

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**YAPAY ZEKA TEKNİKLERİ KULLANILARAK
ÖĞRENCİ DANIŞMANLIK HİZMETİ BİLİŞİM
ALTYAPI TASARIMI**

YÜKSEK LİSANS TEZİ
Sümeyye KAYNAK

Enstitü Anabilim Dalı : BİLGİSAYAR VE BİLİŞİM
MÜHENDİSLİĞİ
Tez Danışmanı : Yrd. Doç. Dr. Hayrettin
EVİRGEN

Haziran 2014

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YAPAY ZEKA TEKNİKLERİ KULLANILARAK
ÖĞRENCİ DANIŞMANLIK HİZMETİ BİLİŞİM
ALTYAPI TASARIMI

YÜKSEK LİSANS TEZİ

Sümeyye KAYNAK

Enstitü Anabilim Dalı : BİLGİSAYAR VE BİLİŞİM
MÜHENDİSLİĞİ

Bu tez 6/6/2014 tarihinde aşağıdaki jüri tarafından Oybirliği ile kabul edilmiştir.

Yrd. Doç. Dr. Hayrettin
AYKIN GÖZALP

Jüri Başkanı

Doç. Dr. Cemil ÖZ

Üye

Prof. Dr. İsmail Hakkı CEDİMOĞLU

Üye

TEŐEKKÜR

Bu tez alıŐmasının tamamlanmasında deęerli katkılarını esirgemeyen danıŐmanım Yrd. Do. Dr. Hayrettin EVİRGEN'e, Yrd. Do. Dr. Halil İbrahim CEBECİ'ye, alıŐmalarım sırasında bana sürekli ve sabırla destek olan eŐim Öğr. Gör. Baran KAYNAK'a, ve aileme teŐekkürü bir bor bilirim.

İÇİNDEKİLER

TEŞEKKÜR.....	ii
İÇİNDEKİLER.....	iii
ŞİMGELER VE KISALTMALAR LİSTESİ.....	vi
ŞEKİLLER LİSTESİ.....	vii
TABLolar LİSTESİ.....	viii
ÖZET.....	x
SUMMARY.....	xi
BÖLÜM 1.	
GİRİŞ.....	1
1.1. Literatür Taraması.....	3
BÖLÜM 2.	
VERİ MADENCİLİĞİ.....	7
2.1. Giriş.....	7
2.2. Veri madenciliği süreci.....	8
BÖLÜM 3.	
VERİ MADENCİLİĞİ YÖNTEMLERİ.....	10
3.1. Sınıflandırma Yöntemler.....	10
3.1.1. Yapay sinir ağları.....	10
3.1.1.1. Yapay sinir hücresi.....	11
3.1.1.2. Yapay sinir ağının yapısı.....	12
3.1.1.3. Yapay sinir ağlarında öğrenme.....	12
3.1.1.4. Yapay sinir ağı modelleri.....	12
3.1.2. Adaptif ağ tabanlı bulanık çıkarım sistemi.....	15
3.1.2.1. ANFIS mimarisi.....	15

3.1.2.2. ANFIS modelinin öğrenme algoritması.....	17
3.1.3. K en yakın komşu algoritması.....	18
3.1.4. Naive bayes algoritması.....	19
3.1.5. Karar ağaçları.....	19
3.1.5.1. Karar ağacının temel yapısı.....	20
3.1.5.2. Karar ağacının oluşturulması.....	20
3.1.5.3. Karar ağacı algoritmaları.....	21
3.2. Kümeleme Yöntemleri.....	21
3.2.1. Apriori algoritmaları.....	21
3.2.2. K ortalamalar yöntemi.....	22
3.2.3. Genetik algoritmalar.....	22
3.2.4. Hiyerarşik kümeleme.....	23
3.3. Performans Değerlendirme Ölçütleri.....	24

BÖLÜM 4.

REGRESYON VE KORELASYON ANALİZİ.....	26
4.1. Regresyon Analizi.....	26
4.1.1. Doğrusal regresyon analizi.....	26
4.1.1.1. Tek değişkenli doğrusal regresyon analizi.....	27
4.1.1.2. Çok değişkenli doğrusal regresyon analizi.....	28
4.1.2. Doğrusal olmayan regresyon analizi.....	28
4.2. Korelasyon Analizi.....	29

BÖLÜM 5.

ÇAPRAZ GEÇERLİLİK YÖNTEMİ.....	30
5.1. Giriş.....	30
5.2. Holdout yöntemi.....	30
5.3. K kat çapraz doğrulama yöntemi.....	31
5.4. Biri hariç çapraz doğrulama yöntemi.....	32
5.5. Rastgele alt-örnekleme yöntemi.....	32

BÖLÜM 6.

UYGULAMA	33
6.1. Verinin Hazırlanması.....	33
6.2. Girdi Parametrelerinin Belirlenmesi.....	35
6.2.1. Korelasyon analiziyle girdi parametrelerinin incelenmesi.....	37
6.2.1.1. BSM429 Korelasyon analizi.....	37
6.2.1.2. BSM421 Korelasyon analizi.....	38
6.2.1.3. BSM425 Korelasyon analizi.....	41
6.2.1.4. BSM435 Korelasyon analizi.....	42
6.2.2. Regresyon analiziyle girdi parametrelerinin incelenmesi.....	43
6.2.2.1. BSM429 Regresyon analizi.....	46
6.2.2.2. BSM421 Regresyon analizi.....	50
6.2.2.3. BSM425 Regresyon Analizi.....	53
6.2.2.4. BSM435 Regresyon Analizi.....	55
6.3. Kullanılan Yöntemler.....	58
6.3.1. Yapay sinir ağları (YSA).....	58
6.3.2. Adaptif ağ tabanlı bulanık çıkarım sistemi (ANFIS).....	61
6.3.3. Karar ağaçları.....	64
6.3.4. Yöntemlerin performanslarının karşılaştırılması.....	69
6.4. YSA Yöntemiyle Seçmeli Derslerin Başarısının Tahminlenmesi.....	71
6.4.1. BSM421.....	74
6.4.2. BSM425.....	76
6.4.3. BSM435.....	78
6.5. Genel Performans Değerlendirmesi.....	80

BÖLÜM 7.

SONUÇ VE ÖNERİLER.....	82
7.1. Sonuç.....	82
7.2. Öneriler.....	83
KAYNAKLAR.....	85
EKLER	93
ÖZGEÇMİŞ	96

SİMGELER VE KISALTMALAR LİSTESİ

YSA : Yapay sinir ađları

ANFIS : Adaptif tabanlı bulanık ıkarım sistemi

ŞEKİLLER LİSTESİ

Şekil 3.1. İleri beslemeli geri yayımlı ağ	13
Şekil 3.2. Genelleştirilmiş regresyon yapay sinir ağı.....	14
Şekil 3.3. 2 girdili-2 kurallı ANFIS mimarisi	16
Şekil 5.1. Holdout yöntemi	31
Şekil 5.2. K kat çapraz doğrulama yöntemi	31
Şekil 5.3. Biri hariç çapraz doğrulama yöntemi.....	32
Şekil 5.4. Rastgele alt-örnekleme yöntemi	32
Şekil 6.1. YSA modelinin çıktı değerleri ile gerçek çıktı değerleri	60
Şekil 6.2. ANFIS modelinin çıktı değerleri ile gerçek çıktı değerleri	62
Şekil 6.3. ANFIS modelinin mimarisi.....	63
Şekil 6.4. Twoing kuralı min leaf size-hata grafiği.....	64
Şekil 6.5. Gini kuralı min leaf size-hata grafiği	65
Şekil 6.6. Gini kuralı ile oluşturulan karar ağacının oluşturduğu çıktı değerleri ile gerçek çıktı değerleri.....	70
Şekil 6.7. Twoing kuralı ile oluşturulan karar ağacının oluşturduğu çıktı değerleri ile gerçek çıktı değerleri.....	68
Şekil 6.8. Farklı öğreticiler tarafından verilen dersin not dağılımı	72
Şekil 6.9. BSM429 dersi not dağılımı	73
Şekil 6.10. BSM421 için YSA modelinin oluşturduğu çıktı değerleri ile gerçek çıktı değerleri.....	76
Şekil 6.11. BSM425 için YSA modelinin oluşturduğu çıktı değerleri ile gerçek çıktı değerleri.....	78
Şekil 6.12. BSM435 için YSA modelinin oluşturduğu çıktı değerleri ile gerçek çıktı değerleri.....	80
Şekil 7.1. SABİS ile entegre not tahmin ekran görüntüsü	84

TABLolar LİSTESİ

Tablo 6.1. BSM429 korelasyon analizi.....	37
Tablo 6.2. BSM421 korelasyon analizi.....	39
Tablo 6.3. BSM421 dersine matematik ve fizik derslerinin etkisi.....	40
Tablo 6.4. BSM425 korelasyon analizi.....	41
Tablo 6.6. BSM435 korelasyon analizi.....	43
Tablo 6.7. Regresyon modeli katsayılar tablosu	44
Tablo 6.8. BSM429 korelasyon katsayısı	47
Tablo 6.9. BSM429 model özeti	48
Tablo 6.10. BSM429 ANOVA tablosu	48
Tablo 6.11. BSM429 katsayılar tablosu	49
Tablo 6.12. BSM429 harici deęişkenler tablosu	49
Tablo 6.14. BSM421 korelasyon katsayısı	50
Tablo 6.16. BSM421 model özeti	51
Tablo 6.17. BSM421 ANOVA tablosu	512
Tablo 6.18. BSM421 katsayılar tablosu.....	52
Tablo 6.19. BSM421 harici deęişkenler tablosu	53
Tablo 6.20. BSM425 korelasyon katsayısı	53
Tablo 6.22. BSM425 model özeti	54
Tablo 6.23. BSM425 ANOVA tablosu	54
Tablo 6.24. BSM425 katsayılar tablosu	55
Tablo 6.25. BSM425 harici deęişkenler tablosu	55
Tablo 6.26. BSM435 korelasyon katsayısı	56
Tablo 6.27. BSM435 model özeti	56
Tablo 6.28. BSM435 ANOVA tablosu	57
Tablo 6.29. BSM435 katsayılar tablosu	57
Tablo 6.31. İleri beslemeli geri yayılım aęının özellikleri.....	58

Tablo 6.32. İleri beslemeli geri yayılım algoritmasının her olası durum için performansı	59
Tablo 6.33. Ağın performansının farklı istatistiksel ölçütlerle değerlendirilmesi	60
Tablo 6.34. ANFIS modelinin her olası durum için performansı	63
Tablo 6.35. Ağın performansının farklı istatistiksel ölçütlerle değerlendirilmesi	64
Tablo 6.36. Gini ve Twoing kurallarıyla oluşturulan karar ağaçlarının her olası durum için performansları	66
Tablo 6.37. Gini kuralıyla oluşturulan karar ağacının performansının farklı istatistiksel ölçütlerle değerlendirilmesi	67
Tablo 6.38. Twoing kuralıyla oluşturulan karar ağacının performansı farklı istatistiksel ölçütlerle değerlendirilmesi	69
Tablo 6.39. Yöntemlerin tahmin performanslarının karşılaştırılması	70
Tablo 6.40. Yöntemlerin performanslarının farklı istatistiksel ölçütlerle değerlendirilmesi	70
Tablo 6.41. BSM421-YSA modeli	75
Tablo 6.42. BSM425-YSA modeli	77
Tablo 6.43. BSM435-YSA modeli	79
Tablo 6.44. Teknik seçmeli derslerin performansları	80
Tablo 6.45. Teknik seçmeli derslerin tahmini çıktı değerlerinin belli bir aralıkta gösterilmesi	81
Tablo 6.46. YSA'nın çıktı değerlerinin farklı aralıklardaki doğruluk oranları	81

ÖZET

Anahtar kelimeler: Yapay Sinir Ağları, Karar Ağaçları, Adaptif Tabanlı Bulanık Çıkarım Sistemi, Karar Destek Sistemleri, Ders Seçim, Danışmanlık Hizmeti

Üniversiteler, öğrencilerin kendi alanlarının farklı dallarında kendilerini geliştirebilmeleri için seçmeli dersler sunarlar. Temel dersleri alan bir öğrenci ders seçim haftalarında bu seçmeli dersler için bir seçim kararı verir. Her öğrencinin bu karardan beklentileri farklıdır. Farklı beklentilerin ortak paydası olarak dersten alınacak başarı puanı gösterilebilir. Öğrencilerin beklentilerinin karşılanmasında ve başarılı olabilecekleri bir alana yönlendirilmesinde danışmanlara büyük sorumluluk düşmektedir.

Danışmanlık hizmeti öğrencilerin doğru yönlendirilebilmesi için önemli bir olgudur ve verim alınabilmesi için danışman öğrenci ilişkisinin kuvvetli olması gerekir. Fakat danışmanların ağır iş yüklerinden, öğrencilerin ilgi alanlarını keşfetmek ve öğrencileri tanımak için zamanın kısıtlı olması, öğrencilerin farklı şehirlerde ikamet etmeleri gibi nedenlerden dolayı öğrencilerle danışmanları arasındaki olması gereken ilişki sağlanamamaktadır.

Bu çalışmada geçmiş öğrenci verilerinden faydalanılarak dönem sonunda öğrencilerin seçmeli dersleri hangi başarı puanı ile tamamlayacağını tahmini yapılmaktadır. Bu çalışmayla öğrencilere danışmanlık hizmeti, danışman ve öğretilere ise bir karar destek sistemi sunulmaktadır.

Çalışma öğrencilere ders seçim haftalarında bir rehberlik sağlar ve öğrencilerin danışmana olan bağımlılıklarını azaltır. Danışmanlara öğrencisinin yeteneklerini, ilgi alanlarını keşfetmesinde ve öğrencilerini kısa zamanda daha iyi tanımada yardım eder. Dersi veren öğretiler, daha önceden sınıfın bilgi düzeyini ölçebilir gerekirse dersiyile ilgili iyileştirmeleri yapabilir, ders içeriğini ve ders akışını öğrenci profiline göre düzenleyebilir. Üniversite yönetimine yeni dönem için yapılacak planlamalarında fayda sağlar. Var olan öğrenci potansiyelleri önceden görülebilir ve derslerin birbirleriyle ilişkisi öğrenilebilir.

STUDENT CONSULTANCY SERVICE INFORMATION INFRASTRUCTURE DESIGN USING ARTIFICIAL INTELLIGENCE TECHNICS

SUMMARY

Key Words: Artificial Neural Network, Decision Tree, Adaptive Neuro-Fuzzy Inference System, Decision Support System, Course Selection, Consultancy Services

Universities offer elective courses to students in order to develop themselves in their different fields. Students who have completed basic courses, makes a choice decision for elective courses in course selection weeks. Each student's expectations from that decision are different. Success point of the course can be shown as the common denominator of different expectations. Consultants have a great responsibility for students in order to satisfy their expectations and directing them to a field that they will able to be successful.

Consulting service is an important phenomenon. To achieve efficiency, counselor and student relationship need to be strong. However, the reasons such as the heavy workload of advisers, lack of time to get to know the them and explore their interests, the students residing in different cities, it is not possible to provide a relationship that need to be between student and counselor.

In this research, the success of a student at the end of semester from elective courses which he/she wants to choose, is estimated by benefiting previous students data. With this study, counseling services to students and a decision support system to counselor and trainers has been presented.

The study provides guidance to students in course selection weeks and reduce students' dependence on consultants. This study provides support to advisor in estimating success of a student in course and provides help in discovering talents of his/her students and better understanding them. Lecturers can observe the capacity of class previously, and make improvements about course, reorganise syllabus and course flow. It helps to the university management for planning the new semesters. Students' potential can be seen before and related lessons can be learned.

BÖLÜM 1. GİRİŞ

Üniversiteler, ulusal ve uluslararası düzeyde çeşitli sektörlerde teknik bilgisini kullanarak çeşitli problemleri başarılı bir şekilde çözebilen, aynı zamanda sosyal, grup içerisinde çalışabilme gibi yeteneklere sahip bireyler yetiştirmeyi amaçlar. Bu amaçlarla öğrenciler için en uygun olan ders programlarını hazırlamayı hedeflerler. Farklı üniversitelerin ders müfredatları incelendiğinde, üniversite eğitiminin ilk yıllarında öğrencilere mesleki temel bilgilerin kazandırılmaya çalışıldığı görülmektedir. Bu amaçla öğrencilere eğitim sürecinin ilk yıllarında zorunlu temel dersler sunulur. Bunun yanında öğrencilere seçmeli derslerde sunulur. Sosyal seçmeli dersler ile öğrencilerin farklı alanlardaki bilgi ve ufuklarını genişletmeyi ve teknik seçmeli dersler ile öğrencilerin ilgi duydukları konularda bilgi ve becerilerini geliştirmeleri amaçlanır.

Üniversitelerin sunduğu seçmeli dersler için öğrenciler bir seçim kararı verirler. Bu kararı verirlerken birçok faktörü dikkate alırlar. Bu faktörler, her öğrencinin ders seçiminden beklentisinin ve ders seçiminden elde edeceği başarı kavramının farklı olmasından dolayı öğrenciye göre değişiklik gösterir. Kardan, Sadeghi, Ghidary ve Sani ders niteliğinin, öğretici karakteristiğinin, öğrencinin iş yükünün ve dersten alacağı tahmini puanın ders seçimini etkileyen faktörler olarak bildirmektedirler. Ayrıca bu faktörler arasında yer alan “öğrencinin dersten alacağı puan” faktörünün, birçok öğrenci için diğer faktörlere nazaran daha büyük etmene sahip olduğunu makalelerinde bildirmektedirler [1].

Öğrenciler, daha iyi bir okulda okumak, daha gelişmiş programlara, üniversitelere kabul edilebilmek, daha iyi bir iş bulabilmek için veya kişisel sebeplerden dolayı yüksek puanlara ihtiyaç duyabilirler. Bu sebeplerden dolayı dersi hangi başarı puanıyla tamamlayacaklarının bilgisi öğrenciler için önemlidir. Öyledir ki bazı

zamanlarda öğrenme onlar için ikincil plan olarak kalmakta ve dersi yüksek puanla tamamlamak birincil amaç olmaktadır [1].

Öğrenciler, ders seçim haftalarında derslerle ilgili (dersin işleyişi, içerikleri, öğreticisi, dersin zorluk derecesi ve sınavları) birbirleriyle bilgi alışverişinde bulunurlar. Fakat bu bilgilerin birçoğu bireysel fikirlerden oluşur ve bu bilgilere dayanarak öğrencinin dersten elde edebileceği başarı hakkında sağlıklı bir çıkarımda bulunulamaz. Bu nedenle bu fikirlere dayanarak atılan adımlar öğrenciyi yanlış yönlendirebilir. Öğrenciyi kendi ilgilerine ve yeteneklerine göre yönlendirebilmek için öğrencinin kendi içerisinde değerlendirilmesi gerekir. Bu aşamada öğrenciye yol gösteren danışmanlara çok önemli görevler düşmektedir.

Danışmanlık hizmeti, öğrencilerin başarısı için önemli bir olgudur. Bir öğrenci danışmanlık hizmetiyle kendi yeteneklerini keşfedebilir ve doğru yönlendirmeye yetenekleri doğrultusunda kendini geliştirebilir.

Danışmanlık sürecinin sağlıklı yürütülebilmesi için iletişimin kuvvetli olması gerekir. Fakat artan öğrenci nüfusu danışman başına düşen öğrenci sayısını arttırmakta ve iletişimi zayıflatmaktadır. Danışmanların ağır iş yüklerinden, öğrencilerin ilgi alanlarını keşfetmek ve öğrencileri tanımak için zamanın kısıtlı olması gibi nedenlerden dolayı öğrencilerle danışmanları arasındaki ilişki sağlanamamaktadır. Bu eksikliği gidermede bir karar destek sistemi yardımcı olabilir.

Bu çalışmada, öğrencilerin seçmeli derslerdeki başarı puanları tahmin edilerek danışmanlık hizmetine destek sunan bir model oluşturulmaktadır. Çalışma, Sakarya Üniversitesi Bilgisayar ve Bilişim Bilimleri Fakültesi Bilgisayar Mühendisliği bölümü 2000-2010 yılları arasında mezun olan öğrencilerin bilgilerine dayanmaktadır.

Çalışma bir karar destek sistemi olmasından dolayı öğrenciye, öğrencinin danışmanına, dersin öğreticisine hatta üniversite yönetimine yardım eder. Öğrencilere ders seçim haftalarında bir rehberlik sağlar. Danışmanlara öğrencisinin

yeteneklerini, ilgi alanlarını keşfetmesinde ve öğrencilerini kısa zamanda daha iyi tanınmasında yardım eder. Dersi veren öğretiler, daha önceden sınıfın bilgi düzeyini ölçebilir gerekirse dersiyle ilgili iyileştirmeleri yapabilir. Öğrencilere ders seçim kararlarında yardım eder. Üniversite yönetimine yeni dönem için yapılacak planlamalarında fayda sağlar. Var olan öğrenci potansiyelleri önceden görülebilir ve derslerin birbirleriyle ilişkisi öğrenilebilir.

Uygulama 5 aşamada yapılmaktadır. Uygulamanın ilk aşaması uygulama için verilerin hazırlanmasıdır. Hazırlanan veriler çok fazla niteliği sahiptir. Bu nedenle tahmini en çok etkileyen niteliklerin belirleme işleminin yapılması gerekir. İkinci aşama modelle verilecek girdi parametrelerinin belirlenmesi işlemidir. Bu aşamayla beraber yöntemlere verilecek veri kümesi hazırlanmış olmaktadır. Bir sonraki aşama en iyi tahmini yapan tekniğin seçilmesidir. Bu aşamada 3 farklı yapay zeka tekniği kullanılmaktadır. En iyi tahmin performansı gösteren model seçilir ve diğer seçmeli derslerin başarı puanının tahminde kullanılır. En son aşama olarak oluşturulan modelin farklı dersler üzerindeki tahmin performansları değerlendirilir.

1.1. Literatür Taraması

Eğitim alanında yapılan birçok çalışmada veri madenciliği yöntemlerinden faydalanılmaktadır. Araştırmacılar, öğrencilerin başarısını artıracak, öğrenciler için daha iyi bir çalışma ortamı ve eğitim hayatı yaşamalarını sağlayacak faktörleri veri madenciliği yöntemleri kullanarak belirlemeye ve analiz etmeye çalışmışlardır. Yalnızca, öğrenciler için değil eğitici, eğitim kurumunun yöneticileri için daha iyi bir eğitim planı ve düzeni sağlayabilmek, eğitimde yaşanan eksikleri giderebilmek için veri madenciliği yöntemlerine başvurulmuştur.

Öğrencinin akademik performansının değerlendirilmesi (SAP) ve belirli bir dersten elde edilecek başarının tahmini üzerine çok fazla çalışma yapılmıştır.

(Taylan and Karagözoğlu 2009) çalışmalarında gerçek 214 adet öğrenci verisini kullanmaktadırlar. Mühendislik ekonomisi dersini alan öğrencilerin ana, ara, final, ödev, kısa sınav notları gibi parametrelerini kullanarak bu dersi hangi başarı puanıyla

tamamlayacağıının tahminini ANFIS yöntemiyle gerçekleştirmektedirler. İstatiksel yöntemlerden elde edilen sonuçlar ile ANFIS yöntemiyle elde edilen sonuç karşılaştırılmaktadır [2]. (Oladokun, Adebajo, and Owaba 2008) tarafından yapılan çalışmada yaş, cinsiyet, üniversite konumu, matematik, ingilizce ve kimya derslerinden alınan puanlar gibi sosyal, okul ve aile bilgilerine dayalı parametrelere bakarak yapay sinir ağları ile üniversiteye kabulü düşünülen bir adayın olası performansı incelenmektedirler [3]. (İbrahim Z 2007) tarafından yapılan çalışmada yapay sinir ağları, karar ağaçları ve doğrusal regresyon yöntemleri kullanılarak öğrencilerin akademik performansının tahmini gerçekleştirilmekte ve tahmin edici değişkenler olarak öğrencilerin demografik profilleri ve lisans seviyesindeki öğrencilerin ilk dönemdeki genel not ortalamaları kullanılmaktadır [4]. (Gedeon and Turner 1993) tarafından yapılan çalışmada lisans öğrencilerinin bilgisayar bilimi dersinden alacağı final notlarının tahmini gerçekleştirilmektedir. Bu tahmin, bilgisayar bilimi dersinin laboratuvar egzersizleri, ara sınav, kısa sınav, ödev puanları kullanılarak ileri beslemeli geri yayılım algoritmasıyla gerçekleştirilmektedir. Çalışmada 153 veri kullanılmaktadır [5]. (ŞENGÜR and TEKİN 2013) çalışmalarında, yapay Sinir Ağları (YSA) ve karar ağaçları (KA) yöntemlerini kullanarak Fırat Üniversitesi Eğitim Fakültesi Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü (BÖTE) öğrencilerinin mezuniyet notlarını tahmin etmektedirler. Tahminde, 127 BÖTE öğrencisinin 4 yıl süresince almış olduğu toplam 49 kültür ve mesleki dersin yılsonu notu kullanılmaktadır [6]. (GÜNER and ÇOMAK 2011) tarafından yapılan çalışmada mühendislik öğrencilerinin matematik 1 dersindeki başarılarının tahmini gerçekleştirilmektedir. Tahmin için destek vektör makineleri kullanılmaktadır. Çalışmada kullanılan veri kümesi 2007 yılı ÖSS sınavı puanları ile Pamukkale Üniversitesi Mühendislik Fakültesinin sekiz farklı bölümüne yerleştirilmiş olan 434 öğrencinin bilgilerini içermektedir. Makine 289 öğrencinin verisi ile eğitilmekte, 145 öğrencinin verisi ile test edilmektedir. Matematik I dersinden geçen öğrencilerin % 86'sı doğru olarak tahmin edilmiştir [7]. (Bozkır, Sezer, and Gök 2009), çalışmalarında veri madenciliği yöntemlerini kullanarak öğrencilerin öğrenci seçme sınavında (OSS) gösterdikleri başarıyı etkileyen faktörleri incelemektedirler. Çalışmada kullanılan veri kümesi ÖSYM'nin internet sitesinde 2008 ÖSS dönemine ilişkin yayınlanmış olan öğrenci bilgi anketini temel almaktadır. Ankette yer alan toplam 80 adet soru, öğrencilerin okulda aldıkları eğitim

ve sosyal durumlarını araştırmaya yönelik dereceleme sorularından oluşmaktadır. Toplam 77 nitelik içeren 9952 adet öğrenci verisinin 8952 adedi karar ağacı ve kümeleme modellerinin eğitimi için kullanılmaktadır [8].

