalama atama, ortanca atama ve mod atama yöntemleri 21 sütun 738 satırdan oluşan veri için incelenecektir.

%10 kayıp ile oluşturulan veri için ortalama ve standart sapma RMSE ve MAE sonuçları tablodaki gibidir. RMSE ve MAE değerleri incelendiğinde Fuzzy 3 ve devamında düşük değerlere ulaştığımız görülmektedir.

RMSE değerinde en iyi performansı BCO'da Fuzzy 5, klasik yöntemlerde ortalama ataması yöntemi vermektedir. MAE değerinde en iyi performansı BCO'da Fuzzy 5, klasik yöntemlerde ortanca ataması yöntemi vermektedir.

Tablo 1.7 %10 Kayıp ile RMSE-MAE (738)

Kümeler ve Atama Yöntemleri	**RMSE**		**MAE**	
	Ortalama	**Standart Sapma**	**Ortalama**	**Standart Sapma**
Fuzzy 1	1.400494	0.070826	1.193108	0.084845
Fuzzy 2	1.369001	0.095756	1.153783	0.115192
Fuzzy 3	0.935332	0.077621	0.736216	0.079018
Fuzzy 4	0.934323	0.060811	0.730945	0.064138
Fuzzy 5	**0.916915**	0.065640	**0.717972**	0.076483
Fuzzy 6	0.929407	0.055892	0.727702	0.061534
Fuzzy 7	0.943989	0.054479	0.744054	0.062558
Fuzzy 8	0.931613	0.060160	0.724189	0.070102
Fuzzy 9	0.925573	0.060168	0.721891	0.063708
Fuzzy 10	0.928127	0.064504	0.724459	0.067877
Fuzzy 11	0.924501	0.055041	0.724189	0.062033
Ortalama Atama	0.978296	0.060401	0.770405	0.064294
Ortanca Atama	0.990609	0.071379	0.752567	0.068143
Mod Atama	1.225227	0.115804	0.849324	0.110193

4.5.6 %10 Kayıp ile RMSE-MAE(369)

Aşağıdaki tabloda bulanık c-ortalamalar 1'den 11'e kadar artan küme sayısı ve klasik kayıp veri bulma yöntemlerinden ortalama atama, ortanca atama ve mod atama yöntemleri 21 sütun 369 satırdan oluşan veri için incelenecektir.

%10 kayıp ile oluşturulan veri için ortalama ve standart sapma RMSE ve MAE sonuçları tablodaki gibidir. RMSE ve MAE değerleri incelendiğinde Fuzzy 3 ve devamında düşük değerlere ulaştığımız görülmektedir.

RMSE değerinde en iyi performansı BCO'da Fuzzy 6, klasik yöntemlerde ortalama ataması yöntemi vermektedir. MAE değerinde en iyi performansı BCO'da Fuzzy 6, klasik yöntemlerde ortalama ataması yöntemi vermektedir.

Tablo 1.8 %10 Kayıp ile RMSE-MAE (369)

Kümeler ve Atama Yöntemleri	**RMSE**		**MAE**	
	Ortalama	**Standart Sapma**	**Ortalama**	**Standart Sapma**
Fuzzy 1	1.394466	0.098042	1.190810	0.120979
Fuzzy 2	1.202739	0.189574	0.973513	0.211384
Fuzzy 3	0.947367	0.087790	0.741351	0.092694
Fuzzy 4	0.941999	0.093106	0.733243	0.098213
Fuzzy 5	0.921373	0.087370	0.718648	0.091881
Fuzzy 6	**0.913274**	0.079136	**0.717567**	0.080796
Fuzzy 7	0.924329	0.082244	0.728648	0.086147
Fuzzy 8	0.936359	0.080867	0.735135	0.089267
Fuzzy 9	0.932764	0.091725	0.729729	0.098239
Fuzzy 10	0.940409	0.078948	0.741891	0.090692
Fuzzy 11	0.916369	0.084228	0.720270	0.086015
Ortalama Atama	0.969887	0.089711	0.760540	0.095421
Ortanca Atama	0.995985	0.110407	0.765675	0.109670
Mod Atama	1.202287	0.151136	0.843783	0.147712

4.5.7 %20 Kayıp ile RMSE-MAE(738)

Aşağıdaki tabloda bulanık c-ortalamalar 1'den 11'e kadar artan küme sayısı ve klasik kayıp veri bulma yöntemlerinden ortalama atama, ortanca atama ve mod atama yöntemleri 21 sütun 738 satırdan oluşan veri için incelenecektir.

%20 kayıp ile oluşturulan veri için ortalama ve standart sapma RMSE ve MAE sonuçları tablodaki gibidir. RMSE ve MAE değerleri incelendiğinde Fuzzy 3 ve devamında düşük değerlere ulaştığımız görülmektedir.

RMSE değerinde en iyi performansı BCO'da Fuzzy 10, klasik yöntemlerde ortalama ataması yöntemi vermektedir. MAE değerinde en iyi performansı BCO'da Fuzzy 10, klasik yöntemlerde ortanca ataması yöntemi vermektedir.

