

**T.C.
ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI**

**LİKERT TİPİ ÖLÇEKLERDE KAYIP VERİLERE DEĞER
ATAMADA YAPAY SİNİR AĞLARININ KULLANIMI**

YÜKSEK LİSANS TEZİ

Münevver Başman

**Ankara
Ocak, 2014**

**T.C.
ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI**

**LİKERT TİPİ ÖLÇEKLERDE KAYIP VERİLERE DEĞER
ATAMADA YAPAY SİNİR AĞLARININ KULLANIMI**

YÜKSEK LİSANS TEZİ

Münevver Başman

Danışman: Prof. Dr. Ezel Tavşancıl

**Ankara
Ocak, 2014**

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼'ne,

Bu alıřma j¼rimiz tarafından ¼lme ve Deęerlendirme Anabilim Dalında
Y¼KSEK LİSANS TEZİ olarak kabul edilmiřtir.

Başkan Prof. Dr. Nizamettin KO

¼ye Prof. Dr. Ezel TAVřANCIL (Danıřman)

¼ye Prof. Dr. řener B¼Y¼K¼ZT¼RK

Onay

Yukarıdaki imzaların, adı geen ¼đretim ¼yelerine ait olduęunu onaylıyorum.

.../.../20..

Enstit¼ M¼d¼r¼

ÖNSÖZ

Veri setlerinde yer alan kayıp veriler nedeniyle yapılan analizlerde zorluklarla karşılaşmaktadır. Tam verilerin analizinden elde edilen sonuçlarla, kayıp veri içeren veri setinin analizinden elde edilen sonuçlara göre daha doğru ve kesin çıkarımlarda bulunulur. Daha doğru kararlar vermek ve daha etkili analiz sonuçları elde etmek adına kayıp veri sorununu çözmek gerekmektedir.

Kayıp verilerin üstesinden gelmek için literatürde çeşitli yöntemler bulunmaktadır. Bunlardan birisi de yapay sinir ağlarıdır. Yapay sinir ağlarının değişkenler arası ilişkinin doğrusal olmadığı durumlarda diğer atama yöntemlerine göre daha gerçekçi sonuçlar üretmesi, herhangi bir ön varsayımın gerektirmemesi, sadece eldeki verilerle sınırlandırılarak sistemin yürütülmesi gibi çeşitli avantajları nedeniyle diğer atama yöntemlerinden daha üstün olduğunun düşünülmesi bu yöntemin kayıp verilere değer atamada da uygun bir yöntem olabileceği fikrini akıllara getirmektedir. Bu araştırmada PISA 2009 öğrenci anketinde yer alan öğrencilerin okumaya ilişkin tutumunu gösteren Türkiye verileri kullanılarak, kayıp verilere değer atama yöntemlerinden yapay sinir ağları yönteminin kullanımı incelenmiştir.

Çalışmamda benden yardımlarını esirgemeyen, fikirlerini aktarıp çalışmamı zenginleştiren sayın hocam Prof. Dr. Ezel TAVŞANCIL'a, kıymetli eleştirileri ile tez jürimde yer alıp beni onurlandıran saygıdeğer hocalarım Prof. Dr. Nizamettin KOÇ'a ve Prof. Dr. Şener BÜYÜKÖZTÜRK'e sonsuz teşekkürlerimi sunar, eşsiz deneyimlerinden daha uzun yıllar faydalanabilmeyi dilerim. Ayrıca bu süreç içerisinde beni cesaretlendiren ve bana güvenen, manevi desteklerini hiçbir zaman esirgemeyen eşim Gökhan BAŞMAN'a ve ağabeyim Onur KAYA'ya, bugünlere gelmemde büyük pay sahibi olan anneme ve babama sonsuz teşekkürlerimi sunarım.

Münevver Başman

ÖZET

LİKERT TİPİ ÖLÇEKLERDE KAYIP VERİLERE DEĞER ATAMADA YAPAY SİNİR AĞLARININ KULLANIMI

Başman, Münevver

Yüksek Lisans, Ölçme ve Değerlendirme Anabilim Dalı

Tez Danışmanı: Prof. Dr. Ezel TAVŞANCIL

Ocak 2014, 94 sayfa

Veri setlerinden daha etkili analiz sonuçları elde etmek ve daha doğru kararlar vermek için veri setlerinde yer alan kayıp verilerin üstesinden gelmek gerekmektedir. Bu amaçla literatürde çeşitli yöntemler bulunmaktadır. Bunlardan birisi de yapay sinir ağlarıdır. Temel araştırma niteliğindeki bu araştırmada PISA 2009 öğrenci anketinde yer alan öğrencilerin okumaya ilişkin tutumunu gösteren Türkiye verileri kullanılarak, kayıp verilere değer atama yöntemlerinden yapay sinir ağları yönteminin kullanımı incelenmiştir. Bu araştırmanın çalışma grubunu PISA 2009 uygulamasına Türkiye'den katılan tam veri setine sahip öğrenciler içerisinden alınan 500 öğrenci oluşturmuştur. Çalışma grubuna dahil edilen 500 öğrencinin 11 maddeye verdiği cevaplardan öncelikle herhangi bir değişkendeki gözenek sayısı, sonrasında herhangi iki değişkendeki gözenek sayısı ve daha sonra herhangi üç değişkendeki gözenek sayısı SPSS programı kullanılarak seçkisiz olarak %5, %15 ve %30 oranlarında kayıp veri elde edilecek şekilde belirlenerek veri setleri oluşturulmuştur. Oluşturulan veri setlerine yapay sinir ağları yöntemiyle kayıp verilere değer ataması MATLAB programı kullanılarak yapılmış, elde edilen tam veri setlerine SPSS programı kullanılarak açımlayıcı faktör analizi yapılmıştır. Ayrıca SPSS ve LISREL programları kullanılarak güvenilirlik analizleri de yapılmıştır. Elde edilen bulgular, kayıp verilere yapay sinir ağlarıyla değer atanması sonucunda oluşan veri setlerinin ölçme aracının faktör yapısını değiştirmedeği, gerçek veri setine yapılan faktör analizindeki gibi 'tek faktörlü' olduğu görülmüştür. Açıklanan varyans oranı, düzeltilmiş madde-toplam korelasyonları ve Cronbach alfa ve omega güvenilirlik katsayılarının da gerçek veri setiyle tutarlı sonuçlar verdiği belirlenmiştir.

SUMMARY

THE USE OF ARTIFICIAL NEURAL NETWORKS IN IMPUTATIONS OF MISSING DATA IN LIKERT TYPE SCALE

Başman, Münevver

Graduate, Department of Educational Measurement and Evaluation

Advisor: Prof. Dr. Ezel TAVŞANCIL

January 2014, 94 pages

Researchers need to find a way to deal with the missing data in data sets in order to obtain better analysis results and to give better decisions. There are several methods developed in literature for this purpose and Artificial Neural Networks is one of these methods. This paper, which is considered to be a fundamental research, aims to examine the effectiveness of the Artificial Neural Networks methods using Türkiye's data showing students' attitudes towards reading located in the PISA 2009 students' questionnaire. The working group of this research study is formed of 500 students who are chosen from complete data among students who attended the PISA 2009 application in Türkiye. Firstly in any one of the variables and then any two of the variables and finally any three of the variables of the 500 students that are included in this working group are selected randomly and data sets are created to obtain missing data at the rate of 5%, 15% and 30% to be used in SPSS program. Missing data in the created data sets are estimated by artificial neural networks using MATLAB. Then we applied an explanatory factor analysis to the completed data sets using SPSS. Furthermore we applied reliability analyses to the completed data sets using SPSS and LISREL. We found out that the data sets obtained after the missing data were assigned by artificial neural networks did not change the factor structure of the measurement tool and it consists of a single factor as it was seen by the factor analysis done for the real data set. It is also seen that the results related to the explained variance rate, corrected item-total correlations and also Cronbach alpha and omega reliability coefficients of the created data sets are consistent with the results of the real data sets.

İÇİNDEKİLER

	<u>Sayfa</u>
JÜRİ ÜYELERİNİN İMZA SAYFASI	i
ÖNSÖZ	ii
ÖZET	iii
SUMMARY	iv
İÇİNDEKİLER	v
ÇİZELGELER LİSTESİ	vii
ŞEKİLLER LİSTESİ	xi
BÖLÜM I	1
GİRİŞ	1
Problem	1
Amaç	22
Önem	22
Sınırlılıklar	23
Kısaltmalar	23
BÖLÜM II	24
İLGİLİ ARAŞTIRMALAR	24
Yurt İçinde Yapılan Araştırmalar	24
Yurt Dışında Yapılan Araştırmalar	27
BÖLÜM III	31
YÖNTEM	31
Araştırma Modeli	31
Çalışma Grubu	31
Verilerin Toplanması	32
Verilerin Analizi	32
Yapay Sinir Ağları	33
Yapay Sinir Ağının Yapısı	35
Yapay Sinir Ağının Tasarımı	39
Çok Katmanlı İleri Beslemeli Yapay Sinir Ağları	44

BÖLÜM IV.....	50
BULGULAR VE YORUMLAR.....	50
Veri Setlerindeki Kayıp Veri Oranına Göre Yapay Sinir Ağlarının Etkililiği.	50
Yapay Sinir Ağları Yöntemiyle Kayıp Verilere Değer Ataması Yapılarak Elde Edilen Tam Veri Setine Yapılan Geçerlik Analizi Bulguları.....	58
Yapay Sinir Ağları Yöntemiyle Kayıp Verilere Değer Ataması Yapılarak Elde Edilen Tam Veri Setine Yapılan Güvenirlik Analizi Bulguları	63
BÖLÜM V.....	69
SONUÇ VE ÖNERİLER.....	69
Sonuç	69
Öneriler.....	70
KAYNAKÇA	72
EKLER.....	79

ÇİZELGELER LİSTESİ

	<u>Sayfa</u>
Çizelge 1. Little'ın TSOK Testi.....	33
Çizelge 2. Biyolojik Sinir Ağları İle Yapay Sinir Ağlarının Karşılaştırılması...35	35
Çizelge 3. Tek Değişkende %5 Kayıp Veri Oranına Göre Hata Matrisi	51
Çizelge 4. Tek Değişkende %15 Kayıp Veri Oranına Göre Hata Matrisi.....	52
Çizelge 5. Tek Değişkende %30 Kayıp Veri Oranına Göre Hata Matrisi.....	53
Çizelge 6. İki Değişkende %5 Kayıp Veri Oranına Göre Hata Matrisi.....	54
Çizelge 7. İki Değişkende %15 Kayıp Veri Oranına Göre Hata Matrisi.....	55
Çizelge 8. İki Değişkende %30 Kayıp Veri Oranına Göre Hata Matrisi.....	56
Çizelge 9. Tek Değişkende Farklı Oranlarda Kayıp Verilere Değer Ataması Yapılmış Veri Setine Ve Gerçek Veri Setine Yapılan Faktör Analizlerine İlişkin Sonuçlar.....	59
Çizelge 10. İki Değişkende Farklı Oranlarda Kayıp Verilere Değer Ataması Yapılmış Veri Setine Ve Gerçek Veri Setine Yapılan Faktör Analizlerine İlişkin Sonuçlar.....	59
Çizelge 11. Üç Değişkende Farklı Oranlarda Kayıp Verilere Değer Ataması Yapılmış Veri Setine Ve Gerçek Veri Setine Yapılan Faktör Analizlerine İlişkin Sonuçlar.....	62
Çizelge 12. Tek Değişkende Farklı Oranlarda Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Ve Gerçek Veri Setindeki Maddelere İlişkin Düzeltilmiş Madde-Toplam Korelasyonu İle Güvenirlik Katsayıları.....	64
Çizelge 13. İki Değişkende Farklı Oranlarda Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Ve Gerçek Veri Setindeki Maddelere İlişkin Düzeltilmiş Madde-Toplam Korelasyonu İle Güvenirlik Katsayıları.....	65
Çizelge 14. Üç Değişkende Farklı Oranlarda Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Ve Gerçek Veri Setindeki Maddelere İlişkin Düzeltilmiş Madde-Toplam Korelasyonu İle Güvenirlik Katsayıları.....	66

Çizelge 15. Tek Değişkende Kayıp Veriye Sahip Veri Setleri İçin Verilerin Düzenlenmesi.....	80
Çizelge 16. İki Değişkende Kayıp Veriye Sahip Veri Setleri İçin Verilerin Düzenlenmesi.....	80
Çizelge 17. Üç Değişkende Kayıp Veriye Sahip Veri Setleri İçin Verilerin Düzenlenmesi.....	81
Çizelge 18. Gerçek Veri Setindeki Maddelere İlişkin Kovaryans Matrisi.....	82
Çizelge 19. Gerçek Veri Setindeki Maddelerin Betimsel İstatistikleri.....	82
Çizelge 20. Gerçek Veri Seti Ölçme Modelleri Uyum İyiliği Değerleri.....	82
Çizelge 21. Tek Değişkende %5 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelere İlişkin Kovaryans Matrisi....	83
Çizelge 22. Tek Değişkende %5 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelerin Betimsel İstatistikleri.....	83
Çizelge 23. Tek Değişkende %5 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Seti Ölçme Modelleri Uyum İyiliği Değerleri.....	83
Çizelge 24. İki Değişkende %5 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelere İlişkin Kovaryans Matrisi....	84
Çizelge 25. İki Değişkende %5 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelerin Betimsel İstatistikleri.....	84
Çizelge 26. İki Değişkende %5 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Seti Ölçme Modelleri Uyum İyiliği Değerleri.....	84
Çizelge 27. Üç Değişkende %5 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelere İlişkin Kovaryans Matrisi....	85
Çizelge 28. Üç Değişkende %5 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelerin Betimsel İstatistikleri.....	85
Çizelge 29. Üç Değişkende %5 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Seti Ölçme Modelleri Uyum İyiliği Değerleri.....	85

- Çizelge 30.** Tek Değişkende %15 Oranında Kayıp Verilere Değer Ataması
Yapılmış Veri Setindeki Maddelere İlişkin Kovaryans Matrisi....86
- Çizelge 31.** Tek Değişkende %15 Oranında Kayıp Verilere Değer Ataması
Yapılmış Veri Setindeki Maddelerin Betimsel İstatistikleri.....86
- Çizelge 32.** Tek Değişkende %15 Oranında Kayıp Verilere Değer Ataması
Yapılmış Veri Seti Ölçme Modelleri Uyum İyiliği Değerleri.....86
- Çizelge 33.** İki Değişkende %15 Oranında Kayıp Verilere Değer Ataması
Yapılmış Veri Setindeki Maddelere İlişkin Kovaryans Matrisi....87
- Çizelge 34.** İki Değişkende %15 Oranında Kayıp Verilere Değer Ataması
Yapılmış Veri Setindeki Maddelerin Betimsel İstatistikleri.....87
- Çizelge 35.** İki Değişkende %15 Oranında Kayıp Verilere Değer Ataması
Yapılmış Veri Seti Ölçme Modelleri Uyum İyiliği Değerleri.....87
- Çizelge 36.** Üç Değişkende %15 Oranında Kayıp Verilere Değer Ataması
Yapılmış Veri Setindeki Maddelere İlişkin Kovaryans Matrisi....88
- Çizelge 37.** Üç Değişkende %15 Oranında Kayıp Verilere Değer Ataması
Yapılmış Veri Setindeki Maddelerin Betimsel İstatistikleri.....88
- Çizelge 38.** Üç Değişkende %15 Oranında Kayıp Verilere Değer Ataması
Yapılmış Veri Seti Ölçme Modelleri Uyum İyiliği Değerleri.....88
- Çizelge 39.** Tek Değişkende %30 Oranında Kayıp Verilere Değer Ataması
Yapılmış Veri Setindeki Maddelere İlişkin Kovaryans Matrisi....89
- Çizelge 40.** Tek Değişkende %30 Oranında Kayıp Verilere Değer Ataması
Yapılmış Veri Setindeki Maddelerin Betimsel İstatistikleri.....89
- Çizelge 41.** Tek Değişkende %30 Oranında Kayıp Verilere Değer Ataması
Yapılmış Veri Seti Ölçme Modelleri Uyum İyiliği Değerleri.....89
- Çizelge 42.** İki Değişkende %30 Oranında Kayıp Verilere Değer Ataması
Yapılmış Veri Setindeki Maddelere İlişkin Kovaryans Matrisi....90

Çizelge 43. İki Değişkende %30 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelerin Betimsel İstatistikleri.....	90
Çizelge 44. İki Değişkende %30 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Seti Ölçme Modelleri Uyum İyiliği Değerleri.....	90
Çizelge 45. Üç Değişkende %30 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelere İlişkin Kovaryans Matrisi....	91
Çizelge 46. Üç Değişkende %30 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelerin Betimsel İstatistikleri.....	91
Çizelge 47. Üç Değişkende %30 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Seti Ölçme Modelleri Uyum İyiliği Değerleri.....	91
Çizelge 48. Üç değişkende %5 kayıp veri oranına göre Hata Matrisi.....	92
Çizelge 49. Üç değişkende %15 kayıp veri oranına göre Hata Matrisi.....	93
Çizelge 50. Üç değişkende %30 kayıp veri oranına göre Hata Matrisi.....	94

ŞEKİLLER LİSTESİ

	<u>Sayfa</u>
Şekil 1. Kayıp Veri Örüntüleri.....	7
Şekil 2. $m=4$ İçin Çoklu Ataması Yapılan Veri Seti.....	17
Şekil 3. Yapay Sinir Hücresinin Yapısı.....	36
Şekil 4. Doğrusal Fonksiyon.....	38
Şekil 5. Adım Fonksiyonu.....	38
Şekil 6. Lojistik Sigmoid Fonksiyonu.....	39
Şekil 7. Hiperbolik Tanjant Fonksiyonu.....	39
Şekil 8. İki Katmanlı İleri Beslemeli Yapay Sinir Ağı Mimarisi.....	40
Şekil 9. Tek Katmanlı Yapay Sinir Ağı.....	44
Şekil 10. Çok Katmanlı Ağ Modeli.....	45
Şekil 11. İleri Beslemeli Ve Geri Beslemeli Ağ Yapıları	46

BÖLÜM I

GİRİŞ

Problem

Araştırmacılar bilgiye ulaşmak ve bilgiyi işlemek için bilimsel süreci izlerler. Karasar (2010) bilimsel süreci; güçlüğün sezilmesi, problemin tanımlanması, çözümün tahmini, gözlenebilir sınıyıcıların belirlenmesi, deneme ve değerlendirmelerin yapılması ve raporlaştırma olarak ele almıştır.

Bilimsel sürecin de aşamalarından biri olan değerlendirme, gözlem sonuçlarının bir ölçüt veya ölçütlerle karşılaştırılarak bir karara varılması işidir (Turgut ve Baykul, 2010). Değerlendirme, ölçme sonuçlarından bir anlam çıkarmak ve ölçülen nesnelere hakkında bir değer yargısına ulaşmaktır. Elde edilen ölçme sonuçlarından bir anlam çıkarmak için söz konusu ölçme sonuçlarının bir ölçüt ile karşılaştırılması gerekir (Tekin, 2010). Tanımdan da anlaşılacağı üzere değerlendirme süreci ölçme sonuçları, ölçüt ve karar olmak üzere üç öğeden oluşmaktadır. Ölçme sonuçları, ölçmeye konu olan nesnelere, olayların veya bireylerin, belirli bir özelliğine karşılık gelen değerlerin ölçüsünün sayı veya sembolleri olarak düşünülebilir. Değerlendirmenin doğruluğu, ölçme sonuçlarının hatalardan mümkün olduğunca arınık olmasına, güvenilir ve geçerli sonuçlar elde edilmesine bağlıdır.

Karar için esas olan gözlem sonuçlarıdır. Eksiksiz gözlem sonuçlarına sahip olmak verilecek kararların güvenilirliği ve geçerliği açısından önemlidir. Bu nedenle bilimsel süreç için kayıp veri büyük bir öneme sahiptir. Rubin

(1987)'e göre kayıp veri, veri vektörlerinin bileşenlerinden bazılarının bütün veri maddeleri için uygun olmaması veya veri tabanında geçerli ve tanımlanmış olmamasıdır. Başka bir anlatımla ilgilenilen olay hakkında bir takım bilginin kaybolma durumudur.

Kayıp veriler istatistiksel yöntemlerin kullanılacağı durumlarda, araştırmacının kalitesini etkilediğinden araştırmacılar için problem teşkil etmektedir. Analizlerde kullanılacak verilerin niteliği ve verilerin tam olup olmaması, analiz sonucunda yapılacak olan çıkarımları etkiler (Abdella ve Marwala, 2005). McKnight, McKnight, Sidani ve Figueredo (2007) kayıp verilerin çalışma sonuçları genelleştirilmek istenildiğinde örneklemin evreni temsilliğini, müdahalelerin gücünü ve uygunluğunu kısıtlayacağını ifade etmişlerdir.

Kayıp veri sistematik gözlemlerin geçerliğini ve güvenilirliğini etkiler. Gözlemlerden çıkarımlarda bulunurken kayıp veri çalışma deseninin gücünü ve değişkenler arasındaki ilişkiler ile ilgili sonuçların geçerliğini yitirmesine neden olabilir. Kayıp verilerin, ölçme araçlarının yapı geçerliğini ve güvenilirliğini nasıl etkilediğine aşağıda kısaca bahsedilmiştir.

McKnight ve diğerleri (2007) yapı geçerliğinin; ölçme sonuçlarının gözlenmek istenilen yapıları (değişkenleri) ne kadar etkili gözler önüne serdiğini ve ölçme sonuçlarındaki puanların ilişkili değişkenleri ne kadar etkili tahmin ettiğini gösterdiğini ifade eder. Araştırmacılar bir yapıyı veya kavramı belirtmek için ölçme aracında tek bir madde veya çoklu maddeler kullanabilir. Ölçmede bu yaklaşım tek işlem (mono-operation) olarak bilinir. Örneğin sigara içme sıklığı son 24 saat içerisinde içilen sigara sayısını belirlemeye yönelik tek bir madde ile ölçülebilir. Bir birey bu madde için kayıp veriye sahipse bu birey için bu yapıya ilişkin bütün bilgi kaybolur. Bu da demektir ki bu durum için bütün bir değişken kayıptır. Modelde kayıp veriyle ilişkili diğer değişkenler olmadıkça kayıp bilgi yerine konulamayabilir. Böylece istatistiksel çıkarımlar için bir problem olur. Eğer araştırmacılar bir değişkeni ölçmek için çoklu maddelerden oluşan tek bir ölçme yaparsa başka problemler ortaya çıkabilir. Maddelerden en az biri için veri kayıpsa ilgilenilen değişkeni gözler önüne sermede ölçmenin gücü zedelenebilir. Maddeler birbirleriyle yüksek

korelasyon gösterdiği zaman kayıp madde verisi yapı geçerliği için zarar verici olmayabilir, çünkü gözlenen maddelerdeki veriler ölçmenin nasıl puanlandığına bağlı olarak kayıp olan verinin yerini tutabilir. Diğer taraftan maddeler birbirleriyle yüksek korelasyona sahip olmadıklarında kayıp veriler ölçme için zarar verici olabilir. Örneğin semptomları ölçen maddelerden oluşan bir depresyon ölçeğinde intihar eğilimi gibi bazı maddeler için kayıp veriler, depresyonun tamamlanmamış bir resmiyle ve çalışmadaki önemli bir değişkenin kusurlu temsili ile sonuçlanmaktadır.

Çoklu ölçmelerde ise amaç yapıya yaklaşımdır. Böylece ölçme süreci yoluyla yapının gözler önüne serilmesi olasılığı artar. Çoklu ölçme kullanma, sadece tek bir ölçme kullanıldığında oluşan kayıp veriden kaynaklanan problemlerden araştırmacıları korur. Aynı yapının çoklu ölçmelerinde veri bir ya da iki ölçmede kayıp olsa bile diğer ölçmeler bu yapı hakkında bilgi sağlayabilir. Kayıp veri problemi tek bir yapının çoklu ölçme kullanıldığı durumlarda da bulunmaktadır. Örneğin ölçme sürecine katılanlar, yasal olmayan maddeleri kullanma gibi kişisel bilgileri elde etmeyi hedefleyen maddelere cevap vermek istemeyebilirler. Madde bağımlılığı ile ilgili ölçme yapılmak istenildiğinde bütün ilişkili ölçmeler bu katılımcılar için bu bilgiyi kaybeder. Çoklu ölçmeler, bir yapının çoklu ölçmesini yapabilmesi ve veri kaybı problemini azaltabilmesi nedeniyle tek bir ölçme kullanılmasına kıyasla daha çok tercih edilirler. Çünkü bu ölçmelerde verinin çoklu kaynağı vardır ve bütün bu kaynakların kaybolması pek olası değildir.

Çoklu maddelerden oluşan tek bir ölçme ile çoklu ölçmeler için kayıp veri, iç tutarlılığın azalmasına neden olabilir. İç tutarlılıktaki azalma maddelerin sayısının azalması veya analiz için yararlı olan bilginin atılmasından (depresyonu ölçmede intihar eğilimi hakkında bilgi olmaması gibi) kaynaklanabilir (McNight ve diğerleri, 2007). Klasik test teorisine göre güvenilirlik indeksi gözlenen puanlarla gerçek puanlar arasındaki korelasyonun karesinin gerçek puanlar varyansının gözlenen puanlar varyansına oranına eşittir (Baykul, 2010).

$$p^2(X,T) = \frac{\sigma^2(T)}{\sigma^2(X)} = 1 - \left(\frac{\sigma^2(E)}{\sigma^2(X)} \right)$$

X: Gözlenen puanlar

T: Gerçek puanlar

E: Hata puanları

Eşitlik dikkate alındığında hata varyansının artmasının güvenilirliği düşürdüğü görülmektedir. Kayıp verilerde de bilgi kaybı hata varyansını artırmakta ve bu da güvenilirliğin düşmesine neden olmaktadır.

Klasik test teorisine göre ölçme aracının iç tutarlılık güvenilirliği madde sayısı ile güçlü bir ilişki içerisinde. Spearman Brown düzeltme formülü, madde sayısının artmasıyla güvenilirliğin artacağını gösterir. Ölçme sonucunda elde edilen kayıp veriler ölçmenin güvenilirliğini de azaltmaktadır. Modern test teorisi ise madde sayıları ile ölçmenin güvenilirliği arasında belirli bir ilişkinin olmadığını ifade eder. Her maddenin gerçek performansı ve maddeler arasındaki ilişki ölçmenin güvenilirliğini belirler. Bu nedenle ölçme aracının iç tutarlılığı madde varyanslarına ve kovaryanslarına yüksek derecede bağımlı iken madde sayısına çok daha az derecede bağımlıdır. Modern test teorisine göre yüksek güvenilir maddeden oluşan az maddeli bir ölçme aracı, çok madde içeren fakat çok zayıf maddelerle yapılan bir ölçme aracından daha güvenilir olmaktadır. Özet olarak, hem klasik test teorisi hem de modern test teorisi ölçme güvenilirliğinin kayıp veri tarafından tehdit altında olduğunu kabul eder.