Öğrencinin ders seçimini tahmin etmede veri madenciliği yöntemlerinin kullanımı üzerine birçok çalışmada mevcuttur. (Kardan et al. 2013) çalışmalarında öğrencilerin seçtiği online derslerle alakalı öğrenci memnuniyetini etkileyen potansiyel faktörlerin tanımlanmasına odaklanırlar. Uygulamada kullanılan veriler öğrencilerden anket yoluyla toplanılmaktadır. Öğrenciler Likert ölçeği ile soruları cevaplandırmışlardır. Sorular ders karakteristiği, öğretici karakteristiği, öğrencinin iş yükü, dersteki puanlama, dersin tipi, dersin zamanı, derslerin çakışması, final sınavının süresi ve öğrenci talepleri şeklinde farklı kategorilere ayrılmaktadır. Uygulamada 3 katmanlı geriye yayılma algoritması kullanılmaktadır. 214 öğrenci ankete katılmaktadır ve 74%'ü eğitim için kullanılmaktadır. Çalışmanın sonucunda, öğrencinin ders seçimine etki eden kategoriler ve bu kategorilerin etkileme dereceleri elde edilmektedir [1] (Lykourantzou et al. 2009) çalışmalarında ileri beslemeli yapay sinir ağı, destek vektör makineleri ve PESFAM tekniklerini kullanarak öğrencilerin dersi bırakıp bırakmayacağını tahminini gerçekleştirirler. Daha doğru bir tahmin gerçekleştirebilmek için 3 teknik birleştirilmekte ve 3 karar şeması kullanılmaktadır. En az bir teknik, öğrencinin bu dersi bırakacağı yönünde karar veriyorsa bu durum karar şeması 1 grubuna girer. En az iki teknik öğrencinin bu dersten ayrılacağı yönünde tahmin ediyorsa bu durum karar şeması 2 grubuna girer. Bütün teknikler bu öğrencinin bu dersten ayrılacağı yönünde karar veriyorsa bu durum karar şeması 3 grubuna girer. Son olarak öğrencinin hangi sınıfa girdiği bu karar şemasına bakılarak tahmin edilir. Bu çalışmada bütün öğrencilerin 75-85% doğru olarak tahmin edilmiştir [9]. (Guo 2010) çalışmasında öğrencinin ders memnuniyetini ölçmek için istatistiksel yöntemleri ve yapay sinir ağlarını kullanır. Veri kümesi 2002 ile 2007 yılları arasında 43 dersten toplanan öğrenci anket sonuçlarından elde edilmektedir ve yapay sinir ağları istatistik yöntemlerden daha başarılı sonuçlar verir [10]. (Herzog 2006) öğrenci profilini tespit etmek için veri madenciliği yöntemlerini kullanırlar. Veri madenciliği yöntemleri olarak lojistik regresyon ve karar destek vektörleri kullanılmakta ve veriler anket yoluyla elde edilmektedir. Lojistik regresyon daha iyi sonuç vermiştir [11]. (Zhang et al. 2007) okulu bırakma potansiyelinde olan

öğrenciyi tahmin etmek için naive bayes, karar ağaçları ve karar destek vektörlerini kullanırlar. Naive bayes en iyi performansı sağlarken, karar ağaçları bu yöntemler içinde en kötü performansa sahiptir [12]. (Hamid n.d.) yaptığı tez çalışmasında lojistik regresyon ve yapay sinir ağlarını kullanarak öğrencilerin performanslarını incelemektedir. Çalışmasında Malezya Utara üniversitesinin mezunlarının bilgilerinden faydalanmaktadır. Yapay sinir ağları daha iyi performans göstermiştir [13].

BÖLÜM 2. VERİ MADENCİLİĞİ

2.1. Giriş

Geçmişten bugüne kadar gelen süre içerisinde veri sayısında hızlı bir artış söz konusudur. Bu büyük veri yığınları gizli ya da çok net olmayan ve önceden bilinmeyen kullanışlı bilgiler içerir. Saklı olan bu bilgilerin ortaya çıkarılması işlemine veri madenciliği adı verilir. Veri madenciliğindeki genel amaç, büyük veri depolarından saklı ya da çok net olmayan yararlı bilgileri ortaya çıkarmak ve kullanım için uygun hale getirmektir.

Çok büyük veri yığınlarından kritik bilgiler çıkarmak, normal şartlar altında uzun zaman alan bir süreçtir. Yapılan çıkarımların doğruluğu da kesin değildir. Veri madenciliği ile kısa sürede doğru bilgiler elde edilebilir. Objektif değerlendirmelerin yapılmasında ve stratejik kararların alınmasında bu bilgiler kullanılabilir. Kısaca, veri madenciliği, çok büyük miktardaki verilerin içindeki ilişkileri inceleyerek aralarındaki bağlantıyı bulmaya yardımcı olan ve veri tabanı sistemleri içerisinde gizli kalmış bilgilerin çekilmesini sağlayan veri analizi tekniğidir [14].

Veri madenciliği sürecinin çeşitli aşamalarında, büyük anlamsız görünen veri toplulukları içerisinde gizli bilgiyi ortaya çıkarmak için çeşitli teknikler kullanılır. Bu teknikler, istatistiksel yöntemler, bellek tabanlı yöntemler, genetik algoritmalar, yapay sinir ağları ve karar ağaçları olarak sıralanabilir. Her geçen gün bu tekniklere yenileri de eklenmektedir. Yıllardır kullanılan klasik yöntemler diyebileceğimiz istatistiksel yöntemler de bir veri madenciliği yöntemidir.

Veri madenciliği, büyük hacimde veri bulunan her yerde kullanılabilir. Günümüzde veri madenciliği yöntemleri karar verme sürecine ihtiyaç duyulan birçok alanda

yaygın olarak kullanılmaktadır. Bankacılık, sağlık, eğitim, mühendislik, endüstri vb. birçok alanda başarılı bir şekilde kullanılmaktadır.

2.2. Veri Madenciliği Süreci

Veri madenciliği bir süreçtir. Büyük bir veri yığınınından anlamlı bir bilginin çıkarılması ve bu bilgi keşfi sürecinde örüntüleri ayrıştırmak, uygulama için hazır hale getirmek veri madenciliği süreçlerinin bir parçasıdır.

Bir veri madenciliği uygulaması için gerekli olan süreçler:

- 1) Problemin tanımlanması,
- 2) Problemi temsil eden verilerin hazırlanması: Verilerin hazırlanması aşaması kendi içinde bir süreç içerir.
 - Veri temizleme: Verilerin hazırlanması işleminde veri kümesi içerisinde yer alan, uygun olmayan veya hatalı girilmiş veriler ayklanır. Eksik veriler uygun değerlerle doldurulur veya eksik veri çok ise bu kayıt veri kümesinden silinebilir. Eksik veri sorunu, eğer yeterince veri mevcutsa değer içeren kayıtlar veri kümesinden atılarak, kayıp verilerin yerine sabit bir değer verilerek giderilebilir.
 - Veri bütünleştirme: Farklı veri tabanlarından ya da veri kaynaklarından elde edilen verilerin bir araya getirilerek birlikte değerlendirmeye alınabilmesi için aynı bilgiyi içeren verilerinin tek tipe dönüştürülmesi işlemidir [15].
 - Veri indirgeme: Veri indirgeme ile veri kümesinin indirgenmiş örneği elde edilir. Veri indirgeme işlemi ile veri madenciliği yöntemleri daha iyi performans sergileyebilirler.
 - Veri dönüştürme: Veri dönüştürme işlemi ile veriler veri madenciliği için uygun formlara dönüştürülür. Veri dönüştürme işlemi ile veri madenciliği yöntemlerinin gösterdikleri performans artabilir. Veri dönüştürme; düzeltme, birleştirme, genelleştirme ve normalleştirme gibi değişik işlemlerden biri veya bir kaçını içerebilir. Veri normalleştirme en sık kullanılan veri dönüştürme işlemlerinden birisidir.
- 3) Modelin kurulması ve değerlendirilmesi: Problemi en iyi temsil eden modelin kurulması işlemidir. Problemi en iyi temsil eden model, birçok deneme sonucunda oluşturulan modeller arasından seçilir. Oluşturulan modeller

değerlendirilir ve en iyi olduğu kabul edilen model problemin temsilinde kullanılır.

- 4) Bilgi üretimi: Kurulan modelle eldeki veri kümesinden bilgilerin üretilmesi aşamasıdır.

BÖLÜM 3. VERİ MADENCİLİĞİ YÖNTEMLERİ

3.1. Sınıflandırma Yöntemleri

Sınıflandırma, gerçek dünyada çokça karşılaşılan önemli bir karar verme eylemidir. Gerçek dünyada sınıflandırmanın birçok örneği mevcuttur. Sanayi, bilim, tıp, mühendislik alanlarındaki birçok problem sınıflandırma problemi olarak ele alınır. Hava durumunun tahmini, tıbbi teşhis [16][17][18][19] el yazısı karakter tanıma [20] [21][22], ses tanıma [23][24], iflas tahmini [25][26][27][28][29][30], kalite kontrol gibi birçok sınıflandırma problemi örneği verilebilir [31].

Genellikle sınıflandırma problemleri iki tiptedir: İkili sınıflandırma problemleri ve çok sınıflı problemler. İkili sınıflandırma problemlerinde sınıflandırılacak nesne 2 sınıftan birine aittir. Fakat çok sınıflı problemlerde ikiden fazla sınıf vardır ve nesne bu sınıflardan birine aittir [32].

Sınıflandırma problemlerine çözüm getirebilmek için birçok yöntem sunulmuştur. İstatiksel yöntemler, yapay sinir ağları, karar ağaçları, destek vektör makineleri (SVM), doğrusal regresyon, lojistik regresyon, diskriminant analizi, genetik algoritma, bulanık mantık, k-en yakın komşu [33][34], ayrıca hibrit teknikleri [25] bu yöntemlerden bazılarıdır.

3.1.1. Yapay sinir ağları

Yapay sinir ağları, sınıflandırma problemlerinde başarılı sonuçlar veren yapay zekâ tekniklerinden biridir [35]. Bu teknik insan beyninden esinlenerek geliştirilen bir yöntemdir. İnsan beyni birçok sinir hücresinden oluşur ve bu sinir hücreleri birbirlerine bağlanarak fonksiyonları yerine getirirler. Yapay sinir ağları insan beynindeki sinir hücrelerine benzer şekilde yapay sinir hücrelerinden ve bu

hücrelerinin birbirlerine olan bağlantılarından oluşur. Yapay sinir ağları doğrusal olmayan bir yapıya sahiptir, kendini veriye göre ayarlar ve eğitir [31].

3.1.1.1. Yapay sinir hücresi

Yapay sinir ağlarında yer alan yapay sinir hücreleri literatürde proses elemanları olarak adlandırılır. İşlemler bu proses elemanlarında icra edilir. Proses elemanının 5 temel elemanı vardır [36].

1. Girdiler: Dış dünyadan veya bir başka proses elemanından gelen bilgilerdir. Bu bilgiler dış dünyaki olayları temsil eden verilerdir.
2. Ağırlıklar: Girdi elemanından alınan bilgiler bir önem derecesine sahiptir. Bilgilere önem derecesine göre hücre içerisinde bir değer verilir. Bu değer negatif olabileceği gibi 0 değeri de olabilir. Negatif değer bu bilginin önemsiz olduğu anlamına gelmez. Gerçek dünya olayını negatif yönde etkilediği anlamına gelir. 0 değeri ise bu bilginin hiçbir etkisinin olmadığını anlamına gelir.
3. Toplama Fonksiyonu: Girdiler önem derecelerine göre bir ağırlık değeriyle çarpılır. Tüm girdiler için elde edilen sonuçlar burada toplanır. Toplama fonksiyonları sinir hücreleri arasında farklılık gösterebilir. Birçok toplama fonksiyonu bulunmaktadır. Bir ağda bulunan tüm sinir hücrelerinin aynı toplama fonksiyonuna sahip olması beklenemez.
4. Aktivasyon Fonksiyonu: Toplama fonksiyonundan elde edilen net girdi işlenerek hücrenin çıkış değerini belirler. Aktivasyon fonksiyonları her hücre için farklı olabilir. En uygun aktivasyon fonksiyonu, tasarımcının denemeleri ile belirlenebilir.
5. Hücrenin Çıktısı: Aktivasyon fonksiyonu ile elde edilen değerdir. Bir sinir hücresinin sadece bir çıktısı vardır. Bu çıktı bir veya birden fazla sinir hücresine girdi olarak verilebilir. Hatta sinir hücresinin çıktısı kendisinin tekrar girdisi olarak kullanılabilir.

3.1.1.2. Yapay sinir ağı yapısı

Yapay sinir ağları, genel olarak 3 katmanlı bir yapıya sahiptir. Bu katmanlar girdi katmanı, ara katmanlar ve çıktı katmanıdır. Bir yapay sinir ağı birden fazla ara katmana sahip olabilir fakat sadece bir girdi ve bir çıktı katmanına sahiptir. Girdi katmanı ile dış dünyadan alınan bilgiler ara katmana iletilir. Ara katmanda işlenen bilgiler çıktı katmanına gönderilir. Çıktı katmanı ara katmandan gelen bilgileri işleyerek ağı nihai çıktısını üretir. Üretilen çıktı bu katmandan dış dünyaya gönderilir.

3.1.1.3. Yapay sinir ağlarında öğrenme

Yapay sinir ağları, öğrenme türlerinden olan öğretmenli öğrenme yoluyla öğrenmesini gerçekleştirir. Sinir ağına, dış dünya olayını temsil eden örnekler verilir. Ağ bu örnekleri işler ve gelen her bilgiye bir ağırlık verir. Ağ, doğru ağırlık değerlerine ulaştığında öğrenmeyi tamamlamış olur. Doğru ağırlık değerlerini bulabilmek için örnekler belli iterasyon sayısınca ağa tekrar tekrar verilir. Her iterasyonda ağırlıklar güncellenir. Güncelleme öğrenme kuralları ile yapılır. Öğrenmeyi tamamlayan bir ağ genelleme yapabilme yeteneğine sahip olur. Genelleme yapabilme yeteneğine adaptif öğrenme denir. Ağ, hiç görmediği örnekler hakkında bu yeteneği sayesinde bir sonuç üretebilir hale gelir. Ağı bu örneklerin girdilerine karşılık ürettikleri çıktılarının doğruluğu ağı öğrenme derecesini gösterir.

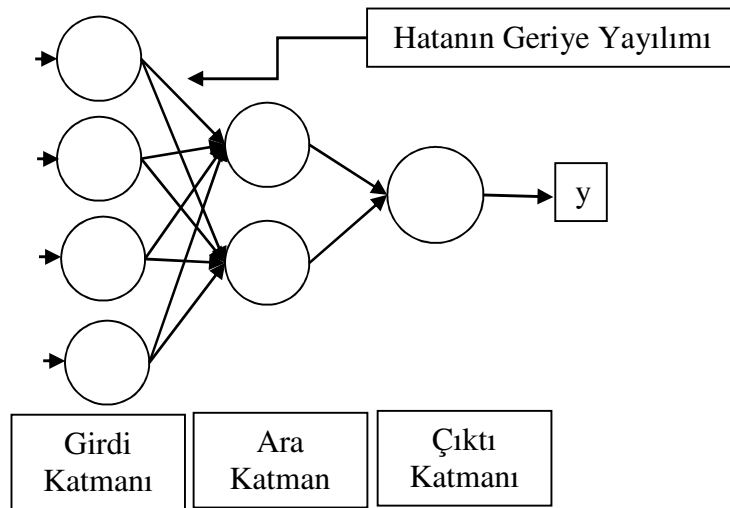
3.1.1.4. Yapay sinir ağı modelleri

Yapay sinir ağı modelleri toplama ve aktivasyon fonksiyonlarına, öğrenme kurallarına ve proses elemanlarının birbirlerine bağlanma şekillerine göre değişir. Bu bölümde pratikte kullanılan ve Matlab aracının arayüzünde yer alan ağ modelleri hakkında bilgi verilecektir.

İleri beslemeli geri yayımlı ağ (FBPNN): İleri beslemeli geri yayımlı ağ, proses elemanlarından ve proses elemanlarının birbirleriyle yaptığı bağlantılardan oluşur. Temel olarak 3 katmandan oluşur: Girdi katmanı, saklı katman ve çıktı katmanı.

Herhangi bir katmanda bulunan proses elemanlarının her biri, kendi katmanlarından bir önceki katmanda yer alan proses elemanlarının tümü ile doğrudan bağlantıya sahiptir. İleri beslemeli bir yapıya sahip olmasından dolayı bilgiler sadece tek yönlü akış halindedir ve ileriye doğrudur. Bağlantılar arasında doğrudan bir döngü yoktur.

İleri beslemeli geri yayımlı ağ öğrenmeyi girdi ve çıktı örneklerine bakarak gerçekleştirir. Alınan örnekler saklı katmanlarda işlenir ve bu girdilere karşılık bir çıktı üretilir. Gerçek çıktı değeri ile ağın ürettiği çıktı değeri arasındaki fark hata değerini verir. Ağ, yapısı gereği hata değerini geriye doğru yayar ve bu hata değerine bakarak proses elemanları arasındaki bağlantı ağırlıklarını günceller. Ağırlıkları genelleştirilmiş delta kuralıyla günceller [37]. Şekil 3.1’de ileri beslemeli geri yayımlı ağ örneği gösterilmektedir.



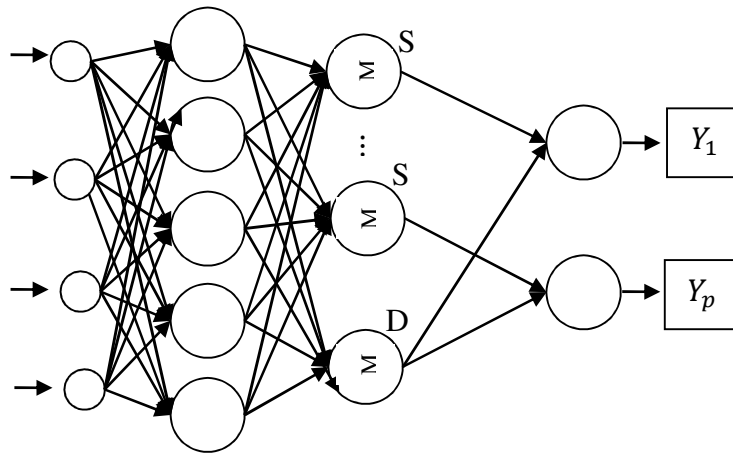
Şekil 3.1. İleri beslemeli geri yayımlı ağ

Radyal esaslı yapay sinir ağı (FBPNN): Radyal tabanlı ağ, fonksiyon yaklaşım teorisinden türemiştir [38]. 3 katmana sahiptir: Girdi katmanı, saklı katman ve çıkış katmanı. Örnekler, radyal temelli fonksiyonlarla saklı katmanda işlenir.

Genelleştirilmiş regresyon yapay sinir ağı (GRNN): Genelleştirilmiş regresyon sinir ağı, radyal temelli bir sinir ağıdır. Az sayıda eğitim örneğinden bile fonksiyon üretebilme yeteneğinden dolayı tercih edilen bir sinir ağıdır [39]. Ağın eğitme işlemi oldukça hızlıdır. Ağ yapısının basitliği nedeniyle görüntü işleminde dâhil olduğu pek çok uygulamada kullanılmaktadır. Genelleştirilmiş regresyon sinir ağları

özellikle sistem modelleme, fonksiyon yaklaşırma ve sınıflandırma uygulamaları için en uygun ileri beslemeli yapay sinir ağlarıdır.

GRNN, giriş katmanı, örüntü katmanı, toplama katmanı, çıkış katmanı olmak üzere 4 katmandan oluşmaktadır. Giriş katmanında giriş parametre sayısı kadar proses elemanı kullanılmaktadır. Örüntü katmanındaki elemanlar, her bir girişin kayıtlı örüntülerden olan uzaklık bilgisini içermektedir [40]. Giriş katmanı ile örüntü katmanı arasında ağırlıklandırılmış bağlantılar mevcuttur. Örüntü katmanındaki her bir eleman toplama katmanındaki S ve D toplama elemanlarına bağlıdır. S toplama elemanı örüntü katmanından gelen çıkışları ağırlıklı olarak toplarken D toplama elemanı ise örüntü tabakasından gelen çıkışları ağırlıksız olarak toplar. Çıkış katmanındaki her bir proses elemanına, sadece bir adet S toplama elemanı ve çıkış katmanındaki her bir elemana ortak olarak bağlanan bir adet D toplama elemanı bağlanır. Ağın çıkışı, S toplamının D toplamına bölünmesiyle elde edilir [41][42]. Şekil 3.2'de genelleştirilmiş regresyon sinir ağına ait bir örnek mimari gösterilmektedir.



Şekil 3.2. Genelleştirilmiş regresyon yapay sinir ağı

3.1.2. Adaptif ağ tabanlı bulanık çıkarım sistemi

Bulanık çıkarım sistemleri, bulanık mantık kullanarak verilen girdiyi ilgili çıktıya haritalar [43]. Bulanık sistemler, if-then kural yapısına bağlı olarak ikiye ayrılır: Mamdani model ve Takagi-Sugeno model [44][45].

Sugeno tipi modelleme 1985 yılında sunulmuştur ve Mamdani tipi modellemenin modifikasyonu şeklindedir. Mamdani tipi bulanık çıkarım sistemi ile Sugeno tipi bulanık çıkarım sistemleri arasındaki temel fark net çıktısının üretilme şeklidir [46][47][48]. Sugeno tipi modellemede çıktı değişkenleri girdilerin bir fonksiyonu şeklinde üyelik fonksiyonlarına sahip olmaktadır ve çıktı üyelik fonksiyonları lineer veya sabit olmak zorundadır [47][49][45]. Çıktı üyelik fonksiyonu sabit ise bulanık çıkarım sistemi zero-order Sugeno modeli olarak adlandırılır. Eğer çıktı üyelik fonksiyonu birinci dereceden bir polinom ise bulanık çıkarım sistemi first-order Sugeno model diye isimlendirilir [50][51][52].

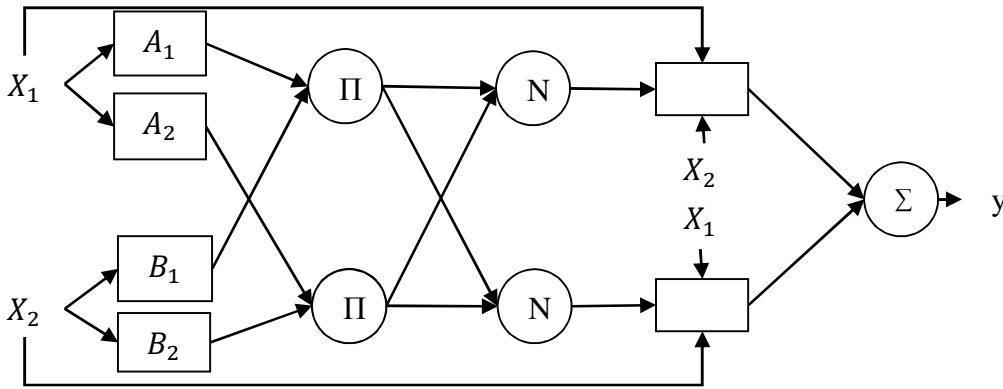
ANFIS yapay sinir ağlarının paralel hesaplayabilme ve öğrenme kabiliyeti ile bulanık mantığın çıkarım özelliğini kullanan melez bir yapay zekâ yöntemidir [53].

3.1.2.1. ANFIS mimarisi

ANFIS yöntemi Sugeno tipi bulanık çıkarım sistemini kullanır. ANFIS mimarisini sunmak için birinci dereceden Sugeno modeline dayalı 2 bulanık kural örneği eşitlik 3.1 'de verilmektedir [51][47][54][55][56][57].

$$\begin{aligned} \text{Eğer } x=A_1, y=B_1 \text{ ise } F_1 &= P_1x + Q_1y + R_1 \\ \text{Eğer } x=A_2, y=B_2 \text{ ise } F_2 &= P_2x + Q_2y + R_2 \end{aligned} \quad (3.1)$$

İki girişli, bir çıkışlı ve 2 kurallı Sugeno tipi ANFIS mimarisi Şekil 3.3'de gösterilmektedir.



Şekil 3.3. 2 girdili-2 kurallı ANFIS mimarisi

Şekilde görüldüğü gibi ANFIS mimarisi 5 katmanlı yapıya sahiptir. Her bir katmanın görevi ve özellikleri aşağıda anlatılmaktadır.

Katman 1: Bu katman bulanıklaştırma katmanıdır [58]. Bu katmanda yer alan her düğüm bir kare düğümdür [59] ve adaptiftir [60]. Adaptif düğümlerin çıktıları bu düğümlerle ilişkili düzenlenebilir parametrelere dayanmaktadır. Bu katmanda bulunan A_1, A_2, B_1, B_2 sözel etiketlerdir. X ve Y bu katmanın girdileridir. Bu katmandaki her düğüm bir düğüm fonksiyonuna sahiptir. Düğüm fonksiyonları bir üyelik fonksiyonudur. Girdiler, üyelik fonksiyonlarının çıktı değerleri kadar bulanık kümelere üyedir. Üçgensel üyelik fonksiyonu, genelleştirilmiş çan eğrisi üyelik fonksiyonu ve gauss üyelik fonksiyonu gibi birçok üyelik fonksiyonu mevcuttur [61]. Jang, ANFIS modelinde Gauss üyelik fonksiyonunu kullanmaktadır. $\mu_{A_i}(x)$ genellikle maksimumu 1 ve minimumu 0 olan bir Gauss eğrisi şeklinde seçilir.