Tablo 1.9 %20 Kayıp ile RMSE-MAE (738)

Kümeler ve Atama Yöntemleri	**RMSE**		**MAE**	
	Ortalama	**Standart Sapma**	**Ortalama**	**Standart Sapma**
Fuzzy 1	1.391394	0.049232	1.197297	0.060708
Fuzzy 2	1.364829	0.095686	1.150608	0.113132
Fuzzy 3	0.932302	0.049571	0.729324	0.049117
Fuzzy 4	0.930583	0.044797	0.727297	0.041568
Fuzzy 5	0.932016	0.045938	0.726959	0.050891
Fuzzy 6	0.937555	0.043428	0.728175	0.048007
Fuzzy 7	0.939813	0.044428	0.731824	0.048283
Fuzzy 8	0.926578	0.044868	0.720675	0.050555
Fuzzy 9	0.927631	0.040970	0.725270	0.043812
Fuzzy 10	**0.925578**	0.044417	**0.718378**	0.050870
Fuzzy 11	0.926960	0.046005	0.723918	0.050380
Ortalama Atama	0.965181	0.041036	0.748513	0.045040
Ortanca Atama	0.990321	0.048948	0.740135	0.049385
Mod Atama	1.213850	0.075345	0.842567	0.072139

4.5.8 %20 Kayıp ile RMSE-MAE(369)

Aşağıdaki tabloda bulanık c-ortalamalar 1'den 11'e kadar artan küme sayısı ve klasik kayıp veri bulma yöntemlerinden ortalama atama, ortanca atama ve mod atama yöntemleri 21 sütun 369 satırdan oluşan veri için incelenecektir.

%20 kayıp ile oluşturulan veri için ortalama ve standart sapma RMSE ve MAE sonuçları tablodaki gibidir. RMSE ve MAE değerleri incelendiğinde Fuzzy 3 ve devamında düşük değerlere ulaştığımız görülmektedir.

RMSE değerinde en iyi performansı BCO'da Fuzzy 8, klasik yöntemlerde ortalama ataması yöntemi vermektedir. MAE değerinde en iyi performansı BCO'da Fuzzy 10, klasik yöntemlerde ortanca ataması yöntemi vermektedir.

Tablo 1.10 %20 Kayıp ile RMSE-MAE (369)

Kümeler ve Atama Yöntemleri	**RMSE**		**MAE**	
	Ortalama	**Standart Sapma**	**Ortalama**	**Standart Sapma**
Fuzzy 1	1.401951	0.076204	1.204459	0.090226
Fuzzy 2	1.177361	0.183718	0.949594	0.191619
Fuzzy 3	0.961891	0.074133	0.750405	0.078031
Fuzzy 4	0.945395	0.066175	0.74	0.069707
Fuzzy 5	0.926951	0.060286	0.725	0.065681
Fuzzy 6	0.939297	0.059716	0.741756	0.067741
Fuzzy 7	0.929309	0.063170	0.725135	0.067777
Fuzzy 8	**0.923129**	0.065007	0.722837	0.067867
Fuzzy 9	0.935156	0.064598	0.730270	0.071685
Fuzzy 10	0.932600	0.067003	**0.722297**	0.074473
Fuzzy 11	0.939874	0.067221	0.740810	0.075138
Ortalama Atama	0.987299	0.061571	0.770405	0.065909
Ortanca Atama	1.002883	0.071323	0.759189	0.069076
Mod Atama	1.216283	0.103742	0.843513	0.095962

4.5.9 %1 Kayıp ile Cronbach Alfa (738&369)

738 ve 369 satırlık tam verilerden %1 kayıp ile oluşturulan veri için ilgili Cronbach Alfa sonuçları tablodaki gibidir. Tam verinin Cronbach Alfa değerine 738 ve 369 satırlık verilerde en çok yaklaşan BCO'da Fuzzy 10, klasik yöntemlerde ortalama ataması yöntemidir.

Tablo 1.11 %1 Kayıp ile Cronbach Alfa (738)

Kümeler ve Atama Yöntemleri	**738 (%1)**		
	Ortalama	**RMSE**	**MAE**
Fuzzy 1	0.905486	0.000177	0.000152
Fuzzy 2	0.905469	0.000197	0.000170
Fuzzy 3	0.905601	0.000083	0.000067
Fuzzy 4	0.905615	0.000081	0.000066
Fuzzy 5	0.905617	0.000073	0.000055
Fuzzy 6	0.905621	0.000078	0.000063
Fuzzy 7	0.905627	0.000066	0.000053
Fuzzy 8	0.905624	0.000068	0.000054
Fuzzy 9	0.905615	0.000065	0.000052
Fuzzy 10	**0.905628**	0.000072	0.000057
Fuzzy 11	0.905625	0.000072	0.000054
Ortalama Atama	0.905572	0.000096	0.000075
Ortanca Atama	0.905570	0.000103	0.000079
Mod Atama	0.905544	0.000142	0.000110

Tablo 1.12 %1 Kayıp ile Cronbach Alfa (369)

Kümeler ve Atama Yöntemleri	**369(%1)**		
	Ortalama	**RMSE**	**MAE**
Fuzzy 1	0.903890	0.000213	0.000180
Fuzzy 2	0.903867	0.000236	0.000189
Fuzzy 3	0.903994	0.000113	0.000090
Fuzzy 4	0.904011	0.000115	0.000086
Fuzzy 5	0.904028	0.000123	0.000095
Fuzzy 6	0.904039	0.000104	0.000081
Fuzzy 7	0.904013	0.000100	0.000079
Fuzzy 8	0.904022	0.000115	0.000087
Fuzzy 9	0.904025	0.000091	0.000075
Fuzzy 10	**0.904045**	0.000109	0.000082
Fuzzy 11	0.904022	0.000095	0.000073
Ortalama Atama	0.903956	0.000133	0.000103
Ortanca Atama	0.903954	0.000143	0.000107
Mod Atama	0.903926	0.000185	0.000132

4.5.10 %5 Kayıp ile Cronbach Alfa (738&369)

738 ve 369 satırlık tam verilerden %5 kayıp ile oluşturulan veri için ilgili Cronbach Alfa sonuçları tablodaki gibidir. Tam verinin Cronbach Alfa değerine 738 satırlık veride BCO'da Fuzzy 7, 369 satırlık veride BCO'da Fuzzy 8 ve klasik yöntemlerde her iki veride Cronbach Alfa değerine yaklaşan yöntem ortalama atamasıdır.