Allison (2002) çeşitli nedenlerden dolayı kayıp verinin, sağlık ve sosyal bilimlerde yaygın bir problem olduğunu ifade eder. Kayıp verinin bu alanlarda problem olmasını da iki nedene bağlar. Bu nedenlerden ilki neredeyse bütün istatistik yöntemlerinin, her durumun analizlerde bulunan bütün değişkenlerdeki bilgilere sahip olduğunu farz etmesidir. Standart istatistiksel yöntemler ve programlar tam bilgi durumu için düzenlendiklerinden eksik veri içeren durumları analiz dışı bırakarak yanlış tahminlere neden olmaktadır. İkinci neden ise istatistik ders kitaplarının büyük bir çoğunluğunun kayıp

veriler hakkında veya kayıp verilerin nasıl üstesinden gelinmesi gerektiği hakkında yeteri kadar bilgi vermemesinden kaynaklanmaktadır.

McKnight ve diğerleri (2007), istatistik literatüründe kayıp veri problemine değinilse de araştırma yöntem bilimcileri ve istatistik uzmanı olmayan sosyal bilimciler için Allison'un (2002) da belirttiği gibi kayıp verinin hala bir bilmece halinde olduğunu belirtmektedir. Bu durumun nedeni olarak da araştırma yöntem bilimcileri ve istatistik uzmanı olmayan sosyal bilimcilerin kayıp veri üzerine istatistik literatürüne yeteri kadar aşına olmamaları gösterilmektedir. Bu aşinalığın olmamasının olası nedenlerinden biri de matematiksel gösterimlerle ifade edilen tanımlar, teoremler ve ispatları içeren teknik literatürü anlamak için gerekli olan matematik ve istatistikte uzmanlık seviyesindeki eksikliklerdir. Diğer bir neden ise kolay ve kullanışlı programların eksikliğidir. Çoğu istatistik programı kayıp verilere değer atama prosedürlerini içerse de bu prosedürlerin kullanımı uzman olmayan veri analizi uzmanları için zordur. Bu nedenle çoğu araştırmacı tarafından kolayca benimsenmemektedir.

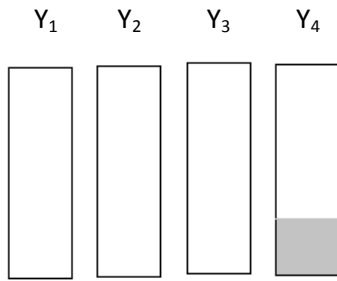
Kayıp verinin miktarı, kayıp veri örüntüsü, kayıp veri mekanizmaları ve kayıp verilere değer atama yöntemlerinin seçimi araştırma sonuçlarının yorumlanması üzerinde önemli etkilere sahiptir (Little ve Rubin, 2002). Kayıp veri probleminin üstesinden gelmek için bu dört durumun da ele alınması gerekmektedir.

Kayıp veri miktarı için Acuna ve Rodriguez (2004) literatüre göre %1'den az veri kaybının genellikle önemsiz olarak düşünüldüğünü, %1-5 arası veri kaybında kayıp verilere değer atamasının yapıp yapılmamasının yorumları çok da etkilemeyeceğini, fakat kayıp verinin %5-15 arasında olduğu durumlarda kayıp verilere değer atama için çeşitli karmaşık yöntemlerin gerektiğini ve %15'den daha fazla olması durumunda ise kayıp verinin yapılacak herhangi bir yorumu çok ciddi şekilde etkileyeceğini belirtmişlerdir. Buradaki miktar kelimesi kayıp veri literatüründe belirsizdir. Kayıp veri miktarı ifadesi tam durum yöntemi olarak bilinen yöntemde tamamlanmamış durumların yüzdesini ifade ederken tam değişken yöntemi olarak bilinen yöntemde kayıp veriye sahip değişkenlerin yüzdesini ifade eder.

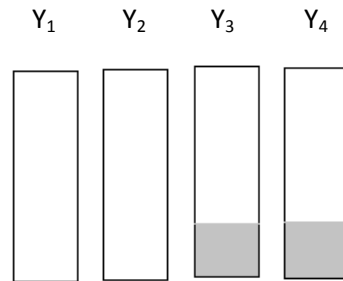
Mevcut durum yöntemi olarak bilinen yöntemde kayıp veri miktarı her bir değişken için kayıp veri yüzdesi anlamına gelirken, Seyrek Matris (Sparse Matrix) yöntemi olarak bilinen yöntemde ise verinin toplam matristeki kayıp verilerin yüzdesi anlamına gelmektedir. Bu dört durumu da göz önünde bulundurmak gerekir (McKnight ve diğerleri, 2007).

Kayıp veri örüntüsü, veri setinde gözlenen ve kayıp değerlerin düzenlenişini ifade eder. Kayıp veri örüntüsü verideki kayıpların yerlerini resmeder; fakat verinin neden kaybolduğunu açıklamaz (Enders, 2010). Örüntü, uygun veri atama yöntemini seçmek için veri setinin yapısı hakkında bazı ön bilgiler verir. Little ve Rubin (2002)'e göre kayıp veri örüntüsü veri setindeki gözlenen değerlerin ve kayıp değerlerin şemasını ifade eder ve uygun veri atama yöntemini seçmek için veri setinin yapısı hakkında bazı ön bilgiler verir.

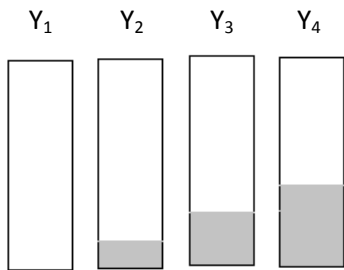
Kayıp veri literatüründe karşılaşılabilecek olan altı kayıp veri örüntüsü Şekil 1'de gösterilmiştir.



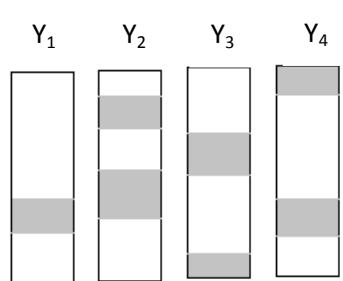
(a) Tek değişken örüntüsü



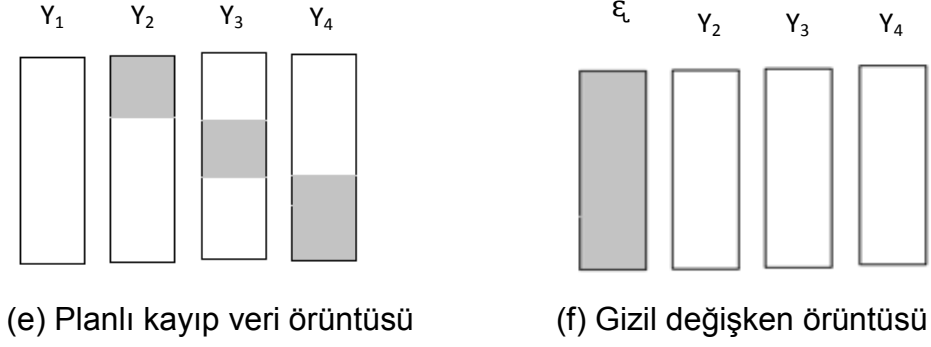
(b) Birim yanıtlamama örüntüsü



(c) Monoton kayıp veri örüntüsü



(d) Genel kayıp veri örüntüsü



Şekil 1. Kayıp veri örüntüleri (Enders, 2010)

Tek değişken örüntüsünde (a) sadece tek bir cevap değişkeninde kayıp verilerin olması durumunu gösterir ve bu örüntüye genelde deneysel çalışmalarda karşılaşılar. Birim yanıtlamama örüntüsü (unit nonresponse pattern) (b) örnekleme çatısındaki her cevaplayıcı için var olan özellikleri gösteren Y_1 ve Y_2 değişkenleri her cevaplayıcı tarafından cevaplanırken Y_3 ve Y_4 değişkenlerini bazı cevaplayıcıların cevaplamayı reddetmeleri durumunu gösterir ve genelde tarama araştırmalarında karşılaşılar. Monoton kayıp veri örüntüsüne (c) ise katılımcıların ayrıldığı ve geri dönmeyecekleri boylamsal çalışmalarda karşılaşılar. Genel kayıp veri örüntüsünde (d) kayıp veriler rastgele olarak veri matrisinde dağılmıştır. Seçkisiz dağılım görüntüsü aldatıcı olabilir ve değerler aslında sistematik olarak kayıp olabilirler. Planlı kayıp veri örüntüsü (e) üç formlu anket deseniyle benzerlik göstermektedir. Bu anket deseninin altında yatan temel düşünce anketleri farklı formlarda dağıtmak ve her bir cevaplayıcı için formların bir alt kümesini uygulamaktır. Planlı kayıp veri örüntüsü cevaplayıcının yükünü azalttığından geniş sayıda anket maddelerini toplamada kullanışlıdır. Gizil değişken örüntüsü (f) yapısal eşitlik modeli gibi gizil değişken analizleri için önemlidir. Gizil değişkenin değerleri bütün örneklem için kayıptır (Enders, 2010).

Kayıp veri örüntülerinde veri kaybının nedenleri seçilecek olan kayıp verilere değer atama yöntemlerini etkiler. Bir veri tabanında kayıp veri çeşitli nedenlerle ortaya çıkabilir. Kayıp verinin üstesinden gelmek için seçilecek olan kabul edilebilir yol verinin nasıl kaybolduğuna bağlıdır (Abdella ve Marwala, 2005). McNight ve diğerleri (2007) bu nedenleri üç esas kategoride sınıflandırmışlardır. Bunlar: çalışmaya katılanlar, çalışmanın dizaynı ve

katılanlarla çalışmanın dizaynının etkileşimidir. Çalışmaya katılanlardan bazıları belirli sorulardan alınganlık göstermiş olup bazı maddeleri cevaplamamış olabilirler (çalışmaya katılanların karakteristiği), çalışma katılımcıların çok fazla zamanını alabilir (dizaynın karakteristiği) ya da katılımcılardan hasta olanlar olabilir ve kendilerine sıkıcı gelen maddeleri cevaplamayabilir (hem katılanın hem de dizaynın karakteristiği). Veri kaybının nedenleriyle ilgili matematiksel ilişki kayıp veri mekanizmaları ile elde edilebilir.

Kayıp veri mekanizmaları da kayıp veri miktarı ve kayıp veri örüntüleri gibi kayıp verinin üstesinden gelmek için kullanılacak uygun yöntemi belirlemede önemli rol oynar. Araştırmacılar kayıp veri örüntüleri ile kayıp veri mekanizmaları ifadelerini bazen birbirlerinin yerine kullansalar da aslında birbirlerinden farklı anlamlara sahiptir. Kayıp veri örüntüleri verideki kayıpların yerlerini resmederken, kayıp veri mekanizması ölçülen değişkenlerle kayıp verinin olabilirliği arasındaki olası ilişkileri betimler (Enders, 2010).

Kayıp veri mekanizmaları kayıp verilere değer atama yöntemlerinin özelliklerinin bu mekanizmalardaki bağımlılığın doğasına güçlü bir şekilde bağlı olması nedeniyle önemlidir. Kayıp değerlerle birlikte veri analizinde mekanizmanın rolü, Rubin (1976; akt. Little ve Rubin, 2002)'in teorisinde seçkisiz değişkenler gibi kayıp veri göstergelerini işleme koyan ve bir dağılıma atayan bir araç yoluyla formüle edilinceye kadar ihmal edilmiştir. Little ve Rubin (2002) kayıp veri mekanizmalarını üç kategoride tanımlamıştır: Tamamıyla seçkisiz olarak kayıp - TSOK (Missing completely at random-MCAR), seçkisiz olarak kayıp - SOK (Missing at random-MAR), ihmal edilemez kayıp - İEK (Nonignorable-NI).

Tamamıyla seçkisiz olarak kayıp mekanizması, kayıp veri olasılığının gözlenen ya da kaybolan değişkenlerin değerleriyle ilişkisi olmadığı durumları ifade eder. Bu da demektir ki kayıp olma durumu ilgilenilen değişkene ve diğer değişkenlerin değerlerine bağlı değildir (Little ve Rubin, 2002). TSOK mekanizmasına örnek olarak anket çalışmasında anketörün herhangi bir soruyu sormayı unutması, bir koşu yarışmasında atletin kural ihlali nedeniyle

bitiş görememesi gibi örnekler verilebilir (Aslan, Yozgatlıgil, İyigün, Batmaz ve Tatlı, 2010).

Seçkisiz olarak kayıp mekanizması, kayıp veri olasılığının kaybolan değişkenin değeriyle ilişkili olmadığı fakat diğer gözlenen değişkenlerle ilişkili olduğu durumları ifade eder. Diğer bir deyişle kayıp veri dışsal etkenlerden dolayı kayıptır, kendi değişkeninden dolayı değildir. SOK mekanizması diğer gözlenen değişkenler tarafından tahmin edilebilir. (Nelwamondo, 2008). Kaplan ve Turner (2012) SOK mekanizması için şöyle bir örnek vermişlerdir: Anket çalışmasında gelir sorusuna ilişkin cevaplar düşünölsün. Yaşlı bireyler gelirlerini doğru cevaplamamak eğiliminde olabilirler. Bu örnekte kişinin geliri sorusu karşılığında verilen cevaplardaki kayıplar kişinin yaşına bağlıdır, gelir miktarına bağlı değildir. Bir diğer örnek, gelir durumunu beyan etmenin medeni duruma bağlı, ama gelirin kendi değeriyle bağlı olmaması durumu verilebilir. (Little ve Rubin, 2002).

İhmal edilemez kayıp mekanizması kayıp veri olasılığı kaybolan değişkenin değeriyle ilişkili olduğu durumları ifade eder (Allison, 2002). Diğer bir deyişle kayıp veri seçkisiz değildir ve kaybolan değişkene bağlıdır. Diğer değişkenler tarafından tahmin edilemez. Örneğin gelir beyanını isteyen bir değişkene yüksek gelirli aileler cevap vermek istemeyebilir ve o değişken için yüksek gelirli ailelerde kayıp veriler bulunur (Little ve Rubin, 2002). Ölçme araçlarının belli ölçme sınırları olabilir. Güneş radyasyonunu ölçen aracın hassasiyet eşliğini aşan radyasyon değerlerinde ölçme yapamamasından kaynaklanan kayıp gözlem örneği bu mekanizma altında değerlendirilebilir. Radyasyon gözlemindeki kayıp verinin kendisi yine radyasyon miktarına bağlı olarak oluşmaktadır (Aslan ve diğerleri, 2010).

Kayıp verinin miktarı, örüntüsü ve kayıp veri mekanizmalarına bağlı olarak **kayıp verilere değer atama yönteminin** seçilmesi gerekebilir. İstatistik paketlerinde kullanılan çeşitli kayıp verilere değer atama yöntemleri vardır (Yansaneh ve diğerleri, 1998; akt. Abdella, 2005). Bazı yöntemler diğerlerinden daha üstün olmasına rağmen hiçbir yöntemin daha üstün olduğu ifade edilemez. Bu nedenle araştırmaların her adımında (dizaynında,

uygulanmasında gibi) kayıp verinin en az olması için büyük bir çaba harcanması ve özveri gösterilmesi gerekmektedir (Allison, 2002).

Kline (1998) kayıp veri probleminin üstesinden gelmek için en çok kullanılan yöntemleri üç kategoride ele almıştır. Bunlar: **Silme Yöntemleri**, **Değiştirme Tabanlı (replacement-based) Yöntemler** ve **Model Tabanlı Yöntemler**dir. Kayıp veri miktarının %5'ten az olduğu durumlarda TSOK, SOK ve İEK mekanizmaları altında silme yöntemleri kullanılır. Kayıp veri miktarı dikkate değer bir oranda ise (%5'ten fazla) TSOK mekanizmaları altında değiştirme tabanlı ya da model tabanlı yöntemler kullanılırken, SOK mekanizması altında daha çok model tabanlı yöntemler kullanılmaktadır (Kaplan ve Turner, 2012; McKnight ve diğerleri, 2007). İEK mekanizması altında kayıp veri miktarı önemsiz değilse (%5'ten fazla) kullanılacak yöntemler kuşku uyandıracak niteliktedir (McKnight ve diğerleri, 2007). Bu yöntemler aşağıda belirtildiği gibi gruplandırılabilir:

1. Silme Yöntemleri
 - 1.1. Dizin Bazında Silme
 - 1.2. Çiftler Bazında Silme
2. Değiştirme Tabanlı Yöntemler
 - 2.1. Ortalama Atama Yöntemi
 - 2.2. Sıfır Atama Yöntemi
 - 2.3. Hot Deck Atama Yöntemi
 - 2.4. Cold Deck Atama Yöntemi
 - 2.5. Regresyon Atama Yöntemi
3. Model Tabanlı Yöntemler
 - 3.1. En Çok Olabilirlik Yöntemi
 - 3.1.1. Beklenti Maksimizasyonu Yöntemi
 - 3.1.2. Tam Bilgi Maksimum Olabilirlik Yöntemi
 - 3.2. Çoklu Atama Yöntemi
 - 3.2.1. Yapay Sinir Ağları

Kline (1998) veri atama yöntemlerinden **silme yöntemini** dizin bazında ve çiftler bazında silme olarak iki şekilde ele almıştır. *Dizin bazında silme yönteminde* eğer herhangi bir durum değişkenlerden herhangi biri için kayıp veriye sahipse analizde bu durum hesaba katılmaz. Sonuçta veri seti artık kayıp veriye sahip olmamaktadır ve bilinen yöntemlerle analiz edilebilir. Bu strateji sosyal bilimlerde durum bazında silme olarak da bilinir. Aynı zamanda tam gözlemler analizi adıyla da karşılaşılmaktadır. Ancak bu yöntemin en büyük dezavantajı örneklemin büyük bir oranını yok saymasıdır (Allison, 2002). Dizin veya durum bazında veri silme kayıp veri problemini çözmenin en kolay ve en basit yoludur fakat aynı zamanda da problemi çözme yöntemlerinin en uygun olmayan seçimidir (Abdella, 2005). Bu yöntem sadece kayıp veri, veri setindeki tam verilerle karşılaştırıldığında çok küçük bir oran içeriyorsa kullanılması uygundur (Little ve Rubin, 2002). Özel olarak eğer veri TSOK ise azaltılan (kayıp veri içeren gözlemlerden arındırılmış) örneklem, orijinal örneklemin seçkisiz bir alt örnekleme olmaktadır. Yani kayıp veri içermeyen tüm veri seti için hesaplamalar yanlı değilse, bu yöntemle kayıp veri içeren gözlemler silindiğinde hesaplamalar yine yansız olmaktadır. Ayrıca, dizin bazında silme yöntemi uygulanmış veri setinden elde edilen standart hatalar ve test istatistikleri tam veri setinden hesaplanan standart hatalar ve test istatistikleri gibi kabul görür. Fakat dizin bazında silme yöntemi uygulanmış veri setinden elde edilen standart hatalar daha az veri kullanıldığından daha büyük olmaktadır (Allison, 2002).

Çiftler bazında silme yöntemi ise eldeki tüm değerleri kullanarak ortalamalar, standart sapmalar ve korelasyon matrisi gibi istatistikleri hesaplamak içindir. Veri seti TSOK ise çiftler bazında silme yöntemi tutarlı parametre hesaplamalarında bulunur ve böylece geniş örneklemelerde yanlı olmayan sonuçlar verir (Allison, 2002).

Veri TSOK olduğunda çiftler bazında silme yöntemi dizin bazında silme yöntemine göre daha fazla bilgi kullandığından daha etkili bir yöntem olması beklenir. Bu yöntemle hesaplamalarda örneklem değişkenliği dizin bazında silme yöntemine göre daha azdır, bu durum tahminlerin daha küçük standart hataya sahip olması olarak da yorumlanabilir. Ancak bu durum her zaman böyle olmayabilir. Doğrusal regresyon modellerinin analitik ve

benzetim çalışmalarında, çiftler bazında silme yöntemi değişkenler arasındaki korelasyonun düşük olduğu durumlarda daha etkili sonuçlar verirken dizin bazında silme yöntemi değişkenler arasındaki korelasyonun yüksek olduğu durumlarda daha etkili sonuçlar vermektedir (Glasser, 1964; Haitovsky, 1968; Kim ve Curry, 1977; akt. Allison, 2002).

Çiftler bazında silme yöntemi ile kovaryans matrisi oluşturulurken, farklı sayıda gözlemlere sahip değişken çiftleri için varyans ve kovaryans değerleri hesaplanmasına bağlı olarak yapısal eşitlik modellemelerinde karşılaşılan belirsizlik problemi (indefiniteness problem) ortaya çıkabilmektedir (Wothke, 1993; akt. Dural, 2010). Belirsizlik problemi, çiftler bazında silme yöntemi kullanılarak iki değişken arasında kabul edilebilir olmayan bir korelasyon değeri (-1'den küçük ya da 1'den büyük) elde edildiğini gösterir. Bu durum çiftler bazında silme yönteminin en önemli sınırlılığı olarak değerlendirilmektedir (Cheung, 2007; akt. Dural, 2010).

Diğer bir yöntem de **değiştirme tabanlı yöntemlerdir**. Değiştirme tabanlı yöntemlerin ana odağı tamamlanmamış veri setini tamamlamaktır. Değiştirme tabanlı yöntemler içerisinde en çok kullanılan yöntemler Ortalama atama, Sıfır atama, Hot Deck atama, Cold Deck atama, Regresyon atama yöntemleridir.

Ortalama atama yöntemi en basit atama yöntemlerinden biridir ve TSOK mekanizması altında çalışır. Araştırmanın özelliğine ve araştırmacının öngörüsüne göre kayıp değerler; durum ortalamasının atanması, alt grup ortalamasının atanması ve toplam ortalamasının atanması ile belirlenebilirler. Durum ortalamasının atanmasında kayıp değerler, cevaplayıcının kayıp değere sahip olmayan maddelerinin ortalamaları ile belirlenir. Alt grup ortalamasının atanmasında kayıp değerler, kayıp değere sahip olan cevaplayıcının içerisinde bulunduğu alt grubun ortalamasıyla belirlenir. Toplam ortalamasının atanmasında ise kayıp değerler, maddeyi cevaplayanların madde ortalamaları ile belirlenir (Tsikriktsis, 2005).

Kayıp verilere kendi grubunun ortalamasıyla değer atama bu gözlemdeki verinin sistematik olarak kendi grubunun özelliklerini taşıması ve

kendi grubu için bir temsilci olarak düşünölebileceđi varsayımına dayanır. Kayıp verileri diđer grubun ortalaması ile atama ise sonuçlar üzerinde mevcut durumu korumayı sağlar ve genel ortalamayı deđiřtirmez (Unnebrink ve Windeler, 2001; akt. Baygöl, 2007).

Ortalama atama yönteminin bazı dezavantajları vardır. Çiftler bazında silme gibi ortalama atama yönteminin de yanlı hesaplamalar üretmesi yüksek olasılıklıdır (Allison, 2002). Ortalama atama yöntemi, kayıp veri sayısının artışıyla daha da artan varyans yanlılığına liderlik eder. Bu yöntemle örnek hacim büyür, deđiřkenlerdeki tüm eksik gözlemlerin yerine ortalama deđer atandıđı için deđiřkenler gerçek dağılımdan uzaklaşır ve varyans olduđundan daha düşük çıkar, deđiřkene ait dağılım ortalamaya eşit olan deđerlerin eklenmesiyle deđiřeceđinden atama sonrasında elde edilen yeni deđiřkenin dağılımı toplumu yeterince yansıtamaz (Little ve Rubin, 2002).

Sıfır atama yöntemi kayıp veri yerine sıfır deđerini koymak olarak tanımlanabilir ve sosyal bilimlerde ve psikometrik çalışmalarda ortak olan bir uygulamadır. Bu yöntemin sıfırın makul deđer olduđu durumlarda kullanılması uygundur. Örneđin bir başarı testinde boş bırakılan maddeler için kullanılması gibi. Sıfırın buradaki anlamı ölçme sonucunda ya da yapıda bireyin zayıf olduđudur. Kayıp veri zayıf çıktıların göstergeleri olduđu zaman sıfır atamanın analiz sonuçları üzerinde etkisi olmamaktadır. Aksi takdirde örneđin bir çalışmada deney grubuna göre kontrol grubundaki bireylerin büyük bir oranı çalışmadan ayrılırsa kontrol grubu için çıktı ölçme sonuçlarının ortalaması sıfır deđerlerinden etkileneceklerdir. Böylece müdahale etkisi olduđundan daha fazla görünebilir (McNight ve diđerleri, 2007).

Hot Deck atama yönteminde kayıp deđer içeren duruma en çok benzeyen durum belirlenir ve kayıp durumun deđer için en benzer durumun deđer atanır (Allison 2002). Kayıp veri TSOK olmadıkça yanlı sonuçlar verecektir.

Hot Deck atama yönteminin avantajları yapısal sadeliđi, sürekli deđiřkenin sürekli olarak kategorik deđiřkenin kategorik olarak kalması

yoluyla deęişkenlerin ölçme seviyelerini muhafaza etmesi ve atama sürecinin sonunda bir tam verinin analiz edilebilirlięi gibi kullanılabilirlięidir (Roth 1994; Rubin 1978; Hu *et al.* 1998; akt: Abdella, 2005). Hot Deck atama yönteminin dezavantajı ise benzerlięi tanımlamadaki zorluktur. Bu bağlamda benzerlięin ne olduęunu tanımlamak için birçok yol olabilir (Hu *et al.* 1998; akt: Abdella, 2005). Hot Deck algoritması tek bir benzer kayıttan daha fazlasını belirleyebilir ve sonrasında kayıp veriye deęer atamak için bunlardan birini seçkisiz olarak seçebilir veya uygun görülen deęerlerin ortalama deęerini kullanabilir (Little ve Rubin, 2002). Cox ve Folsom (1978; akt. Uęuz, 2007), Hot Deck atama yönteminin dizin bazında silme yönteminden daha üstün sonuçlar verdięini ifade etmişlerdir.

Cold Deck atama yönteminde bir maddedeki kayıp deęerler dış kaynaklardan sabit bir deęer alır. Bu deęer önceki gözlemlerdeki deęerlerden olabilir (Nelwamondo, 2008). Bu yöntem de TSOK mekanizması altında çalışır. Bu yöntemin ortalama atama ve Hot Deck yönteminden farkı atanacak deęerin kaynağıdır. Dış kaynaklardan elde edilen deęer, eldeki veriler yardımıyla elde edilen ortalama deęerden daha geçerli olmaktadır. Bu yöntem, ortalama atama yöntemindeki dezavantajlara sahiptir (Alpar, 2003).