Katman 2: Bu katman kural katmanıdır. Bu katmanda yer alan her düğüm daire şeklindedir ve sabit bir düğümdür. Bu katmanın girdilerini ilk katmanın çıktı değerleri oluşturur ve katmanın çıktı değeri 1. Katmandan gelen üyelik derecelerinin çarpımıyla elde edilir. Elde edilen her çıktı, bir kuralın ateşleme seviyesini gösterir. Eşitlik (3.2)'e göre çıktı hesaplanır.

$$W_i = \mu * A_i(x) + \mu * B_i(y), i = 1,2 \quad (3.2)$$

Katman 3: Bu katman normalizasyon katmanıdır. Bu katmanda yer alan her düğüm daire şeklindedir ve sabit bir düğümdür. Katmanın girdilerini kural katmanının çıktı

değerleri oluşturur. Katmandaki her düğüm, her bir kuralın normalleştirilmiş değerini hesaplar [53]. i . düğümün ateşleme seviyesinin tüm düğümlerin ateşleme seviyeleri toplamına oranı katmanın çıktısını verir. Düğümlerin çıkışları, normalize edilmiş ateşleme seviyesi olarak isimlendirilir. Denklem (3.3)'e göre hesaplama gerçekleştirilir.

$$W' = \frac{W_i}{(W_1 + W_2)} \quad (3.3)$$

Katman 4: Bu katman berraklaştırma katmanıdır. Bu katmanda yer alan her düğüm kare düğümdür ve adaptiftir. Katmanda bulunan her bir düğümden verilen bir kuralın ağırlıklandırılmış sonuç değerleri hesaplanır [53]. Düğümün çıkışı denklem (3.4)'e göre hesaplanır. Denklemde ve bu katmanda yer alan parametreler sonuç parametreleri olarak adlandırılır.

$$O_i = W'_i F_i = W_i (P_i X + Q_i X + R_i) \quad (3.4)$$

Katman 5: ANFIS modelinin son katmanıdır. Bu katman toplam katmanıdır. Bu katmanın girdilerini 4. Katmanın çıktıları oluşturur. Çıktısı ise 4. katmanın çıktılarının toplamıdır, denklem (3.5) ile hesaplanır. Katmanda sadece bir düğüm vardır ve bu düğümün çıkışı ANFIS modelinin verdiği nihai sonuçtur.

$$y = \sum_{i=1}^n O_i = W_i (P_i X + Q_i X + R_i) \quad (3.5)$$

3.1.2.2. ANFIS Modelinin Öğrenme Algoritması

ANFIS mimarisinin geliştirildiği zamanlarda, kuralların öğrenilmesi ve optimum kural kümesi elde edebilmek için bir çok yöntem önerilmiştir [62]. Örneğin; (Kumar and Garg 2004) ANFIS eğitimi için Kohonen ağını kullandı [63]. (Mascioli, Varazi and Martinelli 1997) optimum bulanık kümeyi elde etmek için ANFIS modeli ile

Min-max algoritmasını birleştirmeyi önerdi [64]. (Jang 1993), ANFIS yapısının parametrelerini güncellemek için 4 yöntem sundu [65]:

1. Yalnızca GD (Gradyan İniş) : Tüm parametreler yalnızca gradyan inişle güncellenir.
2. En küçük kareler yöntemi sonuç parametrelerinin başlangıç değerlerini almak için eğitimin en başında sadece bir kez uygulanır. Daha sonrasında GD, tüm parametrelerin güncelleştirme işlemini üstlenir.
3. Gradyan iniş ve en küçük kareler yöntemi: Hibrit öğrenmedir.
4. Kalman Filtresi: Tüm parametreleri güncelleştirmek için genişletilmiş Kalman filtresi kullanılır

ANFIS modelinin öğrenme algoritması gradyan metodu ile en küçük kareler yönteminin birleşimidir [66]. Ağın ileriye doğru eğitilmesi safhasında sonuç parametreleri en küçük kareler tahminiyle tespit edilir [67][68]. Hata sinyali çıktı katmanından girdi katmanına doğru, geriye doğru yayılır. Geriye doğru yayılımda Gradyan iniş algoritmasıyla öncül parametreler güncellenir [69].

3.1.3. K en yakın komşu algoritması

Bu teknikte, tüm örnekler bir örüntü uzayında saklanır. Bilinmeyen bir örnek geldiğinde bu örneğin hangi sınıfa ait olduğu belirlenmeye çalışılır. Belirleme işlemi bu örneğe en yakın k nokta incelenerek gerçekleştirilir. Yakınlık Öklid, Manhattan ve Minkowski uzakları ile tanımlanır. Bilinmeyen örnek, en yakın k kadar örneğin hangisine en çok benziyorsa o sınıfa aittir denir.

Öklid uzaklığı: İki birim arasındaki Öklid uzaklığı denklem (3.6) ile hesaplanır [70]:

$$d(i, j) = \sqrt{(X_{i1} - X_{j1})^2 + (X_{i2} - X_{j2})^2 + \dots + (X_{ip} - X_{jp})^2} \quad (3.6)$$

Manhattan uzaklığı: İki birim arasındaki Manhattan uzaklığı denklem (3.7) ile hesaplanır [70]:

$$d(i, j) = (|X_{i1} - Y_{j1}| + |X_{i2} - Y_{j2}| + \dots + |X_{ip} - Y_{jp}|) \quad (3.7)$$

Minkowski uzaklığı: İki birim arasındaki Minkowski uzaklığı denklem (3.8) ile hesaplanır [70]:

$$d(i, j) = [|X_{i1} - X_{j1}|^m + |X_{i2} - X_{j2}|^m + \dots + |X_{ip} - X_{jp}|^m]^{1/m} \quad (3.8)$$

3.1.4. Naive bayes algoritması

Naive bayes, hedef değişkenle bağımsız değişkenler arasındaki ilişkiyi analiz eden tahminci ve tanımlayıcı bir sınıflandırma algoritmasıdır [71] İsmi İngiliz matematikçi Thomas Bayes'ten alır.

Naive bayes algoritmasında sınıf üyeliği bilinmeyen bir veri örneğinin ($x = \{X_1, X_2, \dots, X_n\}$) m adet sınıfın hangisine ait olduğu incelenir. Her sınıf için veri örneğinin o sınıfa ait olma olasılıkları hesaplanır. Olasılık denklemi denklem (3.9)'da verilmektedir.

$$P(C_i | X_1, \dots, X_n) = \frac{P(C_i)P(F_1, \dots, F_n | C_i)}{P(F_1, \dots, F_n)} \quad (3.9)$$

Denklemden yer alan C_i sınıfları temsil etmektedir. Örnek veri, en yüksek olasılığa sahip olan sınıfa aittir.

3.1.5. Karar ağaçları

Karar ağacı yaklaşımı, hedef fonksiyonlarını yaklaşık olarak hesaplamak için kullanılan ve öğrenme fonksiyonunun karar ağacı ile gösterildiği bir yöntemdir [72]. Bir karar ağacı ise ağaç görünümünde tanımlayıcı ve tahmin edici bir modeldir [73]. Karar ağaçları karar noktasını, karar noktasını etkileyen parametreleri, seçenekleri ve bu seçenekleri seçmenin olası sonuçlarını düzenlemede, incelemeye, görmede etkili bir yapıya sahiptir.

Karar ağaçları, karar alıcıya karar alırken hangi faktörlerin göz önüne alınması ve her bir faktörün kararın farklı çıktıları ile geçmişte nasıl ilişkili olduğunun belirlenmesi konularında yardımcı olur [74]. Karar ağaçları sadece kararları göstermez aynı zamanda kararların açıklanmasını da gösterir.

3.1.5.1. Karar ağacının temel yapısı

Bir karar ağacı temel olarak düğüm, dal ve yapraktan oluşur. Düğümler veri kümelerinin özniteliklerini temsil eder. Yapraklar karar ağacı için olası sonuçları temsil etmektedir. Ağacın en üstünde bulunan düğüm ağacın köküdür. Ağacın kökü ile yapraklar arasında bağlantı, bir düğümden başka bir düğüme geçiş dallar aracılığıyla olur. Veri kümesi, ağaç yapısı karar iskeletine göre sınıflandırılır. Veri her karar seçiminde farklı bir dala ulaşır. Yaprğa ulaşana kadar bu işlem gerçekleştirilir. Ulaştığı yaprak verinin sınıfı olur.

3.1.5.2. Karar ağacının oluşturulması

Karar ağaçlarının oluşturulmasında temel prensip verilere ilişkin bir dizi soruların sorulması ve elde edilen cevaplar doğrultusunda hareket edilerek en kısa sürede sonuca gidilmesidir. Karar ağacı sorulara verdiği cevapları toplayarak kuralları oluşturulur. Bu kurallar karar kurallarıdır ve yaprak sayısı kadar kural oluşacaktır [75]. Eğitim verileriyle oluşturulan karar ağaçlarının sınıflandırma performansları test veri kümesiyle değerlendirilir. Test verileri, karar ağacının kökünden başlar ve verdiği karara göre bir alt düğüme geçer bu işlem düğüme ulaşıncaya kadar devam eder. Test verisi bir sınıfa ait olur. Tüm test verileri sınıflandırılır. Sınıflandırılan verilerin sayısı karar ağacının genelleme yeteneğini gösterir. Bir karar ağacı aşağıdaki adımlarla oluşturulur [76].

- 1) Eğer eğitim kümesindeki tüm eğitim nesnelere aynı sonuca sahip ise bu sonuç ile bir yaprak oluşturulur ve dördüncü adıma geçilir.
- 2) Bulgusal değerlendirme fonksiyonunun yardımı ile kökten mevcut düğüme kadar olan yol üzerinde henüz kullanılmamış tüm nitelikler arasından en iyi nitelik bulunur. Seçilen nitelik üzerinde bölünme işlemi ile içsel bir düğüm yaratılır. Sonra eğitim kümesi alt kümelere ayrılır.

- 3) Eğitim nesnelerinin her bir alt kümesi için birinci adıma gidilir.
- 4) Bir düzey yukarı çıkılarak adımlar tekrar edilir.

Karar ağaçları 2 temel işlem gerçekleştirir. İlk temel işlem veri kümesi en iyi şekilde küçük veri kümelerine bölünür. Bölme işlemi yinelemeli bir süreçtir. Tüm veri kümesini içeren kök düğüm ile yineleme başlar. Bu şekilde karar ağacı oluşturulur. İkinci temel işlem budama işlemidir. Budama işlemi karar ağacının ezberlemesini engeller. Budama işlemi ile istenmeyen alt ağaçlar veya düğümler ayıklanır [74].

3.1.5.3. Karar ağacı algoritmaları

Karar ağacı oluşturulmasında önemli olan ağaçtaki dallanmanın hangi kritere göre ya da öznitelik değerine göre olacağını belirlemesidir. Literatürde bu problemin çözümü için geliştirilmiş çeşitli yaklaşımlar vardır. Bunlardan en önemlileri bilgi kazancı ve bilgi kazanç oranı [77], Gini indeksi, Twoing kuralı [78] ve Ki-Kare olasılık tablo istatistiği [79] yaklaşımlarıdır [80].

Gini ve twoing kuralları CART (sınıflandırma ve karar ağacı oluşturma) algoritmalarında kullanılan iki kuraldır ve ikili sınıflandırmada kullanılır.

3.2. Kümeleme Yöntemleri

3.2.1. Apriori algoritmaları

Birliktelik kuralları, geçmiş veriler incelenerek bu veriler içindeki birliktelik davranışlarının tespiti ile geleceğe dönük tahminlerin yapılmasını destekleyen bir yaklaşımdır. Apriori algoritması, birliktelik kuralı çıkarım algoritmaları içerisinde en fazla kullanılan ve bilinen algoritmadır. Algoritmanın gerçekleştirilme aşamaları aşağıda belirtilmektedir:

- 1) Birliktelik çözümlemesinin yapılabilmesi için öncelikle destek ve güven ölçütlerini karşılaştırmak üzere eşik değerler belirlenir. Uygulamadan elde edilen sonuçların bu eşik değerlere eşit ya da büyük olması beklenir.
- 2) Veritabanı taranarak çözülemeye dâhil edilecek her bir ürün için tekrar sayıları, yani destek sayıları hesaplanır. Bu destek sayıları eşik destek sayısı

ile karşılaştırılır. Eşik destek sayısından küçük değerlere sahip satırlar çözümlenmeden çıkarılır ve koşula uygun kayıtlar göz önüne alınır.

- 3) Yukarıdaki adımda seçilen ürünler bu kez ikişerli gruplandırılarak, grupların tekrar sayıları, yani destek sayıları elde edilir. Bu sayılar eşik destek sayıları ile karşılaştırılır. Eşik değerden küçük değerlere sahip satırlar çözümlenmeden çıkarılır.
- 4) Bu kez üçerli, dörderli ve benzeri gruplandırmalar yapılarak bu grupların destek sayıları elde edilir ve eşik değer ile karşılaştırılır, eşik değerlere uygun olduğu sürece işlemlere devam edilir.
- 5) Ürün grubu belirlendikten sonra kural destek ölçütüne bakılarak birliktelik kuralları türetilir ve bu kuralların her birisiyle ilgili olarak güven ölçütleri hesaplanır [81].

3.2.2. K ortalamalar yöntemi

K-ortalama kümeleme, denetimsiz öğrenme algoritmalarından biridir. Daha önceden hiçbir sınıf bilgisi olmayan veri kümesi sadece verilerin özelliklerine bakarak k sayıda kümeye ayrılır. Her bir kümenin merkez değeri ile veri kümesindeki her nesnenin arasındaki farkın kareleri toplamı alınır. Bir gözlem değeri hangi merkeze yakın ise o merkez ile ilgili kümeye dahil edilir. Bu işlem kümeler içerisinde herhangi bir değişiklik olmayıncaya dek sürdürülür. Algoritmanın işleyiş adımları aşağıdaki gibidir:

- 1) Özellik uzayında k tane rastgele başlangıç ağırlık merkezleri seçilir.
- 2) Tüm veriler kendilerine en yakın olan merkeze yerleştirilir.
- 3) Etiketlenen verilerle tekrar küme merkezleri hesaplanır.
- 4) Ağırlık merkezleri değişmeyinceye kadar 2. Adımdan itibaren tekrarlar.

3.2.3. Genetik algoritmalar

Genetik algoritmalar, çok boyutlu arama uzayında bütünsel en iyi çözümü arar. Doğada gözlenen evrimsel sürece benzer şekilde çalışır. Bir çok alanda uygulama imkanı ve uygulamaları olan genetik algoritmaların işleme adımları şöyle açıklanabilir [82]:

- 1) Arama uzayındaki tüm mümkün çözümler dizi olarak kodlanır.
- 2) Genellikle rastsal bir çözüm kümesi seçilir ve başlangıç popülasyonu olarak kabul edilir.
- 3) Her bir dizi için bir uygunluk değeri hesaplanır, bulunan uygunluk değerleri dizilerin çözüm kalitesini gösterir.
- 4) Bir grup dizi belirli bir olasılık değerine göre rastsal olarak seçilip çoğalma işlemi gerçekleştirilir.
- 5) Yeni bireylerin uygunluk değerleri hesaplanarak, çaprazlama ve mutasyon işlemlerine tabi tutulur.
- 6) Önceden belirlenen kuşak sayısı boyunca yukarıdaki işlemler devam ettirilir.
- 7) İterasyon, belirlenen kuşak sayısına ulaşıncaya işlem sona erdirilir.
- 8) Amaç fonksiyonuna göre en uygun olan dizi seçilir.

Genetik algoritmalar bir çözüm uzayındaki her noktayı kromozom adı verilen ikili bit dizisi ile kodlar. Her noktanın bir uygunluk değeri vardır. Tek bir nokta yerine, genetik algoritmalar bir popülasyon olarak noktalar kümesini muhafaza eder. Her kuşakta, genetik algoritma, çaprazlama ve mutasyon gibi genetik operatörleri kullanarak yeni bir popülasyon oluşturur. Birkaç kuşak sonunda, popülasyon daha iyi uygunluk değerine sahip üyeleri içerir [83]. En uygun olan dizi seçilir.

3.2.4. Hiyerarşik kümeleme

Hiyerarşik kümeleme algoritması, kümeleme yöntemlerinden biridir. İki farklı yöntem kullanarak kümeleme işlemi gerçekleştirir. İlk yöntemi eklemeli kümelemedir. Bu yöntemde veri kümesinde her eleman bir küme olarak düşünülür. Bu kümeler arasında birbirlerine en çok benzeyen kümeler birleştirilir. Diğer yönteminde ise tek bir kümeyle başlanır ve birbirine en az benzeyen iki eleman bulunur. Tüm veri kümesi bu iki elemana olan yakınlığına göre iki kümeye ayrılır. Oluşan her alt küme için bu işlemler tekrarlanır. İstenilen küme sayısına ya da önceden belirlenmiş bir toplam benzerlik değerine ulaşıncaya kadar işlemler tekrarlanır.

3.3. Performans Değerlendirme Ölçütleri

Bir modelleme işleminin başarısı (performans), geliştirilen modelin temsil ettiği gerçek sistemin belirli bir giriş işaretine karşılık ürettiği çıkış ile modelin aynı giriş karşılık ürettiği çıkış arasındaki farkın (hata) esas alındığı çeşitli tanımlamalara göre belirlenmektedir [84].

Model performansının ölçümü için kullanılan birçok ölçüt vardır:

- 1) Toplam Karesel Hata (SSE): Toplam karesel hata, gerçek sistem çıkışı ile model çıkışı arasındaki farkın kareleri toplamıdır. Denklem (3.10)'dan faydalanılarak modelin toplam karesel hatası hesaplanabilir. Denklemde yer alan X_i değeri sistemin gerçek çıktı değerini y_i ise modelin çıktı değerini temsil etmektedir.

$$\sum_{i=1}^n (X_i - Y_i)^2 \quad (3.10)$$

- 2) Toplam Karesel Hatanın Ortalaması (MSE): Bu ölçüt de toplam karesel hatanın ortalaması alınır. Denklem (3.11)'de eşitlik verilmektedir.

$$\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)^2 \quad (3.11)$$

- 3) Ortalama Karesel Hatanın Karakökü (RMSE): Modelin performansının ölçümü için kullanılan bir diğer ölçüt denklem (3.12)'de verilmiştir. Bu ölçütte ortalama karesel hatanın karakökü alınır.

$$\sqrt{\left(\frac{1}{n}\right) \sum_{i=1}^n (X_i - Y_i)^2} \quad (3.12)$$

- 4) Mutlak Yüzde Hata (MAE): Modelin performansının ölçümü için kullanılan bir başka ölçüt denklem (3.13)'de verilmiştir.

$$\frac{1}{n} \left(\sum_{i=1}^n |X_i - Y_i| \right) * 100 \quad (3.13)$$

- 5) Ortalama Mutlak Yüzde Hata (MAPE): Ortalama mutlak yüzde hata değerinin bulunabilmesi için denklem (3.14)'de verilen eşitlikten faydalanılmaktadır.

$$\frac{1}{n} \left(\sum_{i=1}^n \left[\left| \frac{Y_i - X_i}{Y_i} \right| \right] \right) * 100 \quad (3.14)$$

Modelin değerlendirilmesi genellikle bu ölçütlerle yapılır. Amaç hata değerini en aza indirgeyen parametreleri bulmaktır. Bu parametrelerin tahmini, hataları en aza indirilmesini esas alan algoritmalarla yapılır [53]

BÖLÜM 4. REGRESYON VE KORELASYON ANALİZİ

4.1. Regresyon Analizi

Regresyon analizi, iki veya daha çok değişken arasındaki ilişkiyi belirlemek için kullanılan ve ilişki gücü hakkında bilgi veren analiz yöntemidir. Regresyon analizinde, değişkenlerden biri bağımlı değişken diğerleri bağımsız değişken olmalıdır. Bağımlı değişken, bağımsız değişkenlerden etkilenir. Bağımsız değişkenler ise bağımlı değişkeni belirli ölçülerde etkiler ve bağımlı değişkenden etkilenmezler. Tek bir bağımsız değişken ve tek bir bağımlı değişken arasındaki ilişkinin incelenmesine tek değişkenli regresyon analizi denir. Eğer bağımsız değişken sayısı birden fazla olursa çok değişkenli regresyon analizi olarak isimlendirilir. Regresyon analizinin amacı bağımlı değişken ve bir veya birden fazla bağımsız değişken arasındaki ilişkiyi açıklayacak veya tanımlayacak matematiksel bir model inşa etmektir [85]. Model, koşullar (bağımsız değişkenlerin değerleri) değiştiğinde, bağımlı değişkenin nasıl değiştiğini sunar. Şehir nüfusu ile çevre kirliliği arasındaki ilişki tek değişkenli regresyon modeline örnek olarak verilebilir.

Regresyon analizi modelin yapısına göre doğrusal regresyon analizi ve doğrusal olmayan regresyon analizi olarak ikiye ayrılır:

4.1.1. Doğrusal regresyon analizi

Doğrusal regresyon analizinde, normal dağılım gösteren ve belli aralıklarla ölçülmüş verilerin bağımlı değişkeni ile bağımsız değişken/değişkenleri arasındaki ilişki doğrusal bir fonksiyonla temsil edilir.

4.1.1.1. Tek deęişkenli doğrusal regresyon analizi

Doğrusal regresyon analizi, bir adet bağımsız deęişken ile bir adet bağımlı deęişken arasındaki ilişkiyi inceliyorsa tek deęişkenli doğrusal regresyon analizi adı verilir. Bağımlı deęişken ile bağımsız deęişken arasındaki doğrusal ilişkiyi açıklayan tek deęişkenli regresyon modeli denklem (4.1)'de verilmektedir [86]

$$y = \alpha + \beta x + \varepsilon, \quad (4.1)$$

y: Bağımlı deęişkendir.

x: Bağımsız deęişkendir.

α : Sabit deęerdir.

β : Regresyon katsayısıdır, x kendi birimi cinsinden 1 birim deęiştğinde Y'de meydana getireceęi deęişim miktarını kendi birimi cinsinden ifade eder.

ε ; gerçek hatadır. (bağımlı deęişkenin gerçek deęeri-gözlenen deęer)

Gerçekte ε tahminlemek, her bir gözlem için farklı deęerler aldığından oldukça zordur. Tahmini regresyon modeli denklem (4.2)'de gösterilmektedir.

$$y' = \alpha' + \beta' x \quad (4.2)$$

y': Tahmin edilen y deęeri (Bağımlı deęişken)

α' : Regresyon sabit deęerinin yansız tahmini

β' : Regresyon eğiminin yansız tahmini

x: Bağımsız deęişken deęeri

En küçük kareler yöntemi kullanılarak a ve b parametreleri tahmin edilir.

$$\beta' = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}, \quad (4.3)$$

$\alpha' = y' - \beta x'$ $y' = y$ deęerlerinin ortalaması $x' = x$ deęerlerinin ortalaması

4.1.1.2. Çok deęişkenli doğrusal regresyon analizi

Çoklu regresyon analizinde, bir adet bağımlı deęişken varken bağımsız deęişken sayısı birden fazladır. Basit regresyon analizindeki eşitlik (4.4) genişletilebilir.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \cdots \beta_k x_k + \varepsilon; \quad i=1,2,\dots,n \quad (4.4)$$

Örnek olarak; 2 bağımsız deęişkenli doğrusal regresyon analizini ele alırsak eđer tahmin edilen regresyon modeli eşitlik (4.5)' de gösterilmektedir.

$$y' = \beta'_0 + \beta'_1 x_1 + \beta'_2 x_2, \quad (4.5)$$

y' = Tahmin edilen y deęeri,

β'_0 = Tahmin edilen regresyon kesim noktası

β'_1, β'_2 = Tahmin edilen eęim katsayıları

4.1.2. Doğrusal olmayan regresyon analizi

Doğrusal olmayan regresyon analizinde, bir adet bağımlı deęişken ile bir veya birden fazla bağımsız deęişken arasındaki ilişki doğrusal bir modelle ifade edilemez. Bu durumda gözlemi yapılan veriler, yaklaşımlar yoluyla kurulan modelle uydurulmaya çalışılır.

4.2. Korelasyon Analizi

Bir ana küleden seçilmiş en az iki veya daha fazla örnek grup alınarak bir katsayı yardımı ile aralarındaki etkileşime bakılır. Bu katsayı korelasyon katsayısıdır. Korelasyon katsayısı r ile gösterilir.

Korelasyon analizinde değişkenler arasında bağımlı veya bağımsız değişken olma şartı aranmaz. İkili olarak incelenen değişkenlerin etkileşimleri bir katsayı yardımı ile yön ve kuvvet olarak bulunur. Korelasyon katsayısı değişkenlerin yönünü ve etkileşimlerini gösterir. Birbirleri arasındaki etkileşimin kuvveti ve yönü -1 ile $+1$ arasında değişen değerler alır. Herhangi bir etkileşimin olmadığı durumlarda ise katsayı 0 değerini alır. Kuvvetli ve artı yönünde bir etkileşim $+1$ ile ifade edilirken kuvvetli ve ters yönlü bir etkileşim -1 ile ifade edilir.

Korelasyon katsayısı eşitlik (4.6)'da verilen denklemlerle hesaplanır.

$$r = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 y_i^2}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (4.6)$$

BÖLÜM 5. ÇAPRAZ GEÇERLİLİK YÖNTEMİ

5.1. Giriş

Çapraz doğrulama bir model değerlendirme yöntemidir. Öğreticinin hiç görmediği veriler hakkında tahmin yapması istenir ve onun bu tahmini ne kadar iyi yapacağını gösterir [87].

Öğreticinin eğitiminde tüm veri kümesi kullanılmaz. Veri kümesi iki altkümeye ayrılır. Eğitim kümesi, model uydurma ve parametrelerin tahmini için kullanılır. Test kümesi modelin öngörü etkinliğinin değerlendirilmesinde kullanılır [30]. Çapraz doğrulama teorisi uygulaması için literatürde çeşitli yöntemler önerilmiştir [88], [89], [90], [91], buna rağmen önerilen yöntemlerin temeli benzerdir [56]. Temel olarak çapraz doğrulama yönteminde, bütün veri kümesi eşit sayıda örneğe sahip k adet gruba bölünür. Ayrılan k adet grubun rastgele seçilen $k-1$ adedi modelin eğitiminde, geriye kalan grup modelin tahmin gücünü ölçmede kullanılır. Bu işlem seçim kombinasyonunun tümü sağlanıncaya kadar gerçekleştirilir. Mühendislik uygulamalarında en çok kullanılan 4 farklı çapraz doğrulama yöntemi aşağıda açıklanmaktadır.