Tablo 1.13 %5 Kayıp ile Cronbach Alfa (738)

Kümeler ve Atama Yöntemleri	**738 (%5)**		
	Ortalama	**RMSE**	**MAE**
Fuzzy 1	0.904986	0.000679	0.000648
Fuzzy 2	0.904926	0.000739	0.000708
Fuzzy 3	0.905471	0.000209	0.000176
Fuzzy 4	0.905516	0.000197	0.000163
Fuzzy 5	0.905548	0.000170	0.000139
Fuzzy 6	0.905570	0.000171	0.000136
Fuzzy 7	**0.905616**	0.000157	0.000124
Fuzzy 8	0.905553	0.000180	0.000139
Fuzzy 9	0.905563	0.000169	0.000134
Fuzzy 10	0.905554	0.000178	0.000115
Fuzzy 11	0.905560	0.000179	0.000148
Ortalama Atama	0.905308	0.000362	0.000327
Ortanca Atama	0.905289	0.000382	0.000347
Mod Atama	0.905130	0.000554	0.000504

Tablo 1.14 %5 Kayıp ile Cronbach Alfa (369)

Kümeler ve Atama Yöntemleri	**369(%5)**		
	Ortalama	**RMSE**	**MAE**
Fuzzy 1	0.903461	0.000666	0.000582
Fuzzy 2	0.903491	0.000660	0.000561
Fuzzy 3	0.903842	0.000328	0.000266
Fuzzy 4	0.903879	0.000292	0.000250
Fuzzy 5	0.903950	0.000255	0.000208
Fuzzy 6	0.903919	0.000252	0.000200
Fuzzy 7	0.903987	0.000217	0.000181
Fuzzy 8	**0.904004**	0.000213	0.000169
Fuzzy 9	0.904000	0.000205	0.000171
Fuzzy 10	0.903956	0.000231	0.000186
Fuzzy 11	0.903988	0.000218	0.000171
Ortalama Atama	0.903755	0.000367	0.000305
Ortanca Atama	0.903735	0.000392	0.000325
Mod Atama	0.903596	0.000551	0.000466

4.5.11 %10 Kayıp ile Cronbach Alfa (738&369)

738 ve 369 satırlık tam verilerde %10 kayıp ile oluşturulan veri için ilgili Cronbach Alfa sonuçları tablodaki gibidir. Tam verinin Cronbach Alfa değerine 738 satırlık veride BCO'da Fuzzy 8, 369 satırlık veride BCO'da Fuzzy 9 ve klasik yöntemlerde her iki veride Cronbach Alfa değerine yaklaşan yöntem ortalama atamasıdır.

Tablo 1.15 %10 Kayıp ile Cronbach Alfa (738)

Kümeler ve Atama Yöntemleri	**738 (%10)**		
	Ortalama	**RMSE**	**MAE**
Fuzzy 1	0.904360	0.001321	0.001275
Fuzzy 2	0.904245	0.001445	0.001389
Fuzzy 3	0.905270	0.000439	0.000370
Fuzzy 4	0.905369	0.000354	0.000308
Fuzzy 5	0.905431	0.000304	0.000247
Fuzzy 6	0.905446	0.000298	0.000245
Fuzzy 7	0.905545	0.000223	0.000178
Fuzzy 8	**0.905547**	0.000264	0.000202
Fuzzy 9	0.905497	0.000244	0.000200
Fuzzy 10	0.905484	0.000238	0.000192
Fuzzy 11	0.905498	0.000253	0.000208
Ortalama Atama	0.904988	0.000692	0.000646
Ortanca Atama	0.904963	0.000723	0.000672
Mod Atama	0.904573	0.001119	0.001061

Tablo 1.16 %10 Kayıp ile Cronbach Alfa (369)

Kümeler ve Atama Yöntemleri	**369(%10)**		
	Ortalama	**RMSE**	**MAE**
Fuzzy 1	0.902769	0.001346	0.001267
Fuzzy 2	0.902902	0.001294	0.001143
Fuzzy 3	0.903631	0.000525	0.000452
Fuzzy 4	0.903710	0.000483	0.000407
Fuzzy 5	0.903735	0.000412	0.000354
Fuzzy 6	0.903784	0.000420	0.000355
Fuzzy 7	0.903887	0.000318	0.000258
Fuzzy 8	0.903886	0.000384	0.000311
Fuzzy 9	**0.903888**	0.000363	0.000284
Fuzzy 10	0.903870	0.000368	0.000290
Fuzzy 11	0.903849	0.000354	0.000284
Ortalama Atama	0.903413	0.000700	0.000632
Ortanca Atama	0.903376	0.000744	0.000670
Mod Atama	0.903078	0.001048	0.000963

4.5.12 %20 Kayıp ile Cronbach Alfa (738&369)

738 ve 369 satırlık tam verilerde %20 kayıp ile oluşturulan veri için ilgili Cronbach Alfa sonuçları tablodaki gibidir. Tam verinin Cronbach Alfa değerine 738 satırlık veride BCO'da Fuzzy 7, 369 satırlık veride BCO'da Fuzzy 8 ve klasik yöntemlerde her iki veride Cronbach Alfa değerine yaklaşan yöntem ortalama atamasıdır.