Regresyon atama yönteminde verilen bir deęişken için tam durum verilerine dayalı bir regresyon eşitlięi geliştirilir. Kayıp veri bağımlı deęişken gibi işlem görürken veri tabanındaki diğer deęişkenler yordayıcı olarak kullanılır. Regresyon eşitlięi ile kayıp deęerler tahmin edilmeye ve yaklaşık bir deęer atamaya çalışılır (Little ve Rubin, 2002). Uygun regresyon atama yöntemine karar verme kayıp deęişkene güçlü şekilde baęlıdır (He, 2006; akt. Nelwamondo, 2008). Kayıp deęişkenlerin doğrusal regresyonu gözlenen veriye dayalı olarak hesaplanır. Kayıp veri doğrusal regresyondan elde edilen deęerlerle atanır. Bu yöntem yine de verinin varyans ve kovaryansını düşük tahmin eder (Little ve Rubin, 2002). Deęişkenler arasındaki ilişkiden yola çıkıldığı için atama sonucunda zaten var olan ilişki daha da güçlenmiş olur, bu yöntemin gereęinden fazla kullanılması elde edilen verinin kendine özgü bir yapıya ulaşmasına neden olur ve genellenebilirlięi azalır (Hill, 1997). Regresyon tahmininde hata terimi de modelin içerisinde yer alabilir ve hata rastgele seçilmiş tam veri setinden elde edilmiş artık deęerler olabilir (Bal,

2003). Bu regresyona stokastik regresyon da denilmektedir. Hata teriminin eklenmediği regresyon eşitliği TSOK mekanizması altında çalışırken, stokastik regresyon TSOK ve SOK mekanizması altında çalışır. Silme ve değiştirme tabanlı yöntemler TSOK mekanizması altında çalışırken model tabanlı yöntemler SOK mekanizması altında da çalışabilmektedir.

Model tabanlı yöntemler içerisinde en çok kullanılan yöntemler en çok olabilirlik ve çoklu atama yöntemleridir.

En çok olabilirlik yöntemi kayıp veriyi elde etmek için en popüler model tabanlı yöntemdir. Model tabanlı yöntemler gözlenen veriden, gözlenen değişkenler arasındaki ilişkilerden ve modelin altında yatan dağılımdan yararlanarak parametre hesaplamalarını oluşturmak için kullanılır. Bu yöntemlere model tabanlı ismini veren bu altında yatan dağılımsal modeldir ve altında yatan dağılımın normal olduğu varsayılır (McNight ve diğerleri, 2007).

En çok olabilirlik yöntemi atama yöntemleri gibi kayıp veri için özellikle dizayn edilmiş bir yöntem değildir. Hiyerarşik doğrusal model ve yapısal eşitlik modeli gibi çeşitli modellerdeki parametreleri hesaplamak için kullanılan bir yöntemdir. Bu yöntemin avantajı kayıp veri TSOK veya SOK mekanizması olduğunda geniş örnekleme yanlı olmayan hesaplamalar üretmesidir (McNight ve diğerleri, 2007).

En çok olabilirlik yönteminin temel prensibi gözlenen veriyi elde etme olasılığını maksimize eden değerlerle hesaplamaları seçmektir. Bilinmeyen parametreler ve verinin bir fonksiyonu olarak, verinin olasılığını hesaplayan bir formül kullanılır. Kayıp veri mekanizması TSOK veya SOK mekanizması olduğunda olabilirlik, kayıp verinin mümkün bütün değerleri üzerindeki olasılıkları hesaplama yoluyla elde edilebilir (Allison, 2002). Olabilirlik yöntemleri örnek verinin yeterince büyük olduğu varsayımını kabul eder. Bu varsayımın altında da yanlı hesaplamalar olmadığı ve hesaplamaların normal dağıldığı kabul edilir. Bu durum da olabilirlik yönteminin dezavantajı olarak gösterilebilir (Nelwamondo, 2008).

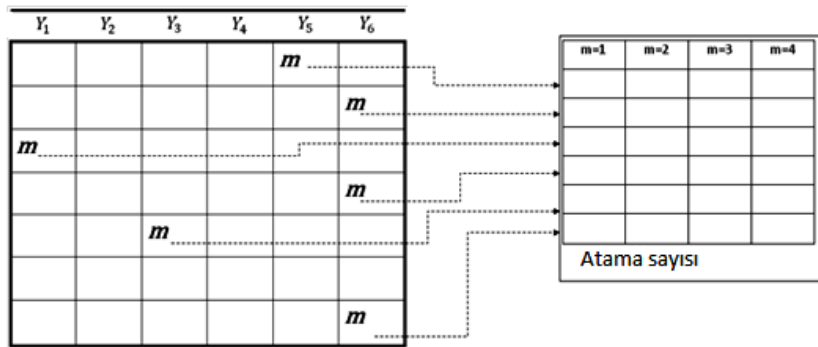
En çok olabilirlik yönteminin beklenti maksimizasyonu ile tam bilgi maksimum olabilirlik yöntemi olmak üzere iki alt yöntemi bulunmaktadır. Beklenti maksimizasyon algoritması Dempster, Liard ve Rubin (1977) tarafından tanıtılmıştır. Beklenti maksimizasyonu yöntemi, istatistiksel yöntemi algoritmik uygulamayla birleştirmiş ve böylece çeşitli kayıp veri problemlerinde daha fazla dikkat çekmeyi başarmıştır. Beklenti maksimizasyonunun sadece tamamıyla seçkisiz olarak kayıp veri mekanizmaları yerine hem tamamıyla seçkisiz olarak kayıp veri hem de seçkisiz olarak kayıp veri mekanizmaları altında uygulanabilir olması nedeniyle silme ve değiştirme tabanlı yöntemlerinden daha üstün olduğu ileri sürülmüştür (Allison, 2002).

Beklenti maksimizasyonun temel adımlarında, ilk olarak en çok olabilirlik fonksiyonu kullanılarak kayıp değerler için değerler atanır. Bu atanmış veriye göre parametreler hesaplanır. Parametre hesaplamalarına dayanarak tekrardan kayıp veri atanır. Atanmış veri setine dayanarak parametreler tekrar hesaplanır. Bu süreç hesaplanan parametreler arasındaki farkın en az olduğu duruma kadar devam eder. Bir başka anlatımla, beklenti maksimizasyonu algoritması tamamlanmamış veri için model kurar ve bir veri modelinin bilinen parametreleri ile kayıp veri arasındaki ilişkiden yararlanır. Eğer kayıp değerler bilinirse sonrasında model parametresini hesaplamak basit olmaktadır. Benzer şekilde veri modelinin parametreleri bilinirse sonrasında kayıp veriler için yansız tahminler elde etmek mümkündür. Model parametreleri ile kayıp değerler arasındaki bu bağımlılık bir tekrarlı yöntem önerir. Bu süreç yakınsama olana kadar tekrarlanır. Basit ifadelerle beklenti maksimizasyonu bir yineleme prosedürü kullanarak çalışır (Little ve Rubin, 2002).

Tam bilgi maksimum olabilirlik yöntemi ise ham maksimum olabilirlik (raw maximum likelihood) olarak da bilinir ve Finkbeiner (1979) tarafından geliştirilmiştir (Enders, 2010). Tam bilgi maksimum olabilirlik yöntemi; dizin bazında silme, çiftler bazında silme ve ortalama atama gibi genelde kullanılan kayıp verinin üstesinden gelme yöntemleriyle üretilen değişkenler arasındaki kovaryans matrisinden ve ortalamalar vektöründen daha gerçekçi

kovaryans matrisi ve ortalamalar vektörü oluşturur. Tam bilgi maksimum olabilirlik yöntemi regresyon modelleri ve yapısal eşitlik modelleri gibi diğer doğrusal modellerle uyum sağlamak için de uygulanabilir. Bu yöntemin kullanımının kolay olması ona büyük bir avantaj sağlar. Diğer bir avantajı ise standart hataların doğrudan hesaplanmasını sağlamasıdır. En büyük dezavantajı ise analizde kullanılan değişkenlerin çok değişkenli normallik varsayımını sağladığını kabul etmesidir (Schafer, 1997; akt. Nelwamondo, 2008). Çiftler bazında silme yönteminde olduğu kadar sık olmamakla birlikte, tam bilgi maksimum olabilirlik yöntemi de bazen belirsizlik problemine yol açabilmektedir (Wothke, 2000; akt: Dural, 2010).

Çoklu atama yöntemi ise Rubin (1987) tarafından geliştirilmiştir ve veri setindeki kayıpların olasılıksal dağılımını yansıtan iki ya da daha fazla kabul edilebilir değerle tamamlanması prosedürüdür (Aslan, 2010). Çoklu atama yöntemi m tekrar sayısı ve $m > 1$ (2 ile 10 arasında) olmak üzere eksik verilerin yerine türetilmiş verilerin kullanıldığı bir Monte Carlo yöntemidir (Molina, Méndez ve Echeverría, 2008). Şekil 2'de $m=4$ için çoklu ataması yapılan veri seti örneği gösterilmiştir.



Şekil 2. $m=4$ için çoklu ataması yapılan veri seti

Çoklu atama yöntemi yoluyla istatistiksel çıkarıma üç adımda gerçekleşir:

Atama: Veri setindeki kayıplar m kez tahmin edilerek m adet tam veri seti oluşturulur.

Analiz etme: m adet tam veri seti standart analiz yöntemleriyle analiz edilir.

Biraraya getirme: m adet analizde elde edilen sonuçlar birleştirilerek çıkarsama yapılır (Aslan ve diğerleri, 2010; Bal, 2003).

Çoklu atama yöntemi genelde Hot Deck atama yöntemiyle ilişkilendirilir. Hot Deck atama yönteminden üstünlüğü tek bir atama değerinden daha fazlasını oluşturma yeteneğidir. Bu da çoklu atama yöntemini, analizlerde kullanılan değişkenlerin normal dağılıma sahip olmama durumlarında diğer yöntemlerden güçlü kılar (He, 2006). Bu yöntemin dezavantajı ise kayıp değerlerin atamasında harcanan zaman miktarıdır (Rubin, 1987). Yapay sinir ağlarıyla atama da bu yöntemin içerisinde yer alır.

Yapay sinir ağları kısaca beyindeki sinirlerin çalışmasından esinlenilerek sistemlere öğrenme, hatırlama, bilgiler arasında ilişkiler oluşturma gibi yetenekleri kazandırmayı amaçlayan bilgi işleme algoritmalarıdır (Taşgetiren, 2006). Günümüzde yapay sinir ağları eksik bilgilerle çalışabilme ve normal olmayan verilere çözüm üretebilme yeteneklerinden dolayı pek çok alanda kullanılabilir. Doğrusal olmayan, çok boyutlu, karmaşık, kesin olmayan, eksik, kusurlu, hata olasılığı yüksek veriler ve problemlerin çözümü için özellikle bir matematiksel model ve algoritmanın bulunmaması durumlarında yaygın halde yapay sinir ağları uygulamaları yapılabilmekte ve başarılı sonuçlar elde edilmektedir. Bu amaçla geliştirilmiş ağlar genel olarak aşağıda belirtilen fonksiyonları yerine getirmektedir (Öztemel, 2003):

Tahmin: Bu amaçla kullanılan yapay sinir ağları, ağa sunulan bilgilerden yararlanarak karşılık gelen çıktı değerlerini tahmin etmektedirler.

Sınıflandırma: Bu amaçla kullanılan yapay sinir ağları, kendilerine verilen bilgileri kategorize etme görevini üstlenmektedirler.

Veri ilişkilendirme: Bu amaçla eğitilen ağlar, ağa sunulan verilerin hatalı olup olmadıklarını belirlemektedirler. Öğrendikleri bilgiler ile eksik bilgileri tamamlamaktadırlar.

Veri filtreleme: Bu amaçla eğitilen ağlar, birçok veri arasından uygun verileri belirleme görevini yerine getirmektedir.

Tanım ve eşleştirme: Değişik şekil ve örüntülerin tanınması, eksik, karmaşık, belirsiz bilgilerin işlenerek eşleştirme ve tanıma fonksiyonlarını gerçekleştirebilmektedir.

Teşhis: Bu amaçla geliştirilen ağlar, sistemlerin olumsuzluklarının ortaya konulması ve problemlerin belirlenmesi işlemini yerine getirmektedirler.

Yorumlama: Bir olay hakkında toplanan örneklerden elde edilen ve eğitim sonucu oluşturulan bilgileri kullanarak yeni olayların yorumlanması işlemleri bu kapsamda düşünülmektedir.

Yapay sinir ağlarının avantajları herhangi bir ön varsayım gerektirmemesi, sadece eldeki verilerle sınırlandırılarak sistemi yürütebilmesidir. Ayrıca paralel işleyen yapısı, seri işlemcilerde meydana gelen herhangi bir birimin yavaşlamasıyla tüm sistemi yavaşlatma etkisini ortadan kaldırır. Böylelikle daha hızlı ve güvenilir sonuçlar elde edilir. Yapay sinir ağlarının genelleştirme gücüne sahip olması sayesinde bilinen değerlerden yararlanarak eğitim sürecinde karşılaşmadığı girdi değerleri için uygun çıktı değerleri üretirler. Tanımlanmamış veriler tahmininde bulunurken genelleme yapabildikleri için güçlü birer örüntü tanımlayıcısı ve sınıflandırıcılarıdır. Böylelikle yapay sinir ağları kayıp veriler için çözüm üretebilmektedirler (Yurtoğlu, 2005; Torun, 2007). Yapay sinir ağları silme yönteminde, değiştirme tabanlı yöntemlerde ve en çok olabilirlik yönteminde karşılaşılan problemlerin üstesinden geldiğinden bu yöntemlere göre daha avantajlı olduğu ifade edilebilir. Örneğin dizin bazında silme yöntemi örneklemin büyük bir oranını yok sayar ve silme yöntemi uygulanmış veri setinden elde edilen standart hatalar daha büyük olur. Çiftler bazında silme yönteminde ise iki değişken arasında kabul edilebilir olmayan bir korelasyon değeri (-1'den küçük ya da 1'den büyük) elde edilebilir. Değiştirme tabanlı yöntemlerden ortalama atamada değişkenler gerçek dağılımdan uzaklaşır ve varyans olduğundan daha küçük çıkar ve dağılım gerçek toplumu yansıtmaz. Hot Deck atama yönteminde durum benzerliği tanımlamak zordur. Model tabanlı yöntemlerden en çok olabilirlik yöntemi ise örnek verinin yeterince büyük olduğunu ve değişkenlerin normal dağıldığını kabul eder. Yapay sinir ağlarının bahsedilen avantajlarının yanı sıra değişkenler arası ilişkinin doğrusal olmadığı durumlarda diğer atama yöntemlerine göre daha gerçekçi

sonuçlar ürettiği ve varsayımlar istememesi gerekçeleriyle diğer yöntemlerden daha üstün olduğu öne sürülmektedir (Little ve Rubin, 2002)

Kayıp verileri yapay sinir ağları yöntemiyle atama çalışmalarında; Kaya, Yeşilova ve Almalı (2010) verilerin seçkisiz olarak kayıp olduğu durumlarda kayıp değerleri atamada yapay sinir ağlarının, beklenti maksimizasyonun ve çoklu atamanın başarılarını incelemişlerdir. Veri seti Yüzüncü Yıl Üniversitesi Beden ve Spor Eğitiminde özel yetenek sınavına katılan 155 bireyin fiziksel özelliklerinden oluşmaktadır. Seçkisiz olarak belirlenen kayıp verilere değer atama için R istatistik yazılımı kullanılmış ve gerçek veri setiyle yapay sinir ağlarıyla, beklenti maksimizasyonu ve çoklu atama ile tamamlanan veri setleri karşılaştırılmıştır. Sonuçta yapay sinir ağlarının gerçek verilerle daha yakın sonuçlar verdiği, beklenti maksimizasyonunun ise yapay sinir ağlarından sonra en yakın sonuç veren yöntem olduğu görülmüştür.

Aslan (2010) çalışmasında, alansal ve zamansal özellikteki meteorolojik zaman serilerinde karşılaşılan kayıp verilerin uygun şekilde tamamlanması için basit aritmetik ortalama (BAO), normal oran (NO) ve korelasyon ağırlıklı normal oran (KANO), çok katmanlı yapay sinir ağları (ÇKYSA) modeli ve beklenti maksimizasyonu tabanlı Monte Karlo Markov Zinciri (BM-MCM) algoritmasını kullanmış ve bu yöntemlerin karşılaştırmasını yapmıştır. Bu yöntemlere ek olarak basit yöntemlerden elde edilen sonuçların yardımcı değişken olarak kullanıldığı modifiye edilmiş BM-MCMC algoritmasını önermiştir. Yöntemlerin karşılaştırılması için hata kareler hesabına dayalı doğruluk hesabı ile Korelasyon Boyutu yöntemi kullanmıştır. Sonuçta uygun referans istasyonunun seçilmesi durumunda BM-MCM ile ÇKYSA yöntemlerinin diğer yöntemlere göre daha güvenilir sonuçlar verdikleri görülmüştür.

Ayrıca araştırmacılar (Gupta ve Lam, 1996; Aydilek ve Arslan, 2012; Abdella ve Marwala, 2005; Nelwamondo, Mohamed ve Marwala, 2007) yapay sinir ağlarıyla birleştirilmiş farklı yöntemleri kullanmışlar ve bunların da diğer yöntemlerden üstün olduğu yönünde iddialarda bulunmuşlardır. Bu araştırmalara ilgili araştırmalar bölümünde yer verilmiştir.

Kayıp değer probleminin üstesinden gelmek için en uygun yöntemi bulma arařtırmalarında, özellikle Sosyal Bilimler ve Eđitim Bilimleri arařtırmalarında sıkça kullanılan Likert tipi ölçeklerden elde edilen verilerde kayıp veri problemine yeterince değinilmediđi görülmüřtür.

Eđitim alanında yapılan arařtırmalarda Likert tipi ölçekler bilgi almak adına oldukça yaygın olarak kullanılmaktadır. Likert ilk olarak tutumları ölçmek amacıyla ölçeđini geliřtirmişse de, tutumlardan başka psikolojik değışkenleri ölçmek için de kullanılmaktadır. Likert tipi ölçekler bireylerin tepkilerine dayandıđından özellikle bireylerden kendileri hakkında bilgi vermeleri istendiđinde bu ölçeklere başvurulur.

Yukarıda belirtilen durumları ařađıdaki gibi özetlemek mümkündür:

- Bilimsel sürecin ařamalarından biri olan değerdendirme sürecinde veri setinde bulunan kayıp veriler analiz sonuçlarını etkilemektedir,
- Kayıp veriler ölçme sonuçlarının geçerliđini ve güvenirliliđini tehdit etmektedir,
- Kayıp veri probleminin üstesinden gelmek için kullanılan yöntemlerden yapay sinir ađlarının diđer yöntemlerden daha avantajlı olduđu ileri sürülmektedir,
- Özellikle eđitim alanında sıklıkla kullanılan Likert tipi ölçeklerde kayıp veri problemine değinmiş arařtırmalar ve yapay sinir ađları yöntemiyle kayıp verilere değerdendirme atama alıřmaları hem yurt içinde hem de yurt dıřında yeterli deđildir.

Belirtilen nedenler dođrultusunda yapay sinir ađı yöntemi ile kayıp verilere değerdendirme atanarak elde edilmiş olan tam veri setine yapılan analizlerin geçerli ve güvenilir sonuçlar verip vermediđinin arařtırılması gerekli bulunmaktadır.

Amaç

Bu araştırmanın genel amacı; PISA 2009 öğrenci anketinde yer alan öğrencilerin okumaya ilişkin tutumunu gösteren Türkiye verileri üzerinde kayıp verilere değer atama yöntemlerinden yapay sinir ağları yönteminin kullanımını incelemektir.

Bu genel amaç doğrultusunda aşağıdaki sorulara yanıt aranmıştır:

1. Yapay sinir ağları yönteminin kullanımı veri setlerindeki kayıp veri oranından etkilenmekte midir?
2. Farklı kayıp veri oranlarında yapay sinir ağları yöntemiyle kayıp verilere değer ataması yapılarak elde edilen tam veri setine yapılan geçerlik analizi ile gerçek veri setine yapılan geçerlik analizi sonuçları tutarlılık göstermekte midir?
3. Farklı kayıp veri oranlarında yapay sinir ağları yöntemiyle kayıp verilere değer ataması yapılarak elde edilen tam veri setine yapılan güvenilirlik analizi ile gerçek veri setine yapılan güvenilirlik analizi sonuçları tutarlılık göstermekte midir?

Önem

Veri setlerinden daha etkili analiz sonuçları elde etmek ve daha doğru kararlar vermek için kayıp verileri uygun bir şekilde ele almak ve eksik veri setini tam veri seti haline getirmek gerekmektedir. Dikkate değer (%5'ten fazla) kayıp verilere sahip veri seti, değişkenler arasındaki ilişkiden yola çıkarak yapılan tahminler sonucunda tam veri seti haline getirilebilir. Böylece yapılacak olan analizlerin daha güvenilir ve geçerli sonuçlar verdiği görülmektedir. Bu sonuçlar doğrultusunda verilecek kararların da daha doğru olacağı düşünülmektedir.

Bu araştırma, veri setinde kayıp verilerin bulunmasından kaynaklanan problemlerin üstesinden gelmek için kullanılması önerilen yapay sinir ağlarının Likert tipi ölçeklerden elde edilen ölçme sonuçlarında da kullanılmasının uygun olup olmamasına yönelik fikir vermesi açısından önemlidir.

Likert tipi ölçeklerde kayıp verilere değer atamada yapay sinir ağlarının kullanımı üzerine az sayıda çalışma bulunmaktadır. Bu çalışma, ilgili literatüre katkı sağlaması ve özellikle kayıp verilere değer atamada yapay sinir ağlarının kullanılıp kullanılmamasına dikkat çekmesi beklentisi ile önemli bulunmaktadır.

Sınırlılıklar

Bu çalışma;

1. PISA 2009 öğrenci anketi okumaya ilişkin tutumunu ölçen maddelere verilen cevapların Türkiye verileriyle,
2. Verilerin kayıp olma durumunun ilgilenilen değişkene ve diğer değişkenlerin değerlerine bağlı olmadığı kayıp veri mekanizması olan Tamamen Seçkisiz Olarak Kayıp mekanizmasıyla,
3. Acuna ve Rodriguez (2004)'in kayıp verinin %5-%15 arasında olduğu durumlarda kayıp verilere değer atama için çeşitli karmaşık yöntemlerin gerektiğini ve %15'ten daha fazla olması durumunda ise kayıp verinin yapılacak herhangi bir yorumu çok ciddi şekilde etkileyeceği görüşü çerçevesinde %5, %15 ve %30 oranlarında kayıp değerlerle sınırlı tutulmuştur.

Kısaltmalar

PISA (Program for International Student Assessment): Uluslararası Öğrenci Başarılarını Değerlendirme Programı.

TSOK (Missing completely at random-MCAR): Tamamıyla Seçkisiz Olarak Kayıp

SOK (Missing at random-MAR): Seçkisiz Olarak Kayıp

İEK (Nonignorable-NI): İhmal Edilemez Kayıp

YSA (Artificial Neural Network – ANN): Yapay Sinir Ağları

BÖLÜM II

İLGİLİ ARAŞTIRMALAR

Bu bölümde yurt içinde ve yurt dışında yapılan çeşitli araştırmalara yer verilmiştir.

Yurt İçinde Yapılan Araştırmalar

Kayıp verilere yapay sinir ağları yöntemiyle değer atama ve Likert tipi ölçeklerde kayıp verilere değer atama ile ilgili yurt içinde yapılan araştırmalar aşağıda özetlenmiştir:

Kaya, Yeşilova ve Almalı (2010) verilerin seçkisiz olarak kayıp olduğu durumlarda kayıp değerleri atamada yapay sinir ağlarının, beklenen maksimizasyonun ve çoklu atamanın başarılarını incelemiştir. Veri seti Yüzüncü Yıl Üniversitesi Beden ve Spor Eğitiminde özel yetenek sınavına katılan 155 bireyin fiziksel özelliklerinden oluşmaktadır. %0-%15 ranjında değişen kayıp değerler içeren veri setine kayıp verilere değer atama için R istatistik yazılımı kullanılmış ve gerçek veri setiyle yapay sinir ağları, beklenti maksimizasyonu ve çoklu atama ile tamamlanan veri setleri karşılaştırılmıştır. Sonuçta yapay sinir ağlarının gerçek verilerle daha yakın sonuçlar verdiği, beklenti maksimizasyonunun ise yapay sinir ağlarından sonra en yakın sonuç veren yöntem olduğu görülmüştür.

Aslan (2010) araştırmasında, 1965-2006 yıllarını kapsayan ve verisi tam olan istasyonların yağış ve sıcaklık gözlemlerini ele almıştır. Eksik veri içermeyen bu istasyon verilerinde çeşitli oranlarda yapay kayıp veri periyotları (%10, %20 ve %50) oluşturmuştur. Alansal ve zamansal özellikteki

meteorolojik zaman serilerinde karşılaşılan kayıp verilerin uygun şekilde tamamlanması için basit aritmetik ortalama (BAO), normal oran (NO) ve korelasyon ağırlıklı normal oran (KANO), çok katmanlı yapay sinir ağı (ÇKYSA) modeli ve beklenti maksimizasyonu tabanlı Monte Karlo Markov Zinciri (BM-MCM) algoritmasını kullanmış ve bu yöntemlerin karşılaştırmasını yapmıştır. Bu yöntemlere ek olarak basit yöntemlerden elde edilen sonuçların yardımcı değişken olarak kullanıldığı modifiye edilmiş BM-MCMC algoritmasını önermiştir. Yöntemlerin karşılaştırılması için hata kareler ortalaması değişim katsayısına dayalı doğruluk ölçümü ile korelasyon boyutu yöntemi kullanmıştır. Sonuçta uygun referans istasyonunun seçilmesi durumunda BM-MCM ile ÇKYSA yöntemlerinin diğer yöntemlere göre daha güvenilir sonuçlar verdikleri görülmüştür.

Aydilek ve Arslan (2011) araştırmalarında kayıp verilere değer atama için k-en yakın komşular yöntemi ile birleştirilmiş yapay sinir ağı yöntemiyle genetik algoritmayla birleştirilmiş yapay sinir ağı yöntemini karşılaştırmışlardır. Bu araştırmada dört farklı veri seti kullanılmış ve yapay kayıp değerler üretilmiştir. Bu kayıp değerlerin oranı dört farklı veri seti için %0-%10 ranjında değişmektedir. Bahsedilen iki yöntemle de kayıp değerler ataması yapılmış; hata oranları, korelasyon katsayısı, tahmin doğruluğu hesaplanıp karşılaştırılmıştır. Sonuçta çalışılan veriler için k-en yakın komşular yöntemi ile birleştirilmiş yapay sinir ağı yönteminin genetik algoritmayla birleştirilmiş yapay sinir ağı yönteminden daha etkili sonuçlar verdiği görülmüştür.