5.2. Holdout Yöntemi

Çapraz doğrulamanın en basit halidir. Belli sayıda örnek sına için kullanılır, geriye kalan örnekler eğitim amaçlı kullanılır. Genelde veri kümesinin $2/3$ 'ü öğrenme, $1/3$ 'ü sına kümesi olarak ayrılır [92]. Hesapsal açıdan çok yük getirmemesi bu yöntemin avantajıdır [80]. Şekil 5.1'de holdout yönteminin bir örneği görülmektedir.



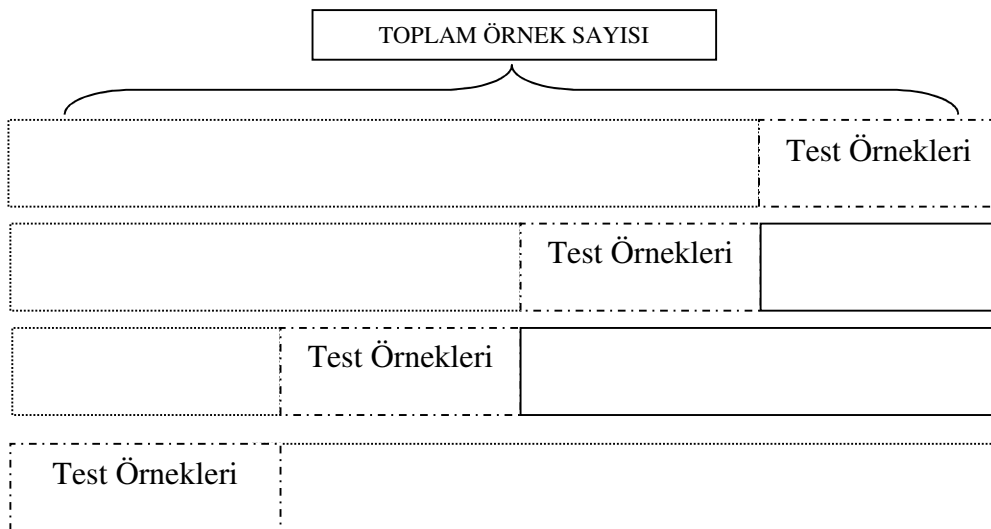
Şekil 5.1. Holdout yöntemi

5.3. K Kat Çapraz Doğrulama Yöntemi

Veri kümesi, k alt kümeye ayrılır ve holdout yöntemi k kez tekrarlanır. Her tekrarlama k adet altkümenin biri test kümesi olarak, diğer $k-1$ adet altküme eğitim kümesi olarak ayrılır. k adet çalışmanın her biri için hata değerleri hesaplanır ve sonuç olarak ortalama bir hata değeri bulunur.

K kat çapraz doğrulama yöntemi hesapsal açıdan çok yük getirir. Veri kümesini k farklı kere eğitim ve test kümesi olarak rastgele bölmesi, kümelerin boyutunun seçimi ve kaç kez tekrarlanacağı konusundaki tercihlerde bağımsız olunması bu yöntemin avantajıdır [93]. Örneklerin hepsi en sonunda hem eğitim amaçlı hem de test amaçlı kullanılır [94].

Şekil 5.2' de k kat çapraz doğrulama yönteminin bir örneği görülmektedir.



Şekil 5.2. K kat çapraz doğrulama yöntemi

5.4. Biri Hariç Çapraz Doğrulama Yöntemi

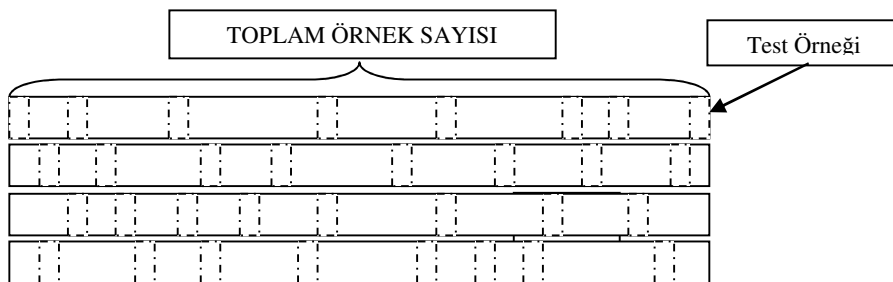
K kat çapraz doğrulama yönteminin özel halidir. N adet örneğin biri test amaçlı diğerleri eğitim amaçlı kullanılır. Bu işlem her örnek 1 kez sınama için kullanılacak şekilde tekrarlanır. Elde edilen her durum için bir hata değeri elde edilir ve sonuç olarak bu hataların ortalaması yöntemin nihai hata değerini verir. Şekil 5.3’de biri hariç çapraz doğrulama yönteminin bir örneği görülmektedir.



Şekil 5.3. Biri hariç çapraz doğrulama yöntemi

5.5. Rastgele Alt-Örnekleme Yöntemi

Bu yöntemde, belli sayıda veri eğitim için, belli sayıda veri de test için seçilir. Veriler rastgele seçilir. Bir veri aynı anda hem eğitim amaçlı hem de test amaçlı kullanılmaz. Şekil 5.4’de rastgele alt-örnekleme yönteminin bir örneği görülmektedir. Şekilde de görüldüğü gibi her deney için bir hata değeri hesaplanır.



Şekil 5.4. Rastgele alt-örnekleme yöntemi

BÖLÜM 6. UYGULAMA

Bu tez çalışmasında öğrencinin talep ettiği dersleri hangi puan aralığıyla tamamlayacağını tahmini incelenmektedir. Yapay sinir ağları, ANFIS ve karar ağaçları yöntemlerinin bu uygulamadaki yeterlilikleri ve performansları örnek bir ders üzerinden araştırılmaktadır. BSM429 (Veri madenciliği) dersi, üniversitenin en çok seçilen dersi olması ve dersin öğreticisinin, ders içeriğinin ve değerlendirme kriterlerinin uzun yıllardır değişmemesi sebepleriyle algoritmaların uygulamadaki etkinliklerinin değerlendirilmesinde ideal ders olarak görülmektedir. Bu değerlendirmede en iyi performansı gösteren modelin diğer dersler üzerinde uygulanabilirliğinin göstergesi olarak BSM421 (Bilgisayar Grafiği), BSM435 (İnternet Mühendisliği) ve BSM425 (Yapay Zeka) teknik seçmeli dersleri örnek dersler olarak seçilmiştir. Bu dersler BSM429 dersi ile benzer kriterlere sahiptirler. Bu derslerin başarı notunun tahmini BSM429 örneğinde en iyi tahmin başarısını gösteren model ile gerçekleştirilecektir.

Uygulamada Bilgisayar ve Bilişim Bilimleri Fakültesi Bilgisayar Mühendisliği Bölümü lisans seviyesindeki mezun öğrencilerin bilgileri kullanılmaktadır. Bu bilgiler Sakarya Üniversitesinin Bilgisayar ve Bilişim Bilimleri Fakültesi Dekanlığı'nın izni ile alınmıştır. Alınan veri kümesi, isim, soy isim, TC numarası gibi öğrencinin kişisel bilgilerini içermemektedir. Veri kümesi C# ile Excel'e aktarılmıştır. Uygulamanın çeşitli aşamalarında Mssql 2012, Excel 2013, SPSS 22, Visual Studio 2013, Matlab R2012b araçlarından faydalanılmaktadır.

6.1. Verinin Hazırlanması

1994 yılında eğitim öğretime başlayan Sakarya Üniversitesi Bilgisayar ve Bilişim Bilimleri Fakültesi Bilgisayar Mühendisliği Bölümü 2013 yılına kadar 1214 öğrenciyi mezun vermiştir. Bu öğrenciler eğitim ve öğretim süreleri boyunca bir

takım zorunlu ve seçmeli dersler almaktadırlar. Seçmeli dersler; teknik seçmeli, sosyal seçmeli ve üniversite ortak seçmeli dersler diye ayrılmaktadır.

Teknik seçmeli dersler, her bölüme özgü derslerdir ve bölüme özgü teknik bilgiler içerir. Bir bölüme ait teknik seçmeli ders, başka bölümde öğrenim gören bir öğrenci tarafından seçilemez. Teknik seçmeli dersler, ait olduğu bölümdeki derslerle ilgili içerikler içerir.

Sosyal seçmeli dersler, fakülteler içinde açılan ortak derslerdir. Fakülteadaki tüm öğrenciler bu dersi seçebilir. Fakülte ile ilgili olabilecek konuları içerir.

Üniversite ortak seçmeli dersler, tüm üniversitedeki öğrenciler tarafından seçilebilir. Dersler herhangi bir bölüm ile ilgili olabilir.

Uygulamanın ilk aşamasında başarı tahminini en iyi gerçekleştiren yöntem araştırılmaktadır. Bu aşamada teknik seçmeli derslerden biri olan “Veri Madenciliği” dersi kullanılmaktadır. 8 yarıyıllık öğretim süresinin 7. yarıyılında verilmekte olan veri madenciliği dersi 2008 yılından beri Sakarya Üniversitesi Bilgisayar Mühendisliği Bölümü tarafından verilmektedir. 2008 yılından itibaren bu ders 444 kadar öğrenci tarafından seçildiği bilinmektedir ve bu bölümün en çok tercih edilen derslerinden biridir.

Veri madenciliği dersini seçen 444 adet öğrenci verisi çeşitli filtreleme işlemlerine tabi tutulmuştur. Bu veri kümesinde bulunan bazı öğrencilerin girdi olarak kullanılacak olan zorunlu derslerin birkaçını veya hiç birini almadığı görülmektedir. Bazı öğrencilerin ise bu derslerden muaf veya yeterli olduğu görülmektedir. Bu özelliklere sahip kayıtlar birtakım işlemlerle düzeltilebilir. Bu problemi gidermek için öğrenci dönemsel ortalamasının, eksik parametrelerin yerine kullanılabilen varsayılabılır. Bir başka yöntem olarak ise bu ders ile ilişkisi olabilecek derslerin başarı puanlarının ortalamaları eksik parametrelerin yerine kullanılabilir. Her iki çözümün de tahmin uygulamasının performansını düşürdüğü görülmektedir. Bu nedenle veri kümesinden bu kayıtlar çıkarılmıştır. Ayrıca derslerden muaf olan veya yeterli görülen öğrencilerin not değerleri belli bir sayısal değerle ifade edilememesi

nedeniyle bu özelliklere sahip öğrenciler de veri kümesinden çıkarılmıştır. Tüm bu veri filtreleme işlemlerinden sonra 444 adet öğrenci verisi 192 adede düşmüştür.

Veri kümesinde girdi parametresi olarak kullanılacak birçok nitelik mevcuttur. Girdi parametrelerinin belirleme işlemleri bölüm 5’de daha ayrıntılı anlatılmaktadır. Girdi parametreleri arasında yer alan cinsiyet bilgisi 1 ve 2 sayısal değeriyle temsil edilir. 1 değeri bayan öğrenciyi, 2 değeri erkek öğrenciyi temsil eder. Öğrencinin dönemsal not ortalaması 0 ile 4 değerleri arasındadır. Sakarya Üniversitesinde öğrencilerin dersteki başarıları AA, BA, BB, CB, CC, DC, DD ve FF harfleriyle gösterilmektedir. AA harfi en iyi notu DD en kötü notu FF ise öğrencinin dersten başarısız olduğunu gösterir. Uygulamada AA harfi ile FF arasındaki değerler 1’den 8’e kadar olan skalada temsil edilmektedir. Tüm bu işlemler sonunda veri kümesi ilk aşama için hazır hale gelir.

Uygulamanın diğer aşaması ise elde edilen ideal yöntem ile diğer seçmeli dersler üzerinde tahminin gerçekleştirimidir. Bu işlem için seçmeli derslerden olan BSM421 (Bilgisayar Grafiği), BSM435 (İnternet Mühendisliği) ve BSM425 (Yapay Zeka) dersleri seçilmiştir. Bu derslerin tercih edilme sebebi veri sayısının çokluğu ve değerlendirmesinde aşırılıklar içermemesidir. Bu dersler için veri madenciliği dersinde kullanılan tüm filtreleme işlemleri tekrar edilmektedir. Bu derslerin verileri veri madenciliği dersinde kullanılan transkript verilerinin şemasıyla aynı özelliklere sahiptir. Bu özellikler arasından girdi parametresi olarak kullanılacak özelliklerin belirlenmesi gerekir. Bu parametreler veri madenciliği dersinde karşılaşılan eksikliklere sahip olabilirler. Bu eksikliklerin giderilmesinde daha önceden kullanılan dönüşümler kullanılmıştır. Filtreleme ve değer dönüşümleri sonucunda seçmeli derslerin verileri yöntemde kullanılabilir hale getirilmiştir.

6.2. Girdi Parametrelerinin Belirlenmesi

Filtreleme işlemlerinden sonra elimizde bulunan veri kümesi, öğrencinin cinsiyeti ve transkript bilgilerini içermektedir. Bu bilgilerin hepsinin analiz için kullanımı model eğitimini zorlaştırır. Modelin eğitimini yanıltır. Bu sebeplerden dolayı tüm bu nitelikler arasında sonucu en çok etkileyenleri bulmak gerekir. İstatistiksel

yöntemlerden korelasyon analizi ve regresyon analizi yöntemleri ile girdi parametresi olarak alınması en uygun parametreler belirlenebilir. İstatistiksel yöntemlerin pratik bir şekilde uygulanabilmesi için SPSS paket programı kullanılmıştır.

Girdi parametrelerinden tahmin edilecek olan dersi etkileyebilecek parametreler aranmaktadır. Bu süreçte incelenecek olan parametreler ilişkili olan diğer dersler, öğrencinin geçmiş dönemdeki not ortalaması, cinsiyet parametreleridir. Korelasyon analizi ile tahmin edilmek istenen değişken ile diğer değişkenlerin her birinin etkisi ikişer ikişer incelenebilir. Ancak bu yaklaşım birden fazla değişkenin birlikte etkilemesi durumlarını göz ardı edecektir. Bu nedenle regresyon analizinin hangi değişkenleri modele aldığı bilgisiyle girdi parametresi olarak kullanılacak parametreler seçilebilir.

Tahmin edilmek istenen derslerin başarı notunu etkileyen girdi parametreleri her ders için aynı olmayabilir. Bu farklılıkları bulabilmek için her ders için ilgili analizlerin ayrı ayrı yapılması gerekmektedir. Bu analizler yapılmadan önce etkisi olduğu düşünülen ders parametreleri eğitim bilgi sistemi üzerinden ders içerikleri ve ders çıktıları araştırılarak belirlenir. Belirlenen bu derslerin ilişki dereceleri ileride analizler yardımı ile araştırılacaktır. Bu uygulamada Sakarya Üniversitesinin eğitim bilgi sisteminden faydalanılmıştır. Her dersin, ders içerikleri, ön koşul bilgileri ve diğer tüm derslerle yakınlığı incelenmiştir ve ilgili olduğu düşünülen dersler analiz için ayrılmıştır.

Kullanılacak olan korelasyon ve regresyon analizleri, ilişki olduğu düşünülen dersler ve diğer parametreleri bağımsız değişkenler olarak, tahmin edilecek dersi ise bağımlı değişken olarak kullanırlar. Örnek olarak BSM429 dersi için bölümdeki diğer derslerin içerikleri incelenmiş olup BSM103, BSM101, BSM102, BSM210 ve IST108 dersleri bu derse en yakın dersler olarak belirlenmiştir. Bu dersler ile diğer parametreler bağımsız değişken olarak alınmış, analiz yapılmıştır. Veri madenciliği, bilgisayar grafiği, internet mühendisliği ve yapay zeka derslerinin her biri için yapılan bu analizlerin sonuç tabloları elde edilmiştir.

6.2.1. Korelasyon Analiziyle Girdi Parametrelerinin İncelenmesi

Korelasyon analizi ilişkili olduğu düşünülen dersleri ve diğer parametreleri bağımsız değişkenler olarak, tahmin edilecek dersi ise bağımlı değişken olarak kullanır.

6.2.1.1. BSM429 Korelasyon Analizi

BSM429 dersi için yapılan korelasyon analizi tablosu Tablo 6.1’de verilmektedir. Bağımsız değişken olarak cinsiyet, ortalamadonem6 ve ilişkili olduğu düşünülen dersler kullanılmaktadır.

Tablo 6.1. BSM429 korelasyon analizi

		Correlation								
		BSM429	Cinsiyet	OD6	BSM103	BSM101	BSM102	BSM210	Mat217	IST108
Pearson Correlation	BSM429	1,000	-,293	-,577	,089	,259	,079	,528	,399	,249
	Cinsiyet	-,293	1,000	,061	,169	,033	,155	-,286	,027	,006
	OD6	-,577	,061	1,000	-,315	-,434	-,164	-,507	-,384	-,380
	BSM103	,089	,169	-,315	1,000	,111	,256	,013	,000	,112
	BSM101	,259	,033	-,434	,111	1,000	-,024	,168	,319	,266
	BSM102	,079	,155	-,164	,256	-,024	1,000	,034	-,019	,024
	BSM210	,528	-,286	-,507	,013	,168	,034	1,000	,399	,141
	Mat217	,399	,027	-,384	,000	,319	-,019	,399	1,000	,137
	IST108	,249	,006	-,380	,112	,266	,024	,141	,137	1,000
Sig. (1- tailed)	BSM429	.	,000	,000	,110	,000	,138	,000	,000	,000
	Cinsiyet	,000	.	,201	,009	,326	,016	,000	,355	,466
	OD6	,000	,201	.	,000	,000	,012	,000	,000	,000
	BSM103	,110	,009	,000	.	,063	,000	,428	,499	,062
	BSM101	,000	,326	,000	,063	.	,370	,010	,000	,000
	BSM102	,138	,016	,012	,000	,370	.	,321	,397	,371
	BSM210	,000	,000	,000	,428	,010	,321	.	,000	,026
	Mat217	,000	,355	,000	,499	,000	,397	,000	.	,029
	IST108	,000	,466	,000	,062	,000	,371	,026	,029	.

Veri madenciliği dersinin analizinden elde edilen korelasyon tablosu incelendiğinde, BSM429 ile cinsiyet parametresi arasındaki ilişki -0.293 değerindedir. Aynı şekilde BSM429, OD6 parametresi ile -0.577 , BSM210 parametresi ile 0.528 , MAT217 parametresi ile 0.399 , BSM103 parametresi ile $.089$, BSM101 parametresi ile $.259$, BSM102 parametresi ile $.079$ ve IST108 parametresi ile $.249$ değerinde bir ilişkiye sahiptir. Buradan geçmiş dönem ortalamasının çok daha önemli olduğu açıkça görülmektedir.

Tüm bu incelemeler sonucunda veri madenciliği dersi için yapılacak tahmin çalışmasında girdi parametreleri olarak cinsiyet, OD6, BSM210 ve MAT217 parametreleri kullanılabilir.

6.2.1.2. BSM421 Korelasyon Analizi

BSM421 dersinin içeriği incelendiğinde BSM207 (Veri Yapıları), BSM102 (Algoritma ve Programlama2), BSM103 (Algoritma ve Programlama3), Mat113 (Lineer Cebir) ve BSM210 (Ayrık İşlemsel Yapılar) derslerinin içerikleri ile uyum sağladığı düşünülmektedir. Bu düşünceyi desteklemek amacı ile bir korelasyon analizi yapılmaktadır. Bu analize cinsiyet ve OrtalamaDonem6 parametreleride dahil edilmektedir.

BSM421 dersinin analizinden elde edilen korelasyon tablosu Tablo 6.2'de gösterilmektedir. Tablolar incelendiğinde BSM421 dersi ile cinsiyet parametresi arasındaki ilişki $-.053$ değerindedir. Aynı şekilde BSM421 dersi OrtalamaDonem6 parametresi ile $-.590$, BSM210 parametresi ile $.193$, MAT113 parametresi ile $.315$, BSM102 parametresi ile $.190$, BSM103 parametresi ile $.175$, BSM 207 parametresi ile $.371$ değerinde bir ilişkiye sahiptir.

Tablo 6.2. BSM421 korelasyon analizi

		BSM421	Cin İyiet	Ortalama Donem6	BSM102	BSM103	BSM207	BSM210	MAT113
Pearson Correlation	BSM421	1,000	-,053	-,590	,190	,175	,371	,193	,315
	Cin İyiet	-,053	1,000	,017	,114	,101	-,163	-,161	,097
	Ortalama Donem6	-,590	,017	1,000	-,268	-,267	-,542	-,461	-,392
	BSM102	,190	,114	-,268	1,000	,382	,170	,145	,055
	BSM103	,175	,101	-,267	,382	1,000	,184	,048	,086
	BSM207	,371	-,163	-,542	,170	,184	1,000	,261	,241
	BSM210	,193	-,161	-,461	,145	,048	,261	1,000	,298
	MAT113	,315	,097	-,392	,055	,086	,241	,298	1,000
Sig. (1-tailed)	BSM421	.	,238	,000	,005	,009	,000	,004	,000
	Cin İyiet	,238	.	,409	,063	,087	,014	,015	,096
	Ortalama Donem6	,000	,409	.	,000	,000	,000	,000	,000
	BSM102	,005	,063	,000	.	,000	,011	,025	,228
	BSM103	,009	,087	,000	,000	.	,006	,261	,125
	BSM207	,000	,014	,000	,011	,006	.	,000	,001
	BSM210	,004	,015	,000	,025	,261	,000	.	,000
	MAT113	,000	,096	,000	,228	,125	,001	,000	.
N	BSM421	183	183	183	183	183	183	183	183
	Cin İyiet	183	183	183	183	183	183	183	183
	Ortalama Donem6	183	183	183	183	183	183	183	183
	BSM102	183	183	183	183	183	183	183	183
	BSM103	183	183	183	183	183	183	183	183
	BSM207	183	183	183	183	183	183	183	183
	BSM210	183	183	183	183	183	183	183	183
	MAT113	183	183	183	183	183	183	183	183

BSM421 için yapılan analiz sonucunda BSM421 dersinin tahmini için OrtalamaDonem6, BSM207 ve Mat113 parametreleri kullanılabilir.

BSM421 dersinin karakteristikleri incelendiğinde bu dersin matematik ve fizik bilimiyle de yakından ilişkili olduğu düşünülebilir. Fakat üniversitede verilen matematik ve fizik derslerinin ders içerikleri temel matematik ve fizik bilgisinin

ötesindedir. Bu içeriklerin bilgisayar grafiği dersine dikkate değer bir katkı sağlamayacağını söyleyebiliriz. Matematik ve fizik derslerinde içinde bulunduğu korelasyon tablosu Tablo 6.3’de gösterilmektedir.

Tablo 6.3. BSM421 dersine matematik ve fizik derslerinin etkisi

		Bsm 421	od6	cin̄iyet	bsm 102	Bsm 103	bsm2 10	Mat 111	Mat 112	Mat 113	Fiz 111	Fiz 112	bsm207
pearson correlation	Bsm 421	1,000	-,554	-,036	,204	,171	,235	,125	,069	,316	,096	,169	,358
	od6	-,554	1,000	,015	-,207	-,214	-,371	-,255	-,329	-,389	-,387	-,424	-,491
	cin̄iyet	-,036	,015	1,000	,120	,101	-,170	-,040	-,040	,097	,108	-,022	-,172
	bsm102	,204	-,207	,120	1,000	,373	,140	,036	-,068	,052	,045	,012	,177
	bsm103	,171	-,214	,101	,373	1,000	,039	,028	,072	,088	,106	,035	,197
	bsm210	,235	-,371	-,170	,140	,039	1,000	,218	,286	,296	,078	,316	,294
	mat111	,125	-,255	-,040	,036	,028	,218	1,000	,341	,237	,237	,323	,212
	mat112	,069	-,329	-,040	-,068	,072	,286	,341	1,000	,137	,316	,370	,229
	mat113	,316	-,389	,097	,052	,088	,296	,237	,137	1,000	,031	,143	,222
	fiz111	,096	-,387	,108	,045	,106	,078	,237	,316	,031	1,000	,465	,123
	fiz112	,169	-,424	-,022	,012	,035	,316	,323	,370	,143	,465	1,000	,218
bsm207	,358	-,491	-,172	,177	,197	,294	,212	,229	,222	,123	,218	1,000	
̄̄̄̄ (1-tailed)	Bsm 421	.	,000	,322	,004	,013	,001	,052	,185	,000	,106	,014	,000
	od6	,000	.	,423	,003	,003	,000	,000	,000	,000	,000	,000	,000
	cin̄iyet	,322	,423	.	,060	,095	,013	,303	,305	,105	,081	,387	,013
	bsm102	,004	,003	,060	.	,000	,035	,322	,190	,250	,279	,437	,011
	bsm103	,013	,003	,095	,000	.	,306	,357	,177	,127	,085	,326	,005
	bsm210	,001	,000	,013	,035	,306	.	,002	,000	,000	,157	,000	,000
	mat111	,052	,000	,303	,322	,357	,002	.	,000	,001	,001	,000	,003
	mat112	,185	,000	,305	,190	,177	,000	,000	.	,038	,000	,000	,001
	mat113	,000	,000	,105	,250	,127	,000	,001	,038	.	,347	,032	,002
	Fiz 111	,106	,000	,081	,279	,085	,157	,001	,000	,347	.	,000	,056
	Fiz 112	,014	,000	,387	,437	,326	,000	,000	,000	,032	,000	.	,002
	bsm207	,000	,000	,013	,011	,005	,000	,003	,001	,002	,056	,002	.

Mat111, Mat112, Fiz111, Fiz112 parametrelerinin korelasyon katsayıları sırasıyla 0.125, 0.069, 0.096, 0.169'dur. Bu parametrelerin girdi parametresi olarak alınmaması daha doğru bir tercih olacaktır.