Tablo 1.17 %20 Kayıp ile Cronbach Alfa (738)

Kümeler ve Atama Yöntemleri	**738 (%20)**		
	Ortalama	**RMSE**	**MAE**
Fuzzy 1	0.903111	0.002585	0.002523
Fuzzy 2	0.902815	0.002866	0.002819
Fuzzy 3	0.905006	0.000729	0.000639
Fuzzy 4	0.905132	0.000588	0.000515
Fuzzy 5	0.905211	0.000521	0.000440
Fuzzy 6	0.905313	0.000503	0.000389
Fuzzy 7	**0.905317**	0.000441	0.000354
Fuzzy 8	0.905262	0.000496	0.000425
Fuzzy 9	0.905199	0.000543	0.000467
Fuzzy 10	0.905249	0.000488	0.000415
Fuzzy 11	0.905283	0.000483	0.000392
Ortalama Atama	0.904431	0.001270	0.001204
Ortanca Atama	0.904355	0.001356	0.001280
Mod Atama	0.903672	0.002027	0.001963

Tablo 1.18 %20 Kayıp ile Cronbach Alfa (369)

Kümeler ve Atama Yöntemleri	**369(%20)**		
	Ortalama	**RMSE**	**MAE**
Fuzzy 1	0.901442	0.002692	0.002594
Fuzzy 2	0.901904	0.002382	0.002133
Fuzzy 3	0.903189	0.000981	0.000852
Fuzzy 4	0.903350	0.000860	0.000718
Fuzzy 5	0.903522	0.000692	0.000579
Fuzzy 6	0.903482	0.000718	0.000599
Fuzzy 7	0.903707	0.000624	0.000507
Fuzzy 8	**0.903727**	0.000544	0.000441
Fuzzy 9	0.903632	0.000613	0.000503
Fuzzy 10	0.903541	0.000749	0.000595
Fuzzy 11	0.903639	0.000614	0.000494
Ortalama Atama	0.902725	0.001414	0.001311
Ortanca Atama	0.902652	0.001491	0.001384
Mod Atama	0.902062	0.002076	0.001974

5
SONUÇ VE ÖNERİLER

Araştırmamızda kullandığımız "Depresyon Anksiyete ve Stres Ölçekleri-21" (DASS-21) veri seti 21 sütun ve 738 satırdan oluşan bir veri setidir. Eksik veri içermeyen bu veri setinden %1, %5, %10 ve %20 oranlarında kayıp veri üretilmiştir. Üretilen kayıpları bulabilmek için bulanık mantık algoritmalarından biri olan bulanık c-ortalamalar (BCO) kullanılmıştır. BCO ile karşılaştırılmak üzere klasik kayıp veri tahmin etme yöntemlerinden ortalama atama, ortanca atama ve mod atama yöntemleri kullanılmıştır.

Bulanık c-ortalamalar (BCO), klasik kayıp veri tahmin etme yöntemlerine alternatif bir yöntem olarak kullanılabilme durumu incelenmiştir. Bu araştırmada BCO 1'den 11'e kadar küme sayısı artırılarak değerler incelenmiştir. Yöntemlerin karşılaştırılabilmesi için "Kök Ortalama Kare Hata (RMSE)" , "Ortalama Mutlak Hata (MAE)" ve "Cronbach Alfa" formülleri kullanılmıştır.

İncelenen diğer bir nokta ise kayıp verileri tahmin ederken, kullanılan yöntemleri örneklem sayısının etkileyip etkilemediğidir. Bu konuda örneklem sayısının etkili olup olmadığını incelemek üzere 21 sütun ve 738 satırdan oluşan veri seti yarıya indirilerek 21 sütun ve 369 satırdan oluşan yeni bir veri oluşturulmuştur. Oluşturulan yeni veri setinden %1, %5, %10 ve %20 oranlarında kayıp veri üretilmiştir. Üretilen veri setlerine aynı yöntemler uygulanmış ve RMSE, MAE ve Cronbach Alfa değerleri bulunmuştur.

Yöntemler karşılaştırılırken her veri setine 100 tekrar yapılmıştır. 100 tekrar sonucu çıkan değerlerin ortalaması alınarak tablolara işlenmiş ve karşılaştırmalar gerçekleştirilmiştir.

Araştırmanın uygulamasında kullanılan kümeler ve atama yöntemleri; fuzzy 1, fuzzy 2, fuzzy 3, fuzzy 4, fuzzy 5, fuzzy 6, fuzzy 7, fuzzy 8, fuzzy 9, fuzzy 10, fuzzy 11, ortalama atama, ortanca atama ve mod atama değerleri R Studio programında bulunmuştur.