Çokluk ve Kayrı (2011), yaklaşık değer atama yöntemleriyle (seriler ortalaması atama, yakın noktaların ortalamasını atama, yakın noktaların medyanını atama, doğrusal değer kestirimi ve noktanın doğrusal eğimi) kayıp verilere değer atamanın, ölçme aracının yapı geçerliği ve güvenilirliği üzerinde etkisi olup olmadığını araştırmışlardır. Kayıp değerlerin olmadığı koşulda ve farklı oranlarda (yaklaşık %15.00-%20.00 ve %0.00-%50.00 ranjında) kayıp değerlerin olduğu koşullarda yaklaşık değer atama yöntemleri ile kayıp verilere değer atama yapılmasının sonucunda elde edilen faktör yapıları, düzeltilmiş madde-toplam korelasyonları ve Cronbach-alfa iç tutarlık katsayıları karşılaştırmalı olarak incelenmiştir. Sonuç olarak kullanılan

Fatalizm ölçeğinin hem gerçek veri seti için hem de kayıp değerlere farklı yöntemlerle yaklaşık değerlerin atandığı durumlar için tek faktörlü yapıya sahip olduğu ortaya çıkmıştır. Diğer taraftan açıklanan varyans oranında, özdeğerlerde ve Cronbach alfa iç tutarlılık katsayılarında atama yöntemi kullanıldığında azalma meydana gelmiştir. Ayrıca araştırmada son yıllarda etkisi incelenen Hot Deck Atama, Beklenti Maksimizasyonu ve Regresyon Yöntemi ile kayıp verilere değer atamanın klasik yöntemlerden daha etkili olduğunu belirten araştırmalara rastlanmakta olduğu belirtilmiş, klasik yöntemlerin yanlı bulgulara yol açabilme olasılıklarının daha yüksek olduğu noktasından hareketle, literatürde önerilen diğer değer atama yöntemlerinin kullanıldığı araştırmaların yapılması ve yöntemlerin karşılaştırmalı olarak incelenmesinin gerektiği üzerine vurguda bulunulmuştur.

Demir (2013) araştırmasında, kayıp verilerin varlığında iki kategorili puanlanan maddelerden oluşan testlerin psikometrik özelliklerini, farklı kayıp veri yöntemleri kullanarak incelemiştir. Testin psikometrik özelliklerine yönelik incelemeler ve kayıp veri analizlerinde, SBS 2011 Matematik Testi A Kitapçığı verilerini kullanmıştır. Kayıp veri yöntemleri olarak silmeye dayalı yöntemlerden izin silme yöntemi, basit atama yöntemlerinden sıfır atama, seri ortalamaları ataması, gözlem birimi ortalaması ataması, yakın noktalar ortalama ataması, yakın noktalar medyan ataması, doğrusal interpolasyon ve doğrusal eğilim noktası ataması yöntemleri, en çok olabilirlik yöntemlerinden regresyon atama, beklenti-maksimizasyon algoritması ve veri çoğaltma yöntemleri, çoklu veri atama yöntemlerinden ise MarkovChain Monte Carlo yöntemi kullanılmıştır. Bulgular doğrultusunda, ihmal edilebilir olmayan kayıp verilerin varlığında, iki kategorili puanlanan maddelerden oluşan testlerin psikometrik özelliklerine yönelik inceleme ve analizlerde, uygun bir kayıp veri yönteminin kullanılmasının gerekli olduğu görülmüştür. En çok olabilirlik ve çoklu veri atama yöntemleri, bu tür verilerde kullanılması en uygun kayıp veri yöntemleri olarak değerlendirilmiştir.

Yurt içinde yapılan araştırmalara bakıldığında araştırmacılar kayıp veri probleminin üstesinden gelmek için çeşitli kayıp verilere değer atama yöntemlerini kullandıkları ve veri setleri için en etkili atamayı yapan yöntemi

belirlemeye çalıştıkları görülmektedir. Araştırmacılar tam veri setlerinde kayıp verileri yapay olarak oluşturmuşlardır. Yapılan araştırmaların çoğunda araştırmacılar veri setlerinde %0 ile %50 ranjında değişen kayıp değerler oluşturmuşlar ve bu veri setleriyle araştırmalarına devam etmişlerdir. Araştırmalarında genelde sürekli verileri kullanmışlardır. Kayıp verilere değer ataması yapılırken kullanılan yöntemlere bakıldığında daha çok model tabanlı yöntemleri kullandıkları ve bu yöntemleri karşılaştırdıkları görülmektedir.

Yurt Dışında Yapılan Araştırmalar

Kayıp verilerin yapay sinir ağları yöntemiyle atanması ve Likert tipi ölçeklerde kayıp verinin atanması ile ilgili yurt dışında yapılan araştırmalar aşağıda özetlenmiştir.

Gupta ve Lam (1996) kayıp değer problemini ele almışlar ve kayıp veri değerini atamada yapay sinir ağlarının performansının diğer istatistiki yöntemlere göre nasıl olduğunu araştırmışlardır. Yapay sinir ağında öğrenme algoritması olarak geri yayılım algoritması kullanılmıştır. Eğitim veri setinde 46 durumdan 18'inde tek kayıp, 6'sında iki kayıp veri oluşturulmuştur. Test verisinde 46 durumdan 20'sinde tek kayıp, 6'sında iki kayıp oluşturulmuştur. Veri setindeki kayıp verilere değer atamada yapay sinir ağları, ortalama atama ve regresyon yöntemi karşılaştırılmıştır. Ayrıca kayıp verilere değer ataması yapıldıktan sonra da diskriminant analiziyle yapay sinir ağlarının sınıflandırma etkililiği incelenmiştir. Yapılan deneysel araştırmada sonuçlar yapay sinir ağlarının diğer yöntemlere göre daha etkili çalıştığını göstermiştir. Ayrıca diskriminant analizinin yapay sinir ağları kadar hassas davranmadığı ve bu araştırmadaki veri seti için sınıflandırmada yeterince güçlü olmadığı ortaya çıkmıştır.

Downey ve King (1998) araştırmalarında Likert tipi ölçeklerde kayıp veri probleminin üstesinden gelmek için kullanılan madde ortalaması ve birey ortalaması yöntemlerinin etkililiğini araştırmışlardır. Minnesota iş memnuniyeti ölçeği ile örgütsel bağlılık ölçekleri kullanılarak veriler elde etmişlerdir. Her iki ölçekten elde edilen tam veri setindeki birey sayısının (834) %5'ini seçip (42) seçilen bireylerde sırasıyla

maddelerin %10, %20, %30, %40, %50, %60 ve %70 oranlarında kayıp veri oluşturmuşlardır. Kayıp veri oluşturma işlemini geri kalan tam veri setindeki birey sayısının (792) %10, %15, %20, %25, %30 ve %35'ini seçip tekrarlanmıştır. Bu kayıp verilere madde ortalamalarını ve birey ortalamalarını atanmışlardır. Gerçek veri setindeki puanlarla atanmış haldeki puanlar arasındaki korelasyona ve alfa güvenirlik katsayısına bakmışlardır. Sonuç olarak cevaplayıcı ve madde sayısının %20 veya %20'den az olduğu durumlar için her ikisinin de etkili sonuçlar verdiği değerlendirilmiştir.

Jönsson ve Wohlin (2004) araştırmalarında k en yakın komşular yöntemini Likert tipi ölçekte kayıp verilere değer atamada kullanmışlardır. Ölçekte 6 madde yer almaktadır. Tam veri setinden yapay olarak kayıp veriler oluşturmuşlardır. Kayıp verilere değeri k en yakın komşular yöntemiyle atanmışlar ve bu atamayı yaparken Likert tipi ölçek kullandıklarından medyan değerlerini almışlardır. Toplam veride %5, %10, %15, %20, %25, %30, %35, %40, %45, %50, %55, %60 oranlarında kayıp veriler oluşturmuşlardır. Sonuç olarak %30 ve daha az kayıp veri olduğu durumlarda k en yakın komşular yönteminin etkili sonuçlar verdiği görülmüştür.

Abdella ve Marwala (2005) araştırmalarında genetik algoritmayla sinir ağlarının birleşimini kullanarak kayıp verilere değer atamada yeni bir yöntem olarak tanıtmışlardır. Önerilen yöntemde genetik algoritma, otomatik ilişkili sinir ağından üretilen hata fonksiyonunu en aza indirmek için kullanılmıştır. Önerilen yöntemin tahmin doğruluğunun kayıp verilerin sayısının artmasıyla azalıp azalmayacağı da araştırılmıştır. 198 durum ve 14 değişkenden oluşan veri setinde durumlarda sırasıyla 1, 2, 3, 4, ve 5 kayıp veri oluşturularak kayıp verilere değer atamada yöntemin etkililiği incelenmiştir. Araştırmalarında ayrıca çok katmanlı perseptron ile Merkezci tabanlı fonksiyon (Radial Basis Function) ağlarını, sinir ağlarını eğitmek için kullanmışlardır. Sonuç olarak önerilen yöntemin tahmin doğruluğunda kayıp veri sayısının artmasıyla manidar bir fark olmadığı ortaya çıkmış ve Merkezci tabanlı fonksiyon kullanılarak elde edilen sonuçlar çok katmanlı perseptron ile elde edilen sonuçlardan daha etkili olduğunu ortaya koymuşlardır.

Nelwamondo, Mohamed ve Marwala (2007) beklenti maksimizasyonu ile genetik algoritmayla birleştirilmiş otomatik ilişkili sinir ağlarının (NN-GA) endüstriyel enerji santrali, endüstriyel bobin sarımı ve HIV seroplansı araştırma veri setlerini kullanarak karşılaştırmışlardır. Veri setlerinde sürekli ve süreksiz veriler bulunmaktadır. Veri setlerinde %10'luk kayıp veri oluşturmuş ve kayıp verilere değer ataması beklenti maksimizasyonu ve NN-GA yöntemleri ile yapılmıştır. Ataması yapılan değerlerle gerçek değerler arasındaki korelasyona bakılmış ve sonuçta beklenti maksimizasyonun girdi değişkenleri arasında ilişki olmama ya da az derecede ilişki olduğu durumlarda daha etkili sonuçlar verdiği, NN-GA verilen değişkenlerin bazıları arasında doğrusal olmayan ilişkiler olduğu zaman daha uygun olduğu görülmüştür.

Leite ve Beretvas (2010) Likert tipi maddelerde kayıp veriler için çoklu atama yönteminin etkililiğini incelemişlerdir. 3'lü, 5'li ve 7'li Likert tipi ölçekler kullanılmış, TSOK ve SOK mekanizmaları altında %10, %30 ve %50 oranlarında kayıp veriler oluşturulmuştur. Kayıp veri oranı %10 olduğu durumlarda her iki mekanizmada da çoklu atama yönteminin uygun sonuçlar verdiği, %30 kayıp veri oranında sadece TSOK mekanizması altında çoklu atama yönteminin etkili sonuçlar verdiği görülmüştür. %50 kayıp veri oranında ise her iki mekanizmada da yanlış sonuçlar vermiştir.

Yurt dışında yapılan araştırmalar bir bütün olarak değerlendirildiğinde araştırmacıların tam veri setlerine kayıp verileri yapay olarak oluşturdukları ve uygun atama yöntemini belirlemek amacıyla yöntemleri karşılaştırma yoluna gittikleri görülmektedir. Belirtilen araştırmalarda genelde hem sürekli hem de süreksiz veriler kullanılmıştır. Kayıp verilere değer ataması yapılırken kullanılan yöntemlere bakıldığında daha çok model tabanlı yöntemleri kullandıkları ve bu yöntemleri karşılaştırdıkları görülmektedir.

Özetle, yurt içinde ve yurt dışında yapılan araştırmalarda kayıp veri probleminin üstesinden gelmek amacıyla kayıp verilere değer atama yöntemlerinin karşılaştırılmasına gidildiği görülmektedir. Araştırmalarda hem sürekli hem de süreksiz verilerle çalışılmıştır. Bu araştırmalar incelendiğinde atama yöntemlerinden yapay sinir ağları yöntemine daha çok değinildiği

görülmektedir. Ayrıca yapay sinir ağlarının çeşitli yöntemlerle birleştirildiği (k-en yakın komşular yöntemi ile birleştirilmiş yapay sinir ağı yöntemi, genetik algoritmayla birleştirilmiş yapay sinir ağı) yeni yöntemlerin de geliştirildiği ve etkililiğinin incelendiği görülmektedir.

Sonuç olarak, hem yurt içinde hem de yurt dışında yapılan çalışmalarda yapay sinir ağları yönteminin ve farklı algoritmalarla birleştirilmiş yapay sinir ağları yöntemlerinin kayıp verilere değer atamada etkili yöntemler olduğu görülmektedir.

BÖLÜM III

YÖNTEM

Bu bölümde araştırma modeli, çalışma grubu, verilerin toplanması ve verilerin analizi açıklanmıştır.

Araştırma Modeli

Bu araştırmada, PISA 2009 öğrenci anketinde yer alan öğrencilerin okumaya ilişkin tutumunu gösteren Türkiye verileri üzerinde kayıp verilere değer atama yöntemlerinden yapay sinir ağı yönteminin kullanımı incelendiğinden araştırma, bilgi üretmeye yönelik temel araştırma niteliğindedir.

Çalışma Grubu

Çalışma grubunu PISA 2009 uygulamasına Türkiye’de katılan 15 yaş grubu 4996 öğrenciden okumaya ilişkin tutumunu ölçen maddeleri eksik cevaplayan öğrenciler çıkartılarak elde edilen 4735 tam veri setine sahip öğrenciler içerisinden alınan 500 öğrenci oluşturmuştur. Çalışma grubunun 500 kişi olarak belirlenmesinin nedeni ilgili literatürde örneklem büyüklüğünün değişkenlerden daha fazla olması gerektiği, faktör analizine tabi tutulacak madde sayısının en az 5 katı büyüklükte olması ve 100 kişiden az olmaması gerektiğinin belirtilmesidir (Bryman ve Cramer, 2005).

Verilerin Toplanması

Bu arařtırmada PISA 2009 öğrenci anketinde yer alan öğrencilerin okumaya ilişkin tutumu ölçen 11 maddeyi içeren 24. maddenin verileri kullanılmıştır. Likert tipi ölçek 'kesinlikle katılmıyorum (1)', 'katılmıyorum (2)', 'katılıyorum (3) ve 'kesinlikle katılıyorum (4)' şeklinde puanlanmaktadır. 11 maddenin içerisinde ters puanlanması gereken maddeler de bulunmaktadır.

Uluslararası Öğrenci Başarılarını Değerlendirme Programı (PISA) Ekonomik İşbirliği ve Kalkınma Teşkilâtı (OECD) tarafından düzenlenen; öğrencilerin matematik, fen ve okuma becerileri alanlarındaki bilgi ve becerilerinin değerlendirildiği uluslararası en büyük eğitim arařtırmalarından biridir. Üç yılda bir yapılan PISA her dönem, öğrencilerin üç temel konu (okuma becerileri, matematik okuryazarlığı ve fen okuryazarlığı) alanlarından birine ağırlık vermektedir. PISA 2009'da okuma becerileri alanına ağırlık verilmiş olup öğrencilerin bilgi ve becerilerini değerlendirmenin yanında, öğrenme stratejileri, problem çözme becerileri ve değerlendirilen konu alanına yönelik ilgi ve tutumları da arařtırılmıştır (Milli Eğitim Bakanlığı, 2010).

Hazır verilerin kullanılması planlandığı için PISA 2009 verileri tercih edilmiştir ve PISA 2009 yılına ait veri dosyası PISA web sayfasından (www.PISA.oecd.org) elde edilmiştir.

Verilerin Analizi

Çalışma grubuna dahil edilen 500 öğrencinin 11 maddeye verdiği cevaplardan SPSS programı kullanılarak herhangi bir değişkendeki (6.madde seçkisiz olarak alınmıştır) gözenek sayısı seçkisiz olarak %5, %15 ve %30 oranlarında kayıp veri elde edilecek şekilde veri setleri oluşturulmuştur. İki değişkendeki (5. ve 6. maddeler) ve üç değişkendeki (5., 6. ve 7. maddeler) gözenek sayıları da seçkisiz olarak %5, %15 ve %30 oranlarında kayıp veri elde edilecek şekilde veri setleri oluşturulmuştur. Yani 500 öğrencinin seçkisiz olarak seçilen 25'inin (%5), 75'inin (%15) ve 150'sinin (%30) 6. maddeye verdikleri cevaplar; 5. ve 6. maddelere verdikleri cevaplar; 5., 6. ve 7. maddelere verdikleri cevaplar silinerek kayıp veri haline getirilmiştir. Öncelikle veri setlerinin örüntülerine bakılmıştır. Veri setlerinin hangi kayıp

veri mekanizmasına sahip olduğunu belirlemek için bu veri setleri Little'ın TSOK testine tabi tutulmuştur. Little'ın TSOK testi SPSS yardımıyla yapılmış, manidarlığı incelenmiştir.

Veri setlerinde kayıp veri oluşturma sürecinde seçkisizlik olsa da bunu test etme amacıyla ve kayıp verilerin mekanizmasını belirlemek amacıyla Çizelge 1'de Little'ın TSOK testinin bulguları verilmiştir.

Çizelge 1. Little'ın TSOK testi

Kayıp veri oranı	Kayıp madde	χ^2	sd	p
%5	6.madde	11.775	10	.300
	5. ve 6. madde	10.902	9	.289
	5., 6. ve 7. madde	9.309	8	.317
%15	6.madde	14.462	10	.153
	5. ve 6. madde	12.348	9	.194
	5., 6. ve 7. madde	10.814	8	.212
%30	6.madde	13.461	10	.199
	5. ve 6. madde	13.300	9	.150
	5., 6. ve 7. madde	12.775	8	.120

Çizelge 1'de görüldüğü üzere kayıp veri mekanizması her kayıp veri oranı ve her kayıp değişken sayısı için manidar değildir ($p>.05$). Kayıp veri setlerinin tamamen seçkisiz olarak kayıp olduğu ifade edilebilir.

Yapay sinir ağları yöntemi ile kayıp verilere değer atama için MATLAB bilgisayar programı kullanılmıştır. Yapay sinir ağlarının MATLAB programında kayıp verilere değer atamasına geçmeden önce aşağıda yapay sinir ağlarından bahsedilmiştir.

Yapay Sinir Ağları

Bilgi işleme, günümüzde büyük ölçüde bilgisayarlarla gerçekleştirilmektedir. Bilgisayarlar sayısal işlemlerde insan beyniyle kıyaslanamayacak kadar hızlı olmalarına rağmen ses ve görüntü algılama ve işleme, plan yapma ve öğrenme gibi işlemleri insan beyni çok daha kısa sürede gerçekleştirir. Bilgisayarlar ise bu işlemleri insan beynine kıyasla

oldukça fazla zaman ve hafıza kullanarak, çok az miktarda ve çok daha az başarıyla sonuçlandırır (Baş, 2006).

Bilgisayarlar ilk zamanlarında sayısal makineler olarak dikkate alınsa da, daha sonra insanların problem çözme gibi bilişsel aktivitelerinin davranışını sergilemek için sembollerini işleyebilir olduğunun farkına varılmıştır. Bu farkındalık araştırmacıları insanların bilişsel becerilerini taklit edebilen programları araştırmaya yöneltmiştir. Böylece, 1950'lerde söz konusu programların tasarımı ve uygulanmasıyla ilgili bilgiler ortaya çıkmaya başlamıştır. Bu sergilenen davranış insan beyninde değil de makinede olduğundan bu davranışa yapay zeka denilmektedir (Aziz ve Dar, 2004).

Yapay zeka sistemlerinin en temel özelliklerinin, olaylara ve problemlere çözümler üretirken veya çalışırken, bilgiye dayalı olarak karar verebilme özelliklerinin olması ve eldeki bilgiler ile olayları öğrenerek, sonraki olaylar hakkında kararlar verebilmeleri olduğu ifade edilebilir (Öztemel, 2003). Elmas (2003) yapay zekanın uzman sistemler, bulanık mantık, yapay sinir ağları ve genetik algoritma gibi alt dallarının yaygın bir şekilde araştırma ve uygulama alanı bulunduğunu ifade etmiştir. Yapay zeka yöntemlerinde, öğrenme çeşitli özel donanımlarda ya da çeşitli yazılımlar aracılığıyla bilgisayarlarda meydana gelmektedir. Bu nedenle, yapay zeka yöntemlerindeki öğrenmeye makine öğrenmesi denilmektedir. Makine öğrenmesinin bir çeşidi olan yapay sinir ağları, beyin faaliyetlerinin matematiksel modelleridir (Shah ve Murtaza, 2000). Beynin bütün davranışlarını modelleyebilmek için fiziksel bileşenlerinin doğru olarak modellenmesi gerektiği düşüncesi ile çeşitli yapay sinir hücreleri ve ağ modelleri geliştirilmiştir. Böylece, "Yapay Sinir Ağları" denen günümüz bilgisayarlarının algoritmik hesaplama yöntemlerinden farklı bir çalışma disiplini ortaya çıkmıştır (Saraç, 2004).

Kohonen (1987) yapay sinir ağlarını biyolojik sinir sisteminde olduğu gibi gerçek yaşam nesnelileriyle etkileşmeyi amaçlayan basit elemanların ve onların hiyerarşik düzenlemelerinin paralel, içice bağlantılı olan ağlar olarak ifade eder. Haykin (1998)'e göre ise yapay sinir ağı, deneysel bilgi saklama ve kullanıma hazır hale getirme için doğal bir eğilime sahip basit işlem

birimlerinden oluşan yoğun bir şekilde paralel dağıtılmış bir işlemcidir. Bu işlemci iki şekilde beyin ile benzerlik gösterir:

- Bilgi, bir öğrenme süreci yoluyla çevresinden ağ tarafından elde edilir,
- Sinaptik ağırlıklar olarak da bilinen nöronlar arası bağlantı güçleri elde edilen bilgileri depolamak için kullanılır (Yurtoğlu, 2005).

Yapay Sinir Ağının Yapısı

Yapay sinir ağları her ne kadar beyinle benzer yapıda olsa da beyin çok daha karmaşık ve yoğun bağlantılı bir yapıya sahiptir. Yapay sinir ağlarını oluşturmak için kullanılan yapay nöronlar, beyindekilere kıyasla oldukça ilkel sayılırlar da genel yapı olarak beyinle tutarlıdır. Diğer bir ifadeyle YSA'lar, beyin sadece en temel elemanlarını kopyalamaya çalışmaktadır. Çizelge 2'de gösterilen biyolojik sinir ağlarıyla yapay sinir ağlarının karşılaştırılmasıyla bu iki ağın benzerlikleri gösterilmiştir.

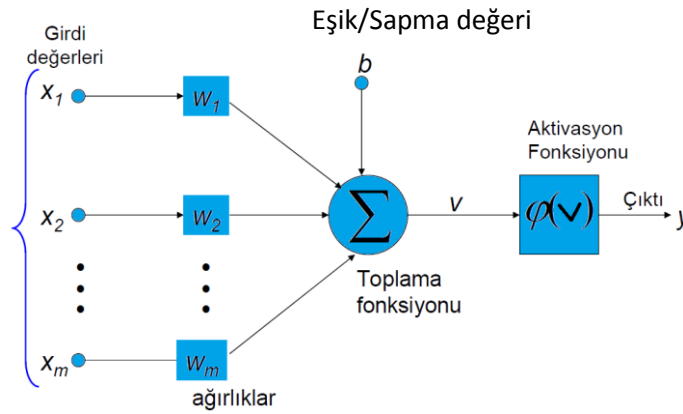
Çizelge 2. Biyolojik Sinir Ağları ile Yapay Sinir Ağlarının Karşılaştırılması

Biyolojik Sinir Ağı	Yapay Sinir Ağı
Sinir Sistemi	Sinirsel Hesaplama Sistemi
Sinir Hücresi (Nöron)	İşlemci Eleman (Yapay Sinir Hücresi, Düğüm)
Sinaps	İşlemci elemanlar arasındaki bağlantı ağırlıkları
Dendrit	Toplama Fonksiyonu
Hücre Gövdesi	Aktivasyon Fonksiyonu
Akson	İşlemci Eleman çıktısı

Kaynak: Baş, 2006

Yapay sinir ağlarında bilgi işleme, yapısı Şekil 3'te gösterilen sinir hücresi adı verilen birçok basit elemanda gerçekleşmektedir. Bir sinir

hücresinden gelen sinyal, sinir hücreleri arasındaki ilişkiyi sağlayan bağlantılarla iletilmektedir. Her bir bağlantının bir ağırlık değeri vardır ve girdiyi önemine göre ağırlıklandırılarak geçişini sağlamaktadır. Sinir ağı içindeki her bir sinir hücresine ayrı bir aktivasyon fonksiyonu uygulanmaktadır (bu fonksiyon genelde doğrusal olmayan bir fonksiyondur) ve bu fonksiyonun çıkış değeri sayesinde sinir hücresinin çıkış sinyali hesaplanmaktadır. Herhangi bir yapay sinir ağı; sinir hücreleri arasındaki bağlantının bir modeli yani mimarisi, bağlantılar arasındaki ağırlıkların hesaplanması (bu hesaplama, öğrenme kuralı ya da öğrenme algoritması olarak da adlandırılır) ve aktivasyon fonksiyonu ile tanımlanabilir (Aydın, 2002).



Şekil 3. Yapay Sinir Hücresinin Yapısı

Şekil 3'te girdi değerler x_i matematiksel sembolü ile gösterilmiştir ve bu gösterimde $i = 1, 2, \dots, m$ değerlerini almaktadır. Bu girdi değerlerin her biri bir bağlantı ağırlığıyla çarpılmaktadır. Bu ağırlıklar ise w_i ile gösterilmektedir. En basit yapıda, bu çarpımlar toplanır ve bir transfer fonksiyonuna gönderilerek sonuç üretilir. Bu sonuç daha sonra bir çıktıya dönüştürülür. Yapay nöron yapısı değişik toplama fonksiyonları ve transfer fonksiyonları kullanabilir.

Yapay sinir ağının temel hesaplama birimi sinir hücresidir. Yapay sinir hücresi kavramı, 1940 yılında McCulloch ve Pitts tarafından insan beynindeki sinir hücrelerini taklit etme çabasıyla geliştirilmiştir. Bir yapay sinir ağında yer

alan her bir sinir hücresi, sırasıyla şu görevleri yerine getirmektedir (DeTienne, DeTienne ve Joshi, 2003):

- Diğer sinir hücrelerinden sinyaller alınır (x_1, x_2, \dots, x_n).
- Sinyaller, sinir hücresinin ağırlıklarıyla çarpılır ($w_1x_1, w_2x_2, \dots, w_nx_n$).
- Ağırlıklandırılmış sinyaller toplanır (Toplam = $w_1x_1 + w_2x_2 + \dots + w_nx_n$).
- Hesaplanan toplam, aktivasyon fonksiyonu tarafından dönüşüme tabi tutulur [$f(\text{Toplam})$].
- Dönüştürülmüş toplam, diğer sinir hücrelerine gönderilir.

Sapma değeri, x vektörüne $x_0 = 1$ değeri eklenerek oluşturulabilir.