6.2.1.3. BSM425 Korelasyon Analizi

BSM425 dersinin ders içeriği incelendiğinde BSM103 (Ayrık İşlemsel Yapılar), BSM207 (Veri Yapıları), BSM306 (Yazılım Mühendisliği) ve MAT217 (Sayısal Analiz) dersleriyle uyumlu olduğu görülmektedir. Bu düşünceyi desteklemek amacı ile bir korelasyon analizi yapılmaktadır. Bu korelasyon analizine cinsiyet ve OrtalamaDonem6 parametreleri dahil edilerek bu parametrelerin BSM425 dersine olan etkileri araştırılmaktadır.

BSM425 dersinin analizinden elde edilen korelasyon tablosu Tablo 6.4 ve Tablo 6.5'de gösterilmektedir. Tablo incelendiğinde, BSM425 parametresi ile cinsiyet parametresi arasındaki ilişki -.008 değerindedir. Cinsiyet parametresinin BSM425 bağımlı değişkenine neredeyse hiç etkisi yoktur. Aynı şekilde BSM425 parametresi OrtalamaDönem6 parametresi ile -.475, BSM103 parametresi ile .162, MAT217 parametresi ile .056, BSM306 parametresi ile .284, Cinsiyet parametresi ile -.008, BSM207 parametresi ile .215 değerinde bir ilişkiye sahiptir.

Tablo 6.4. BSM425 korelasyon analizi

		BSM425	BSM103	BSM207	BSM306	MAT217
Pearson Correlation	BSM425	1,000	,162	,215	,284	,056
	BSM103	,162	1,000	,117	,269	,178
	BSM207	,215	,117	1,000	,036	,134
	BSM306	,284	,269	,036	1,000	,306
	MAT217	,056	,178	,134	,306	1,000
	Cinsiyet	-,008	,009	-,138	-,003	,086
	OrtalamaDonem6	-,475	-,509	-,368	-,404	-,492
□□ (1-tailed)	BSM425	.	,074	,027	,005	,310

Tablo 6.5. BSM425 korelasyon analizi (Devamı)

	BSM103	,074	.	,150	,008	,056
	BSM207	,027	,150	.	,374	,117
	BSM306	,005	,008	,374	.	,003
	MAT217	,310	,056	,117	,003	.
	Cinsiyet	,471	,469	,110	,488	,223
	OrtalamaDonem6	,000	,000	,000	,000	,000

BSM425 için yapılan analiz sonucunda diğer analizlerde olduğu gibi OrtalamaDonem6'nın yüksek bir korelasyon katsayısına sahip olduğu gözlenmiştir. Bu parametreyi BSM207 parametresi ve BSM306 parametresi takip etmektedir. BSM425 parametresi BSM207 parametresi ile .215, BSM306 parametresi ile .284 değerinde bir ilişkiye sahiptir. En yüksek korelasyon katsayısına sahip parametreler girdi parametreleri olarak kullanılmaktadır.

6.2.1.4. BSM435 Korelasyon Analizi

BSM435 (İnternet Mühendisliği) dersinin temeli veri iletişimi ve bilgisayar ağları derslerinin konularına dayanmaktadır. Bu nedenle internet mühendisliği dersi için bu dersler girdi parametreleri olarak alınabilir. Diğer zorunlu derslerin içerikleri de incelenmiş olup uyum sağlayabilecek herhangi bir ders içeriği bulunamamıştır. Cinsiyet, ortalamadonem6, BSM305 ve BSM304 parametreleri internet mühendisliği dersinin başarı puanının tahmininde girdi parametresi olarak alınabilir. Korelasyon analizi ile bu parametrelerin etki dereceleri incelenmiştir. Korelasyon katsayısı çok düşük olan parametre veya parametreler girdi parametresi olmaktan çıkarılmış, yüksek korelasyon katsayısına sahip parametreler girdi parametresi olarak kullanılmıştır. Cinsiyet, ortalamadonem6, BSM305 ve BSM304 parametreleri kullanılarak yapılan korelasyon analizi Tablo 6.6'da gösterilmektedir.

Tablo 6.6. BSM435 korelasyon analizi

		bsm435	Cinsiyet	bsm305	bsm304	ortalamadonem6
Pearson correlation	BSM435	1,000	,108	,394	,397	-,660
	Cinsiyet	,108	1,000	,016	-,117	,030
	BSM305	,394	,016	1,000	,309	-,493
	BSM304	,397	-,117	,309	1,000	-,490
	ortalamadonem6	-,660	,030	-,493	-,490	1,000
Sig. (1-tailed)	BSM435	.	,100	,000	,000	,000
	Cinsiyet	,100	.	,426	,084	,364
	BSM305	,000	,426	.	,000	,000
	BSM304	,000	,084	,000	.	,000
	ortalamadonem6	,000	,364	,000	,000	.

Korelasyon analizi incelendiğinde Cinsiyet, BSM305, BSM304 ve ortalamadonem6 parametreleri sırasıyla 0.108, 0.394, 0.397 ve -0.660 korelasyon katsayısına sahip olduğu görülmektedir. ortalamadonem6 parametresi en yüksek korelasyon katsayısına sahiptir. Bu parametreyi BSM304 ve BSM305 parametreleri takip eder. cinsiyet parametresi en düşük korelasyon katsayısına sahiptir. Bu nedenle cinsiyet parametresi girdi parametreleri kümesinden çıkarılmıştır.

6.2.2. Regresyon Analiziyle Girdi Parametrelerinin İncelenmesi

Cinsiyet, dönem ortalaması ve ilişkili olan dersler için yapılan analizler sonucunda regresyon analizinin korelasyon analizi ile uyumluluk göstermediği görülmüştür. Regresyon analizi modeli incelendiğinde verilmiş olan tüm bağımsız değişkenlerin modele eklenmeye çalışıldığı ve en çok etkisi olan ortalamanın diğer parametrelerin önem derecesini ciddi anlamda düşürdüğü gözlenmiştir.

Tablo 6.7’de verilen regresyon modeli katsayılar tablosunda öğrencinin dönem ortalamasının p değerinin sıfır olduğu ancak ilişkili olduğu düşünülen derslerin p değerlerinin kabul edilen yüzde 95 güven aralığının dışında kaldığı gözlenmektedir.

Tablo 6.7. Regresyon modeli katsayılar tablosu

Model	Unstandardized Coefficient		Standardized Coefficient	T	Sig.	
	B	Std. Error	Beta			
1	(Constant)	5,016	1,162		4,317	,000
	BSM103	-,088	,070	-,144	-1,259	,212
	BSM207	,038	,090	,046	,427	,671
	BSM306	,120	,080	,162	1,497	,139
	MAT217	-,176	,073	-,273	-2,390	,019
	Cinsiyet	,102	,303	,033	,335	,738
	OrtalamaDonem6	-1,123	,265	-,600	-4,242	,000

Regresyon analizi ile tahmin edilecek olan seçmeli derse etki eden diğer dersler araştırılmaktadır. Bağımlı değişken olarak tahmin edilecek ders, bağımsız değişkenler olarak ise geçmişte alınan dersler seçilmiştir. Modele etki edebilecek not ortalaması, cinsiyet gibi diğer parametreler dersler arasındaki ilişkinin araştırılmasına engel olduğu için aşağıda yorumlanacak olan modellerin bağımsız değişkenleri arasında ele alınmamıştır.

Bağımsız değişkenlerin bağımlı değişkene olan etkileri araştırılırken bağımlı değişkenlerin tümünün regresyon modeline girmesini zorlamak modelin başarısını azaltacaktır. Çünkü tüm derslerin birden tahmin edilecek olan derse etkisi olmayabilir. Bu nedenle bağımsız değişkenlerin bağımlı değişkene olan etkisini farklı bir yöntem ile araştırmak daha doğru bir yöntem olacaktır. Regresyon analizinde bağımlı değişkenlerin modele alınmaları veya modelden çıkarılmalarının yöntemleri vardır. Bu yöntemler ile sadece tolerans kriterini sağlayan parametreler modele alınarak bu parametreler ile bağımlı değişkeni en iyi açıklayabilen bağımsız değişkenlerin modele alınması hedeflenmiştir.

Bu çalışmada stepwise regresyon metodu kullanılmıştır. Stepwise metodu her bir adımda F olasılığının p değerine bakarak bağımsız değişkenleri modele ekler veya modelden çıkarır. SPSS en düşük p değerine sahip olan değişkenden başlayarak modele ekler. Modele önceden eklenen değişkenlerin p değeri düşerse modelden çıkarılır. Herhangi bir eklenecek ya da çıkarılacak bir değişken kalmayınca kadar

bu işlem sürdürülür. Bu çalışmada F olasılığının p değeri 0.05 olarak kabul edilmiştir.

Stepwise regresyon metodu ile yapılan bu araştırma ile sadece en güvenilir olan değişkenler alınarak etkisi az olan veya olmayan değişkenlerin modeli etkilememesi sağlanmıştır. En çok ilişkili derslerin belirlenebilmesi açısından stepwise metodu diğer metodlara göre daha anlamlı çıktılar üretmiştir. Enter, backward, forward gibi diğer metodlar ile oluşturulan modeller bu araştırma için anlamlı çıktılar üretmekte stepwise kadar başarılı olamamıştır.

İlişkili olduğu düşünülen dersler bağımsız değişken olarak alındığında elde edilen veriler için yapılmış olan korelasyon ve regresyon analizlerinin çıktı tabloları sırasıyla alt başlıklarda incelenmektedir.

Korelasyon ve regresyon analizlerinin çıktı tablolarının yorumlanabilmesi için bu tabloların açıklamaları aşağıda verilmiştir.

Betimsel istatistiksel: Temel betimsel istatistikler, veri setindeki değişkenler hakkında bilgi verir. Bağımlı ve bağımsız değişkenlerin her biri için ortalama, standart sapma ve gözlem sayısı değerlerini verir.

Bu çalışmada araştırılmakta olan BSM429, BSM421, BSM435 ve BSM425 derslerinden elde edilen betimsel istatistik tabloları EK A'da verilmektedir.

Korelasyon katsayısı: Korelasyon analizi iki farklı değişken arasındaki ilişkinin yönü ve şiddeti hakkında bilgi verir. Korelasyon neden sonuç ilişkisinin göstergesi değildir. Korelasyon matrisinde gösterilen değerler r'yi ifade eder. r, -1 ile 1 arasında değer alır. İlişkinin yönünü r'nin işareti, derecesini ise katsayının büyüklüğü belirler. Bir değişken artarken diğeri azalıyor ise eksi değerle, her iki değişken birlikte azalış veya artış gösteriyorsa artı değerle gösterilir.

Model özeti: Model özeti tablosu çoklu regresyon uygulamalarında en önemli çıktılarından biridir. Model özeti tablosunda R değeri tahmin modelinin sonucu ne

kadar iyi temsil ettiğinin ölçüsüdür. Fakat daha doğru ölçüm yapabilmek için R^2 'nin alınması gerekir. R^2 , belirleme katsayısıdır. Bağımlı değişkenin ne kadarının regresyon doğrusu tarafından açıklanabildiğini gösterir.

ANOVA: ANOVA bağımsız değişkenlerin kendi aralarında nasıl etkileşime girdiklerini ve bu etkileşimlerin bağımlı değişken üzerindeki etkilerini analiz etmek için kullanılır. Bağımlı ve bağımsız değişkenlerin $p < 0,05$ düzeyinde ilişkinin istatistiksel olarak anlamlı olup olmadığı Sig. (Anlamlılık) sütunundaki değerden okunabilir.

Katsayılar (Coefficient): Regresyon denklemi için kullanılan regresyon katsayılarını ve bunların anlamlılık düzeylerini gösterir. B değeri çoklu regresyon formülünün katsayılarını verir. Beta sütunu ise bağımsız değişkenlerin bir birim artması durumunda modellenin ne kadar etkileneceği bilgisini verir. Sig. bağımsız değişkenlerin her birinin modelde bağımlı değişkene olan ilişkisinin istatistiksel olarak anlamlı olup olmadığını gösterir. Sig. değerinin 0.05'ten küçük olması anlamlı olduğunu ifade eder. Stepwise metodu bu değere bakarak değişkenlerin modele alınıp alınmayacağına karar verir.

6.2.2.1. BSM429 Regresyon Analizi

BSM429 dersi için yapılan regresyon analizinden elde edilen tablolar aşağıda verilmektedir.

Korelasyon katsayısı:

Tablo 6.8'deki korelasyon matrisinde BSM429'u etkileyen dersler korelasyon katsayısı değerlerine göre sırasıyla verilmektedir. BSM210 0,528, MAT217 0,399, BSM101 0,259, IST108 0,249, BSM102 0,079 ve BSM103 0,089 değerinde bir korelasyon katsayısına sahip olduğu görülmektedir. Bu çıktıya göre tahmini yapacak olan metoda, ilişkili ders girdileri olarak BSM210 ve MAT217 verilebilir.

Tablo 6.8. BSM429 korelasyon katsayısı

		BSM429	BSM103	BSM101	BSM102	BSM210	MAT217	IST108
Pearson Correlation	BSM429	1,000	,089	,259	,079	,528	,399	,249
	BSM103	,089	1,000	,111	,256	,013	,000	,112
	BSM101	,259	,111	1,000	-,024	,168	,319	,266
	BSM102	,079	,256	-,024	1,000	,034	-,019	,024
	BSM210	,528	,013	,168	,034	1,000	,399	,141
	MAT217	,399	,000	,319	-,019	,399	1,000	,137
	IST108	,249	,112	,266	,024	,141	,137	1,000
Sig. (1-tailed)	BSM429	.	,110	,000	,138	,000	,000	,000
	BSM103	,110	.	,063	,000	,428	,499	,062
	BSM101	,000	,063	.	,370	,010	,000	,000
	BSM102	,138	,000	,370	.	,321	,397	,371
	BSM210	,000	,428	,010	,321	.	,000	,026
	MAT217	,000	,499	,000	,397	,000	.	,029
	IST108	,000	,062	,000	,371	,026	,029	.
N	BSM429	192	192	192	192	192	192	192
	BSM103	192	192	192	192	192	192	192
	BSM101	192	192	192	192	192	192	192
	BSM102	192	192	192	192	192	192	192
	BSM210	192	192	192	192	192	192	192
	MAT217	192	192	192	192	192	192	192
	IST108	192	192	192	192	192	192	192

Model özeti: BSM429 dersinin model özeti tablosundaki (tablo 6.9) R^2 değerinin 0,346 olduğu görülmektedir. Bu değere bakılarak, regresyon modelinin bağımlı değişkeni çok iyi açıkladığı söylenemez. Bu nedenle yapılan regresyon analizinin doğrudan tahmin için kullanılması hatalı olur. Uygulamada regresyon analizi tahmin yöntemi olarak kullanılmayacaktır, sadece karar vermede yardımcı bir araç olacaktır.

Tablo 6.9. BSM429 model özeti

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	,528 ^a	,279	,275	1,505	,279	73,456	1	190	,000
2	,566 ^b	,321	,314	1,464	,042	11,697	1	189	,001

ANOVA: BSM429 dersinin analizi sonucu elde edilen ANOVA tablosu (Tablo 6.10) incelendiğinde Sig. sütunu .000 değerinde olduğu görülmektedir. Bağımlı değişkenin bağımsız değişkenlerle olan ilişkisi $p < 0,05$ düzeyinde istatistiksel olarak anlamlı olduğu söylenebilir.

Tablo 6.10. BSM429 ANOVA tablosu

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	166,278	1	166,278	73,456	,000 ^b
	Residual	430,092	190	2,264		
	Total	596,370	191			
2	Regression	191,345	2	95,672	44,644	,000 ^c
	Residual	405,025	189	2,143		
	Total	596,370	191			
3	Regression	206,385	3	68,795	33,164	,000 ^d
	Residual	389,985	188	2,074		
	Total	596,370	191			

a. Dependent Variable: BSM429

b. Predictors: (Constant), BSM210

c. Predictors: (Constant), BSM210, MAT217

d. Predictors: (Constant), BSM210, MAT217, IST108

Katsayılar (Coefficient): BSM429 dersi için yapılan analiz sonucunda elde edilen regresyon modelleri Tablo 6.11’de gösterilmiştir. 3 adet model oluşturulmuş ve her model için bağımsız değişkenlerin değerleri gösterilmiştir. Regresyon doğrusunun katsayı değerleri B sütunu altında verilmektedir. Model 3’e göre BSM210, MAT217 ve IST108 parametrelerinin regresyon modelinde etkili olduğu görülür. Bu sonuçlara göre; çoklu doğrusal regresyon analizi Eşitlik 6.1’de gösterilmektedir. Sig sütunu incelendiğinde ise BSM210, MAT217 ve IST108 parametrelerinin anlamlılık gösterdiği görülür. Diğer niteliklerin bağımlı değişken olan BSM 429 kodlu veri

madenciligi dersiyle olan ilişkisinin modelde anlamlılık göstermediği görülür. Tablo 6.12 ve Tablo 6.13’de modele dahil edilmeyen değişkenler olarak ayrılmıştır.

Eşitlik 6.1’de modele giren bağımsız parametreler BSM210, MAT217 ve IST108 değişkenleridir. Ancak korelasyonda IST108 dersinin etkisi 0.249 değerindedir ve BSM101 dersinden daha düşük bir korelasyon katsayısına sahiptir. Bu nedenle regresyon modeline girmiş olsa da girdi parametresi olarak alınmayacaktır.

Tablo 6.11. BSM429 katsayılar tablosu

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	1,850	,267		6,922	,000	1,323	2,377
	BSM210	,582	,068	,528	8,571	,000	,448	,716
2	(Constant)	1,454	,285		5,109	,000	,893	2,015
	BSM210	,484	,072	,439	6,712	,000	,342	,626
	MAT217	,213	,062	,224	3,420	,001	,090	,336
3	(Constant)	1,052	,317		3,313	,001	,425	1,678
	BSM210	,466	,071	,422	6,535	,000	,325	,606
	MAT217	,198	,061	,208	3,223	,001	,077	,319
	IST108	,160	,059	,161	2,693	,008	,043	,277

a. Dependent Variable: BSM429

Tablo 6.12. BSM429 harici değişkenler tablosu

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	BSM103	,082 ^b	1,333	,184	,097	1,000
	BSM101	,175 ^b	2,859	,005	,204	,972
	BSM102	,061 ^b	,993	,322	,072	,999
	MAT217	,224 ^b	3,420	,001	,241	,841
	IST108	,178 ^b	2,919	,004	,208	,980

Tablo 6.13. BSM429 harici deęişkenler tablosu (Devamı)

2	BSM103	,083 ^c	1,389	,166	,101	1,000
	BSM101	,127 ^c	2,028	,044	,146	,896
	BSM102	,069 ^c	1,143	,255	,083	,998
	IST108	,161 ^c	2,693	,008	,193	,972
	BSM103	,066 ^d	1,115	,266	,081	,987
3	BSM101	,093 ^d	1,460	,146	,106	,847
	BSM102	,065 ^d	1,101	,272	,080	,997
	BSM103	,083 ^c	1,389	,166	,101	1,000

$$1.052 + 0.466 \times BSM210 + 0.198 \times MAT217 + 0.160 \times IST108 \quad (6.1)$$

6.2.2.2. BSM421 Regresyon Analizi

BSM421 dersi için yapılan regresyon analizinden elde edilen tablolar ařaęıda verilmektedir.

Korelasyon analizi: Ařaęıdaki korelasyon matrisinde BSM421'i etkileyen dersler korelasyon katsayısı deęerlerine gre sırasıyla verilmektedir. BSM207 0,371, MAT113 0,315, BSM210 0,193, BSM102 0,190 ve BSM103 0,175 deęerinde bir korelasyon katsayısına sahip olduęu grlmektedir. Bu çıktıya gre tahmini yapacak olan metoda, iliřkili ders girdileri olarak BSM207 ve MAT113 verilebilir.

Tablo 6.14. BSM421 korelasyon katsayısı

		BSM421	BSM102	BSM103	BSM207	BSM210	MAT113
Pearson Correlation	BSM421	1,000	,190	,175	,371	,193	,315
	BSM102	,190	1,000	,382	,170	,145	,055
	BSM103	,175	,382	1,000	,184	,048	,086
	BSM207	,371	,170	,184	1,000	,261	,241
	BSM210	,193	,145	,048	,261	1,000	,298
	MAT113	,315	,055	,086	,241	,298	1,000

Tablo 6.15. BSM421 korelasyon katsayısı (Devamı)

Si□ (1-tailed)	BSM421	.	,005	,009	,000	,004	,000
	BSM102	,005	.	,000	,011	,025	,228
BSM103	,009	,000	.	,006	,261	,125	
BSM207	,000	,011	,006	.	,000	,001	
BSM210	,004	,025	,261	,000	.	,000	
MAT113	,000	,228	,125	,001	,000	.	

Model özeti: BSM421 dersinin model özeti tablosundaki R^2 değerinin 0,192 olduğu görülmektedir. Bu değere bakılarak, regresyon modelinin bağımlı değişkeni çok iyi açıkladığı söylenemez. Bu nedenle yapılan regresyon analizinin doğrudan tahmin için kullanılması hatalı olur. Uygulamada regresyon analizi tahmin yöntemi olarak kullanılmayacaktır, sadece karar vermede yardımcı bir araç olacaktır.

Tablo 6.16. BSM421 model özeti

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	,371a	,137	,133	1,3808	,137	28,851	1	181	,000
2	,438b	,192	,183	1,3404	,054	12,077	1	180	,001

a. Predictors: (Constant), BSM207

b. Predictors: (Constant), BSM207, MAT113

ANOVA: BSM421 dersinin analizi sonucu elde edilen ANOVA tablosu incelendiğinde Sig. sütunu 0.000 değerinde olduğu görülmektedir. Bağımlı değişkenin bağımsız değişkenlerle olan ilişkisi $p < 0,05$ düzeyinde istatistiksel olarak anlamlı olduğu söylenebilir.

Tablo 6.17. BSM421 ANOVA tablosu

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	55,009	1	55,009	28,851	,000b
	Residual	345,100	181	1,907		
	Total	400,109	182			
2	Regression	76,708	2	38,354	21,347	,000c
	Residual	323,402	180	1,797		
	Total	400,109	182			

a. Dependent Variable: BSM421

b. Predictors: (Constant), BSM207

c. Predictors: (Constant), BSM207, MAT113

Katsayılar (Coefficient): BSM421 dersi için yapılan analiz sonucunda elde edilen regresyon modelleri Tablo 6.18’de gösterilmiştir. 2 adet model oluşturulmuş ve her model için bağımsız değişkenlerin değerleri gösterilmiştir. Regresyon doğrusunun katsayı değerleri B sütunu altında verilmektedir. Model 2’ye göre BSM207 ve MAT113 parametrelerinin regresyon modelinde etkili olduğu görülür. Sig sütunu incelendiğinde ise BSM207 ve MAT113 parametrelerinin anlamlılık gösterdiği görülür. Diğer niteliklerin bağımlı değişken olan BSM421 parametresiyle olan ilişkisinin modelde anlamlılık göstermediği görülür ve Tablo 6.19’da modele dahil edilmeyen değişkenler olarak ayrılmıştır. Eşitlik 6.2’de modele giren bağımsız parametreler BSM207 ve MAT113’dür.

Tablo 6.18. BSM421 katsayılar tablosu

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	1,555	,266		5,848	,000	1,030	2,080
	BSM207	,352	,066	,371	5,371	,000	,223	,482
2	(Constant)	,981	,307		3,198	,002	,376	1,585
	BSM207	,297	,066	,313	4,532	,000	,168	,427
	MAT113	,206	,059	,240	3,475	,001	,089	,322

Tablo 6.19. BSM421 harici deęişkenler tablosu

Model	Beta In	T	Sig.	Partial Correlation	Collinearity Statistics			
					Tolerance	VIF	Minimum Tolerance	
1	BSM102	,131b	1,880	,062	,139	,971	1,030	,971
	BSM103	,110b	1,572	,118	,116	,966	1,035	,966
	BSM210	,103b	1,445	,150	,107	,932	1,073	,932
	MAT113	,240b	3,475	,001	,251	,942	1,062	,942
2	BSM102	,127c	1,884	,061	,139	,971	1,030	,917
	BSM103	,100c	1,469	,143	,109	,964	1,037	,915
	BSM210	,045c	,633	,528	,047	,873	1,145	,873

a. Dependent Variable: BSM421

b. Predictors in the Model: (Constant), BSM207

c. Predictors in the Model: (Constant), BSM207, MAT113

$$0.981 + 0.297 \times BSM207 + 0.206 \times MAT113 \quad (6.2)$$

6.2.2.3. BSM425 Regresyon Analizi

BSM425 dersi için yapılan regresyon analizinden elde edilen tablolar ařaęıda verilmektedir.

Korelasyon analizi: Ařaęıdaki korelasyon matrisinde BSM425'i etkileyen dersler korelasyon katsayısı deęerlerine gre sırasıyla verilmektedir. BSM207 0,215, BSM306 .284, MAT217 .056 ve BSM103 0,162 deęerinde bir korelasyon katsayısına sahip olduęu grlmektedir. Bu çıktıya gre tahmini yapacak olan metoda, iliřkili ders girdileri olarak BSM207 ve BSM306 verilebilir.

Tablo 6.20. BSM425 korelasyon katsayısı

		BSM425	BSM103	BSM207	BSM306	MAT217
Pearson Correlation	BSM425	1,000	,162	,215	,284	,056
	BSM103	,162	1,000	,117	,269	,178
	BSM207	,215	,117	1,000	,036	,134
	BSM306	,284	,269	,036	1,000	,306
	MAT217	,056	,178	,134	,306	1,000

Tablo 6.21. BSM425 korelasyon katsayısı (Devamı)

Si \square (1-tailed)	BSM425	.	,074	,027	,005	,310
	BSM103	,074	.	,150	,008	,056
	BSM207	,027	,150	.	,374	,117
	BSM306	,005	,008	,374	.	,003
	MAT217	,310	,056	,117	,003	.
N	BSM425	81	81	81	81	81
	BSM103	81	81	81	81	81
	BSM207	81	81	81	81	81
	BSM306	81	81	81	81	81
	MAT217	81	81	81	81	81

Model Özeti: BSM425 dersinin model özeti tablosundaki R^2 değerinin 0,081 olduğu görülmektedir.