Bulanık C-Ortalamaların, klasik kayıp veri tahmin etme yöntemlerine alternatif bir yöntem olarak kullanabilme konusu incelendiğinde; BCO değerleri ile klasik yöntemlerin değerlerinin karşılaştırılmasında kök ortalama kare hata (RMSE) ve ortalama mutlak hata (MAE), tahminlerin gerçek değerlere göre ne kadar doğru olduğunu değerlendirmek için kullanılan yaygın hata ölçümleridir. Daha düşük RMSE ve MAE değerleri, daha iyi tahminlerin işaretidir. Bulduğumuz değerlerde düşük olan değerler bizim için en iyi sonuçları vermektedir. %1, %5, %10 ve %20 oranlarında kayıp içeren verilerin değerleri incelendiğinde bulanık c-ortalamalar (BCO)'da en iyi sonuçları fuzzy 3 ve fuzzy 3'ten büyük küme sayılarının verdiği görülmektedir. BCO için tüm veriler incelendiğinde küme sayısı arttıkça değerlerin tam verinin değerlerine yakınsadığı görülmektedir. Fuzzy küme sayısı 3'ten fazla olması durumu iyileştirmektedir.

Yaptığımız araştırmada, veri setinin karmaşıklığına bağlı olarak küme sayısının seçilmesinin önemi görülmektedir. Daha fazla küme sayısı, veri setinin daha iyi bir şekilde temsil edebileceği anlamına gelmektedir.

Ortalama atama, ortanca atama ve mod atama yöntemleri ile yapılan karşılaştırmalı analiz, alışagelmiş yöntemlerin bulanık c-ortalamalar ile ayrıştığını bulanık c-ortalamalar yönteminin daha iyi sonuçlar verdiğini göstermektedir. Klasik yöntemlerde en iyi sonucu ortalama ataması yöntemi vermektedir.

Kayıpların tahmin edildiği veri setinin Cronbach Alfa değeri tam veri setinin Cronbach Alfa değerine ne kadar yakınsa verinin içsel tutarlılığına katkıda bulunduğunu gösterir. Cronbach Alfa değerleri incelendiğinde, kayıpsız verinin Cronbach Alfa değeri 0,906'dır. Kayıplarını tamamladığımız verilerin en iyi sonuç vermesi için eksiksiz verinin Cronbach Alfa değerine yakınsamasını istiyoruz. İncelenen sonuçlarda Cronbach Alfa değerleri eksiksiz verinin alfa değerine oldukça yakın çıkmıştır.

Nihai sonuç olarak BCO algoritmasının eksik veri tamamlama için etkili bir yöntem olduğunu görmekteyiz. Bu algoritma, veri setinin belirli bir küme sayısı ile temsil edilmesine ve eksik verilerin tahmin edilmesine yardımcı olmuştur. BCO ile eksik veri tamamlama konusundaki bulguları, bu yöntemin veri setlerinde etkili bir şekilde kullanılabileceğini ve farklı yöntemlerle karşılaştırıldığında rekabetçi

sonuçlar verdiğini göstermektedir. Bulanık C-Ortalamalar kayıp verileri tahmin etmek için alternatif bir yöntem olarak kullanılabilir.

Çalışmanın geliştirilmesi daya iyi sonuçların alınabilmesi için yapılan çalışmaya ek olarak ileri ki çalışmalarda veri setinin büyüklüğü ve çeşitliliği, eksik veri tahmin yöntemlerinin performanslarını etkileyebilir. Daha büyük ve farklı veri setleri üzerinde testler yaparak, bulanık c-ortalamalar ve diğer yöntemlerin genel uygulanabilirliğini ve performansları daha iyi anlaşılabilir.

BCO veya diğer eksik veri tahmin etme yöntemlerinin pratik uygulamalarda veya endüstriyel senaryolarda nasıl kullanılabileceği daha ayrıntılı olarak incelenebilir. Bu yöntemlerin gerçek dünya problemlerinde nasıl kullanılabileceği konusunda fikir verebilir.

Farklı eksik veri tahmin yöntemleri ile bulanık c-ortalamalar algoritması karşılaştırılarak daha kapsamlı sonuçlara ulaşılabilir. Bu, farklı yöntemlerin avantajları ve dezavantajlarının daha iyi anlamamıza yardımcı olacaktır.

İlerleyen çalışmalarda simülasyon çalışması yapılabilir. Yapılan simülasyon çalışması daha ayrıntılı ve kapsamlı gözle görülebilir sonuçlara ulaştırabilir.

KAYNAKÇA

[1] A. Karadoğan, A. B. (2001). Bulanık Küme Teorisinin Yeraltı Üretim Yöntemi Seçiminde Kullanılabilirliği . Tumaks, Isbn 975-395-416-6 .

[2] Afifi, A. A. (1966). Missing Observations İn Multivariate Statistics: I. Review Of The Literature. Journal Of The American Statistical Association, 61(315), 595.

[3] Afifi, A. A. (1967). Missing Observations İn Multivariate Statistics: Iı. Point Estimation İn Simple Linear Regression. Journal Of The American Statistical Association, 62(317), 10.

[4] Afifi, A. A. (1969(A)). Missing Observations İn Multivariate Statistics. Iıı: Large Sample Analysis Of Simple Linear Regression. . Journal Of The American Statistical Association, 64(325), 337.

[5] Afifi, A. A. (1969(B)). Missing Observations İn Multivariate Statistics. Iv: A Note On Simple Linear Regression. Journal Of The American Statistical Association, 359.

[6] Alavala, C. R. (2008). Fuzzy Logic And Neural Networks Basic Concepts & Applications. India: New Age International Publications.

[7] Alkan, N. (2012). Kayıp Verili Cox Regresyon Yöntemine Bayesci Bir Yaklaşım. Doktora Tezi. Samsun: Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü.