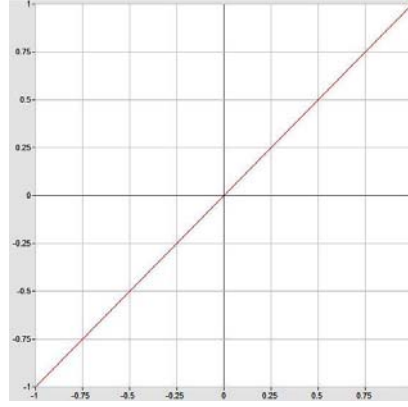
$$x = (1, x_1, \dots, x_i, \dots, x_n)$$

Sapma aynı diğer ağırlıklar gibi davranır, yani $w_{0j} = b_j$ dir. y_j birimi için ağ girdisi şu şekilde hesaplanır:

$$y_j = w_{0j} + \sum_{i=0}^n w_{ij}x_i = b_j + \sum_{i=0}^n w_{ij}x_i$$

Aktivasyon fonksiyonu, toplama fonksiyonundan gelen girdiyi işleyerek yapay sinir hücresinin çıkışı belirler. Doğrusal ve doğrusal olmayan aktivasyon fonksiyonlarının, YSA yapılarında kullanılması, karmaşık ve çok farklı problemlere uygulanmasını sağlamıştır. Yapay sinir ağında hangi aktivasyon fonksiyonunun kullanılacağı probleme bağlı olarak değişmektedir (Saygılı, 2008).

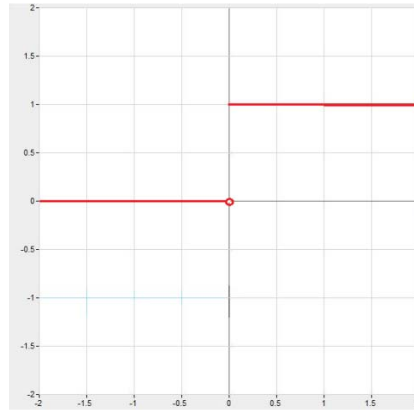
Yaygın olarak kullanılan çeşitli aktivasyon fonksiyonları doğrusal fonksiyon, adım fonksiyonu ve sigmoid fonksiyonudur. Şekil 4'te gösterilen doğrusal fonksiyon, hücrenin net girdisini hücrenin çıkışı olarak veren doğrusal problemleri çözme amacıyla kullanılan fonksiyondur. Matematiksel ifadesi $f(x) = ax$ şeklinde tanımlanabilir. a sabit değeri eğimi verir.



Şekil 4. Doğrusal fonksiyon

Şekil 5'te gösterilen adım fonksiyonu, gelen net girdi değerinin belirlenen bir eşik değerinin altında ya da üstünde olmasına göre hücrenin çıktısı 0 veya 1 değerlerini almaktadır. Matematiksel ifadesi şu şekildedir:

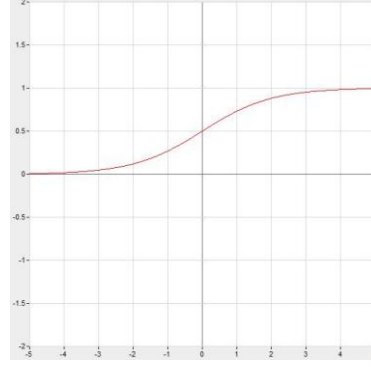
$$f(x) = \begin{cases} 0, & x < \theta \\ 1, & x \geq \theta \end{cases}$$



Şekil 5. Adım Fonksiyonu

Şekil 6'da gösterilen sigmoid fonksiyonu türevi alınabilir, sürekli ve doğrusal olmayan bir fonksiyon olması nedeniyle uygulamada en çok kullanılan aktivasyon fonksiyonudur. Sigmoid fonksiyonları içerisinde en yaygın olanları lojistik ve hiperbolik tanjant fonksiyonlarıdır. Lojistik sigmoid fonksiyonu net girdinin her değeri için 0 ile 1 arasında bir değer üretmektedir. Matematiksel gösterimi

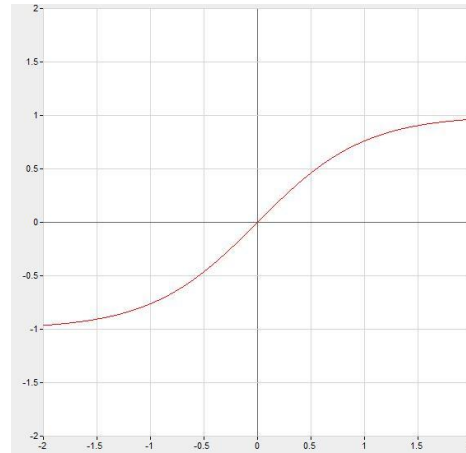
$$f(x) = \frac{1}{1+e^{-x}} \text{ şeklindedir.}$$



Şekil 6. Lojistik Sigmoid Fonksiyonu

Şekil 7’de gösterilen hiperbolik tanjant fonksiyonu, gelen net girdinin tanjant fonksiyonundan geçirilmesi ile hesaplanmaktadır ve sigmoid aktivasyon fonksiyonunun farklı bir çeşididir. Lojistik sigmoid aktivasyon fonksiyonunda çıktı 0 ile 1 arasında bir değer alırken, hiperbolik tanjant fonksiyonunda çıktı -1 ile 1 arasındadır. Matematiksel ifadesi şu şekildedir:

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

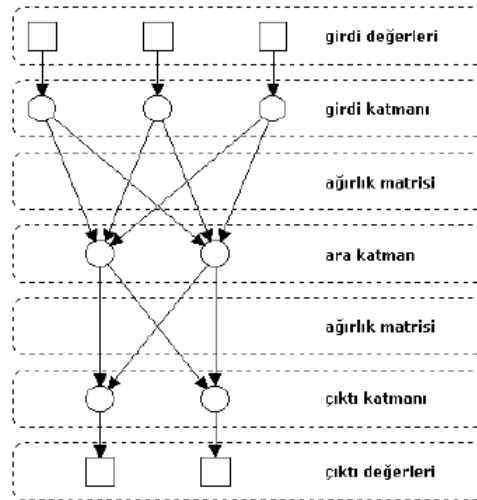


Şekil 7. Hiperbolik Tanjant Fonksiyonu

Yapay Sinir Ağının Tasarımı

Yapay sinir hücreleri farklı şekillerde bir araya gelerek yapay sinir ağını meydana getirmektedir. Yapay sinir hücrelerinin genellikle birbiriyle bağlantılı girdi katmanı, ara katman ve çıktı katmanı olarak adlandırılan katmanlar halinde oluştukları görülür. Girdi katmanı dış ortamdan bilgileri alarak ara

katmana, gizli katman olarak da adlandırılan ara katman da girdi katmanından gelen bilgileri işleyerek çıktı katmanına iletir. Ara katman birden fazla olabilmektedir. Ara katmanlar çok sayıda yapay sinir hücresi içermektedirler ve bu hücreler yapay sinir ağı içindeki diğer hücrelerle bağlantılıdır. Çıktı katmanı, ara katmandan gelen işlenmiş bilgileri dış ortama aktarır (Baş, 2006). Bu katmanları anlayabilmek için Şekil 8'de yapay sinir ağının bir örneği verilmiştir.



Şekil 8. İki Katmanlı İleri Beslemeli Yapay Sinir Ağı Mimarisi (Fröhlich, 1997; akt. Baş, 2006)

Bir yapay sinir ağında, işlemci elemanların bağlanması, işlemci elemanların sahip oldukları toplama ve aktivasyon fonksiyonları, kullanılan öğrenme yöntemi ve öğrenme kuralı ağın modelini belirlemektedir. Yapay sinir ağı uygulamasının başarısı, modelin oluşturulması aşamasının en doğru şekilde yürütülmesi ile yakından ilgilidir. Bunun için yapay sinir ağı tasarımcısının, ağın yapısına ve işleyişine ilişkin şu kararları vermesi gerekmektedir (Saraç, 2004):

- Ağ mimarisinin seçimi ve yapısal özelliklerinin belirlenmesi (katman sayısı ve katmandaki işlemci eleman sayısı gibi),
- İşlemci elemanların kullandığı fonksiyonların karakteristik özelliklerinin belirlenmesi,
- Öğrenme algoritması ve parametrelerinin belirlenmesi,
- Eğitim ve test setinin oluşturulması.

Yapay sinir ağı modeli tasarlamak için ağda kullanılacak algoritmaya karar vermek gerekir. Yapay sinir ağının öğrenme algoritmasının seçimiyle mimarisi de seçilmiş olmaktadır. Örneğin eğer geri yayılım algoritması seçilirse ileri beslemeli ağ mimarisi de otomatik olarak seçilmiş olmaktadır. Model tasarlamakta diğer adımlar, çok katmanlı bir ağ modelinin seçilmesi durumunda, ağdaki ara katman sayısına karar vermek ve her katmandaki işlemci eleman sayısını belirlemektir. Katman sayısını ve işlemci elemanların sayısını belirlemenin en etkili yolu, deneme-yanılmayla ağın performansına bağlı olarak en uygun katman ve işlemci eleman sayısına karar vermektir. Bir yapay sinir ağında işlemci eleman sayısının, olması gerekenden daha az olması ağın öğrenme yeteneğini azaltırken, gereğinden çok olması genelleme yeteneğini azaltmaktadır. İşlemci elemanların karakteristik özelliklerinin belirlenmesi de yapay sinir ağının tasarımında önemli kararlardan biridir. Toplama ve aktivasyon fonksiyonlarının belirlenmesi, büyük ölçüde verilerin özelliklerine ve ağın neyi öğrenmesinin istendiğine bağlıdır. Aktivasyon fonksiyonları içinde en çok kullanılanlar, sigmoid ve hiperbolik tanjant fonksiyonlarıdır. Eğer ağın, bir modelin ortalama davranışını öğrenmesi isteniyorsa sigmoid, ortalamadan sapmasını öğrenmesi isteniyorsa hiperbolik tanjant fonksiyonlarının kullanımı önerilir (Saraç, 2004).

Yapay sinir ağlarında katman sayısını, nöronları bağlayan ağırlıklı bağlantı sayısı belirler. Bağlantı sayısı bir taneyle tek katmanlı, birden fazla ise çok katmanlı yapay sinir ağı olarak adlandırılır.

Yapay sinir ağlarında çok çeşitli ağ yapıları ve modelleri vardır. Bunlardan en çok kullanılanlara aşağıda kısaca anlatılmıştır.

Hebb net: Yapay sinir ağı modellerinin ilk bilinenidir ve Hebb (1949) tarafından tanımlanmıştır. Hebb'in koymuş olduğu kural şudur: Eğer bir hücre diğer hücreden girdi alırsa ve her iki hücre aynı aktivasyonu gösteriyorsa (matematiksel olarak aynı işaretliyse) hücreler arasındaki bağlantının ağırlığı arttırılmalıdır, eğer iki hücrenin aktivasyonu farklı ise girdi örüntü tipine göre aralarındaki bağlantının ağırlığı azaltılmalıdır ya da değiştirilmemelidir.

Örneğin girdi örüntü tipi ikili veri (1 veya 0) ise hücreler farklı aktiflikte olsa da ağırlıklar değiştirilmez (Fausett,1994).

Perseptron: Hebb net modelinden sonra F.Rosenblatt tarafından 1959'da tanıtılan perseptron modelinde eşikleme fonksiyonundan geçirilme işlemi söz konusudur. Perseptron modelinde en önemli etken eşik değeridir. Eşik değeri problemin karakteristiğine göre belirlenir. Eşik değeri yardımıyla perseptron modeli Hebb net'e göre daha başarılı bir sınıflandırma yapabilmektedir. Perceptron en yalın şekilde bir yapay sinir hücresinin birden fazla girdiyi alarak tek bir çıktı üretmesi temeline dayanmaktadır. Ağın çıktısı 1 veya 0 olmakta ve çıktı değerinin hesaplanması adım fonksiyonu kullanılarak yapılmaktadır. Perseptronun öğrenme algoritması 'hata' üzerine dayandırılmıştır. Eğer sistem hata ile karşılaşılırsa öğrenmektedir. Başka bir ifadeyle perseptron hatalardan ders çıkararak, sistemi en az hataya indirgemeyi hedeflemektedir (Lippmann,1987).

Uyarlanabilir doğrusal elemanlar (Adaptive Linear Element-ADALINE): Widrow ve Hoff tarafından 1960 yılında geliştirilmiştir. En küçük ortalama kareler (least mean square) yöntemine dayanmaktadır. Öğrenme kuralına delta kuralı denilmektedir ve bu öğrenme kuralında en küçük hata fonksiyonunu verecek ağırlıklar dereceli azalma (gradient descent) yöntemi ile bulunur. Hata, her bir ağırlık değerinin hataya bağlı kısmi türevlerinin negatif (-) değerleri kullanılarak azaltılır.

Kohonen ağı: Özörgütlemeli Özellik Haritası (Self Organization Feature Map Network - SOM) olarak da adlandırılan Kohonen ağı, 1972 yılında Kohonen tarafından geliştirilmiştir. Genel olarak sınıflandırma yapmak için kullanılan Kohonen ağının, girdi vektörlerini sınıflandırma ve girdi vektörlerinin dağılımını öğrenebilme yetenekleri oldukça yüksektir.

Uyarlanabilir yankılaşım teorisi (Adaptive Resonance Theory-ART): Grossberg'in beynin fonksiyonlarını açıklamaya yönelik araştırmalarının sonucunda 1976 yılında ortaya çıkmıştır. Bu model, biyolojik sisteme ait normalizasyon, ayrıştırabilme ve ayrıntıların saklandığı kısa dönemli hafıza temelleri üzerine kurulmuştur (Kröse ve van der Smagt, 1996).

Beynin bu özelliklerinden yola çıkarak beynin kullandığı sezgisel yaklaşımları matematiksel modele dönüştüren Grossberg ve arkadaşları ART ağlarını oluşturmuşlardır. Sınıflandırma problemlerinin çözümünde kullanılmak üzere geliştirilmiş geri beslemeli ağlardır. ART yapay sinir ağlarında işlemci elemanların çıktıları, katmanlar arasında sürekli olarak ileri geri hareket etmekte ve bu sırada örnek girdi belirlenmiş bir sınıfa uyuyorsa ağ kararlı hale gelmektedir (Gülseçen, 1993).

Hopfield kuralı: 1982 yılında bir John Hopfield tarafından geliştirilmiştir. Hebb Kuralına benzemektedir. Aralarındaki farklılık Hopfield kuralında bağlantı ağırlığında yapılacak değişikliğin büyüklüğünün belirlenmesidir. Buna göre, girdi ve istenilen çıktının ikisi de aktifse veya ikisi de aktif değilse, bağlantı ağırlığı öğrenme oranı kadar artırılır, aksi durumda ise öğrenme oranı kadar azaltılır. Kolay bir yapıya sahip olması ve matematiksel tutarlılığı nedeniyle yapay sinir ağları çalışmalarında önemli yeri olan ve tercih edilen bir modeldir. Her probleme uygulanamamasına rağmen, daha karmaşık modelleri anlamak için giriş bilgisi sağlamaktadır (Gülseçen, 1993). Hopfield ağı tek katmanlı ve geri beslemeli bir yapay sinir ağıdır. İşlemci elemanlar, birbirleriyle tam bağlantı halindedir ve bağlantılar çift yönlüdür (bilgi her iki yönde de akmaktadır) (Baş, 2006).

Vektör nicelemeyi öğrenme (Learning Vector Quantization-LVQ): 1984 yılında Tuevo Kohonen tarafından geliştirilmiştir. Öğrenme kuralı olarak Kohonen öğrenme kuralını kullanmaktadır. Buradaki amaç, bir vektörün belirli sayıda vektör ile gösterilmesidir. LVQ ağının öğrenmesi ile girdi vektörünün hangi vektör seti tarafından temsil edilmesi gerektiğinin bulunması kastedilmektedir. Bu vektör setine referans vektörleri denirse LVQ ağının görevi, öğrenme yolu ile bu referans vektörleri belirlemektir. Yani, girdi vektörlerinin üyesi olabilecekleri vektör sınıfını belirlemektir (Öztemel, 2003).

Çok katmanlı perseptron (Multilayered Perceptron-MLP): Hebb Net, perseptron ve Adaline gibi tek katmanlı yapay sinir ağı modellerine göre doğrusal olmayan problemlere çözüm üretebildiğinden daha geniş kullanım alanları bulmaktadır. Çok katmanlı perseptron, en küçük kareler yöntemine dayalı olan Delta Kuralı'nın genelleştirilmiş şekli .olan Genelleştirilmiş Delta

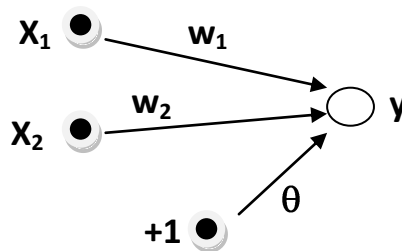
Kuralı adı verilen öğrenme kuralını kullanır. Minsky ve Papert, 1969'da doğrusal olmayan bir ilişki gösteren XOR problemini örnek göstererek tek katmanlı perceptronların doğrusal olmayan elementlerin sınıflandırılmasında kullanılamayacağını göstermişlerdir. Bu olay tek katmanlı ağların periyodunun sonu olmuştur. Buna rağmen birkaç araştırmacı araştırmalarına devam etmiştir. Rumelhart, Hinton ve Williams'ın 1986 yılında geri yayılım algoritmasını geliştirerek XOR problemini çözmesiyle birlikte çalışmalar tekrar hız kazanmıştır (Kröse ve van der Smagt, 1996).

Bu ağ yapıları içerisinde en yaygın kullanım alanı bulan ve bu çalışmada da kullanılan çok katmanlı ileri beslemeli yapay sinir ağlarıdır.

Çok Katmanlı İleri Beslemeli Yapay Sinir Ağları

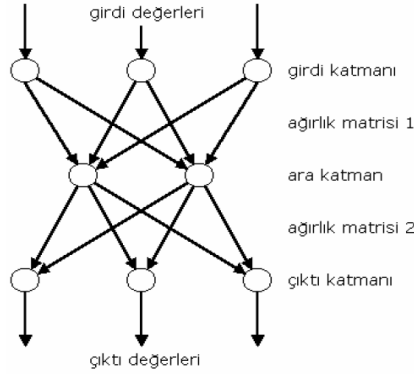
Haykin (1999) yapay sinir ağlarının mimarisini tek katmanlı ileri beslemeli ağ, çok katmanlı ileri beslemeli ağ ve geri beslemeli ağlar olarak sınıflandırmıştır.

Tek katmanlı yapay sinir ağında bir tane ağırlıklı bağlantı katmanı bulunur. Her tek katmanlı yapay sinir ağının bir ya da birden fazla girdisi ve tek bir çıktısı vardır. Her girdi bir ağırlık değeriyle çıktıya bağlıdır. En basit tek katmanlı yapay sinir ağı Şekil 9'da gösterildiği gibi iki girdi ve tek çıktıya sahip ağıdır (Baş, 2006). Tek katmanlı YSA örnek sınıflandırma, tanıma, örnek ilişkilendirme gibi doğrusal olarak ayrılabilen verilerle problemlerin çözülmesinde kullanılabilir. Örneğin, örnek sınıflandırma problemlerinde, her giriş vektörü (örnek, numune) belirli sınıflara ait olabilir ya da olmayabilir. Çıkış birimi için +1 cevabının alınmasıyla örneğin o sınıfa üye olduğu, -1 cevabı alınırsa, örneğin o sınıfa üye olmadığı belirlenir. Bu tip durumlarda, her bir sınıf için bir çıkış birimi vardır (Ocakoğlu,2006).



Şekil 9. Tek katmanlı yapay sinir ağı

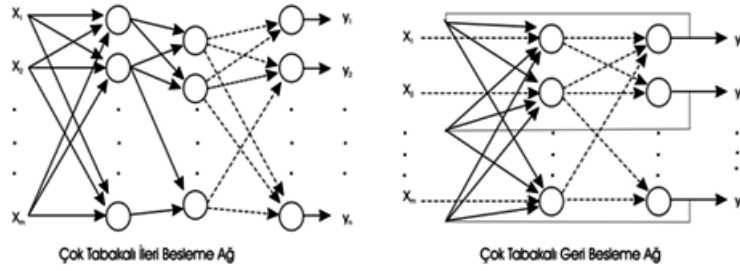
Tek katmanlı sinir ağı doğrusal olmayan problemlerin çözümünde yetersiz kalmakta ve çok katmanlı sinir ağı bu problemin üstesinden gelebilmektedir. Çok katmanlı sinir ağı tek katmanlıya göre çok daha karmaşık problemleri çözebilir; fakat eğitimleri çok daha zor olabilir (Fausett, 1994). Şekil 10'da çok katmanlı yapay sinir ağı modelinin bir örneği gösterilmektedir.



Şekil 10. Çok Katmanlı Ağ Modeli (Fröhlich, 1997; Baş, 2006)

Çok katmanlı ağa eğitim sırasında hem girdiler hem de o girdilere karşılık gelen çıktılar gösterilir. Çok katmanlı ağda bilgiler girdi katmanından ağa sunulur, ara katmanlardan geçerek çıktı katmanına ulaşır ve bu katman tarafından ağın cevabı olarak dış ortama iletilir. Bilgi akışı ileri doğru olduğu için çok katmanlı ağlar ileri beslemeli yapay sinir ağı sınıfına girmektedir (Baş 2006).

Şekil 11'de ileri ve geri beslemeli ağ yapıları gösterilmektedir. İleri beslemeli sinir ağlarında işlem elemanları arasındaki bağlantılar bir döngü oluşturmazlarken geri beslemeli ağlarda bağlantılar döngü oluştururlar. İleri beslemeli ağlar girdi veriye genellikle hızlı bir şekilde karşılık üretir. Geri beslemeli ağlarda ise her döngüde yeni veri kullanılabildiğinden geri beslemeli ağlar girdinin karşılığını yavaş bir şekilde oluştururlar. Bu nedenle geri beslemeli ağların eğitim süreci daha uzun olmaktadır (Yurtoğlu, 2005).



Şekil 11. İleri Beslemeli ve Geri Beslemeli Ağ Yapıları (Yurtoğlu, 2005)

Geri yayılım algoritması, yapay sinir ağları içinde en yaygın olarak kullanılan bir öğrenme algoritmasıdır. Geriye yayılım öğrenme algoritması, benzer fikirler daha önce Werbos (1974), Parker (1985), Cun (1985) gibi bilim adamlarının yayınlamış olduğu makalelerde görünse de 1986 yılında Rumelhart, Hinton ve Williams tarafından geliştirilerek duyurulmuştur. Geri yayılım algoritması çok katmanlı ve ileri beslemeli ağ yapılarının eğitimi amacıyla geliştirilmiştir (Mehrota ve diğerleri, 1997; akt. Saygılı, 2008). Özellikle öngörü ve sınıflandırma problemlerindeki üstün başarısı, kullanım alanlarını yaygınlaştırmıştır. Geriye yayılım öğrenme algoritmasında sigmoid aktivasyon fonksiyonları yaygın olarak kullanılır. Bunun nedeni fonksiyonun belirli bir noktadaki değeri ile onun türevinin değeri arasındaki ilişki öğrenme zamanındaki hesap yükünü azaltır. Lojistik fonksiyon, değerleri 0 ile 1 arasında değişen bir sigmoid fonksiyonudur ve sinir ağları için aktivasyon fonksiyonu olarak çoğunlukla kullanılır (Bishop, 1995). Geri yayılım algoritmasında öğrenme her bir örnek için ağ çıktısı ile hedef çıktısının karşılaştırılması ve hataları ağa geri besleme yoluyla geri vermesi ile olur. Geri yayılım algoritmasında hata, ağdaki ağırlıkların bir fonksiyonu olarak görülür ve hataların kareleri toplamı dereceli azaltma yöntemi kullanılarak, minimize edilmeye çalışılmaktadır. Hata kareleri toplam değerini azaltmak için ağırlıklandırılmalar değiştirilir. Bu algoritma, hataları çıkıştan girişe geriye doğru azaltmaya çalışmasından dolayı geri yayılım ismini almıştır (Curram ve Minger, 1994).

Geriye yayılım algoritmasının temelinde aşağıdaki aşamalar uygulanır:

- Girdi eğitim örneklerinin ileri beslemesi,
- Birleşmiş hatanın hesaplanması ve geriye yayılması,
- Ağırlıkların güncellenmesi.

Eğitimden sonra ağ uygulaması sadece ağın ileri besleme aşamasını gerçekleştirir ve hesaplamaları yapar (Fausett, 1994).

Bu çalışmada tek katmanlı yapay sinir ağlarının doğrusal olmayan problemlerin çözümünde yetersiz kalması, çok katmanlı yapay sinir ağlarının bu problemin üstesinden gelebilmesi ve tek katmanlı sinir ağlarına göre çok daha karmaşık problemleri çözebilmesi nedeniyle çok katmanlı ileri beslemeli yapay sinir ağları kullanılmıştır. Yapay sinir ağı modeli kurulurken öncelikle veriler programa tanıtılmak için uygun hale getirilmiştir. Örneğin sadece 6.maddedeki kayıp değerler için, eğitim setinde kullanılan verilerde hedef değer 1 ise [1 0 0 0], 2 ise [0 1 0 0], 3 ise [0 0 1 0] ve 4 ise [0 0 0 1] matrisleri oluşturulmuştur. Verilerin düzenlenmesi Ek 1'de verilmiştir.

Acuna ve Rodriguez (2004)'in kayıp verinin %5-%15 arasında olduğu durumlarda kayıp verilere değer atama için çeşitli karmaşık yöntemlerin gerektiğini ve %15'ten daha fazla olması durumunda ise kayıp verinin yapılacak herhangi bir yorumu çok ciddi şekilde etkileyeceği görüşü çerçevesinde bu çalışmada %5, %15 ve %30 kayıp verili veri setleri oluşturulmuştur. İlk olarak veri setleri, tek bir değişkende (6.madde seçkisiz olarak alınmıştır) gözenek sayısı seçkisiz olarak %5, %15 ve %30 oranlarında kayıp veri elde edilecek şekilde oluşturulmuştur. Sonuçta tek bir değişkendeki kayıp için 10 girdi (değişken), hedef değer için 4 çıktı oluşacak hale getirilmiştir. Ağ fonksiyonlarından örüntü tanıma ve sınıflandırmada güçlü olan ve doğrusal olmayan problemlerde perseptronlardan daha başarı gösteren 'patternnet' işleme konulmuştur. Doğrusal olmayan transfer fonksiyonuna sahip çok tabakalı ağlar, girdi ve çıktı vektörleri arasındaki doğrusal olmayan ilişkiyi öğrenmelerine odaklandığından ve örüntü tanıma için kullanılan standart ağ olan iki katmanlı ileri beslemeli ağ ve sigmoid fonksiyonu olan hiperbolik tanjant aktivasyon fonksiyonu kullanılmıştır. Veri seti seçkisiz olarak üç alt sete ayrılmıştır. %70'i eğitim seti, eğimi hesaplamak

ve ağın ağırlıkları ve eşik değerini güncellemek için kullanılmıştır. Eğitim sırasında bu veriler ağa sunulmuştur ve hataya göre ağın düzeltilmesi sağlanmıştır. %15'i geçerlik seti olarak kullanılmıştır. Bu set, genelleştirme gelişmeyi durdurduğunda eğitimin durması ve ağın genelleştirilmesini ölçmek için kullanılmıştır. Aşırı uyum olmaması için ağ durdurur. Geçerlik setindeki hata eğitim sürecinde izlenilip eğitimin ilk aşamasında düşmesiyle eğitim setinin hatasını verir. Eğitimin ilk aşamasında geçerlik hatası normal olarak azalır. Ağ aşırı uyum sağlamaya başladığında geçerlik setindeki hata yükselmeye başlar. Ağ ağırlıkları ve eşikleri geçerlik seti hatasının minimumunda kaydedilir. %15'i test seti olarak kullanılmıştır. Test setinin eğitimde hiçbir etkisi yoktur ve ağ genelleştirilmesinin bağımsız testi için kullanılmıştır. Test seti hatası farklı modelleri karşılaştırmak için kullanılmıştır.