Tablo 6.22. BSM425 model özeti

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	,284a	,081	,069	1,1883	,081	6,951	1	79	,010

a. Predictors: (Constant), BSM306

ANOVA: BSM425 dersinin analizi sonucu elde edilen ANOVA tablosu incelendiğinde Sig. sütunu 0.010^b değerinde olduğu görülmektedir. Bağımlı değişkenin bağımsız değişkenlerle olan ilişkisi $p < 0,05$ düzeyinde istatistiksel olarak anlamlı olduğu söylenebilir.

Tablo 6.23. BSM425 ANOVA tablosu

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	9,814	1	9,814	6,951	,010b
	Residual	111,544	79	1,412		
	Total	121,358	80			

a. Dependent Variable: BSM425

b. Predictors: (Constant), BSM306

Katsayılar (Coefficient): BSM425 dersi için yapılan analiz sonucunda elde edilen regresyon modelleri Tablo 6.24'de gösterilmiştir. 1 adet model oluşturulmuştur.

Regresyon doğrusunun katsayı değerleri B sütunu altında verilmektedir. Modele göre sadece BSM306 parametresinin regresyon modelinde etkili olduğu görülür. Sig sütunu incelendiğinde ise sadece BSM306 parametresinin anlamlılık gösterdiği görülür. Diğer niteliklerin bağımlı değişken olan BSM425 parametresiyle olan ilişkisinin modelde anlamlılık göstermediği görülür ve Tablo 6.25’de modele dahil edilmeyen değişkenler olarak ayrılmıştır. Eşitlik 6.3’de modele giren bağımsız parametreler BSM306’dır. Ancak BSM207 dersi 0,215 korelasyon katsayısı ile BSM306 dersine çok yakındır ve regresyon analizinde 0,057 sig. değeri 0,007 fark ile modele girememiştir. Bu sebepten dolayı BSM207 dersi uygulamada girdi parametresi olarak alınacaktır.

Tablo 6.24. BSM425 katsayılar tablosu

Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.	95,0% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
1 (Constant)	1,538	,351		4,381	,000	,839	2,236
BSM306	,210	,080	,284	2,636	,010	,052	,369

Tablo 6.25. BSM425 harici değişkenler tablosu

Model	Beta In	T	Sig.	Partial Correlation	Collinearity Statistics		
					Tolerance	VIF	Minimum Tolerance
1 BSM103	,093b	,824	,412	,093	,928	1,078	,928
BSM207	,205b	1,930	,057	,213	,999	1,001	,999
MAT217	-,034b	-,300	,765	-,034	,907	1,103	,907

a. Dependent Variable: BSM425

b. Predictors in the Model: (Constant), BSM306

$$1.538 + 0.210 \times BSM306 \quad (6.3)$$

6.2.2.4. BSM435 Regresyon Analizi

BSM435 dersi için yapılan regresyon analizinden elde edilen tablolar aşağıda verilmektedir.

Korelasyon katsayısı: Aşağıdaki korelasyon matrisinde BSM435'i etkileyen dersler korelasyon katsayısı değerlerine göre sırasıyla verilmektedir. BSM304 0,421, ve BSM305 0,414 değerinde bir korelasyon katsayısına sahip olduğu görülmektedir. Bu çıktıya göre tahmini yapacak olan metoda, ilişkili ders girdileri olarak BSM304 ve BSM305 verilebilir.

Tablo 6.26. BSM435 korelasyon katsayısı

		BSM435	BSM304	BSM305
Pearson Correlation	BSM435	1,000	,421	,414
	BSM304	,421	1,000	,304
	BSM305	,414	,304	1,000
Sig. (1-tailed)	BSM435	.	,000	,000
	BSM304	,000	.	,000
	BSM305	,000	,000	.
N	BSM435	137	137	137
	BSM304	137	137	137
	BSM305	137	137	137

Model Özeti: BSM435 dersinin model özeti tablosundaki R^2 değerinin 0,267 olduğu görülmektedir.

Tablo 6.27. BSM435 model özeti

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	,421a	,177	,171	1,63411	,177	29,064	1a	135	,000
2	,517b	,267	,256	1,54792	,090	16,453	1b	134	,000

ANOVA: BSM435 dersinin analizi sonucu elde edilen ANOVA tablosu incelendiğinde Sig. sütunu 0.010^b değerinde olduğu görülmektedir. Bağımlı değişkenin bağımsız değişkenlerle olan ilişkisi $p < 0,05$ düzeyinde istatistiksel olarak anlamlı olduğu söylenebilir.

Tablo 6.28. BSM435 ANOVA tablosu

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	77,609	1	77,609	29,064	,000b
	Residual	360,493	135	2,670		
	Total	438,102	136			
2	Regression	117,032	2	58,516	24,422	,000c
	Residual	321,070	134	2,396		
	Total	438,102	136			

a. Dependent Variable: BSM435

b. Predictors: (Constant), BSM304

c. Predictors: (Constant), BSM304, BSM305

Katsayılar (Coefficient): BSM435 dersi için yapılan analiz sonucunda elde edilen regresyon modelleri Tablo 6.29'da gösterilmiştir. 2 adet model oluşturulmuştur. Regresyon doğrusunun katsayı değerleri B sütunu altında verilmektedir. Modele göre BSM304 ve BSM305 parametrelerinin regresyon modelinde etkili olduğu görülür. Eşitlik 6.4'de modele giren bağımsız parametreler BSM304 ve BSM305'dir.

Tablo 6.29. BSM435 katsayılar tablosu

Model		Unstandardized Coefficients		Standardized	t	Sig.
		B	Std. Error	Coefficients		
1	(Constant)	2,158	,462		4,670	,000
	BSM304	,506	,094	,421	5,391	,000
2	(Constant)	1,030	,519		1,987	,049
	BSM304	,391	,093	,325	4,186	,000
	BSM305	,355	,088	,315	4,056	,000

Tablo 6.30. BSM435 katsayılar tablosu(devamı)

Model		Correlations			Collinearity Statistics	
		Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)					
	BSM304	,421	,421	,421	1,000	1,000
2	(Constant)					
	BSM304	,421	,340	,310	,907	1,102
	BSM305	,414	,331	,300	,907	1,102

$$1.030 + 0.391 \times \text{BSM304} + 0.355 \times \text{BSM305}$$

.(6.4)

6.3. Kullanılan Yöntemler

Bu çalışmada, öğrencinin seçmeli dersi hangi puan aralığıyla tamamlayacağını tahmini incelenmektedir. Tahmin bir sınıflandırma çalışmasıdır. Sınıflandırma yöntemlerinden olan ve sınıflandırmada başarılı sonuçlar veren yapay sinir ağları, adaptif ağ tabanlı bulanık çıkarım sistemleri ve karar ağaçları yöntemleri kullanılmaktadır. Yöntemlerin performansları örnek bir ders üzerinden incelenir. Uygulamada en başarılı tahminlerde bulunan yöntem ile diğer seçmeli derslerin tahmini gerçekleştirilir. Yöntemlerin gerçekleştirilmesinde Matlab R2012b aracından faydalanılmaktadır. Matlab aracının esnekliği ve model oluşturulmasında farklı birçok seçenek sunması uygulamada kolaylık sağlamaktadır.

6.3.1. Yapay Sinir Ağları (YSA)

Öncelikle 192 adet öğrenci verisi ağın eğitiminde ve testinde kullanılmak üzere 2 kümeye ayrılır. Uygulamada k kat çapraz geçerlilik yöntemi kullanıldığından tüm veri 6 parçaya bölünür ve rastgele 5 parça modelin eğitiminde, geriye kalan 1 parça modelin testinde kullanılır. Her parça 32 adet veri içerir. Her parçaya fold denir. Tüm veri kümesinin 0.83'ü eğitim amaçlı 0.17'si test amaçlı kullanılır. Eğitim amaçlı ayrılan veri kümesi öncelikle Matlab aracının komut penceresinde normalize edilir. Matlab, farklı yapay sinir ağları oluşumuna izin veren farklı seçenekler sunar. Uygulamada ileri beslemeli geri yayılım ağı seçilmiştir. İleri beslemeli geri yayılım algoritması uygulamadaki görüş açısı ve basitliği bakımından en popüler algoritmalarından biridir [95], [96]. Bu ağ, birçok farklı özellik seçeneklerine sahiptir. Uygulamada bu ağın farklı seçenekleri denenmiş ve en iyi performansı veren özelliklere sahip ağ kullanılmıştır. Ağ iki katmanlı bir yapıdadır. Bu ağın eğitiminde kullanılan fonksiyonlar ve ağ bilgileri Tablo 6.31.'de gösterilmektedir.

Tablo 6.31. İleri beslemeli geri yayılım ağının özellikleri

Trainin□ Function	Adaption learnin□function	performance function	tran□fer function	number of layer
TRAINLM	LEARNGDM	MSE	TANSIG	2

Oluşturulan ileri beslemeli geri yayılım ağı normalize edilmiş eğitim verileriyle eğitilir.

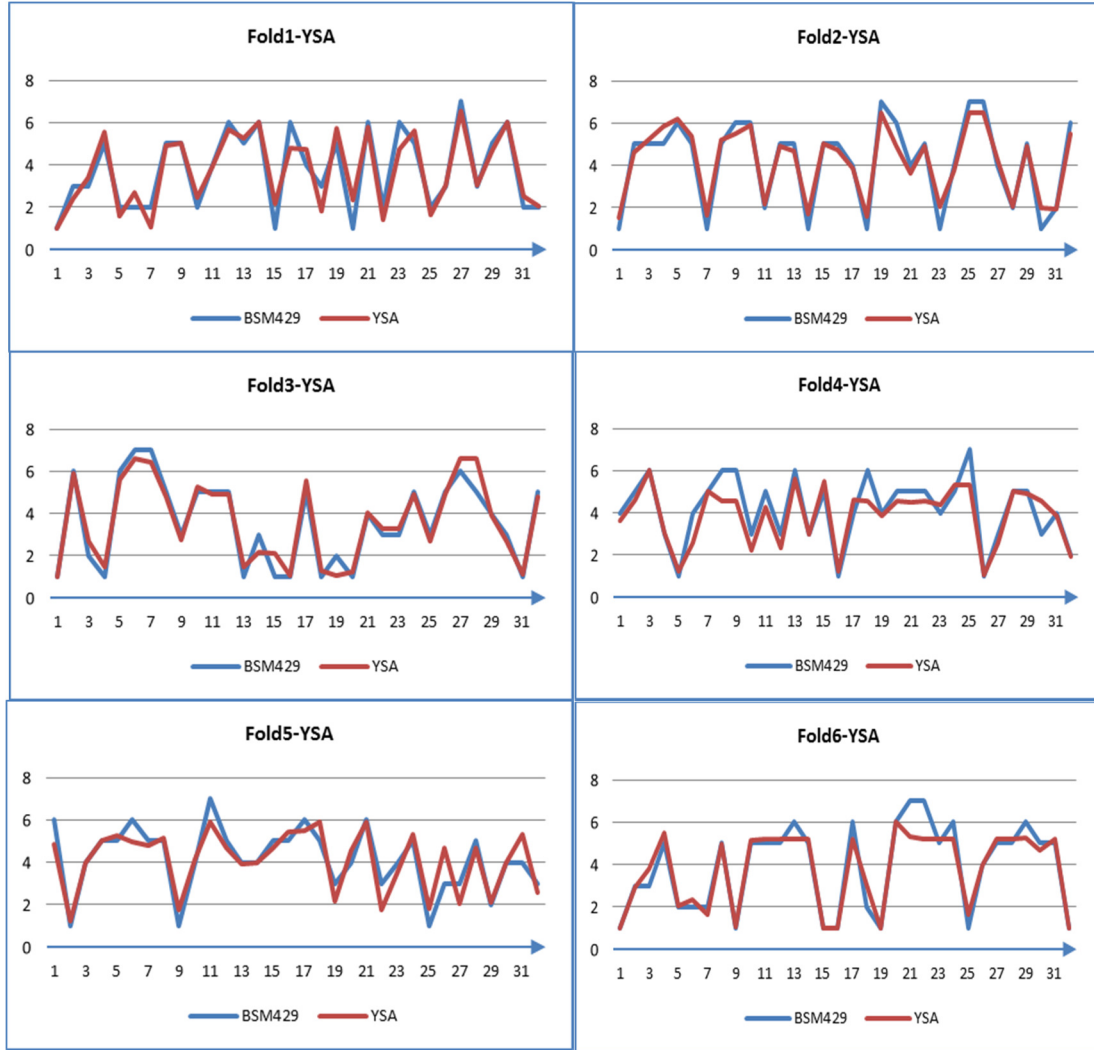
Tablo 6.32’de eğitimde ve modelin testinde kullanılan foldlar, eğitim performansları, tahmin değerlerinin gerçek değerlerden sapma miktarları ve bu sapma miktarlarının frekansları her satır bir olası durumu temsil edecek şekilde gösterilmektedir.

Tablo 6.32. İleri beslemeli geri yayılım algoritmasının her olası durum için performansı

FOLDS USED IN TRAINING PHASES	FOLDS USED IN TESTING PHASES	TRAINING PERFORMANCE	Sapma değerleri ve bu değerlerin frekansları
1,2,3,4,5	6	0.0403627	+/-1=>14 +/-2=>0 Doğru Tahmin Sayısı: 18
1,2,3,4,6	5	0.0382234	+/-1=>13 +/-2=>0 Doğru Tahmin Sayısı: 19
1,2,3,5,6	4	0.0443766	+/-1=>7 +/-2=>1 Doğru Tahmin Sayısı: 24
1,2,4,5,6	3	0.0423543	+/-1=>11 +/-2=>1 Doğru Tahmin Sayısı: 20
1,3,4,5,6	2	0.0437529	+/-1=>12 +/-2=>1 Doğru Tahmin Sayısı: 19
2,3,4,5,6	1	0.0425249	+/-1=>7 +/-2=>2 Doğru Tahmin Sayısı: 23

Tablo 6.34’de görüldüğü gibi 192 adet veriden ileri beslemeli geri yayımlı ağ, toplamda 123 adedini doğru tahminde bulunmuştur. 64 adedini bir birim sapmayla, 5 adedini de iki birim sapma ile tahmin etmiştir.

Şekil 6.1’de her fold için gerçek değer ile yapay sinir ağı modelinin tahmin ettiği değerler aynı grafik üzerinde gösterilmektedir.



Şekil 6.1. YSA modelinin çıktı değerleri ile gerçek çıktı değerleri

Yapay sinir ağları, en fazla 2 sapma ile tahmini gerçekleyebilmiştir. Farklı istatistiksel ölçütlerle ağın performans değerlendirilmesi yapılmıştır. Yapılan değerlendirme sonuçları her bir fold için Tablo 6.33'de gösterilmektedir. Ayrıca bu tabloda YSA modelinin uygulamada ki ortalama performans değeri bilgisi de yer almaktadır.

Tablo 6.33. Ağın performansının farklı istatistiksel ölçütlerle değerlendirilmesi

Değerlendirme Ölçütü	Fold1	Fold2	Fold3	Fold4	Fold5	Fold6	Ortalama
RMSE	0,66143	0,63737	0,5863	0,6846	0,7071	0,6602	0.6561
MAPE	18,2812	22,2767	10,133	12,4032	15,6547	9,5982	14,724
MAD	0,4375	0,40625	0,2812	0,4375	0,4375	0,3437	0.3906

6.3.2. Adaptif Ağ Tabanlı Bulanık Çıkarım Sistemi (ANFIS)

ANFIS modelinin kurulumu, ileri beslemeli geri yayılım ağı gibi farklı özellik seçeneklerine sahiptir. Uygulamada farklı birçok model oluşturulmuş, en iyi performans veren model uygulamada kullanılmıştır.

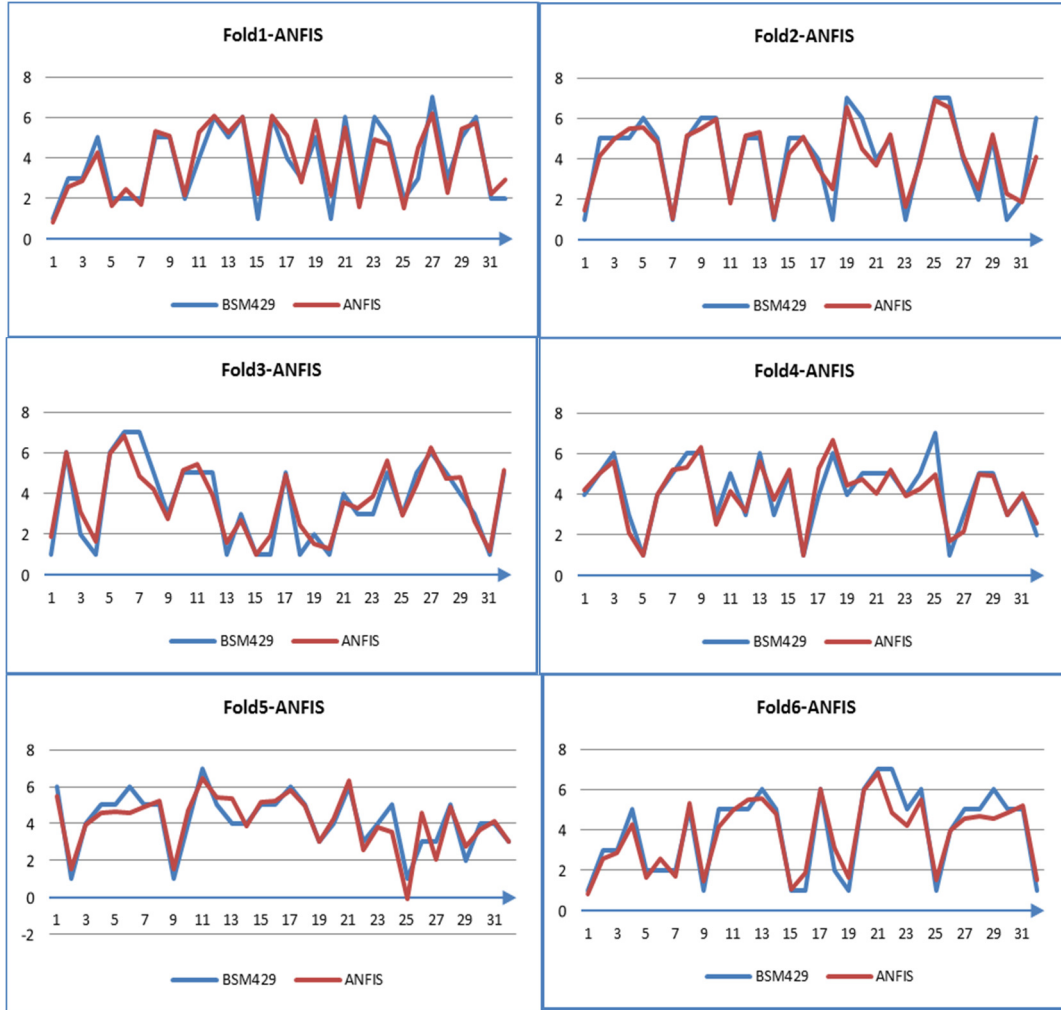
ANFIS modelinin oluşturulması için öncelikle her bir girdi için üyelik fonksiyonları atanmıştır. Girdi kümesinin karakteristiğine bağlı olarak farklı üyelik fonksiyonları seçilebilir. Atanan üyelik fonksiyonlarıyla girişlerin her biri bulanıklaştırılır. Bulanık girişlerin her biri ANFIS modelinin katmanlarında işlem görür. Bu işlemler kurallar tarafından gerçekleştirilir. İşlemlerin sonucunda her kural bir çıktı oluşturur. Bu çıktılar çıktı üyelik fonksiyonlarının girdilerini oluşturur. Çıktı üyelik fonksiyonları 2 farklı tiptedir:

- First-order çıktı üyelik fonksiyonu: $ax + by + c$ tipinde doğrusal bir fonksiyondur.
- Zero-order çıktı üyelik fonksiyonu: Bu üyelik fonksiyonunda a ve b değişkenleri 0 değerindedir. Bu nedenle c sabitine eşit olup, bu fonksiyona sabit çıktı üyelik fonksiyonu adı verilir.

ANFIS modelinin cinsiyet girdisi için 1, 6. yarıyıl sonundaki genel not ortalaması için 2, BSM 210 ve Mat 217 için 3 adet üyelik fonksiyonu atanmıştır. Girdi üyelik fonksiyonu olarak Gbellmf üyelik fonksiyonu, çıktı üyelik fonksiyonu olarak da lineer üyelik fonksiyonu seçilmiştir. Öğrenme kuralı olarak ise, gradyan düşümü ve en küçük kareler yöntemlerinin kombinasyonu olan hybrid öğrenme kuralı seçilmiştir. Epoch sayısı 50 olarak alınmıştır.

Uygulamada kullanılacak toplam 192 adet öğrenci verisi ağın eğitiminde ve testinde kullanılmak üzere 2 kümeye ayrılır. Çapraz geçerlilik yöntemi kullanıldığından tüm veri 6 parçaya bölünür ve rastgele 5 parça modelin eğitiminde, geriye kalan 1 parça modelin testinde kullanılır. Her parça 32 adet veri içerir. Her parçaya fold denir. Tüm veri kümesinin 0.83'ü eğitim amaçlı 0.17'si test amaçlı kullanılır. Eğitim amaçlı ayrılan veri kümesi ile ANFIS modeli eğitilir. Eğitim sonucunda ANFIS modeli 18 kural oluşturmaktadır. ANFIS modelinin eğitim sonrasındaki mimari yapısı Şekil 6.3'de verilmektedir. Eğitilen modele, eğitimde kullanılmayan denetleme verileri

girdi olarak verilir. ANFIS bu verilerle bir çıktı oluşturur. Denetleme verilerinin gerçek çıktı değerleri ile ANFIS modelinin oluşturduğu çıktı değerleri arasında ki fark modelin hatasını vermektedir. Şekil 6.2’de her fold için gerçek çıktı değerleri ile ANFIS modelinin verdiği çıktı değerleri aynı grafik üzerinde gösterilmektedir.

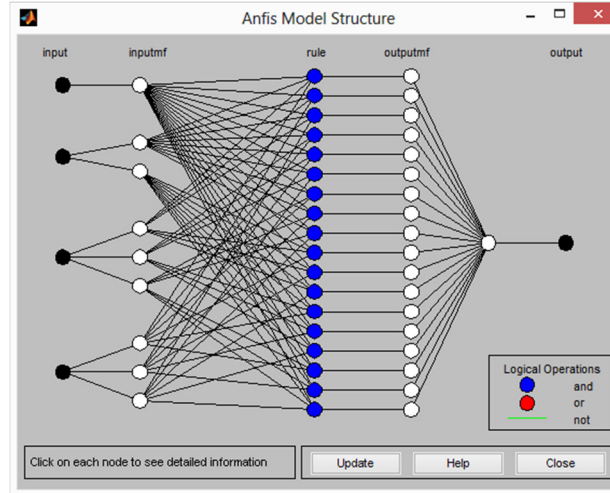


Şekil 6.2. ANFIS modelinin çıktı değerleri ile gerçek çıktı değerleri

ANFIS modelinin girdilerinin her biri için önceden tanımlanan üyelik fonksiyonu sayısı kural katmanındaki kuralların sayısını doğrudan etkilemektedir.

Şekil 6.3’de görüldüğü gibi girdilerin üyelik fonksiyon sayısının çarpımı sonucunda 18 kural ortaya çıkmıştır. ANFIS modelinin uzman tarafından düzenlenebilmesi bu ekran aracılığı ile mümkün hale gelebilmektedir. Bir uzman, kuralların üzerine tıklayarak ilgili kuralı düzenleyebilir. Uzman tarafından çok iyi bilinen

uygulamalarda bu işlem ile daha düzgün sonuçlar elde edilebilir. Bu çalışmada herhangi bir kural düzenlemesine gidilmemiştir.



Şekil 6.3. ANFIS modelinin mimarisi

Tablo 6.34’de eğitimde ve modelin testinde kullanılan foldlar, tahmin değerlerinin gerçek değerlerden sapma miktarları ve bu sapma miktarlarının frekansları her satır bir olası durumu temsil edecek şekilde gösterilmektedir.

Tablo 6.34. ANFIS modelinin her olası durum için performansı

FOLD ÜÇÜNDE KULLANILAN VERİLERİN SAYISI	FOLD ÜÇÜNDE KULLANILAN VERİLERİN SAYISI	Sapma değerleri ve bu değerlerin frekansları
1,2,3,4,5	6	+/-1=>12 +/-2 => 1 Doğru Tahmin değeri: 19
1,2,3,4,6	5	+/-1=>11 +/-2 => 1 Doğru Tahmin değeri: 20
1,2,3,5,6	4	+/-1=>12 +/-2 => 1 Doğru Tahmin değeri: 19
1,2,4,5,6	3	+/-1=>12 +/-2 => 1 Doğru Tahmin değeri: 19
1,3,4,5,6	2	+/-1=>10 +/-2 => 1 Doğru Tahmin değeri: 21
2,3,4,5,6	1	+/-1=>14 +/-2 => 1 Doğru Tahmin değeri: 17

Tablo 6.34’de görüldüğü gibi ANFIS modeli 192 adet verinin 116 adedini doğru tahminde bulunmuştur. 70 adedini bir birim sapma ile, 6 adedini de iki birim sapma

ile tahmin etmiştir. Farklı istatistiksel ölçülerle ağın performans değerlendirilmesi yapılmıştır. Farklı istatistiksel ölçütlerle yapılan değerlendirme sonuçları her bir fold için Tablo 6.35’de gösterilmektedir. Ayrıca bu tabloda ANFIS modelinin uygulamadaki ortalama performans değeri bilgisi de yer almaktadır.