[8] Allan, F. E. (1930). A Method Of Estimating The Yield Of A Missing Plot İn Field Experimental Work. The Journal Of Agricultural Science, 399–406.

[9] Allison, P. D. (2003). Missing Data Techniques For Structural Equation Modeling. Journal Of Abnormal Psychology , 545-557.

[10] Alpar, R. (2003). Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş 1. Ankara: Nobel.

[11] Alpar, R. (2013). Uygulamalı Çok Değişkenli İstatistik Yöntemler. Detay Yayıncılık Dördüncü.

[12] Altaş, D. D. (Temmuz 1999). Bulanık Mantık : Bulanıklılık Kavramı. Bilesim Yayıncılık A.Ş, 80-85.

[13] Anderson, T. W. (1957). Maximum Likelihood Estimates For A Multivariate Normal Distribution When Some Observations Are Missing. Journal Of The American Statistical Association, 52(278), 200.

[14] Avcı, D. (2017). Bulanık C Ortalamalar Kümeleme Tabanlı Hedef Eko Sinyal Ortamlarındaki Gerçek Hedef Sayısının Tespiti. Fırat Üniv. Mühendislik Bilimleri Dergisi , 29(1), 73-82.

[15] Baygül, A. (2007). Kayıp Veri Analizinde Sıklıkla Kullanılan Etkin Yöntemlerin Değerlendirilmesi. İstanbul, 195017: Yüksek Lisans Tezi,İstanbul Üniversitesi Sağlık Bilimleri Üniversitesi.

[16] Bose, S. S. (1938). On Estimating Individual Yields İn The Case Of Mixed-Up Yields Of Two Or More Plots İn Field Experiment. Indian Statistical Institute.

[17] Buck, S. F. (1960). A Method Of Estimation Of Missing Values İn Multivariate Data Suitable For Use With An Electronic Computer. Journal Of The Royal Statistical Society. Series B (Methodological). Wileyroyal Statistical Society.

[18] Cool, A. (2000). A Review Of Methods For Dealing With Missing Data (Rapor). Annual Meeting Of The Southwest Educational Resarch Association. Dallas.

[19] Çakar, T. (2020). Bulanık Çok Ölçütlü Karar Verme Yöntemleri. İstanbul: İstanbul Gelişim Üniversitesi Yayınları.

[20] Dear, R. E. (1959). A Principal Component Missing Data Method For Multiple Regression Models. California: System Development Corporation.

[21] Dempster, A. P. (1977). Maximum Likelihood From Incomplete Data Via The Em Algorithm. Journal Of The Royal Statistical Society. Series B (Methodological). Wileyroyal Statistical Society.

[21] Eberle, W. A. (1999). Paper: Handling Of Missing Values İn Statistical Software Packages For Windows. Sonderforschungsbereich:-Institut Für Statistik.

[22] Edgett, G. L. (1956). Multiple Regression With Missing Observations Among The Independent Variables. Journal Of The American Statistical Association, 51(273), 122.

[23] Enders, C. K. (2001). The Relative Performance Of Full İnformation Maximum Likelihood Estimation For Missing Data İn Structural Equation Models. Structural Equation Modeling: A Multidisciplinary Journal.

[24] Enders, C. K. (2010). Applied Missing Data Analysis. New York: The Guilford Press.

[25] Ergül Demir, B. P. (2012). Türkiye'de Eğitim Araştırmalarında Kayıp Veri Sorunu. Eğitimde Ve Psikolojide Ölçme Ve Değerlendirme Dergisi, 230-241.

[26] Erilli, N. A. (Temmuz 2014). Bulanık Kümeleme Analizi İle İstatistiki Bölge Birimlerinin (İbbs) Mali Değişkenlere Göre Sınıflandırılması. Cumhuriyet Üniversitesi,İibf, Ekonometri Bölümü, Cilt: 4 Sayı: 2.

[27] Ertuğrul, İ. (2010). Akademik Performans Değerlendirmede Bulanık Mantık Yaklaşımı . Pamukkale Üniversitesi İibf İşletme Bölümü , 155 - 176.

[28] Evren Sezgin, Y. Ç. (2013). Veri Madenciliğinde Kayıp Veriler Đçin Kullanılan Yöntemlerin Karşılaştırılması . Akademik Bilişim Konferansı Bildirileri (S. 569-572). Antalya: Akdeniz Üniversitesi.

[29] F. Martin Mcneill, E. T. (1994). Fuzzy Logic: A Practical Approach. London: Academic Press.

[30] Fırat Artuğer, F. Ö. (2021). Görüntü Sıkıştırma Algoritmalarının Performans Analizi İçin Değerlendirme . Ijpas, 104.

[31] G., M. G. (2005). Models For Discrete Longitudinal Data. New York: Springer.

[32] Glasser, M. (1964). Linear Regression Analysis With Missing Observations Among The Independent Variables. Journal Of The American Statistical Association, 59(307), 834.

[33] Haitovsky, Y. (1968). Missing Data İn Regression Analysis. Journal Of The Royal Statistical Society. Series B (Methodological). Wileyroyal Statistical Society.

[34] Hartley, H. O. (1958). Maximum Likelihood Estimation From Incomplete Data. Biometrics, 174.

[35] Hartley, H. O. (1971). The Analysis Of Incomplete Data. Biometrics, 27(4), 783.

[36] Hasan Yılmaz, M. E. (2023). Bulanık Mantık Kavramına Genel Bir Bakış. Takvim-İ Vekayi, 94-129.