Ağ için gizli katmandaki nöron sayısına da karar vermek gerekir. Bunun için deneme-yanılma yoluyla en uygun nöronu bulmak için 2 ile 50 arasındaki değerler denenmiştir. 1000 iterasyon yapılmış ve hatanın minimum olduğu değerdeki nöron sayısı alınmıştır. Çıktı nöronunun sayısı hedef vektörlerinin elementlerinin sayısına eşittir.

Yapay sinir ağının etkililiğini incelemek amacıyla hata kareleri ortalaması ve hata matrisi (confusion matrix) dikkate alınmıştır.

$$F=HKO \text{ (Hata Kareleri Ortalaması)} = \frac{1}{n} \sum_{i=1}^n (t_i - a_i)^2$$

t_i = hedef değer

a_i = çıktı değer

Bu süreç 6. maddesi silinmiş olan öğrencilerin ikinci bir değişkene (5.madde) verdikleri cevaplar da silinerek tekrarlanmıştır. İki değişkende kayıp veri olma durumu için kullanılan verilerin düzenlenme şekli Ek 1'de verilmiştir. Örneğin, 5. ve 6. maddelerdeki kayıp değerlerin her ikisi için de hedef değerler 1 ([1 1]) ise [1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0] matrisi oluşturulmuştur. Bu şekilde toplamda 16 matris kullanılmıştır. Üçüncü bir değişken olarak da 7.madde silinmiş ve Ek 1'de görüldüğü üzere toplamda 64 matris kullanılarak süreç tekrar edilmiştir. Analizler sonucunda %5, %15

ve %30 kayıp verili veri setleri için yapay sinir ağı atama yönteminin uygunluğunda farklılık olup olmadığı belirlenmiştir.

Oluşturulan en uygun yapay sinir ağına kayıp verilere değer ataması için girdi değerleri verilip çıktı değerleri elde edilmiştir. Bu yöntemle kayıp verilere değer ataması yapıldıktan sonra tutum ölçeklerinin geçerliği ve güvenilirliği için yapılan analizler SPSS ve LISREL programlarında yapılmıştır. Geçerlilik için, farklı koşullar altında ölçme aracının faktör yapısında meydana gelen değişimleri belirlemek amacıyla açıklayıcı faktör analizi yapılmıştır. Faktör analizi sonucunda açıklanan varyans, düzeltilmiş madde toplam korelasyonları ve faktör yapısı incelenmiştir. Güvenirlik analizi için maddelerin yapısına bakılmıştır. Konjenerik ölçme modellerinde maddelerin ortalamaları, standart sapmaları ve kovaryansları birbirinden farklıdır (Yurdugül, 2006). Diğer taraftan kullanılacak ölçme modeline karar verirken Konjenerik modelin, eşbiçimli/eşdeğer modelin ve paralel modelin uyumları test edilir ve uyum istatistikleri elde edilir. Uyum istatistikleri sonuçları önce konjenerik modelle eşbiçimli/eşdeğer modeller karşılaştırılır. Eşbiçimli/eşdeğer modelin uyum istatistikleri sonuçları da paralel modelle karşılaştırılır. Uyum istatistiklerindeki düşüşün yeteri kadar fazla olup olmaması ve bu düşüşün anlamlı olup olmamasına göre kullanılacak ölçme modeli seçilir (Graham, 2006). Ek 2'de görüldüğü üzere maddelere ilişkin ortalamaların, standart sapmaların ve kovaryansların farklı olduğu görülmüştür. Ayrıca ölçme modellerinin uyum iyiliği değerleri incelendiğinde (Ek 2) sonuçların benzer olduğu görülmektedir. Bu nedenle hem Cronbach alfa hem de omega (ω) güvenilirlik katsayıları hesaplanmıştır. Cronbach alfa güvenilirlik katsayısı için SPSS programı, omega güvenilirlik katsayısı için LISREL 9.1 programı kullanılmıştır. Gerçek veri seti için de güvenilirlik ve geçerlik çalışmaları yapılmış ve bu yöntemle elde edilen veri setleriyle yapılan analiz sonuçlarıyla gerçek veri seti analizi sonuçları karşılaştırılmıştır.

BÖLÜM IV

BULGULAR VE YORUMLAR

Bu bölümde, araştırma verilerinin amaçlara uygun olarak analiz edilmesi sonucunda elde edilen bulgulara ve yorumlara yer verilmiştir. Öncelikle yapay sinir ağlarının kayıp veri oranlarına göre etkililiğini gösteren hata matrisleri ve hata kareleri ortalamalarına yer verilmiştir. Sonrasında kayıp verinin olmadığı gerçek veri setiyle, yapay sinir ağıyla kayıp verilere değer atanarak elde edilen veri setlerinde ortaya çıkan faktörler, özdeğerler, açıklanan varyans oranları ve maddelere ilişkin faktör yük değerleri karşılaştırmalı olarak incelenmiştir. Ardından düzeltilmiş madde-toplam korelasyonları ve Cronbach alfa ve omega güvenirlik katsayılarının karşılaştırılmasına ilişkin bulgular sunulmuştur.

Veri Setlerindeki Kayıp Veri Oranına Göre Yapay Sinir Ağlarının Etkililiği

Kayıp değerlere sahip veri setlerinde bu değerlere en uygun şekilde değer atanması için yapay sinir ağları oluşturulmuş ve oluşturulan ağlarının hata matrisleri aşağıda sunulmuştur.

Çizelge 3. Tek değişkende (6.maddede) %5 kayıp veri oranına göre Hata Matrisi

Hata Matrisi

Çıktı Sınıfı	1	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
	2	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
	3	9 1.9%	16 3.4%	148 31.2%	16 3.4%	78.3% 21.7%
	4	4 0.8%	2 0.4%	9 1.9%	271 57.1%	94.8% 5.2%
		0.0% 100%	0.0% 100%	94.3% 5.7%	94.4% 5.6%	88.2% 11.8%
	1	2	3	4		
		Hedef Sınıfı				

Çizelge 3'te görüldüğü üzere tek bir değişkende (6.madde) %5 kayıp veri olduğu takdirde yapay sinir ağı toplamda %88.2 doğrulukla kayıp veriye değer atama gücüne sahiptir. Yeşil alanlar doğruluk yüzdesini gösterirken kırmızı alanlar yanlış sınıflandırılma yüzdesini göstermektedir. Örneğin, hedef (gerçek) değeri 3 olan verilerin 148 tanesi (tüm verilerin %31.2'si) doğru, 9 tanesi ise yanlış tahmin edilmiş ve 4 değerini vermiştir. Mavi alan toplamda gösterdiği doğru sınıflandırma yüzdesini vermektedir. Yapay sinir ağı 1 (4.satır 1.sütunda gösterildiği üzere) ile 2 (4.satır 2.sütunda gösterildiği üzere) hedef değerleri için hiçbir doğru sınıflandırma yapamazken 3 (4.satır 3.sütunda gösterildiği üzere) hedef değeri için yapay sinir ağının verdiği değerler arasında %94.3 ve 4 (4.satır 4.sütunda gösterildiği üzere) hedef değeri için %94.4 oranlarında doğru sınıflandırma yapmıştır. 1. hedef değeri için 9 tane 3 ve 4 tane 4 değeri vererek tamamen yanlış sınıflandırmalarda bulunmuştur. 2. hedef değeri için 16 tane 3 ve 2 tane 4 değeri vererek tamamen yanlış sınıflandırmalarda bulunmuştur. 3. hedef değeri için 9 tane 4

değeri vererek yanlış sınıflandırmada bulunurken 148 tane 3 değeri vererek doğru sınıflandırmada bulunmuştur. Bu da 157 tane 3 değerinin 148'ini doğru tahminde bulunarak %94.3 doğru tahminde bulunduğunu göstermektedir. Aynı şekilde 4. hedef değeri için 16 tane 3 değeri vererek yanlış sınıflandırmada bulunurken 271 tane 4 değeri vererek doğru sınıflandırmada bulunmuştur. Bu da 287 tane 4 değerinin 271'ini doğru tahminde bulunarak %94.4 doğru tahminde bulunduğunu göstermektedir.

Çizelge 4. Tek değişkende (6.maddede) %15 kayıp veri oranına göre Hata Matrisi

Hata Matrisi

Çıktı Sınıfı	1	4 0.9%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	2	0 0.0%	3 0.7%	0 0.0%	0 0.0%	100% 0.0%
	3	1 0.2%	9 2.1%	116 27.3%	15 3.5%	82.3% 17.7%
	4	7 1.6%	5 1.2%	19 4.5%	246 57.9%	88.8% 11.2%
		33.3% 66.7%	17.6% 82.4%	85.9% 14.1%	94.3% 5.7%	86.8% 13.2%
	1	2	3	4		
	Hedef Sınıfı					

Çizelge 4'te görüldüğü üzere tek bir değişkende (6.madde) %15 kayıp veri olduğu takdirde yapay sinir ağı toplamda %86.8 doğrulukla kayıp veriye değer atama gücüne sahiptir. Yapay sinir ağı 1 hedef değeri için verdiği değerler arasında %33.3, 2 hedef değeri için %17.6, 3 hedef değeri için %85.9 ve 4 hedef değeri için %94.3 oranlarında doğru sınıflandırma yapmıştır.

Çizelge 5. Tek değişkende (6.maddede) %30 kayıp veri oranına göre Hata Matrisi

Hata Matrisi

Çıktı Sınıfı	1	4 1.1%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	2	0 0.0%	4 1.1%	0 0.0%	0 0.0%	100% 0.0%
	3	3 0.9%	4 1.1%	104 29.7%	14 4.0%	83.2% 16.8%
	4	3 0.9%	5 1.4%	8 2.3%	201 57.4%	92.6% 7.4%
		40.0% 60.0%	30.8% 69.2%	92.9% 7.1%	93.5% 6.5%	89.4% 10.6%
	1	2	3	4		
	Hedef Sınıfı					

Çizelge 5'te görüldüğü üzere tek bir değişkende (6.madde) %30 kayıp veri olduğu takdirde yapay sinir ağı toplamda %89.4 doğrulukla kayıp veriye değer atama gücüne sahiptir. Yapay sinir ağı 1 hedef değeri için verdiği değerler arasında %40, 2 hedef değeri için %30.8, 3 hedef değeri için %92.9 ve 4 hedef değeri için %93.5 oranlarında doğru sınıflandırma yapmıştır.

Bu değerler modelin tek değişkende %5, %15 ve %30 kayıp değerler için oldukça uygun ve gerçek değerlere yakın değerler verdiğini göstermektedir. Yapay sinir ağları yöntemi %5 kayıp veri oranında (%88.2 doğruluk) %15 kayıp veri oranından (%86.8 doğruluk) daha etkili sonuç verirken %15 kayıp veri oranında %30 kayıp veri oranından (%89.4 doğrulukla) daha az etkili sonuç vermiştir. Yapay sinir ağları yönteminin tek değişkende kayıp değerlerin oranının değişmesiyle orantılı olarak tahmin gücünde farklılık göstermediği ifade edilebilir. Yani kayıp değer oranı arttıkça yapay sinir ağları yönteminin etkililiği azalmamakta ya da artmamaktadır.

Çizelge 6. İki değişkende (5. ve 6.maddede) %5 kayıp veri oranına göre Hata Matrisi

Hata Matrisi

Çıktı Sınıfı	Hedef Sınıfı															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	2 0.4%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	50.0%
2	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN%
3	0 0.0%	0 0.0%	2 0.4%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	36.7%
4	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN%
6	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN%
7	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN%
8	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN%
9	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN%
10	2 0.4%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	33.3%
11	1 0.2%	0 0.0%	0 0.0%	1 0.2%	1 0.2%	4 0.8%	31 6.5%	9 1.9%	0 0.0%	9 1.9%	86 18.1%	22 4.6%	0 0.0%	0 0.0%	5 1.1%	8 1.7%
12	3 0.6%	0 0.0%	4 0.8%	8 1.7%	0 0.0%	0 0.0%	8 1.7%	19 4.0%	1 0.2%	2 0.4%	8 1.7%	113 23.8%	1 0.2%	0 0.0%	5 1.1%	26 5.5%
13	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN%
14	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN%
15	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	2 0.4%	0 0.0%
16	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	3 0.6%	1 0.2%	0 0.0%	0 0.0%	1 0.2%	14 2.9%	0 0.0%	1 0.2%	1 0.2%	66 13.9%
	25.0%	0.0%	33.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	8.3%	30.5%	75.8%	0.0%	0.0%	15.4%	56.0%
	75.0%	100%	66.7%	100%	100%	100%	100%	100%	100%	91.7%	9.5%	24.2%	100%	100%	84.6%	34.0%

Çizelge 6'da görüldüğü üzere aynı bireylerin iki değişkende (5. ve 6. Maddelerinde) %5'er kayıp veri olduğu takdirde yapay sinir ağı toplamda %57.3 doğrulukla kayıp veriye değer atama gücüne sahiptir. Yapay sinir ağı [1 1] hedef değeri için verdiği değerler arasında %25, [1 3] hedef değeri için %33.3, [3 2] hedef değeri için %8.3, [3 3] hedef değeri için %90.5, [3 4] hedef değeri için %75.8, [4 3] hedef değeri için %15.4 ve [4 4] hedef değeri için %66 oranlarında doğru sınıflandırırken; [1 2], [1 4], [2 1], [2 2], [2 3], [2 4], [3 1], [4 1] ve [4 2] hedef değerleri için hiçbir doğru sınıflandırmada bulunamamıştır. Diğer taraftan yapay sinir ağının bu iki maddeden en az birini doğru sınıflandırma yüzdesi 89.1'dir. Çizelgedeki köşegen değerleri her iki kayıp verinin de doğru tahmin edildiği durumu

göstermektedir; ancak kayıp verilerin sadece birinin doğru tahmin edildiği durumlar köşegen değerlerinde görülmektedir. Örneğin, Çizelge 6'daki matrisin 11. sütununda görüldüğü üzere hedef değeri 11 (kayıp veri değerleri [3 3]) olan 95 veri bulunmaktadır ve 11. sütunun 11. satır ile kesişiminde görüldüğü üzere 86 tanesi tam olarak doğru tahmin edilmiştir. Bunun yanında hedef değeri 11 olan bu 95 verinin sekiz tanesi 12 (kayıp veri değerleri [3 4]), bir tanesi de 16 (kayıp veri değerleri [4 4]) olarak tahmin edilmiştir. Buna göre bu 95 verinin 12 tanesinde iki kayıp veriden sadece biri (5. madde) doğru tahmin edilmiştir. Benzer şekilde diğer verilere bakılarak en az bir değer doğru tahmin edilme yüzdeleri hesaplanmış ve en az bir değer doğru tahmin edilme yüzdesi 89.1 olarak bulunmuştur.

Çizelge 7. İki değişkende (5. ve 6.maddede) %15 kayıp veri oranına göre Hata Matrisi

Hata Matrisi

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
1	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN%
2	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN%
3	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN%
4	0 0.0%	0 0.0%	0 0.0%	3 0.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN%
6	1 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	33.3%
7	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	1 0.2%	16 3.8%	0 0.0%	0 0.0%	1 0.2%	4 0.9%	2 0.5%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	59.3%
8	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	10 2.4%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	33.3%
9	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100%
10	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN%
11	2 0.5%	0 0.0%	1 0.2%	2 0.5%	1 0.2%	2 0.5%	13 3.1%	9 2.1%	0 0.0%	9 2.1%	63 14.8%	15 3.5%	0 0.0%	1 0.2%	3 0.7%	10 2.4%	48.1%
12	2 0.5%	0 0.0%	1 0.2%	2 0.5%	0 0.0%	0 0.0%	5 1.2%	6 1.4%	1 0.2%	1 0.2%	14 3.3%	110 25.9%	1 0.2%	0 0.0%	5 1.2%	13 3.1%	58.3%
13	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN%
14	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN%
15	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN%
16	2 0.5%	0 0.0%	1 0.2%	1 0.2%	0 0.0%	0 0.0%	1 0.2%	1 0.2%	0 0.0%	1 0.2%	2 0.5%	9 2.1%	0 0.0%	0 0.0%	3 0.7%	66 15.5%	75.9%
	0.0%	NaN%	0.0%	37.5%	0.0%	25.0%	44.4%	38.5%	50.0%	0.0%	75.0%	80.9%	0.0%	0.0%	0.0%	72.5%	53.5%
	100%	NaN%	100%	62.5%	100%	75.0%	55.6%	61.5%	50.0%	100%	25.0%	19.1%	100%	100%	100%	27.5%	36.5%
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	

Hedef Sınıfı

Çizelge 8'de görüldüğü üzere aynı bireylerin iki değişkende (5. ve 6. Maddelerinde) %30'ar kayıp veri olduğu takdirde yapay sinir ağı toplamda %59.7 doğrulukla kayıp veriye değer atama gücüne sahiptir. Yapay sinir ağı [1 1] hedef değeri için verdiği değerler arasında %40, [1 4] hedef değeri için %37.5, [2 3] hedef değeri için %43.8, [2 4] hedef değeri için %25, [3 2] hedef değeri için %50, [3 3] hedef değeri için %83.8, [3 4] hedef değeri için %72 ve [4 4] hedef değeri için %60.5 oranlarında doğru sınıflandırırken; [1 3], [2 1], [2 2], [3 1], [4 1], [4 2] ve [4 3] hedef değerleri için hiçbir doğru sınıflandırmada bulunamamıştır. [1 2] değeri için veri setinde bulunmadığından doğru sınıflandırma gibi bir durum da olmamıştır. Diğer taraftan yapay sinir ağının iki değişkenden (5. ve 6. maddelerinden) en az birini doğru sınıflandırma yüzdesi 89.3'tür.

Yukarıda belirtilen değerler modelin iki değişkende (5. ve 6. maddeler) %5, %15 ve %30 kayıp değerler için oldukça uygun ve gerçek değerlere yakın değerler verdiğini göstermektedir. Yapay sinir ağları yöntemi %5 kayıp veri oranında (%89.1 doğruluk) %15 kayıp veri oranından (%86.5 doğruluk) daha etkili sonuç verirken %15 kayıp veri oranında %30 kayıp veri oranından (%89.3 doğrulukla) daha az etkili sonuç vermiştir. Yapay sinir ağları yönteminin iki değişkende kayıp değerlerin oranının değişmesiyle orantılı olarak tahmin gücünde farklılık göstermediği ifade edilebilir. Yani kayıp değer oranı arttıkça yapay sinir ağları yönteminin etkililiği azalmamakta ya da artmamaktadır.

Üç değişkende (5., 6. ve 7. maddelerde) %5, %15 ve %30 kayıp veri oranına göre Hata Matrislerine, boyutunun büyük olması nedeniyle Ek 3'te yer verilmiştir. Aynı bireylerin üç değişkende (5., 6. ve 7. maddelerinde) %5'er kayıp veri olduğu takdirde yapay sinir ağı toplamda %36.8, üç değişkenden (5., 6. ve 7. Maddelerinden) en az bir maddesini %91.4 doğrulukla kayıp veriye değer atama gücüne sahip olduğu görülmüştür. %15'er kayıp veri olduğu takdirde yapay sinir ağı toplamda %39.3, üç değişkenden (5., 6. ve 7. maddelerinden) en az bir maddesini %91.3 doğrulukla kayıp veriye değer atama gücüne sahip olduğu görülmüştür. %30'ar kayıp veri olduğu takdirde ise yapay sinir ağı

toplamda %42.6, üç deęişkenden (5., 6. ve 7. maddelerinden) en az bir maddesini %88.9 doęrulukla kayıp veriye deęer atama gücüne sahip olduęu görölmüştür.

Yukarıda belirtilen deęerler modelin üç deęişkende %5, %15 ve %30 kayıp deęerler için oldukça uygun ve gerçek deęerlere yakın deęerler verdięini göstermektedir. Yapay sinir aęları yöntemi %5 kayıp veri oranında (%91.4 doęruluk), %15 kayıp veri oranından (%91.3 doęruluk) ve %30 kayıp veri oranından (%88.9 doęrulukla) daha etkili sonuç vermiştir. Yapay sinir aęları yöntemi %30 kayıp veride daha az bir doęrulukla olsa da kayıp deęerlerin oranıyla orantılı olarak tahmin gücünde farklılık göstermemektedir.

Yapay Sinir Aęları Yöntemiyle Kayıp Verilere Deęer Ataması Yapılarak Elde Edilen Tam Veri Setine Yapılan Geçerlik Analizi Bulguları

Oluşturulan yapay sinir aęlarından elde edilen kayıp verilerin deęerleri veri setinde yerleştirdikten sonra faktör analizine tabi tutulmuştur. Veri setlerinin faktör analizine uygun olup olmadıęını belirlemek amacıyla Kaiser-Meyer-Olkin (KMO) katsayısı ve Barlett testi incelenmiştir. Faktör analizinde hem gerçek veri seti için hem de kayıp veriye deęer ataması yapılan veri setleri için Kaiser-Meyer-Olkin (KMO) katsayısı .80'dan büyük ve Barlett testi anlamlı çıkmıştır ($p < .01$). Böylelikle veri setlerinin faktörleştirilebilir oldukları görölmektedir. Yapılan açımlayıcı faktör analizi sonucunda elde edilen gerçek veri setinin sonuçlarıyla kayıp veriye deęer ataması yapılmış veri setinin sonuçları karşılaştırılmıştır. Faktör analizi sonucunda açıklanan varyans ve faktör yapısı karşılaştırılmıştır.

Gerçek veri setine yapılan faktör analizine göre ölçme aracının tek faktör altında toplanan 10 (2. madde binişiklik gösterdięinden çıkartılmıştır) maddeden oluştuęu görölmektedir.

Çizelge 9. Tek Değişkende Farklı Oranlarda Kayıp Verilere Değer Ataması Yapılmış Veri Setine ve Gerçek Veri Setine Yapılan Faktör Analizlerine İlişkin Sonuçlar

Faktör Yük Değerleri				
Maddeler	Gerçek Veri Seti	%5 Kayıp	%15 Kayıp	%30 Kayıp
1	.592	.592	.590	.594
2				
3	.710	.708	.706	.702
4	.635	.636	.642	.645
5	.600	.600	.599	.593
6	.626	.627	.633	.623
7	.645	.644	.646	.647
8	.700	.664	.702	.705
9	.531	.702	.536	.533
10	.573	.535	.570	.570
11	.655	.572	.652	.654
Açıklanan Toplam Varyans (%)	39.568	39.641	39.576	39.523

Çizelge 9'da görüldüğü üzere, gerçek veri setiyle yapılan faktör analizine göre maddelerin faktör yük değerleri .531 ile .710 arasında değişmektedir. En yüksek faktör yük değerine 3. madde sahipken en düşük faktör yük değerine 9. madde sahiptir. Bu koşul altında faktöre ilişkin özdeğer 3.957 ve açıklanan varyans oranı ise %39.568'dir.

Tek değişkende %5'lik kayıp veri setine yapılmış atama yoluyla elde edilen veri seti için yapılan faktör analizine göre de ölçme aracının tek faktör altında toplanan 10 (2. madde binişiklik gösterdiğinden çıkartılmıştır) maddeden oluştuğu görülmektedir. Bu durumda maddelerin faktör yük değerleri .535 ile .708 arasında değişmektedir. En yüksek faktör yük değerine 3. madde sahipken en düşük faktör yük değerine 10. madde sahiptir. Bu koşul altında faktöre ilişkin özdeğer 3.964 ve açıklanan varyans oranı ise %39.641'dir.

Tek değişkende %15'lik kayıp veri setine yapılmış atama yoluyla elde edilen veri seti için yapılan faktör analizine göre de ölçme aracının tek faktör altında toplanan 10 (2. madde binişiklik gösterdiğinden çıkartılmıştır)

maddeden oluştuğu görülmektedir. Bu durumda maddelerin faktör yük değerleri .536 ile .706 arasında değişmektedir. En yüksek faktör yük değerine 3. madde sahipken en düşük faktör yük değerine 9. madde sahiptir. Bu koşul altında faktöre ilişkin özdeğer 3.958 ve açıklanan varyans oranı ise %39.576'dır.

Tek değişkende %30'luk kayıp veri setine yapılmış atama yoluyla elde edilen veri seti için yapılan faktör analizine göre de ölçme aracının tek faktör altında toplanan 10 (2. madde binişiklik gösterdiğinden çıkartılmıştır) maddeden oluştuğu görülmektedir. Bu durumda maddelerin faktör yük değerleri .533 ile .705 arasında değişmektedir. En yüksek faktör yük değerine 8. madde sahipken en düşük faktör yük değerine 9. madde sahiptir. Bu koşul altında faktöre ilişkin özdeğer 3.952 ve açıklanan varyans oranı ise %39.523'dir.

Çizelge 10. İki Değişkende Farklı Oranlarda Kayıp Verilere Değer Ataması Yapılmış Veri Setine ve Gerçek Veri Setine Yapılan Faktör Analizlerine İlişkin Sonuçlar

Maddeler	Faktör Yük Değerleri			
	Gerçek Veri Seti	%5 Kayıp	%15 Kayıp	%30 Kayıp
1	.592	.589	.586	.616
2				.706
3	.710	.705	.706	.723
4	.635	.642	.647	.679
5	.600	.568	.577	.604
6	.626	.619	.603	.609
7	.645	.638	.644	.659
8	.700	.703	.709	.704
9	.531	.540	.536	.485
10	.573	.571	.563	
11	.655	.655	.652	.639
Açıklanan Toplam Varyans (%)	39.568	39.106	39.032	41.713

Çizelge 10'da görüldüğü üzere iki değişkende %5'lik kayıp veri setine yapılmış atama yoluyla elde edilen veri seti için yapılan faktör analizine göre de ölçme aracının tek faktör altında toplanan 10 (2. madde, 5. madde ve 10.maddeler binişiklik göstermiş olup bu maddelerin yük değerleri arasındaki

farkın en az olduđu 2.madde analizden ıkartılıp tekrardan faktör analizine tabi tutulduğunda 5. ve 10. maddeler binişiklik göstermeden faktörler altında toplanmışlardır) maddeden oluştuđu görülmektedir. Bu durumda maddelerin faktör yük değerleri .540 ile .705 arasında deđişmektedir. En yüksek faktör yük değerine 3. madde sahipken en düşük faktör yük değerine 9. madde sahiptir. Bu koşul altında faktöre ilişkin özdeđer 3.911 ve açıklanan varyans oranı ise %39.106'dir.