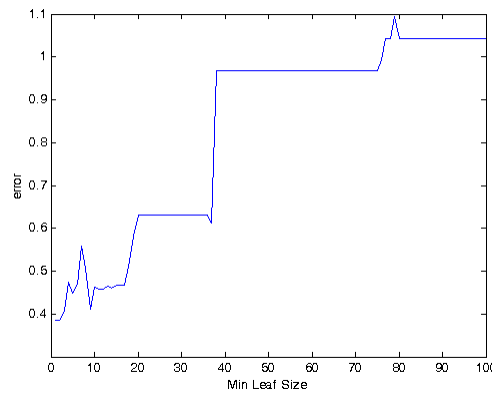
Tablo 6.35. Ağın performansının farklı istatistiksel ölçütlerle değerlendirilmesi

Değerlendirme Ölçütü	Fold1	Fold2	Fold3	Fold4	Fold5	Fold6	Ortalama
RMSE	0,75	0,66143	0,707107	0,70710	0,6846	0,707106	0.7029
MAPE	18,98	15,67708	22,40327	12,79018	17,7381	20,05952	17,9413
MAD	0,5	0,375	0,4375	0,4375	0,4375	0,4375	0,4375

6.3.3. Karar Ağaçları

Karar ağaçlarıyla analizin gerçekleştirilmesinde Gini ve Twoing kuralları kullanılmıştır. Tüm veri kümesi çapraz doğrulama yönteminin gerektirdiği gibi 6 altkümeye bölünmüştür. Foldların biri diğer iki yöntemde olduğu gibi ağacın performans değerlendirilmesinde kullanılmıştır. Geri kalan foldlardaki veri ile ağaç oluşturulmuştur. Eğitim, karar ağacının optimum olduğu derinlikte durdurulmuştur. Optimum karar ağacının tespit edilebilmesi için Matlab komutlarıyla bir fonksiyon yazılmıştır. İki ayrı kural için fonksiyon çalıştırılmıştır. Fonksiyon Ek B’de verilmektedir.

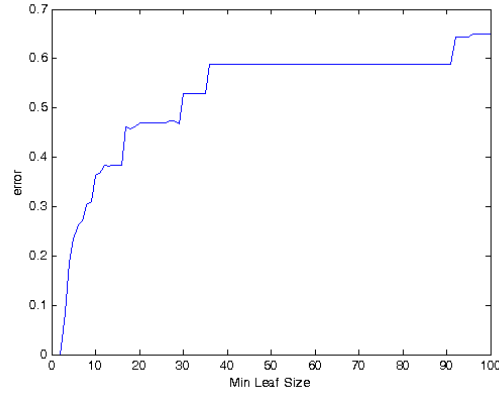
Twoing kuralıyla elde edilen min leaf size-hata grafiği Şekil 6.4’de verilmektedir.



Şekil 6.4. Twoing kuralı min leaf size-hata grafiği

Şekil 6.4’de grafik incelendiğinde genel olarak yaprak sayısının artması hatayı arttırdığı gözlenmektedir. Minleaf parametresi bir ve civarında bir değer aldığı anda hatanın en aza indiği görülmektedir. Bu nedenle minleaf parametresi uygulamada varsayılan değerinde (1) bırakılmıştır.

Gini kuralıyla elde edilen min leaf size-hata grafiği Şekil 6.5’de verilmektedir.



Şekil 6.5. Gini kuralı min leaf size-hata grafiği

Fonksiyon gini kuralıyla çalıştırıldığında Şekil 6.5’de görüldüğü gibi bir grafik elde edilmiştir. Min leaf size değerinin artması hata değerini de arttırmaktadır. Bu nedenle parametrenin değeri varsayılan değerinde (1) bırakılmıştır.

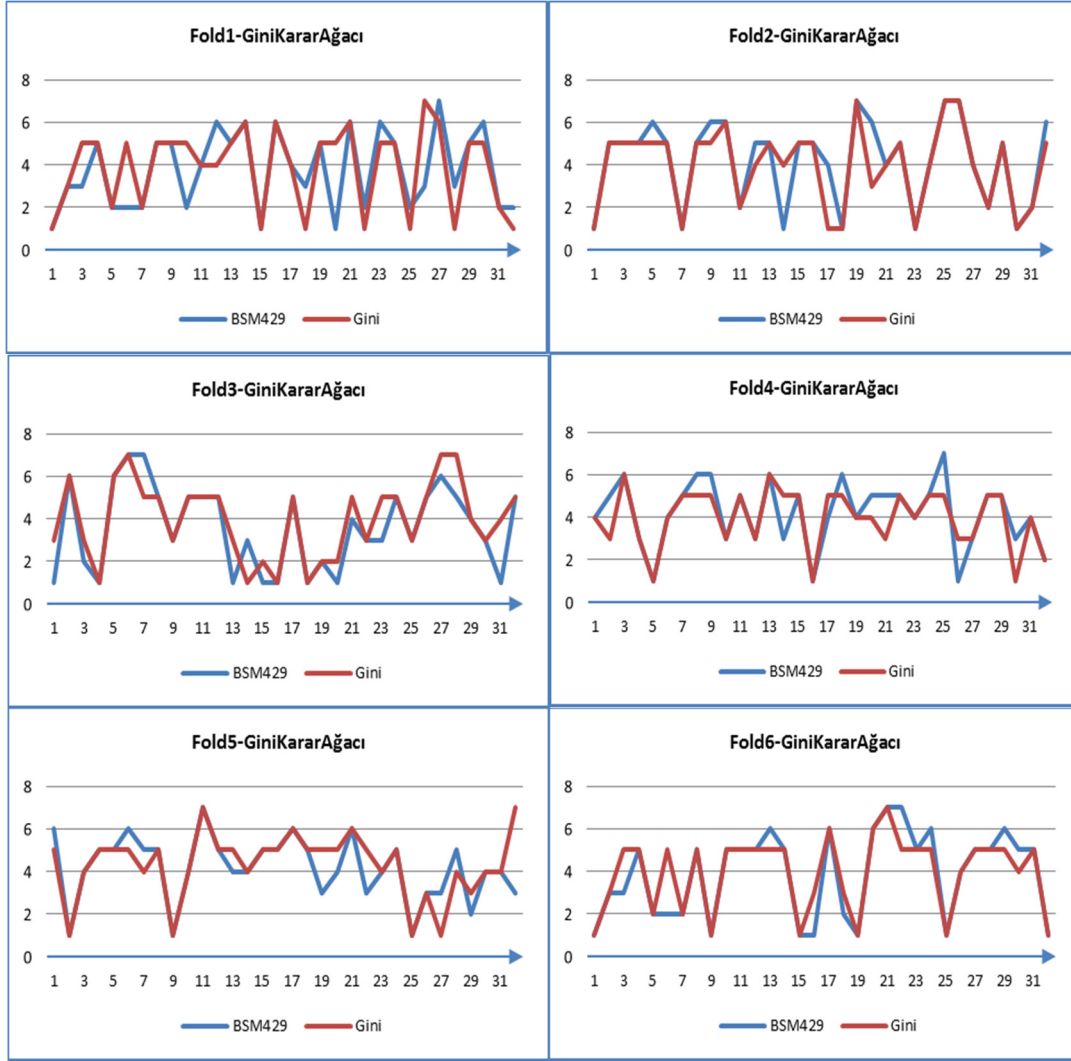
Gini ve Twoing kurallarıyla oluşturulan karar ağaçlarının performansları test foldları ile değerlendirilmiştir.

Tablo 6.36’de eğitimde ve modelin testinde kullanılan foldlar, tahmin değerlerinin gerçek değerlerden sapma miktarları ve bu sapma miktarlarının frekansları Gini ve Twoing kurallarının her biri için her satır bir olası durumu temsil edecek şekilde gösterilmektedir.

Tablo 6.36. Gini ve Twoing kurallarıyla oluşturulan karar ağaçlarının her olası durum için performansları

FOLDS USED IN TRAINING PHASES	FOLDS USED IN TESTING PHASES	Gini	Twoing
1,2,3,4,5	6	+/-1 =>5 +/-2=>3 +/-3=>1 +/-4=>0 Doğru Tahmin Sayısı: 23	+/-1 =>5 +/-2=>4 +/-3=>1 +/-4=>1 Doğru Tahmin Sayısı: 21
1,2,3,4,6	5	+/-1 =>7 +/-2=>3 +/-3=>0 +/-4=>1 Doğru Tahmin Sayısı: 21	+/-1 =>8 +/-2=>2 +/-3=>0 +/-4=>1 Doğru Tahmin Sayısı: 21
1,2,3,5,6	4	+/-1 =>5 +/-2=>6 +/-3=>0 +/-4=>0 Doğru Tahmin Sayısı: 21	+/-1 =>6 +/-2=>4 +/-3=>0 +/-4=>1 Doğru Tahmin Sayısı: 21
1,2,4,5,6	3	+/-1 =>5 +/-2=>6 +/-3=>1 +/-4=>0 Doğru Tahmin Sayısı: 20	+/-1 =>6 +/-2=>5 +/-3=>2 +/-4=>0 Doğru Tahmin Sayısı: 19
1,3,4,5,6	2	+/-1 =>4 +/-2=>0 +/-3=>3 +/-4=>0 Doğru Tahmin Sayısı: 25	+/-1 =>5 +/-2=>0 +/-3=>2 +/-4=>2 Doğru Tahmin Sayısı: 23
2,3,4,5,6	1	+/-1 =>6 +/-2=>4 +/-3=>2 +/-4=>2 Doğru Tahmin Sayısı: 18	+/-1 =>4 +/-2=>4 +/-3=>1 +/-4=>3 Doğru Tahmin Sayısı: 20

Gini algoritması toplamda 192 verinin 128 birimini doğru olarak tahmin etmiştir. 29 örneği bir birim sapma ile 22 adet örneği 2 birim sapma ile 15 adet örneği 3 birim sapma ile 3 adet örneği 4 birim sapma ile tahminlemiştir. Şekil 6.6'da her fold için, gerçek değer ve gini kuralı ile oluşturulan karar ağacı yönteminin gerçekleştirdiği tahmin sonuçları aynı grafik üzerinde gösterilmektedir.



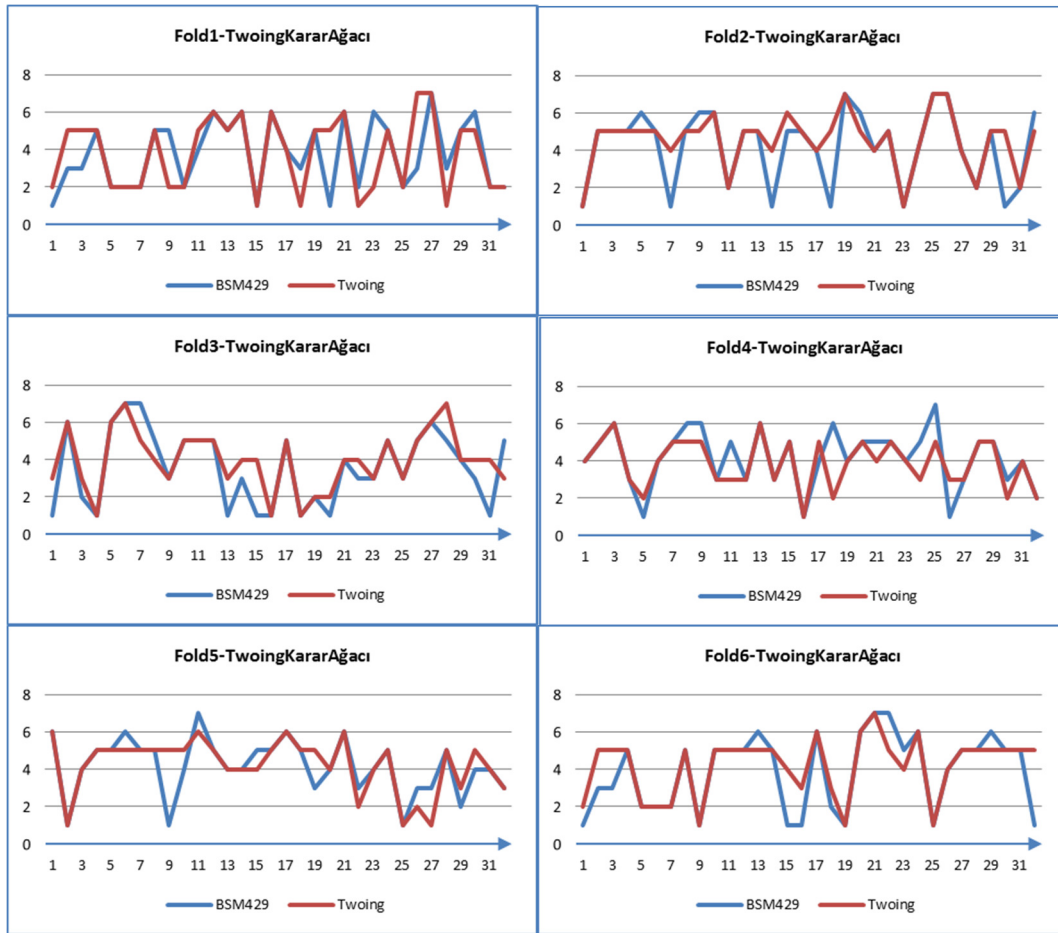
Şekil 6.6. Gini kuralı ile oluşturulan karar ağacının oluşturduğu çıktı değerleri ile gerçek çıktı değerleri

Gini algoritması kullanılarak oluşturulan karar ağacının farkı istatistiksel ölçütlerle yapılan değerlendirme sonuçları her satır bir foldu temsil edecek şekilde Tablo 6.37’de gösterilmektedir. Ayrıca bu tabloda modelinin uygulamadaki ortalama performans değeri bilgisi de yer almaktadır.

Tablo 6.37. Gini kuralıyla oluşturulan karar ağacının performansının farklı istatistiksel ölçütlerle değerlendirilmesi

Değerlendirme Ölçütü	Fold1	Fold2	Fold3	Fold4	Fold5	Fold6	Ortalama
RMSE	1,5	0,984251	1,08972	0,95197	1,0458	0,90138	1,0788
MAPE	39,50	15,46875	37,2991	16,77827	15,8333	17,6636	23,758
MAD	0,875	0,40625	0,625	0,53125	0,53125	0,4375	0,5677

Twoing kuralıyla oluşturulan karar ağacı toplam 192 adet verinin 125 adedini doğru olarak tahmin etmiştir. 23 adet örneği bir birim sapma ile 19 adet örneği 2 birim sapma ile 6 adet örneği 3 birim sapma ile 8 adet veriyi 4 birim sapma ile tahminlemiştir. Şekil 6.7’de her fold için, gerçek değer ve twoing kuralı ile oluşturulan karar ağacı yönteminin gerçekleştirdiği tahmin sonuçları aynı grafik üzerinde gösterilmektedir.



Şekil 6.7. Twoing kuralı ile oluşturulan karar ağacının oluşturduğu çıktı değerleri ile gerçek çıktı değerleri

Twoing kuralı ile oluşturulan karar ağacının farkı istatistiksel ölçütlerle yapılan değerlendirme sonuçları her satır bir foldu temsil edecek şekilde Tablo 6.38’de gösterilmektedir. Ayrıca bu tabloda modelinin uygulamada ki ortalama performans değeri bilgisi de yer almaktadır.

Tablo 6.38. Twoing kuralıyla oluşturulan karar ağacının performansı farklı istatistiksel ölçütlerle değerlendirilmesi

Değerlendirme Ölçütü	Fold1	Fold2	Fold3	Fold4	Fold5	Fold6	Ortalama
RMSE	1,5512	1,31101	1,1726	1,08972	1	1,19895	1,2205
MAPE	34,947	46,458	43,080	18,3407	23,467	39,5386	34,3055
MAD	0,8437	0,59375	0,6875	0,5625	0,5	0,625	0,6354

6.3.4. Yöntemlerin Performanslarının Karşılaştırılması

Uygulamada öğrencinin seçmek istediği dersleri hangi başarı derecesiyle tamamlayacağını tahmini incelenmektedir. Yapay sinir ağları, karar ağaçları ve anfis yöntemlerinin tahmindeki performansı örnek bir ders üzerinden değerlendirilmektedir. Örnek ders olarak veri madenciliği dersi seçilmiş olup en iyi performansı veren ağ ile teknik seçmeli derslerden olan yapay zeka, bilgisayar grafiği, internet mühendisliği derslerinden öğrencilerin alacağı not tahmin edilmektedir.

Yapay sinir ağları yöntemi, uygulamada en fazla 2 birimlik bir sapma hatası vermiştir. 10 yıllık bir küme içerisinde sadece 5 öğrenciyi 2 birimlik bir sapmayla tahmin etmiştir. 62 adet öğrencinin alacağı notu 1 birimlik sapmayla, 123 adet öğrencinin alacağı notu doğru bir şekilde tahmin etmiştir.

ANFIS yöntemi, uygulamada en fazla 2 birimlik bir sapma hatası vermiştir. 70 adet öğrenciyi 1 birimlik sapma ile, 6 adet öğrenciyi ise 2 birimlik bir sapmayla tahmin etmiştir. 192 adet verinin 116 adedini doğru bir şekilde tahmin etmiştir.

Karar ağaçları yönteminde, iki farklı kural kullanılmaktadır. Gini kuralıyla oluşturulan karar ağacı toplamda 192 adet verinin 128 adedini doğru bir şekilde tahmin etmiştir. 29 adedini 1 birimlik bir sapma ile, 22 adedini 2 birimlik bir sapma ile, 15 adedini 3 birimlik sapma ile 3 adedini 4 birimlik bir sapma ile tahmin etmiştir. Twoing kuralıyla oluşturulan karar ağacı toplamda 192 adet verinin 125 adedini doğru bir şekilde tahmin etmiştir. 23 adedini 1 birimlik bir sapma ile 19 adedini 2 birimlik bir sapma ile 6 adedini 3 birimlik bir sapma ile 8 adedini 4 birimlik bir sapma ile tahmin etmiştir.

Yöntemler incelendiğinde Gini kuralıyla oluşturulan karar ağaçları yönteminin doğru tahmin sayısı diğer yöntemlerden fazladır. 192 öğrenci verisinden 128 adet öğrenci verisini doğru bir şekilde tahmin etmiştir. Fakat tahmin sonuçlarında çok sayıda sapmalar mevcuttur ve sapma miktarları fazladır. 4 birimlik, 3 birimlik sapmalarla cevap verdiği görülmüştür. Sapma miktarının çok fazla olması performansı çok fazla yarıltmaktadır. Doğru tahmin etme sayısı ise bu olumsuzluğu azaltacak derecede diğer yöntemlerden fazla değildir.

ANFIS yönteminde elde edilen doğru tahmin etme sayısı ysa yöntemiyle elde edilen doğru tahmin sayısı arasında çok fark yoktur. ANFIS yöntemi en fazla 2 birimlik bir sapma hatası vermiştir. Sapma miktarları ysa yöntemiyle aynıdır fakat sapma gösteren veri sayısı ysa yönteminden daha fazladır.

Yöntemlerin tahmin değerlerinin gerçek değerlerden sapma miktarları ve bu sapma miktarlarının frekansları Tablo 6.39’da verilmektedir.

Tablo 6.39. Yöntemlerin tahmin performanslarının karşılaştırılması

	Yapay İir ağları	ANFIS	Gini karar ağacı	Twoin□karar ağacı
Doğru tahmin [ayı□	123	113	128	125
1 birimlik [apma	62	71	29	23
2 birimlik [apma	5	8	22	19
3 birimlik [apma	0	0	15	6
4 birimlik sapma	0	0	3	8

Tablo 6.40’da YSA, ANFIS ve karar ağacı yöntemlerinin performansları RMSE, MAD ve MAPE ölçütleriyle değerlendirilmiştir.

Tablo 6.40. Yöntemlerin performanslarının farklı istatistiksel ölçütlerle değerlendirilmesi

Yöntemler	RMSE	MAD	MAPE
YSA	0.65615	14,7246	0.3906
ANFIS	0.702901	17,9413	0,4375
Karar ağaçları (Gini)	1,0788	23,7586	0.5677
Karar ağaçları (Twoin□)	1,2205	34,3055	0,6354

Tablo 6.40 incelendiğinde, YSA yönteminin daha iyi performans gösterdiği görülmektedir.

6.4. YSA Yöntemiyle Seçmeli Derslerin Başarısının Tahminlenmesi

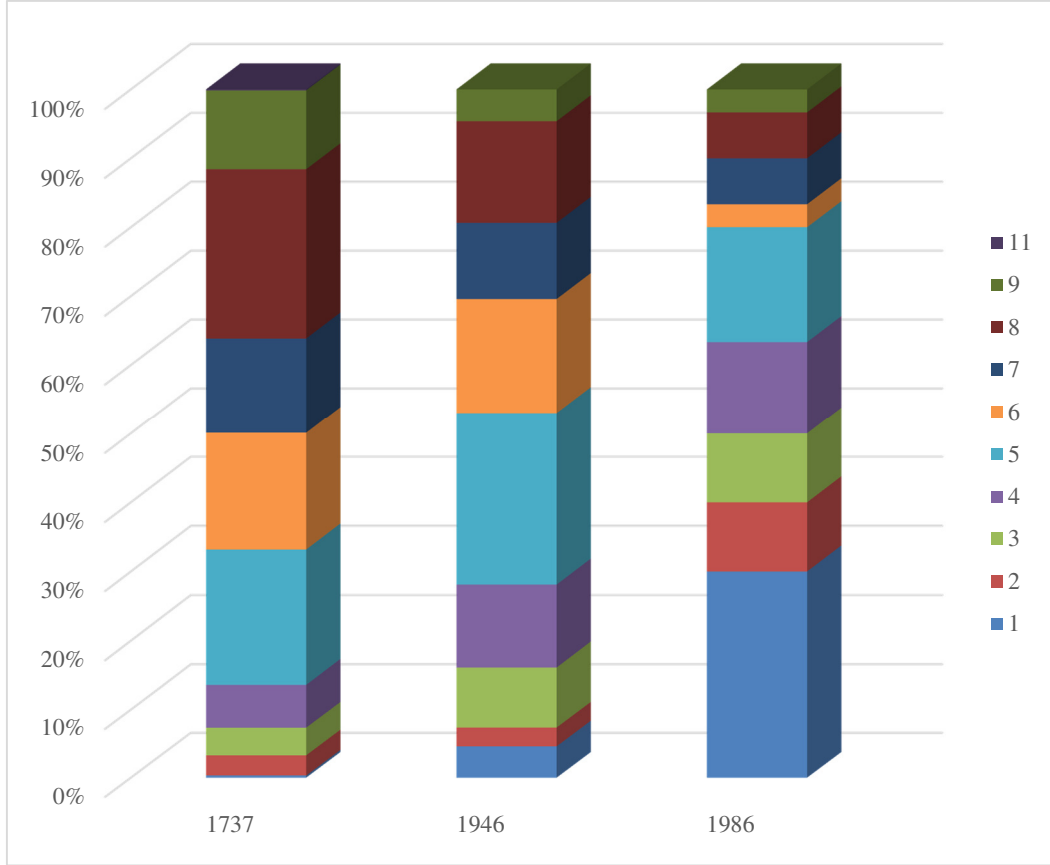
Uygulamada öğrencilerin veri madenciliği dersini hangi başarı puanıyla tamamlayacağını tahmini 3 farklı yöntem kullanılarak yapılmıştır ve yapay sinir ağları yönteminin diğer yöntemlerden daha başarılı olduğu görülmüştür. Bu nedenle öğrencilerin seçmek istedikleri dersleri hangi başarı puanıyla tamamlayacağını tahmininde yapay sinir ağları kullanılacaktır.

Sakarya Üniversitesi Bilgisayar ve Bilişim Bilimleri Fakültesinin 7. ve 8. Yarıyılında teknik seçmeli dersler mevcuttur. Bu dersler aynı kriterlere sahip değildirler. Bazı dersler uzun bir eğitim öğretim geçmişine sahiptir, bazıları sadece birkaç yıllık bir geçmişe sahiptir. Bazı dersler, uzun bir zaman diliminde aynı öğreticiye sahiptir, bazı derslerin ise öğreticisi sıklıkla değişmiştir. Aynı şekilde ders içerikleri büyük oranda değişen, değerlendirme kriterleri değişen birçok ders mevcuttur. Tüm bu farklılıklar, tahmini olumlu yönde de olumsuz yönde de etkileyebilir.

Uzun bir eğitim öğretim geçmişine sahip dersler, yapay sinir ağının eğitiminde yeterli veri sayısına sahipken, birkaç yıllık bir geçmişe sahip derslerin veri sayısı yapay sinir ağının eğitimi için yetersiz kalabilir. Örneğin; BSM449 (Tıbbi istatistik ve tıp bilimine giriş) dersi 7. Yarıyıl teknik seçmeli dersidir ve sadece 2013 yılının 1. dönemine ait veri mevcuttur. BSM437 (ERP sistemleri), BSM443 (Gömülü sistemler), BSM439 (İleri sayısal sistem tasarımı ve sentezleme), BSM445 (Kuyruk teorisi), BSM447 (Mobil uygulama geliştirme), BSM431 (Sayısal işaret işleme), BSM441 (Sistem yöneticiliği) dersleri eğitim için yeterli veri sayısına sahip değildirler.

Öğretici kriteri öğrencinin ve dersin performansını etkilemektedir. Şekil 6.8'de aynı dersin farklı bir öğretici tarafından verilmesi durumu gösterilmektedir. Grafik incelendiğinde öğreticinin değişmesi durumunda sınıfın değişen not dağılımı görülmektedir. Bu durum algoritmaları yanıltabilir, yanlış eğitime yol açabilir.

Aynı problem dersin içeriğinin kapsamlı bir şekilde değişmesinde kaynaklanabilir.