[37] Healy, M. &. (1956). Missing Values İn Experiments Analysed On Automatic Computers. Applied Statistics, 5(3), 203.

[38] Horton, N. &. (2007). Much Ado About Nothing: A Comparison Of Missing Data Methods. American Statistical Association, 79-90.

[39] Hüseyin Akbulut, C. G. (15 Haziran 2019). Bulanık Mantık Yöntemi İle Sinyalize Kavşaklarda Trafik Işığı Süresi Belirlenmesi. Journal Of Innovations İn Civil Engineering And Technology, 41-56.

[40] Işıklı, Ş. (2007). Bulanık Mantık Ve Bulanık Teknolojiler. Ankara Üniversitesi Dil Ve Tarih-Coğrafya Fakültesi Felsefe Bölümü Dergisi, 105-126.

[41] Jones, M. P. (1996). Indicator And Stratification Methods For Missing Explanatory Variables İn Multiple Linear Regression. Journal Of The American Statistical Association, 91(433), 222.

[42] Juhola, M. &. (2013). Missing Values: How Many Can They Be To Preserve Classification Reliability? Artificial Intelligence Review, 231-245.

[43] Karataş, İ. (29 Ekim 2018). Bulanık Mantık İle Klasik Ve Sembolik Mantık İlişkisi (Karşılaştırılması). Avrupa Eğitim Ve Sosyal Bilimler Dergisi, 3(2), 144-163.

[44] Kermack, W. O. (1927). A Contribution To The Mathematical Theory Of Epidemics. Royal Society, 700-721.

[45] Kermack, W. O. (1932). Contributions To The Mathematical Theory Of Epidemics. Iı. The Problem Of Endemicity. Royal Society.

[46] Kermack, W. O. (1933). Contributions To The Mathematical Theory Of Epidemics. Iıı. Further Studies Of The Problem Of Endemicity. Royal Society.

[47] Kiki Aristiawati, T. S. (2019). Missing Values Imputation Based On Fuzzy C-Means Algorithm For Classification Of Chronic Obstructive Pulmonary Disease (Copd. Aıp Conference Proceedings, 2192, 060003.

[48] Kırğıl, E. N. (2022). Makine Öğrenmesi Teknikleriyle Yazılım Uyum Metriklerinin Tahmini. Yüksek Lisans Tezi. Ankara: Yök.

[49] L.A., Z. (1990). The Birth And Evolution Of Fuzzy Logic.

[50] Little, R. J. (1987). Statistical Analysis With Missing Data. Wiley.

[51] Little, R. J. (1992). Regression With Missing X's: A Review. Journal Of The American Statistical Association, 87(420), 1227.

[52] Longford, N. T. (2004). Missing Data And Small-Area Estimation İn The Uk Labour Force Survey. Journal Of The Royal Statistical Society Ser. 341-373.

[53] Lord, F. M. (1955). Estimation Of Parameters From Incomplete Data. Journal Of The American Statistical Association, 50(271), 870.

[54] Louis, T. A. (1982). Finding The Observed Information Matrix When Using The Em Algorithm. Journal Of The Royal Statistical Society. Series B (Methodological). Wileyroyal Statistical Society.

[55] M. Sarı, Y. Ş. (15 Aralık 2005). Bulanık Modelleme Yaklaşımı Ve Uygulamaları. Dumlupınar Üniversitesi Bilim Ve Teknoloji Dergisi, 77-92.

[56] Mahmut Hekim, U. O. (Şubat 2011). Bulanık C-Means Kümeleme Yöntemine Çıkarımlı Yaklaşım. İtü Dergisi, Cilt:10, Sayı:1, 11-17.

[57] Matthai, A. (1951). Estimation Of Parameters From Incomplete Data With Application To Design Of Sample Surveys. The Indian Journal Of Statistics (1933- 1960). Indian Statistical Institute.

[58] Mckendrick, A. G. (1926). Applications Of Mathematics To Medical Problems. Proceedings Of Edinburgh Mathematical Society, (S. 98-130).

[59] Melike Şişeci, B. C. (2012). Traverten Plaka Taşlarda Sınıfların K-Ortalamalar Ve Bulanık C-Ortalamalar Kümeleme Yöntemleri İle Belirlenmesi. Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 238-247.

[60] Merve Şahin Kürşad, Z. N. (2015). Kayıp Veri Sorununun Çözümünde Kullanılan Farklı Yöntemlerin Ölçeklerin Geçerlik Ve Güvenirliği Bağlamında Karşılaştırılması. Eğitimde Ve Psikolojide Ölçme Ve Değerlendirme Dergisi, 6(2); 254-267.

[61] Mustafa Furkan Keskenler, E. F. (29 Haziran 2017). Bulanık Mantığın Tarihi Gelişimi. Takvim-İ Vekayi, Cilt: 5 No: 1 Sayfa: 1-10 .

[62] Naciye Aydin, G. K. (Aralık 2021). Jensen Shannon Mesafesi Temelli Uyarlanmış Bulanık C Ortalamalar Kümeleme Yöntemi. Avrupa Bilim Ve Teknoloji Dergisi, S. 58-64.

[63] Nazife Baykal, T. B. (2004). Bulanık Mantık İlke Ve Temelleri. Ankara: Bıçaklar Kitabevi.

[64] Nazire Memmedova, İ. K. (2009). Hayvancılıkta Bulanık Mantık Uygulamaları. Selçuk Tarım Ve Gıda Bilimleri Dergisi, 89-95.