%15'lik kayıp veri setine yapılmış atama yoluyla elde edilen veri seti için yapılan faktör analizine göre de ölçme aracının tek faktör altında toplanan 10 (2. madde binişiklik gösterdiğinden ıkartılmıştır) maddeden oluştuđu görülmektedir. Bu durumda maddelerin faktör yük değerleri .536 ile .709 arasında deđişmektedir. En yüksek faktör yük değerine 8. madde sahipken en düşük faktör yük değerine 9. madde sahiptir. Bu koşul altında faktöre ilişkin özdeđer 3.903 ve açıklanan varyans oranı ise %39.032'dir.

%30'luk kayıp veri setine yapılmış atama yoluyla elde edilen veri seti için yapılan faktör analizine göre de ölçme aracının tek faktör altında toplanan 10 (2. madde ve 10.maddeler binişiklik göstermiş olup bu maddelerin yük değerleri arasındaki farkın en az olduđu 10.madde analizden ıkartılıp tekrardan faktör analizine tabi tutulduğunda 2. madde binişiklik göstermeden faktör altında toplanmıştır) maddeden oluştuđu görülmektedir. Bu durumda maddelerin faktör yük değerleri .485 ile .723 arasında deđişmektedir. En yüksek faktör yük değerine 3. madde sahipken en düşük faktör yük değerine 9. madde sahiptir. Bu koşul altında faktöre ilişkin özdeđer 4.171 ve açıklanan varyans oranı ise %41.713'dir.

Çizelge 11. Üç Değişkende Farklı Oranlarda Kayıp Verilere Değer Ataması Yapılmış Veri Setine ve Gerçek Veri Setine Yapılan Faktör Analizlerine İlişkin Sonuçlar

Maddeler	Faktör Yük Değerleri			
	Gerçek Veri Seti	%5 Kayıp	%15 Kayıp	%30 Kayıp
1	.592	.586	.583	.606
2				.713
3	.710	.706	.713	.733
4	.635	.640	.657	.676
5	.600	.582	.540	.597
6	.626	.628	.639	.599
7	.645	.634	.640	.686
8	.700	.704	.702	.704
9	.531	.537	.521	.456
10	.573	.572	.560	
11	.655	.652	.654	.646
Açıklanan Toplam Varyans (%)	39.568	39.227	38.952	41.770

Çizelge 11'de görüldüğü üzere üç değişkende %5'lik kayıp veri setine yapılmış atama yoluyla elde edilen veri seti için yapılan faktör analizine göre de ölçme aracının tek faktör altında toplanan 10 (2. madde binişiklik gösterdiğinden çıkartılmıştır) maddeden oluştuğu görülmektedir. Bu durumda maddelerin faktör yük değerleri .537 ile .706 arasında değişmektedir. En yüksek faktör yük değerine 3. madde sahipken en düşük faktör yük değerine 9. madde sahiptir. Bu koşul altında faktöre ilişkin özdeğer 3.923 ve açıklanan varyans oranı ise %39.227'dir.

%15'lik kayıp veri setine yapılmış atama yoluyla elde edilen veri seti için yapılan faktör analizine göre de ölçme aracının tek faktör altında toplanan 10 (2. madde ve 10.maddeler binişiklik göstermiş olup bu maddelerin yük değerleri arasındaki farkın en az olduğu 2.madde analizden çıkartılıp tekrardan faktör analizine tabi tutulduğunda 10. madde binişiklik göstermeden faktör altında toplanmıştır) maddeden oluştuğu görülmektedir. Bu durumda maddelerin faktör yük değerleri .521 ile .713 arasında değişmektedir. En yüksek faktör yük değerine 3. madde sahipken en düşük

faktör yük değerine 9. madde sahiptir. Bu koşul altında faktöre ilişkin özdeğer 3.895 ve açıklanan varyans oranı ise %38.952'dir.

%30'luk kayıp veri setine yapılmış atama yoluyla elde edilen veri seti için yapılan faktör analizine göre de ölçme aracının tek faktör altında toplanan 10 (2. madde, 6. madde ve 10.maddeler binişiklik göstermiş olup bu maddelerin yük değerleri arasındaki farkın en az olduğu 10.madde analizden çıkartılıp tekrardan faktör analizine tabi tutulduğunda 2. ve 6. maddeler binişiklik göstermeden faktörler altında toplanmışlardır) maddeden oluştuğu görülmektedir. Bu durumda maddelerin faktör yük değerleri .456 ile .733 arasında değişmektedir. En yüksek faktör yük değerine 3. madde sahipken en düşük faktör yük değerine 9. madde sahiptir. Bu koşul altında faktöre ilişkin özdeğer 4.117 ve açıklanan varyans oranı ise %41.770'dir.

Bulgular genel olarak değerlendirildiğinde, bu çalışma kapsamında yapı geçerliği incelenen ölçme aracının, kayıp değerlerin olmadığı orijinal veri setiyle elde edilen "tek faktörlü" yapısının, farklı oranlarda kayıp değerlere yapay sinir ağları yönteminin değerler atadığı durumlar için de "tek faktörlü" olarak elde edildiği görülür. Farklı oranlarda kayıp değer içeren veri setleri ile gerçekleştirilen temel bileşenler analizi yöntemine dayalı açıklayıcı faktör analizlerine ilişkin bulgular genel olarak değerlendirildiğinde, en düşük (9.madde) ve en yüksek (3.madde) faktör yük değeri veren maddelerin neredeyse tüm koşullarda tutarlılık gösterdiği ifade edilebilir. Açıklanan varyans oranlarına bakıldığında dikkate değer bir değişiklik olmadığı, iki ve üç değişkenli %30 oranında kayıp veriye değer ataması yapılmış veri setlerinde açıklanan varyansın biraz daha yüksek çıktığı ifade edilebilir.

Yapay Sinir Ağları Yöntemiyle Kayıp Verilere Değer Ataması Yapılarak Elde Edilen Tam Veri Setine Yapılan Güvenirlik Analizi Bulguları

Faktör analizi sonucunda öncelikle madde analizi için düzeltilmiş madde toplam korelasyonlarına bakılmıştır. Güvenirlik analizi için Cronbach α ve omega (ω) güvenirlik katsayıları hesaplanmıştır. Madde toplam korelasyonları ile güvenirlik katsayıları aşağıda verilmiştir.

Çizelge 12. Tek Değişkende Farklı Oranlarda Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki ve Gerçek Veri Setindeki Maddelere İlişkin Düzeltilmiş Madde-Toplam Korelasyonu ile Güvenirlik Katsayıları

Düzeltilmiş madde toplam korelasyonları ile güvenirlik katsayıları				
Maddeler	Gerçek Veri Seti	%5 Kayıp	%15 Kayıp	%30 Kayıp
1	.496	.496	.494	.496
2	.598	.599	.598	.599
3	.628	.627	.625	.622
4	.538	.539	.543	.545
5	.493	.492	.492	.487
6	.514	.515	.520	.511
7	.552	.551	.553	.554
8	.590	.591	.591	.593
9	.404	.406	.407	.404
10	.460	.459	.457	.457
11	.549	.548	.547	.549
Cronbach α güvenirlik katsayısı	.847	.847	.847	.846
Omega (ω) güvenirlik katsayısı	.848	.848	.848	.849

Çizelge 12’de görüldüğü üzere gerçek veride, düzeltilmiş madde-toplam korelasyonları .404 ile .628 arasında değişmektedir. Cronbach α güvenirlik katsayısı .847, omega güvenirlik katsayısı .848’dir. Tek değişkende %5’lik kayıp veri setine yapılmış atama yoluyla elde edilen veri seti için düzeltilmiş madde-toplam korelasyonları .406 ile .627 arasında değişmektedir. Cronbach α güvenirlik katsayısı .847, omega güvenirlik katsayısı .848’dir. Tek değişkende %15’lik kayıp veri setine yapılmış atama yoluyla elde edilen veri seti için düzeltilmiş madde-toplam korelasyonları .407 ile .625 arasında değişmektedir. Cronbach α güvenirlik katsayısı .847, omega güvenirlik katsayısı .848’dir. Tek değişkende %30’luk kayıp veri setine yapılmış atama yoluyla elde edilen veri seti için düzeltilmiş madde-toplam

korelasyonları .404 ile .622 arasında değişmektedir. Cronbach α güvenilirlik katsayısı .846, omega güvenilirlik katsayısı .849'dir. Yapay sinir ağırları yöntemiyle tek değişkende kayıp verilere değer ataması yapılmış veri setlerinde madde toplam korelasyonunda ve güvenilirlik katsayılarında dikkate değer değişiklik meydana gelmemiş, gerçek veri setiyle tutarlı sonuçlar elde edilmiştir.

Çizelge 13. İki Değişkende Farklı Oranlarda Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki ve Gerçek Veri Setindeki Maddelere İlişkin Düzeltilmiş Madde-Toplam Korelasyonu ile Güvenirlik Katsayıları

Düzeltilmiş Madde toplam korelasyonları ile güvenilirlik katsayıları				
Maddeler	Gerçek Veri Seti	%5 Kayıp	%15 Kayıp	%30 Kayıp
1	.496	.492	.489	.496
2	.598	.594	.596	.593
3	.628	.623	.624	.625
4	.538	.543	.547	.555
5	.493	.458	.467	.515
6	.514	.505	.489	.528
7	.552	.546	.550	.551
8	.590	.590	.596	.600
9	.404	.409	.405	.399
10	.460	.457	.451	.464
11	.549	.549	.545	.548
Cronbach α güvenilirlik katsayısı	.847	.844	.844	.849
Omega (ω) güvenilirlik katsayısı	.848	.846	.846	.851

Çizelge 13'te görüldüğü üzere iki değişkende %5'lik kayıp veri setine yapılmış atama yoluyla elde edilen veri seti için düzeltilmiş madde-toplam korelasyonları .409 ile .623 arasında değişmektedir. Cronbach α güvenilirlik katsayısı .844, omega güvenilirlik katsayısı .846'dır. İki değişkende %15'lik kayıp veri setine yapılmış atama yoluyla elde edilen veri seti için düzeltilmiş madde-toplam korelasyonları .405 ile .624 arasında değişmektedir. Cronbach α güvenilirlik katsayısı .844, omega güvenilirlik katsayısı .846'dır. İki değişkende %30'luk kayıp veri setine yapılmış atama yoluyla elde edilen veri

seti için düzeltilmiş madde-toplam korelasyonları .399 ile .625 arasında değişmektedir. Cronbach α güvenilirlik katsayısı .849, omega güvenilirlik katsayısı .851'dir. Yapay sinir ağları yöntemiyle iki değişkende kayıp verilere değer ataması yapılmış veri setlerinde madde toplam korelasyonunda ve güvenilirlik katsayılarında dikkate değer değişiklik meydana gelmemiş, gerçek veri setiyle tutarlı sonuçlar elde edilmiştir.

Çizelge 14. Üç Değişkende Farklı Oranlarda Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki ve Gerçek Veri Setindeki Maddelere İlişkin Düzeltilmiş Madde-Toplam Korelasyonu ile Güvenirlik Katsayıları

Düzeltilmiş Madde toplam korelasyonları ile güvenilirlik katsayıları				
Maddeler	Gerçek Veri Seti	%5 Kayıp	%15 Kayıp	%30 Kayıp
1	.496	.490	.486	.489
2	.598	.595	.597	.599
3	.628	.625	.629	.634
4	.538	.541	.556	.555
5	.493	.472	.428	.500
6	.514	.514	.526	.516
7	.552	.540	.546	.571
8	.590	.592	.589	.604
9	.404	.407	.393	.378
10	.460	.458	.448	.454
11	.549	.546	.547	.555
Cronbach α güvenilirlik katsayısı	.847	.844	.843	.848
Omega (ω) güvenilirlik katsayısı	.848	.846	.844	.850

Çizelge 14'te görüldüğü üzere üç değişkende %5'lik kayıp veri setine yapılmış atama yoluyla elde edilen veri seti için düzeltilmiş madde-toplam korelasyonları .407 ile .625 arasında değişmektedir. Cronbach α güvenilirlik katsayısı .844, omega güvenilirlik katsayısı .846'dır. Üç değişkende %15'lik kayıp veri setine yapılmış atama yoluyla elde edilen veri seti için düzeltilmiş madde-toplam korelasyonları .393 ile .629 arasında değişmektedir. Cronbach α güvenilirlik katsayısı .843, omega güvenilirlik katsayısı .844'tür. Üç değişkende %30'luk kayıp veri setine yapılmış atama yoluyla elde edilen veri

seti için düzeltilmiş madde-toplam korelasyonları .378 ile .634 arasında değişmektedir. Cronbach α güvenilirlik katsayısı .848, omega güvenilirlik katsayısı .850'dir. Yapay sinir ağları yöntemiyle üç değişkende kayıp verilere değer ataması yapılmış veri setlerinde madde toplam korelasyonunda ve güvenilirlik katsayılarında dikkate değer değişiklik meydana gelmemiş, gerçek veri setiyle tutarlı sonuçlar elde edilmiştir.

Bulgular genel olarak değerlendirildiğinde, bu çalışma kapsamında güvenilirliği incelenen ölçme aracının, kayıp değerlerin olmadığı orijinal veri setiyle elde edilen güvenilirlik katsayıları ile madde-toplam korelasyonları, farklı oranlarda kayıp değerlere yapay sinir ağları yönteminin değerler atadığı durumlar için de tutarlı sonuçlar verdiği ifade edilebilir.

Elde edilen bulgularla bu konuda benzer çalışmalar karşılaştırıldığında aşağıdaki sonuçların elde edildiği görülmektedir.

Acuna ve Rodriguez (2004) literatüre göre %1'den az veri kaybının genellikle önemsiz olarak düşünüldüğünü, %1-5 arası veri kaybında kayıp verilere değer atamanın yapılıp yapılmaması yorumları çok da etkilemeyeceğini, fakat kayıp verinin %5-15 arasında olduğu durumlarda kayıp verilere değer atama için çeşitli karmaşık yöntemlerin gerektiğini ve %15'den daha fazla olması durumunda ise yapılacak herhangi bir yorumu çok ciddi şekilde etkileyeceğini belirtmişlerdir. Downey ve King (1998) çalışmasında Minnesota iş memnuniyeti ölçeği ile örgütsel bağlılık ölçekleri kullanılarak veriler elde etmişlerdir. Kayıp verilere madde ortalamalarını ve birey ortalamalarını atanmışlardır. Gerçek veri setindeki puanlarla atanmış haldeki puanlar arasındaki ortalamaya ve standart hataya bakmışlardır. Sonuç olarak cevaplayıcı ve madde sayısının %20 veya %20'den az olduğu durumlar için her iki atama yönteminin de etkili sonuçlar verdiğine değinmişlerdir. Yapay sinir ağları ile yapılan bu çalışmada ise %5, %15 ve %30'luk veri kayıplarında yapay sinir ağları yönteminin etkililiğinin kayıp veri miktarına bağlı olmadığı görülmüştür. Gerçek veri setinin analizinden çıkartılan sonuçlar ve yorumlar, veri ataması yapılmış olan veri setlerinin analizlerinden çıkartılan sonuçlar ve yorumlarla oldukça benzerdir.

Kaya, Yeşilova ve Almalı (2010) yaptıkları çalışmada yapay sinir ağlarının tahmin gücünün yüksek olduğunu vurgulamış ve veriye değer atamada kullanılması gerektiği fikrini öne sürmüştür. Aslan (2010) çalışmasında, alansal ve zamansal özellikteki meteorolojik zaman serilerinde karşılaşılan kayıp verilerin uygun şekilde tamamlanması için çok katmanlı yapay sinir ağlarının daha güvenilir sonuçlar verdiğini ifade etmiştir. Gupta ve Lam (1996) kayıp veriye değer atamada öğrenme algoritması olarak geri yayılım yönteminin daha etkili olduğu sonucuna varmışlar ve çok değişkenli analizlerde yapay sinir ağlarının kayıp veriye değer atamada kullanışlı bir yöntem olduğunu ifade etmişlerdir. Bu çalışmada da geri yayılım öğrenme algoritmasına sahip çok katmanlı yapay sinir ağının kayıp veriler değer atamada oldukça etkili olduğu görülmüştür.

Çokluk ve Kayrı (2011) yaptıkları araştırmada Fatalizm ölçeğinin hem gerçek veri seti için hem de kayıp değerlere farklı yöntemlerle (seriler ortalaması atama, yakın noktaların ortalamasını atama, yakın noktaların medyanını atama, doğrusal değer kestirimi ve noktanın doğrusal eğimi yöntemleri) yaklaşık değerlerin atandığı durumlar için tek faktörlü yapıya sahip olduğunu; fakat açıklanan varyans oranında, özdeğerlerde ve Cronbach alfa iç tutarlılık katsayılarında atama yöntemi kullanıldığında azalma meydana geldiğini bulmuşlardır. Kullanılan atama yöntemlerinin yanlı bulgulara yol açabilme olasılıkları daha yüksektir. Bu çalışmada, yapay sinir ağlarıyla ataması yapılmış veri setlerine uygulanan açımlayıcı faktör analizi sonucunda ölçeğin tek faktörlü yapısının korunduğu görülmüştür. Diğer taraftan yapay sinir ağlarıyla ataması yapılmış veri setleri gerçek veri setleriyle hem açıklanan varyans oranında hem de güvenilirlik katsayılarında oldukça benzer sonuçlar üretmiştir.

BÖLÜM V

SONUÇ VE ÖNERİLER

Bu bölümde, araştırmadan elde edilen bulgulara dayalı sonuçlar ve öneriler sunulmuştur.

Sonuç

Yapay sinir ağlarının kayıp verilere değer atamada kullanımı incelendiğinde elde edilen sonuçlar aşağıda özetlenmiştir.

Yapay sinir ağlarının tek değişkende %5, %15 ve %30'luk oranlarında kayıp veri bulunan durumlar için yüksek değerler (sırasıyla %88.2, %86.8, %89.4) verdiği görülmüştür. İki değişkende %5, %15 ve %30'luk oranlarda kayıp veri bulunan durumlar için ise yeterli olmayan düzeyde değerler (sırasıyla %57.3, %63.5, %59.7) vermiştir. Yalnız bu değerler yapay sinir ağlarının her iki değişkeni de doğru tahmin etme yüzdesidir. Yapay sinir ağının bu iki değişken için en az birini tahmin etme yüzdeleri sırasıyla %89.1, %86.5, %89.3 olarak bulunmuştur ve bu değerler de yapay sinir ağlarının doğru tahmin etme yüzdesinin yüksek olduğunu gösterir. Üç değişkende %5, %15 ve %30'luk oranlarda kayıp veri bulunan durumlar için ise yapay sinir ağlarının etkililiği sırasıyla %36.8, 39.3 ve %42.6 olarak bulunmuş ve bu değerler her üç değişkenin tahmin edilme yüzdeleridir. Yapay sinir ağının bu üç değişken için en az birini tahmin etme yüzdeleri sırasıyla %91.4, %91.3 ve %88.9 olarak bulunmuştur. Genel olarak sonuçlara bakıldığında yapay sinir ağları yönteminin %85 ve yukarısı doğrulukla kayıp veri tahmininde bulunduğu görülmüştür. Yapay sinir ağları

yöntemi psikolojik özellikleri ölçen bir araçtan elde edilen ölçme sonuçları üzerinden tahmine gittiğinden bu oran oldukça yüksektir.

Kayıp verilerin yapay sinir ağlarıyla atanması sonucu oluşan veri setlerine yapılan faktör analizi sonucunda ölçme aracının gerçek veri setine yapılan faktör analizindeki gibi 'tek faktörlü' olduğu görülmüştür.

Maddelerin faktör yük değerlerine bakıldığında genellikle en düşük faktör yük değeri 9, en yüksek yük değeri 3 olarak bulunmuş ve hemen hemen hepsinde tutarlı sonuçlar vermiştir. Faktör yük değeri, maddelerin faktörlerle olan ilişkisini açıklayan bir katsayıdır (Büyüköztürk, 2009). Bu durumda faktörü en az açıklayan 9.madde iken en fazla açıklayan da 3.maddedir. Açıklanan varyans oranlarına bakıldığında dikkate değer bir değişiklik olmadığı, iki ve üç değişkende %30 oranında kayıp verilere değer ataması yapılmış veri setlerinde açıklanan varyansın biraz daha yüksek çıktığı görülmüştür. Büyüköztürk (2009) açıklanan varyansın yüksek olması ilgili kavram ya da yapının o denli etkili ölçüldüğünün göstergesi olarak yorumlandığını ifade eder. Güvenirlik katsayıları ile madde-toplam korelasyonları için de tutarlı sonuçlar vermiştir. Böylelikle yapay sinir ağları yönteminin kayıp verilere değer atamadaki etkililiğinin yüksek olduğu ve gerçek veri setiyle oldukça benzer sonuçlar verdiği görülmüştür.

Öneriler

Araştırmanın sonuçları doğrultusunda öneriler aşağıda sunulmuştur.

1. Araştırma sonucunda yapay sinir ağları yöntemi kayıp verilere değer atamada kullanıldığında farklı koşullarda yüksek oranda etkili sonuçlar elde ettiği görülmektedir. Kayıp verili veri setlerinde bilgi kaybını olabildiğince en aza indirmek ve tam veri setiyle analizleri yapmak, daha tutarlı ve gerçekçi çıkarımlar elde edeceğinden gerçeğe yakın değerler atayan yapay sinir ağlarıyla kayıp değerlerin atanması önerilir.
2. Bu araştırmada yapay sinir ağları modellemesinde iki katmanlı ileri beslemeli geri yayımlı ağ ve sigmoid fonksiyonu olan hiperbolik tanjant aktivasyon fonksiyonu kullanılmıştır. Geri yayımlı ağlar

dışındaki ağ yapıları, modeller oluşturulup denenebilir ve sonuçlar karşılaştırılabilir.

3. Yapay sinir ağının sürekli verilerdeki veya ikili verilerdeki kayıp verilere değer atamadaki kullanımı da incelenebilir.
4. Kayıp veriye değer atamada kullanılan çeşitli yöntemlerle yapay sinir ağı yönteminin etkililiği karşılaştırılabilir. Böylelikle yöntemlerin kullanımlarına katkı sağlanabilir ve yaygınlaştırılabilir.

KAYNAKÇA

- Abdella, M. I. (2005). *The use of genetic algorithms and neural networks to approximate missing data in database*. Published master's thesis, University of the Witwatersrand, Johannesburg.
- Abdella, M. I. and Marwala, T. (2005). The use of genetic algorithms and neural networks to approximate missing data in database. *Computing and Informatics*, 24, 577–589.
- Acuna, E. and Rodriguez, C. (2004). The treatment of missing values and its effect in the classifier accuracy. In D. Banks, L. House, F.R. McMorris, P. Arabie, W. Gaul (Eds). *Classification, Clustering and Data Mining Applications*, Springer-Verlag Berlin-Heidelberg, 639-648.
- Allison, P.D. (2002). *Missing data*. Thousand Oaks, California: Sage.
- Alpar, R. (2003). *Uygulamalı çok değişkenli istatistiksel yöntemlere giriş 1*. İstanbul: Nobel yayınevi.
- Aslan, S. (2010). *Comparison of missing value imputation methods for meteorological time series data*. Yayınlanmamış yüksek lisans tezi, Ortadoğu Teknik Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Aslan, S., Yozgatlıgil, C., İyigün, C., Batmaz, İ. ve Tatlı, H. (2010). Meteorolojik zaman serilerinde kayıp veri tahmin yöntemlerinin başarımlarının korelasyon boyutu analiziyle karşılaştırılması. *İstatistik araştırma dergisi*, 8 (2), 55-67
- Aydın, B. (2002). *Basit ve geri yayımlı yapay sinir ağları, uygulama alanları*. Yayınlanmamış yüksek lisans tezi, Çanakkale Onsekiz Mart Üniversitesi, Fen Bilimleri Enstitüsü, Çanakkale.
- Aydilek, İ. B. ve Arslan, A. (2012). A novel hybrid approach to estimating missing values in databases using k-nearest neighbors and neural Networks. *International journal of Innovative computing, information*

and control. 8 (7), 4705-4717. Web: <http://www.ijicic.org/ijicic-11-02062.pdf> adresinden 23 Mart 2013'te alınmıştır.

- Aziz, M. A. ve Dar, H. A. (2004). Predicting corporate bankruptcy: Whither do we stand? *3rd annual meeting of the european economics and finance society world economy and european integration*, University of Gdansk.
- Bal, C. (2003). *Çok gruplu veri setlerinde eksik gözlem sorununun çözümlenmesi ve sağlık alanında bir uygulama*. Yayınlanmamış doktora tezi, Osmangazi Üniversitesi Sağlık Bilimleri Enstitüsü, Eskişehir.
- Baş, N. (2006). *Yapay sinir ağları yaklaşımı ve bir uygulama*. Yayınlanmamış yüksek lisans tezi, Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Baygül, A. (2007). *Kayıp veri analizinde sıklıkla kullanılan etkin yöntemlerin değerlendirilmesi*. Yayınlanmamış yüksek lisans tezi, İstanbul Üniversitesi Sağlık Bilimleri Enstitüsü, İstanbul.
- Baykul, Y. (2010). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: Pegem Akademi.
- Bishop, C. M. (1995). *Neural Networks For Pattern Recognition*. Oxford: Clarendon Press.
- Bryman, A. and Cramer, D. (2005). *Quantitative Data Analysis with SPSS Release 12 and 13: A Guide for Social Scientist*. Hove: Routledge
- Büyüköztürk, Ş. (2009). *Sosyal Bilimler İçin Veri Analizi El Kitabı*, Ankara: Pegem Akademi
- Çokluk, Ö. ve Kayrı, M. (2011). Kayıp değerlere yaklaşık değer atama yöntemlerinin ölçme araçlarının geçerlik ve güvenilirliği üzerindeki etkisi. *Kuram ve Uygulamada Eğitim Bilimleri*, 11 (1), 289-309.