[65] Necdet Tekin, S. K. (10 Ocak 1995). Fuzzy Mantık (Bulanık Mantık). Öneri Dergisi, 3-15.

[66] Nicholson, G. E. (1957). Estimation Of Parameters From Incomplete Multivariate Samples. Journal Of The American Statistical Association, 52(280), 523.

[67] Nuran Peker, C. K. (2021). Boyut Azaltmanın Bulanık C-Ortalama Kümeleme Teknikleri Üzerindeki Etkisi . Veri Bilimi Dergisi, 1-7.

[68] Oğuzlar, A. (2001). Alan Araştırmalarında Kayıp Değer Problemi Ve Çözüm Önerileri. Adana: V. Ulusal Ekonometri Ve İstatistik Sempozyumu,Çukurova Üniversitesi İibf Ekonometri Bölümü.

[69] Orchard, T. &. (1972). A Missing İnformation Principle: Theory And Applications. Proceedings Of The Sixth Berkeley Symposium On Mathematical Statistics And Probability, Volume 1: Theory Of Statistics. 1-19.

[70] Ödük, M. N. (2019). Bulanık Mantık Yöntemi Ve Uygulamaları. Ankara: Iksad Publications.

[71] Ömay Çokluk, M. K. (2011). Kayıp Değerlere Yaklaşık Değer Atama Yöntemlerinin Ölçme Araçlarının Geçerlik Ve Güvenirliği Üzerindeki Etkisi. Kuram Ve Uygulamada Eğitim Bilimleri, 289-309.

[72] Peng, C. J. (2007). Advances İn Missing Data Methods And Implications For Educational Research. In S. S.Sawilowsky (Ed.). Real Data Analysis, 31-78. Information Age.

[73] Pınar Özyılmaz Küçükyağcı, M. O. (2019). Bulanık Mantık Yönteminin Kentsel Alan Çalışmalarında Kullanımı. Kent Kültürü Ve Yönetimi Hakemli Elektronik Dergisi, 299-308.

[74] Prof. Dr. Necdet Tekin, A. G. (1995 Ocak). Fuzzy Mantık (Bulanık Mantık). Öneri Dergisi, 3-15.

[75] Roderick J.A. Little, D. B. (1987). Statistical Analysis With Missing Data. New York: John Wiley&Sons.

[76] Rubin, D. B. (1976). Inference And Missing Data. Biometrika, 63(3), 581.

[77] Sait Çüm, E. K. (2018). Kayıp Veriler Yerine Yaklaşık Değer Atamak İçin Kullanılan Gelişmiş Yöntemlerin Farklı Koşullar Altında Karşılaştırılması. Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi, 230-249.

[78] Scheffer, J. (2002). Dealing With Missing Data. Auckland: Res. Lett. Inf. Math. Sci.

[79] Sıddıkoğlu, D. (2019). Kayıp Veri Problemi Ve Farklı Değer Atama Yaklaşımlarının Whodas-2.0 Ölçeğinin Psikometrik Özellikleri Üzerindeki Etkisinin İncelenmesi, Doktora Tezi. Ankara: Ankara Üniversitesi Sağlık Bilimleri Enstitüsü.

[80] Sturge, M. D. (1911). On Some Of The Biological And Statistical Errors In The Work On Parental Alcoholism By Miss Elderton And Professor Karl Pearson. F.R.S. The British Medical Journal. Bmj.

[81] Şen, Z. (2020). Bulanık Mantık İlkeleri Ve Modelleme. İstanbul: Su Vakfı.

[82] Turan Paksoy, N. Y. (2013). Bulanık Küme Teorisi. Ankara: Nobel Akademik Yayıncılık.

[83] Türkbey, O. (2003). Makina Sıralama Problemlerinde Çok Amaçlı Bulanık Küme. Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi, 18(2) : 63-77.

[84] Türkşen, İ. B. (2015). Dereceli (Bulanık) Sistem Modelleri. İstanbul: Abaküs.

[85] Wilkinson, G. N. (1957). The Analysis Of Covariance With Incomplete Data. Biometrics, 363.

[86] Wilkinson, G. N. (1958). The Analysis Of Variance And Derivation Of Standard Errors For Incomplete Data. Biometrics, 360.

[87] Wilks, S. S. (1932). Moments And Distributions Of Estimates Of Population Parameters From Fragmentary Samples. The Annals Of Mathematical Statistics. Institute Of Mathematical Statistics.

[88] Wisniewski S.R., L. A. (2006). Prevention Of Missing Data İn Clinical Research Studies. Biol Psychiatry.

[89] Yates, F. (1933). The Analysis Of Replicated Experiments When The Field Results Are İncomplete. The Analysis Of Replicated Experiments When The Field Results Are İncomplete., 1(2), 129–142.

[90] Yates, F. (1936). Incomplete Randomized Blocks. . Annals Of Eugenics, 121–140.

[91] Yılancı, V. (2010). Bulanık Kümeleme Analizi İle Türkiye'deki İllerin Sosyoekonomik Açıdan Sınıflandırılması. Süleyman Demirel Üniversitesi İktisadi Ve İdari Bilimler Fakültesi Dergisi, S.453-470.

[92] Zeynel Cebeci, F. Y. (2015). Bulanık C-Ortalamalar Algoritmasının Farklı Küme Büyüklükleri İçin Hesaplama Performansı Ve Kümeleme Geçerliliğinin Karşılaştırılması. . 9. Ulusal Zootekni Bilim Kongresi, Bildiriler Kitabı, 227-239.