- Demir, E. (2013). *Kayıp verilerin varlığında iki kategorili puanlanan maddelerden oluşan testlerin psikometrik özelliklerinin incelenmesi*. Yayınlanmamış yüksek lisans tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- DeTienne, K. B., DeTienne, D.H. and Joshi, S.A. (2003). Neural Networks as Statistical Tools for Business Researchers, *Organizational Research Methods*, 6 (2), 236-265.
- Downey, R.G. and King, C. (1998). Missing data in Likert ratings: A comparison of replacement methods. *Journal of General Psychology*, 125, 175-191.
- Dural, S. (2010). *Farklı kayıp veri tekniklerinin çok göstergeli örtük gelişme modelleri üzerindeki etkisi*. Yayınlanmamış doktora tezi, Ege Üniversitesi Sosyal Bilimler Enstitüsü, İzmir.
- Enders, C. (2010). *Applied Missing Data Analysis*. New York: The Guilford Press.
- Elmas, Ç. (2003). *Yapay sinir ağları (kuram, mimari, eğitim, uygulama)*. Ankara: Seçkin yayıncılık.
- Fausett, L. V. (1994). *Fundamentals of Neural Networks: Architecture, Algorithms and Applications*. New Jersey: Prentice Hall.
- Graham, M. J. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability what they are and how to use them. *Educational and Psychological Measurement*, 66 (6). 930-944
- Gupta, A. and Lam, M.S. (1996). Estimating missing values using neural networks. *Journal of the Operational Research Society*, 47, 229-238.
- Gülseçen, S.(1993). *Yapay sinir ağları işletme alanında uygulanması ve bir örnek çalışma*. Yayınlanmamış doktora tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.

- Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*, New Jersey: Prentice Hall International.
- He, Y. (2006). *Missing Data Imputation for Tree-Based Models*, Unpublished Ph.D thesis, University of California, Los Angeles.
- Hill, M. (1997). *SPSS missing value analysis 7.5*. Chicago: SPSS Inc. Web: <http://www.bf.lu.lv/grozs/Datorlietas/SPSS/SPSS%20Missing%20Value%20Analysis%207.5.pdf> adresinden 17 Şubat 2013'te alınmıştır.
- Jönsson, P. and Wohlin, C. (2004). An Evaluation of k-nearest Neighbour Imputation Using Likert Data", Proceedings 10th International Symposium on Software Metrics, 108-118, Chiaco, USA.
- Kaplan, D. and Turner, L. (2012). Statistical Matching of PISA 2009 and TALIS 2008 data in Iceland. OECD Education Working Papers, 78, Paris: OECD Publishing. Web:[http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=EDU/WKP\(2012\)15&docLanguage=En](http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=EDU/WKP(2012)15&docLanguage=En) adresinden 12 Şubat 2013'te alınmıştır.
- Karasar, N. (2010). *Bilimsel araştırma yöntemi*. Ankara: Nobel Yayınları.
- Kaya, Y., Yeşilova, A. ve Almalı, M.N. (2010). An application of expectation and maximization, multiple imputation and neural network methods for missing value. *World applied sciences journal*, 9 (5), 561-566.
- Kline, R.B. (1998). *Principles and practice of structural equation modelling*. Guilford Press, New York.
- Kröse, B. and van der Smagt, P. (1996). An introduction to neural Networks, The university of Amsterdam. Web: <http://lia.univ-avignon.fr/chercheurs/torres/livres/book-neuro-intro.pdf> adresinden 16 Mart 2013'te alınmıştır.
- Leite, W. and Beretvas, S. N. (2010). The performance of multiple imputation for likert-type items with missing data. *Journal of modern applied*

statistical methods. 9 (1), 64-74. Web :
<http://digitalcommons.wayne.edu/jmasm/vol9/iss1/8> adresinden 10
 Ekim 2013'te alınmıştır.

Lippman, R.P. (1987).An introduction to computing with neural nets. *IEEE ASSP Magazine*, 4(2), 4-22. Web:
<http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=01165576>
 adresinden 4 Nisan 2013'te alınmıştır.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*,
 New Jersey: John Wiley and Sons.

McKnight, P.E., McKnight, K.M., Sidani, S. and Figueredo, A.J.
 (2007). *Missing Data: A Gentle Introduction*. New York: Guilford.

Milli Eğitim Bakanlığı. (2010). *PISA 2009 Uluslararası Öğrenci Değerlendirme Projesi, Ulusal Ön Rapor*. Ankara: Eğitim Araştırma ve Geliştirme Dairesi Yayınları. Web:
<http://www.earged.meb.gov.tr/pdf/PISA2009rapor.pdf> adresinden 12
 Şubat 2013'te alınmıştır.

Molina, T. L., Méndez, A. P. and Echeverría, F. R. (2008). Missing values imputation techniques for neural networks patterns. *12th WSEAS International Conference on SYSTEMS, Heraklion, Greece*.

Nelwamondo, F.V. (2008). *Computational intelligence techniques for missing data imputation*. Unpublished Ph.D. thesis, University of the Witwatersrand, Johannesburg.

Nelwamondo, F.V., Mohamed, S. and Marwala,T. (2007). Missing data: a comparison of neural network and expectation maximisation techniques. *Current Science*, 93 (11), 1514-1521.

Ocakoğlu, G. (2006). *Lojistik regresyon analizi ve yapay sinir ağları tekniklerinin sınıflama özelliklerinin karşılaştırılması ve bir uygulama*.

Yayınlanmamış yüksek lisans tezi, Uludağ Üniversitesi Sağlık Bilimleri Enstitüsü, Bursa.

- Öztemel, E. (2003). *Yapay sinir ağları*. Papatya Yayıncılık, İstanbul.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*, New York: Wiley.
- Saraç, T. (2004). *Yapay sinir ağları seminer projesi*, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Saygılı, Y. S. (2008). *İstatistiksel yöntemlerle yapay sinir ağları uygulamalarının karşılaştırılması: Milli savunma bakanlığı bütçesinin öngörülmesi*. Yayınlanmamış yüksek lisans tezi, Kara Harp Okulu Savunma Bilimleri Enstitüsü, Ankara.
- Shah, J. R. and Murtaza, M.B. (2000). A Neural Network Based Clustering Procedure for Bankruptcy Prediction, *American Business Review*.18 (2).
- Taşgetiren, M.F. (2006). Çok katmanlı yapay sinir ağları. Web: <http://www.bilimbilmek.com/tr/cok-katmanli-yapay-sinir-aglari.html> adresinden 25 Şubat 2013'te alınmıştır.
- Tekin, H. (2010). *Eğitimde ölçme ve değerlendirme*. Ankara: Yargı Yayınevi
- Torun, T. (2007). *Finansal başarısızlık tahmininde geleneksel istatistikî yöntemlerle yapay sinir ağlarının karşılaştırılması ve sanayi işletmeleri üzerinde uygulama*. Yayınlanmamış doktora tezi, Erciyes Üniversitesi Sosyal Bilimler Enstitüsü, Kayseri.
- Tsikriktsis, N. (2005). A review of techniques for treating missing data in OM survey research. *Journal of Operations Management*, 24, 53–62
- Turgut, M.F. ve Baykul, Y. (2010). *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem-Akademi.

Uğuz, B. (2007). *An analysis of a multiple imputation model for the missing values in selected variables of TDHS-2003 data: the case of anthropometric measures*. Yayınlanmamış yüksek lisans tezi, Hacettepe Üniversitesi Nüfus Etütleri Enstitüsü, Ankara.

Yurdugül, H. (2006). Paralel, eşdeğer ve konjenerik ölçmelerde güvenilirlik katsayılarının karşılaştırılması. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 39(1), 15-37.

Yurtoğlu, H. (2005). *Yapay sinir ağları metodolojisi ile öngörü modellemesi: bazı makroekonomik değişkenler için Türkiye örneği*. Uzmanlık Tezi, Yayın No: Dpt: 2683, Ankara.

EKLER

Ek 1. Verilerin Düzenlenmesi

Ek 2. Ölçme Modelinin Belirlenmesi

Ek 3. Hata Matrisi

EK 1.

Verilerin Düzenlenmesi

Çizelge 15. Tek değişkende kayıp veriye sahip veri setleri için verilerin düzenlenmesi

1	[1 0 0 0]
2	[0 1 0 0]
3	[0 0 1 0]
4	[0 0 0 1]

Çizelge 16. İki değişkende kayıp veriye sahip veri setleri için verilerin düzenlenmesi

[1 1]	[1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
[1 2]	[0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
[1 3]	[0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0]
[1 4]	[0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0]
[2 1]	[0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0]
[2 2]	[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0]
[2 3]	[0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0]
[2 4]	[0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0]
[3 1]	[0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0]
[3 2]	[0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0]
[3 3]	[0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0]
[3 4]	[0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0]
[4 1]	[0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0]
[4 2]	[0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0]
[4 3]	[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1]
[4 4]	[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1]

EK 2

Ölçme Modelinin Belirlenmesi

Çizelge 18. Gerçek Veri Setindeki Maddelere İlişkin Kovaryans Matrisi

	1	2	3	4	5	6	7	8	9	10	11
1	.726	.290	.233	.309	.163	.210	.218	.296	.170	.126	.199
2	.290	.673	.386	.302	.231	.174	.315	.252	.116	.182	.295
3	.233	.386	.726	.278	.250	.179	.338	.306	.168	.239	.371
4	.309	.302	.278	.736	.164	.219	.221	.347	.224	.116	.223
5	.163	.231	.250	.164	.624	.206	.244	.175	.157	.205	.260
6	.210	.174	.179	.219	.206	.511	.164	.214	.234	.193	.157
7	.218	.315	.338	.221	.244	.164	.653	.247	.149	.188	.248
8	.296	.252	.306	.347	.175	.214	.247	.737	.260	.178	.337
9	.170	.116	.168	.224	.157	.234	.149	.260	.735	.180	.181
10	.126	.182	.239	.116	.205	.193	.188	.178	.180	.588	.267
11	.199	.295	.371	.223	.260	.157	.248	.337	.181	.267	.799

Çizelge 19. Gerçek Veri Setindeki Maddelerin Betimsel İstatistikleri

	1	2	3	4	5	6	7	8	9	10	11
Ortalama	3.19	2.77	2.75	2.86	2.97	3.50	2.93	2.89	3.01	3.14	2.99
Standart Sapma	.852	.820	.852	.858	.790	.715	.808	.858	.857	.767	.894

Çizelge 20. Gerçek Veri Seti Ölçme Modelleri Uyum İyiliği Değerleri

	sd	X ²	RMSEA	AGFI	NFI	NNFI	CFI	GFI
Konjenerik	44	253.13	.098	.87	.93	.93	.94	.92
Eşdeğer	54	328.73	.101	.87	.91	.93	.93	.89
Paralel	64	369.02	.098	.88	.90	.93	.92	.88

Çizelge 21. Tek Değişkende %5 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelere İlişkin Kovaryans Matrisi

	1	2	3	4	5	6	7	8	9	10	11
1	.726	.290	.233	.309	.163	.203	.218	.296	.170	.126	.199
2	.290	.673	.386	.302	.231	.170	.315	.252	.116	.182	.295
3	.233	.386	.726	.278	.250	.169	.338	.306	.168	.239	.371
4	.309	.302	.278	.736	.164	.215	.221	.347	.224	.116	.223
5	.163	.231	.250	.164	.624	.199	.244	.175	.157	.205	.260
6	.203	.170	.169	.215	.199	.479	.155	.214	.237	.185	.147
7	.218	.315	.338	.221	.244	.155	.653	.247	.149	.188	.248
8	.296	.252	.306	.347	.175	.214	.247	.737	.260	.178	.337
9	.170	.116	.168	.224	.157	.237	.149	.260	.735	.180	.181
10	.126	.182	.239	.116	.205	.185	.188	.178	.180	.588	.267
11	.199	.295	.371	.223	.260	.147	.248	.337	.181	.267	.799

Çizelge 22. Tek Değişkende %5 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelerin Betimsel İstatistikleri

	1	2	3	4	5	6	7	8	9	10	11
Ortalama	3.19	2.77	2.75	2.86	2.97	3.51	2.93	2.89	3.01	3.14	2.99
Standart Sapma	.852	.820	.852	.858	.790	.692	.808	.858	.857	.767	.894

Çizelge 23. Tek Değişkende %5 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Seti Ölçme Modelleri Uyum İyiliği Değerleri

	sd	X ²	RMSEA	AGFI	NFI	NNFI	CFI	GFI
Konjenerik	44	262.47	.100	.87	.93	.93	.94	.91
Eşdeğer	54	342.98	.104	.86	.91	.92	.92	.89
Paralel	64	369.02	.098	.88	.90	.93	.92	.88

Çizelge 24. İki Değişkende %5 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelere İlişkin Kovaryans Matrisi

	1	2	3	4	5	6	7	8	9	10	11
1	.726	.290	.233	.309	.140	.192	.218	.296	.170	.126	.199
2	.290	.673	.386	.302	.197	.161	.315	.252	.116	.182	.295
3	.233	.386	.726	.278	.214	.164	.338	.306	.168	.239	.371
4	.309	.302	.278	.736	.159	.216	.221	.347	.224	.116	.223
5	.140	.197	.214	.159	.589	.191	.216	.153	.160	.182	.236
6	.192	.161	.164	.216	.191	.479	.141	.208	.231	.184	.152
7	.218	.315	.338	.221	.216	.141	.653	.247	.149	.188	.248
8	.296	.252	.306	.347	.153	.208	.247	.737	.260	.178	.337
9	.170	.116	.168	.224	.160	.231	.149	.260	.735	.180	.181
10	.126	.182	.239	.116	.182	.184	.188	.178	.180	.588	.267
11	.199	.295	.371	.223	.236	.152	.248	.337	.181	.267	.799

Çizelge 25. İki Değişkende %5 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelerin Betimsel İstatistikleri

	1	2	3	4	5	6	7	8	9	10	11
Ortalama	3.19	2.77	2.75	2.86	2.99	3.52	2.93	2.89	3.01	3.14	2.99
Standart Sapma	.852	.820	.852	.858	.768	.692	.808	.858	.857	.767	.894

Çizelge 26. İki Değişkende %5 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Seti Ölçme Modelleri Uyum İyiliği Değerleri

	sd	X ²	RMSEA	AGFI	NFI	NNFI	CFI	GFI
Konjenerik	44	260.02	.099	.87	.93	.92	.94	.91
Eşdeğer	54	349.84	.105	.86	.91	.92	.92	.89
Paralel	64	369.02	.098	.88	.90	.93	.92	.88

Çizelge 27. Üç Değişkende %5 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelere İlişkin Kovaryans Matrisi

	1	2	3	4	5	6	7	8	9	10	11
1	.726	.290	.233	.309	.141	.199	.200	.296	.170	.126	.199
2	.290	.673	.386	.302	.211	.166	.303	.252	.116	.182	.295
3	.233	.386	.726	.278	.230	.165	.328	.306	.168	.239	.371
4	.309	.302	.278	.736	.160	.219	.210	.347	.224	.116	.223
5	.141	.211	.230	.160	.597	.191	.232	.161	.153	.197	.240
6	.199	.166	.165	.219	.191	.483	.151	.218	.241	.185	.145
7	.200	.303	.328	.210	.232	.151	.626	.236	.139	.179	.240
8	.296	.252	.306	.347	.161	.218	.236	.737	.260	.178	.337
9	.170	.116	.168	.224	.153	.241	.139	.260	.735	.180	.181
10	.126	.182	.239	.116	.197	.185	.179	.178	.180	.588	.267
11	.199	.295	.371	.223	.240	.145	.240	.337	.181	.267	.799

Çizelge 28. Üç Değişkende %5 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelerin Betimsel İstatistikleri

	1	2	3	4	5	6	7	8	9	10	11
Ortalama	3.19	2.77	2.75	2.86	2.98	3.51	2.94	2.89	3.01	3.14	2.99
Standart Sapma	.852	.820	.852	.858	.773	.695	.791	.858	.857	.767	.894

Çizelge 29. Üç Değişkende %5 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Seti Ölçme Modelleri Uyum İyiliği Değerleri

	sd	X ²	RMSEA	AGFI	NFI	NNFI	CFI	GFI
Konjenerik	44	270.19	.101	.87	.92	.92	.94	.91
Eşdeğer	54	351.85	.105	.86	.91	.92	.92	.89
Paralel	64	369.02	.098	.88	.90	.93	.92	.88

Çizelge 30. Tek Değişkende %15 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelere İlişkin Kovaryans Matrisi

	1	2	3	4	5	6	7	8	9	10	11
1	.726	.290	.233	.309	.163	.202	.218	.296	.170	.126	.199
2	.290	.673	.386	.302	.231	.171	.315	.252	.116	.182	.295
3	.233	.386	.726	.278	.250	.168	.338	.306	.168	.239	.371
4	.309	.302	.278	.736	.164	.239	.221	.347	.224	.116	.223
5	.163	.231	.250	.164	.624	.203	.244	.175	.157	.205	.260
6	.202	.171	.168	.239	.203	.499	.167	.219	.247	.183	.151
7	.218	.315	.338	.221	.244	.167	.653	.247	.149	.188	.248
8	.296	.252	.306	.347	.175	.219	.247	.737	.260	.178	.337
9	.170	.116	.168	.224	.157	.247	.149	.260	.735	.180	.181
10	.126	.182	.239	.116	.205	.183	.188	.178	.180	.588	.267
11	.199	.295	.371	.223	.260	.151	.248	.337	.181	.267	.799

Çizelge 31. Tek Değişkende %15 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelerin Betimsel İstatistikleri

	1	2	3	4	5	6	7	8	9	10	11
Ortalama	3.19	2.77	2.75	2.86	2.97	3.51	2.93	2.89	3.01	3.14	2.99
Standart Sapma	.852	.820	.852	.858	.790	.706	.808	.858	.857	.767	.894

Çizelge 32. Tek Değişkende %15 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Seti Ölçme Modelleri Uyum İyiliği Değerleri

	sd	X ²	RMSEA	AGFI	NFI	NNFI	CFI	GFI
Konjenerik	44	267.98	.101	.87	.93	.92	.94	.91
Eşdeğer	54	343.26	.104	.86	.91	.92	.92	.89
Paralel	64	369.02	.098	.88	.90	.93	.92	.88

Çizelge 33. İki Değişkende %15 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelere İlişkin Kovaryans Matrisi

	1	2	3	4	5	6	7	8	9	10	11
1	.726	.290	.233	.309	.131	.181	.218	.296	.170	.126	.199
2	.290	.673	.386	.302	.202	.156	.315	.252	.116	.182	.295
3	.233	.386	.726	.278	.222	.149	.338	.306	.168	.239	.371
4	.309	.302	.278	.736	.161	.222	.221	.347	.224	.116	.223
5	.131	.202	.222	.161	.549	.178	.223	.161	.128	.173	.235
6	.181	.156	.149	.222	.178	.479	.144	.213	.243	.163	.130
7	.218	.315	.338	.221	.223	.144	.653	.247	.149	.188	.248
8	.296	.252	.306	.347	.161	.213	.247	.737	.260	.178	.337
9	.170	.116	.168	.224	.128	.243	.149	.260	.735	.180	.181
10	.126	.182	.239	.116	.173	.163	.188	.178	.180	.588	.267
11	.199	.295	.371	.223	.235	.130	.248	.337	.181	.267	.799

Çizelge 34. İki Değişkende %15 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelerin Betimsel İstatistikleri

	1	2	3	4	5	6	7	8	9	10	11
Ortalama	3.19	2.77	2.75	2.86	3.01	3.52	2.93	2.89	3.01	3.14	2.99
Standart Sapma	.852	.820	.852	.858	.741	.692	.808	.858	.857	.767	.894

Çizelge 35. İki Değişkende %15 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Seti Ölçme Modelleri Uyum İyiliği Değerleri

	sd	X ²	RMSEA	AGFI	NFI	NNFI	CFI	GFI
Konjenerik	44	268.36	.101	.87	.92	.92	.94	.91
Eşdeğer	54	365.50	.108	.86	.90	.91	.92	.88
Paralel	64	369.02	.098	.88	.90	.93	.92	.88

Çizelge 36. Üç Değişkende %15 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelere İlişkin Kovaryans Matrisi

	1	2	3	4	5	6	7	8	9	10	11
1	.726	.290	.233	.309	.114	.198	.191	.296	.170	.126	.199
2	.290	.673	.386	.302	.179	.180	.304	.252	.116	.182	.295
3	.233	.386	.726	.278	.209	.181	.325	.306	.168	.239	.371
4	.309	.302	.278	.736	.150	.238	.235	.347	.224	.116	.223
5	.114	.179	.209	.150	.533	.179	.209	.134	.110	.151	.207
6	.198	.180	.181	.238	.179	.499	.155	.219	.234	.188	.161
7	.191	.304	.325	.235	.209	.155	.592	.226	.115	.164	.238
8	.296	.252	.306	.347	.134	.219	.226	.737	.260	.178	.337
9	.170	.116	.168	.224	.110	.234	.115	.260	.735	.180	.181
10	.126	.182	.239	.116	.151	.188	.164	.178	.180	.588	.267
11	.199	.295	.371	.223	.207	.161	.238	.337	.181	.267	.799

Çizelge 37. Üç Değişkende %15 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelerin Betimsel İstatistikleri

	1	2	3	4	5	6	7	8	9	10	11
Ortalama	3.19	2.77	2.75	2.86	3.01	3.49	2.95	2.89	3.01	3.14	2.99
Standart Sapma	.852	.820	.852	.858	.730	.706	.770	.858	.857	.767	.894

Çizelge 38. Üç Değişkende %15 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Seti Ölçme Modelleri Uyum İyiliği Değerleri

	sd	X ²	RMSEA	AGFI	NFI	NNFI	CFI	GFI
Konjenerik	44	259.77	.099	.87	.93	.92	.94	.91
Eşdeğer	54	353.79	.105	.86	.90	.92	.92	.89
Paralel	64	369.02	.098	.88	.90	.93	.92	.88

Çizelge 39. Tek Değişkende %30 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelere İlişkin Kovaryans Matrisi

	1	2	3	4	5	6	7	8	9	10	11
1	.726	.290	.233	.309	.163	.201	.218	.296	.170	.126	.199
2	.290	.673	.386	.302	.231	.163	.315	.252	.116	.182	.295
3	.233	.386	.726	.278	.250	.141	.338	.306	.168	.239	.371
4	.309	.302	.278	.736	.164	.234	.221	.347	.224	.116	.223
5	.163	.231	.250	.164	.624	.175	.244	.175	.157	.205	.260
6	.201	.163	.141	.234	.175	.454	.159	.212	.225	.173	.144
7	.218	.315	.338	.221	.244	.159	.653	.247	.149	.188	.248
8	.296	.252	.306	.347	.175	.212	.247	.737	.260	.178	.337
9	.170	.116	.168	.224	.157	.225	.149	.260	.735	.180	.181
10	.126	.182	.239	.116	.205	.173	.188	.178	.180	.588	.267
11	.199	.295	.371	.223	.260	.144	.248	.337	.181	.267	.799

Çizelge 40. Tek Değişkende %30 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelerin Betimsel İstatistikleri

	1	2	3	4	5	6	7	8	9	10	11
Ortalama	3.19	2.77	2.75	2.86	2.97	3.53	2.93	2.89	3.01	3.14	2.99
Standart Sapma	.852	.820	.852	.858	.790	.674	.808	.858	.857	.767	.894

Çizelge 41. Tek Değişkende %30 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Seti Ölçme Modelleri Uyum İyiliği Değerleri

	sd	X ²	RMSEA	AGFI	NFI	NNFI	CFI	GFI
Konjenerik	44	271.09	.102	.87	.93	.92	.94	.91
Eşdeğer	54	356.03	.106	.86	.91	.92	.92	.89
Paralel	64	369.02	.098	.88	.90	.93	.92	.88

Çizelge 42. İki Değişkende %30 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelere İlişkin Kovaryans Matrisi

	1	2	3	4	5	6	7	8	9	10	11
1	.726	.290	.233	.309	.145	.218	.218	.296	.170	.126	.199
2	.290	.673	.386	.302	.209	.166	.315	.252	.116	.182	.295
3	.233	.386	.726	.278	.247	.159	.338	.306	.168	.239	.371
4	.309	.302	.278	.736	.198	.241	.221	.347	.224	.116	.223
5	.145	.209	.247	.198	.545	.168	.240	.184	.154	.211	.248
6	.218	.166	.159	.241	.168	.462	.154	.236	.207	.196	.156
7	.218	.315	.338	.221	.240	.154	.653	.247	.149	.188	.248
8	.296	.252	.306	.347	.184	.236	.247	.737	.260	.178	.337
9	.170	.116	.168	.224	.154	.207	.149	.260	.735	.180	.181
10	.126	.182	.239	.116	.211	.196	.188	.178	.180	.588	.267
11	.199	.295	.371	.223	.248	.156	.248	.337	.181	.267	.799

Çizelge 43. İki Değişkende %30 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelerin Betimsel İstatistikleri

	1	2	3	4	5	6	7	8	9	10	11
Ortalama	3.19	2.77	2.75	2.86	3.00	3.53	2.93	2.89	3.01	3.14	2.99
Standart Sapma	.852	.820	.852	.858	.738	.680	.808	.858	.857	.767	.894

Çizelge 44. İki Değişkende %30 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Seti Ölçme Modelleri Uyum İyiliği Değerleri

	sd	X ²	RMSEA	AGFI	NFI	NNFI	CFI	GFI
Konjenerik	44	269.38	.101	.87	.93	.93	.94	.91
Eşdeğer	54	346.68	.104	.86	.91	.92	.93	.89
Paralel	64	369.02	.098	.88	.90	.93	.92	.88

Çizelge 45. Üç Değişkende %30 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelere İlişkin Kovaryans Matrisi

	1	2	3	4	5	6	7	8	9	10	11
1	.726	.290	.233	.309	.118	.211	.201	.296	.170	.126	.199
2	.290	.673	.386	.302	.215	.166	.304	.252	.116	.182	.295
3	.233	.386	.726	.278	.250	.159	.340	.306	.168	.239	.371
4	.309	.302	.278	.736	.168	.243	.224	.347	.224	.116	.223
5	.118	.215	.250	.168	.517	.154	.243	.192	.106	.193	.245
6	.211	.166	.159	.243	.154	.459	.157	.214	.201	.175	.155
7	.201	.304	.340	.224	.243	.157	.575	.246	.099	.171	.252
8	.296	.252	.306	.347	.192	.214	.246	.737	.260	.178	.337
9	.170	.116	.168	.224	.106	.201	.099	.260	.735	.180	.181
10	.126	.182	.239	.116	.193	.175	.171	.178	.180	.588	.267
11	.199	.295	.371	.223	.245	.155	.252	.337	.181	.267	.799

Çizelge 46. Üç Değişkende %30 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Setindeki Maddelerin Betimsel İstatistikleri

	1	2	3	4	5	6	7	8	9	10	11
Ortalama	3.19	2.77	2.75	2.86	3.01	3.51	2.96	2.89	3.01	3.14	2.99
Standart Sapma	.852	.820	.852	.858	.719	.677	.758	.858	.857	.767	.894

Çizelge 47. Üç Değişkende %30 Oranında Kayıp Verilere Değer Ataması Yapılmış Veri Seti Ölçme Modelleri Uyum İyiliği Değerleri

	sd	X ²	RMSEA	AGFI	NFI	NNFI	CFI	GFI
Konjenerik	44	276.43	.103	.86	.93	.93	.94	.91
Eşdeğer	54	372.43	.109	.85	.90	.92	.92	.88
Paralel	64	369.02	.098	.88	.90	.93	.92	.88