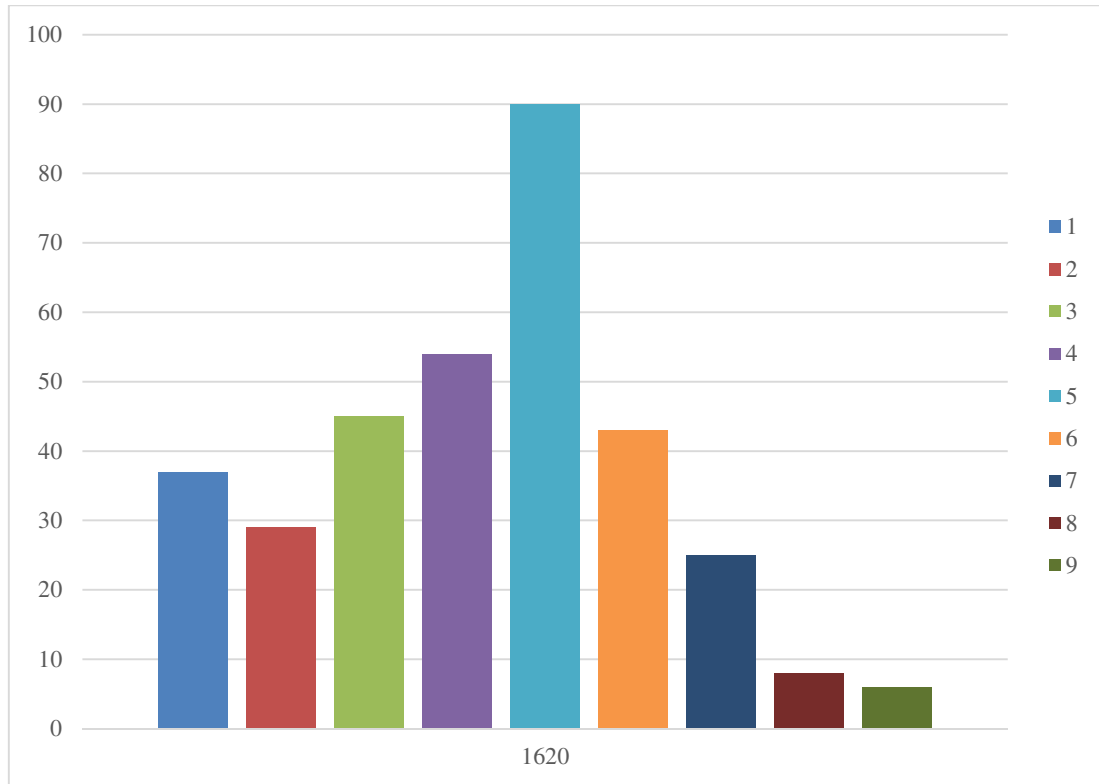


Şekil 6.8. Farklı öğretim elemanları tarafından verilen dersin not dağılımı

Şekil 6.8'de BSM309 dersinin 3 farklı öğretim elemanı tarafından verilmesi durumunda öğrencilerin aldığı başarı puanlarının dağılımı yüzde olarak verilmektedir. Yatay eksen öğretim elemanları dikey eksen ise öğrencilerin aldığı başarı notunun dağılımı yüzde olarak gösterilmektedir. Başarı notları 1'den (AA) 8'e (FF) kadar en başarılı nattan başarısız nota kadar sıralanmıştır. 9 değeri öğrencinin sınava girmediğini, 11 değeri ise öğrencinin dersten yeterli olduğunu gösterir. Bu 3 grafik 2007-2013 yılları arasında temsil etmektedir. Bu yıllar arasında bu dersi 3 farklı öğretim görevlisinin verdiği ve değerlendirme ve ders anlatımlarının birbirlerinden bağımsız olduğu bilinmektedir. Şekil 6.8'de görüldüğü üzere AA (1) alan öğrenci sayısının öğretim görevlisine göre çok büyük bir değişim gösterdiği açıkça görülmektedir. Diğer başarı notlarının dağılımları da aynı şekilde farklılık arz etmektedir. Bu durumda bir dersin farklı öğretim görevlileri tarafından verilmesi,

öğrencinin alacağı başarı notunu önemli derecede etkilediği görülmektedir. Bu nedenle bir dersten bir öğrencinin alacağı başarı notunun tahmin edilebilmesi için geçmiş verilerin analizi yapılırken mutlaka öğretim görevlisi kriterinin de ele alınması gerektiği görülmektedir.

Şekil 6.9’da BSM429 dersinde öğrencilerin yıllara göre not dağılımı sunulmaktadır.



Şekil 6.9. BSM429 dersi not dağılımı

Şekil 6.9’da BSM429 dersinin 2007-2013 yılları arasındaki başarı notunun dağılımı gösterilmektedir. Bu dersin bu tarihler arasında tek bir öğretim görevlisi tarafından verildiği bilinmektedir. Grafiğe bakıldığında normal dağılıma büyük oranda uyduğu görülebilir.

Öğretici her dönem öğrencileri belirli kıstaslara göre değerlendirir. Değerlendirme yöntemlerinin değişmesi o dönemdeki verilerin değişmesine yol açabilir. Örneğin; öğrencilerin vize ve final sınavlarının yerine ödev alması ve öğrencilerin bu ödevlere göre değerlendirilmesi o dönemdeki öğrencilerin performanslarında (öğrenci

verilerinde) dalgalanmalara neden olabilir. Fakat bu deęişikliklerin tespit edilmesi ve takibi mümkün deęildir. Bu nedenle bu deęişim ihmal edilecektir.

Uygulama için BSM421, BSM435 ve BSM425 dersleri seçilmiştir. Sakarya Üniversitesi'nin 7. yarıyılında verilen bu teknik seçmeli derslerin seçilme nedeni uzun yıllardır müfredatta yer almaları ve öğreticilerinin deęişmemesidir.

Bu bölümde BSM429 dersinin tahmininde başarı sonuçlar veren YSA modeli ile bu dersleri öğrenciler hangi başarı puanı ile tamamlayacaklarının tahmini incelenmektedir.

6.4.1. BSM421

BSM421 dersi temel grafik donanım ve yazılım bilgisine ihtiyaç duyar. Dersin hedefi ise öğrenciye 2D ve 3D modelleme ve dönüşümler yapabilme, dönüşüm gözlemlene, izdüşümü, grafiksel yazılım paketleri ve grafik sistemlerini kullanma becerisi kazandırmaktır. Ayrıca, renk ve oyun geliştirme derste öğrenciye verilmek istenen kavramlardır.

Zorunlu derslerin ders içerikleri incelendiğinde MAT113 (Lineer cebir), BSM103 (Algoritma ve Programlama 1), BSM102 (Algoritma ve Programlama 2), BSM207 (Veri yapıları) ve BSM210 (Ayrık İşlemsel Yapılar) derslerinin BSM421 dersi ile ilişkili olduğu söylenebilir. Bu nedenle bu derslerin başarı puanlarının ayrıca cinsiyet ve öğrencinin 6. dönem sonundaki genel not ortalaması bilgilerinin BSM421 dersinin başarı puanının tahmininde etkili olduğu düşünülmektedir. Bu düşüncelerin desteklenmesi adına Bölüm 4.2'de regresyon ve korelasyon analizleri yapılmıştır. Yapılan analizler sonucunda BSM207 ve MAT113 derslerinin diğer derslere nazaran BSM421 dersi ile daha uyumlu olduğu görülmüştür. BSM421 dersi ile cinsiyet bilgisinin ilişkisi yeterli görülmemiş, öğrencinin 6. dönem sonundaki genel not ortalaması bu ilişkide etkin ve anlamlı bulunmuştur. BSM421 dersi için girdi parametreleri olarak BSM207, MAT113 ve öğrencinin 6. dönem sonundaki genel not ortalaması bilgileri alınmıştır. Bu girdi parametrelerinden oluşan veri kümesi Bölüm

4.1’de verilen çeşitli filtreleme işlemlerine tabi tutulmuştur. Filtreleme işlemlerinden sonra veri kümesinde 196 adet veri kalmıştır.

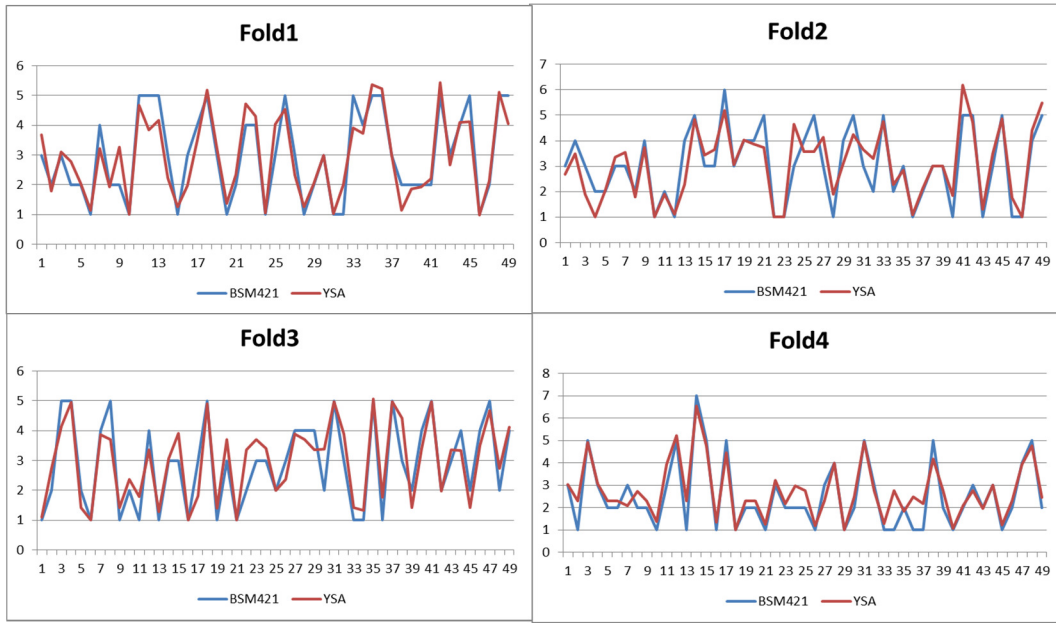
Uygulamada çapraz geçerlilik yöntemi kullanılmaktadır. Bu nedenle 196 adet veri 4 kümeye ayrılmaktadır. Rastgele 4 küme ile model eğitilmekte, geri kalan küme ile modelin performansı değerlendirilmektedir.

Tablo 6.41’de eğitimde ve modelin testinde kullanılan foldlar, eğitim performansları, tahmin değerlerinin gerçek değerlerden sapma miktarları ve bu sapma miktarlarının frekansları her satır bir olası durumu temsil edecek şekilde gösterilmektedir.

Tablo 6.41. BSM421-YSA modeli

FOLDS USED IN TRAINING PHASES	FOLDS USED IN TESTING PHASES	TRAINING PERFORMANCE	Sapma değerleri ve bu değerlerin frekansları
2,3,4	1	0.0525218	+/-1 =>17 +/-2 =>0 Doğru Tahmin Sayısı: 32
1,3,4	2	0.047082	+/-1=>17 +/-2 =>2 Doğru Tahmin Sayısı: 30
1,2,4	3	0.0725232	+/-1=>23 +/-2 =>0 Doğru Tahmin Sayısı: 26
1,2,3	4	0.0687517	+/-1=>13 +/-2 =>1 Doğru Tahmin Sayısı: 35

Tablo 6.41’de görüldüğü gibi 196 adet veriden ileri beslemeli geri yayımlı ağ, toplamda 123 adedini doğru tahminde bulunmuştur. 70 adedini bir birim sapmayla, 3 adedini de iki birim sapma ile tahmin etmiştir. Şekil 6.10’da her fold için gerçek değer ile yapay sinir ağı modelinin tahmin ettiği değerler aynı grafik üzerinde gösterilmektedir.



Şekil 6.10. BSM421 için YSA modelinin oluşturduğu çıktı değerleri ile gerçek çıktı değerleri

6.4.2. BSM425

BSM425 dersi öğrenme teknikleri, arama teknikleri, yapay zeka teknikleri hakkında öğrenciye bilgi verir. Öğrencinin problem çözme yeteneğini geliştirmeyi amaçlar. Temel algoritma ve matematik bilgisine ihtiyaç duyar.

Zorunlu derslerin ders içerikleri incelendiğinde MAT217 (Sayısal Analiz), BSM103 (Algoritma ve Programlama 1), BSM207 (Veri yapıları) ve BSM306 (Yazılım Mühendisliği) derslerinin BSM425 dersi ile ilişkili olduğu söylenebilir. Bu nedenle bu derslerin başarı puanlarının ayrıca cinsiyet ve öğrencinin 6. dönem sonundaki genel not ortalaması bilgilerinin BSM425 dersinin başarı puanının tahmininde etkili olduğu düşünülmektedir. Bu düşüncelerin desteklenmesi adına bölüm 4.2’de regresyon ve korelasyon analizleri yapılmıştır. Yapılan analizler sonucunda BSM207 ve BSM306 derslerinin diğer derslere nazaran BSM421 dersi ile daha uyumlu olduğu görülmüştür. BSM421 dersi ile cinsiyet bilgisinin ilişkisi yeterli görülmemiş, öğrencinin 6. dönem sonundaki genel not ortalaması bu ilişkiye etkin ve anlamlı bulunmuştur. BSM425 dersi için girdi parametreleri olarak BSM207, BSM306 ve öğrencinin 6. dönem sonundaki genel not ortalaması bilgileri alınmıştır. Bu girdi parametrelerinden oluşan veri kümesi Bölüm 4.1’de verilen çeşitli filtreleme

işlemlerine tabi tutulmuştur. Filtreleme işlemlerinden sonra veri kümesinde 149 adet veri kalmıştır.

Uygulamada çapraz geçerlilik yöntemi kullanılmaktadır. Bu nedenle 149 adet veri 4 kümeye ayrılmaktadır. Rastgele 4 küme ile model eğitilmekte, geri kalan küme ile modelin performansı değerlendirilmektedir.

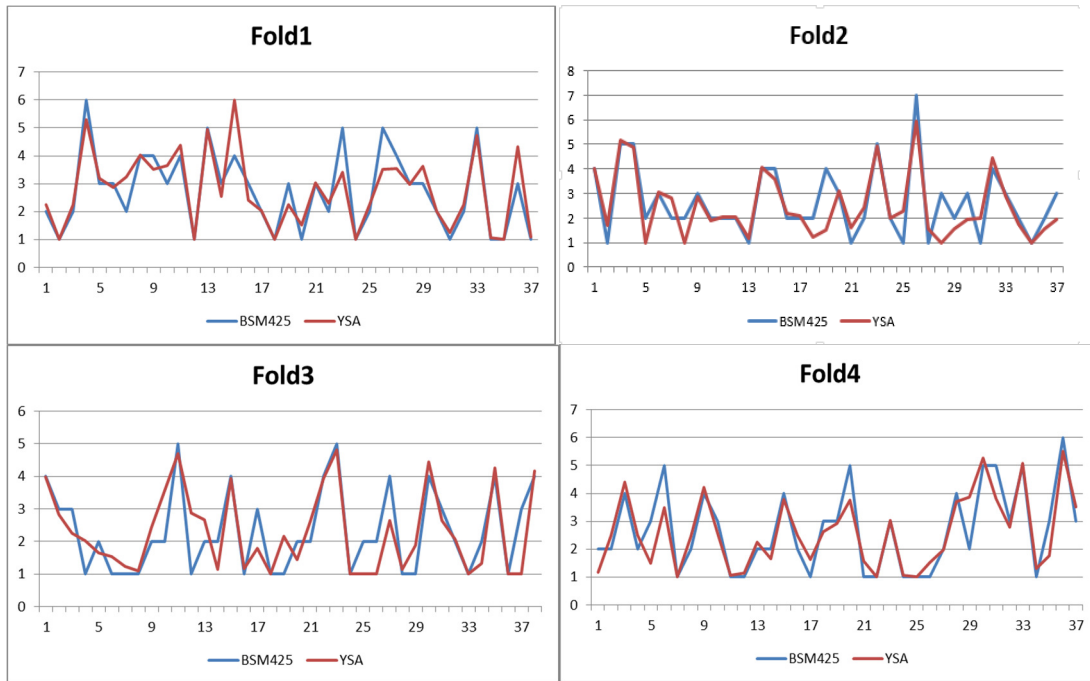
Tablo 6.42’de eğitimde ve modelin testinde kullanılan foldlar, eğitim performansları, tahmin değerlerinin gerçek değerlerden sapma miktarları ve bu sapma miktarlarının frekansları her satır bir olası durumu temsil edecek şekilde gösterilmektedir.

Tablo 6.42. BSM425-YSA modeli

FOLDS USED IN TRAINING PHASES	FOLDS USED IN TESTING PHASES	TRAINING PERFORMANCE	Sapma değerleri ve bu değerlerin frekansları
2,3,4	1	0.0670068	+/-1 =>9 +/-2 => 2 Doğru Tahmin Sayısı: 26
1,3,4	2	0.0652444	+/-1=>12 +/-2 => 2 Doğru Tahmin Sayısı: 23
1,2,4	3	0.0594858	+/-1=>14 +/-2 => 3 Doğru Tahmin Sayısı: 21
1,2,3	4	0.0465493	+/-1=>9 +/-2 => 2 Doğru Tahmin Sayısı: 26

Tablo 6.42’de görüldüğü gibi 149 adet veriden ileri beslemeli geri yayımlı ağ, toplamda 96 adedini doğru tahminde bulunmuştur. 44 adedini bir birim sapmayla, 9 adedini de iki birim sapma ile tahmin etmiştir.

Şekil 6.11’de her fold için gerçek değer ile yapay sinir ağı modelinin tahmin ettiği değerler aynı grafik üzerinde gösterilmektedir.



Şekil 6.11. BSM425 için YSA modelinin oluşturduğu çıktı değerleri ile gerçek çıktı değerleri

6.4.3. BSM435

BSM435 dersinde bir öğrenci RFC standartlarına göre TCP/IP protokol mimarisi içerisindeki çekirdek protokoller (TCP, UDP, IP, ICMP, ARP) ve uygulama katmanında hizmet veren protokoller hakkında bilgi sahibi olur. Öğrencinin bu dersten başarılı olabilmesi için temel ağ bilgisine ve veri iletişimi konusunda temel bilgilere sahip olması gerekir.

BSM435 dersinin ders içeriği incelendiğinde bu içerikle uyum gösteren sadece 2 dersin olduğu görülmektedir. Bu dersler BSM305 ve BSM304 dersleridir. BSM435 dersinin ders içeriği bakımından diğer derslerin ders içeriği ile bir uyum gösterememesi nedeniyle bu iki ders doğrudan girdi parametreleri olarak alınabileceği söylenebilir. Daha önceden yapılan regresyon ve korelasyon analizleri sonucunda bu düşüncelerin desteklendiği görülmektedir. Sonuç olarak BSM305, BSM304 ve ortalamadonem6 parametreleri girdi parametreleri olarak kabul edilmiştir. Bu girdi parametrelerinden oluşan veri kümesi Bölüm 4.1'de verilen çeşitli filtreleme işlemlerine tabi tutulmuştur. Filtreleme işlemlerinden sonra veri kümesinde 137 adet veri kalmıştır.

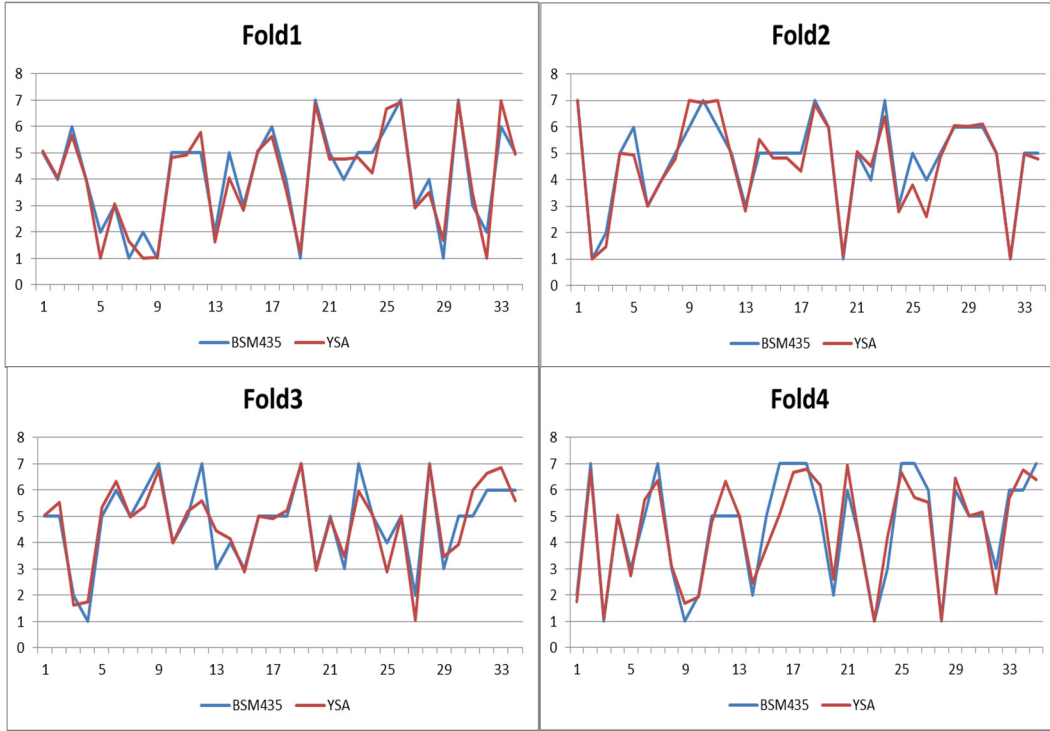
Uygulamada çapraz geçerlilik yöntemi kullanılmaktadır. Bu nedenle 137 adet veri 4 kümeye ayrılmaktadır. Rastgele 4 küme ile model eğitilmekte, geri kalan küme ile modelin performansı değerlendirilmektedir.

Tablo 6.43’de eğitimde ve modelin testinde kullanılan foldlar, eğitim performansları, tahmin değerlerinin gerçek değerlerden sapma miktarları ve bu sapma miktarlarının frekansları her satır bir olası durumu temsil edecek şekilde gösterilmektedir.

Tablo 6.43. BSM435-YSA modeli

FOLDS USED IN TRAINING PHASES	FOLDS USED IN TESTING PHASES	TRAINING PERFORMANCE	Sapma değerleri ve bu değerlerin frekansları
2,3,4	1	0.0345129	+/-1 =>11, +/-2 => 0 Doğru Tahmin Sayısı: 23
1,3,4	2	0.0583385	+/-1=>10 +/-2 => 0 Doğru Tahmin Sayısı: 24
1,2,4	3	0.043563	+/-1=>12 +/-2 => 0 Doğru Tahmin Sayısı: 22
1,2,3	4	0.0443367	+/-1=>13 +/-2 => 1 Doğru Tahmin Sayısı: 21

Tablo 6.43’de görüldüğü gibi 137 adet veriden ileri beslemeli geri yayımlı ağ, toplamda 90 adedini doğru tahminde bulunmuştur. 46 adedini bir birim sapmayla, 1 adedini de iki birim sapma ile tahmin etmiştir. Şekil 6.12’de her fold için gerçek değer ile yapay sinir ağı modelinin tahmin ettiği değerler aynı grafik üzerinde gösterilmektedir.



Şekil 6.12. BSM435 için YSA modelinin oluşturduğu çıktı değerleri ile gerçek çıktı değerleri

6.5. Genel Performans Değerlendirmesi

Yapay sinir ağları ile 4 farklı teknik seçmeli ders için tahmin yapılmaktadır. Derslerin gösterdiği performans Tablo 6.44'de verilmektedir.

Tablo 6.44. Teknik seçmeli derslerin performansları

Teknik Seçmeli ders	Veri Sayısı	Doğru tahmin Sayısı	1 birim Sapma	2 birim Sapma
BİM429	192	123	64	5
BİM421	196	123	70	3
BİM425	149	96	44	9
BİM435	137	90	46	1

Tablo 6.44 incelendiğinde derslerden elde edilecek başarı puanının tahmininde en fazla 2 sapma mevcuttur. Sapma gösteren tahminler incelendiğinde yapay sinir ağı modelinin verdiği tahmini çıktılar öğrencilerin bu derste gösterecekleri başarı

potansiyellerini gösterdiği ve modelin anlamlı sonuçlar çıkardığı görülmektedir. Yapay sinir ağı BSM429 dersi için 0.64, BSM421 için 0.63, BSM425 için 0.64 ve BSM435 için 0.65 oranında doğru tahminde bulunmuştur.

Yapay sinir ağlarının çıktı değerlerini belli bir aralık içinde yorumlarsak eğer Tablo 6.45’de verilen bilgilere ulaşılabilir. (AA-BB=1, CB-CC=2, DC-DD=3)

Tablo 6.45. Teknik seçmeli derslerin tahmini çıktı değerlerinin belli bir aralıkta gösterilmesi

Teknik Seçmeli ders	Veri Sayısı	Doğru Tahmin Sayısı
BSM429	192	162
BSM421	196	171
BSM425	149	136
BSM435	137	123

Yapay sinir ağının çıktı değerleri belli bir aralık içine çekildiğinde elde edilen sonuçların doğruluk oranı Tablo 6.46’da görüldüğü gibi artmıştır. Yapay sinir ağları BSM429 dersi için 0.84, BSM421 için 0.87, BSM425 için 0.91, BSM435 için 0.89 oranında doğru tahminde bulunmuştur.

Aralık içine alınarak yapılan tahmin değerleri ile aralığa alınmadan yapılan tahmin değerlerinin performansları Tablo 6.46’da verilmektedir.

Tablo 6.46. YSA’nın çıktı değerlerinin farklı aralıklardaki doğruluk oranları

Teknik Seçmeli ders	Aralığa alınmadan yapılan tahmin	Aralık içine alınarak yapılan tahmin
BSM429	0.64	0.84
BSM421	0.62	0.87
BSM425	0.64	0.91
BSM435	0.65	0.89

BÖLÜM 7.SONUÇ VE ÖNERİLER

7.1. Sonuç

Üniversiteler, temel ve mesleki bilgilerle donatılmış, bunun yanında kişisel gelişimine ve sosyal sorumluluklarına yatırım yapan bireyler yetiştirmeyi amaçlar. Bu amaçlarla üniversiteler öğrencilere, meslekleri ile ilgili temel bilgileri öğrenebilmeleri, farklı alanlarda ufuklarını genişletebilmeleri ve ilgi duydukları alanda kendilerini geliştirebilmeleri için birçok imkan sunar. Eğitim süreleri boyunca öğrencilere sosyal ve teknik seçmeli derslerin sunulması bu imkanlardan sadece biridir.

Teknik seçmeli dersler temel mesleki dersleri alan bir öğrencinin kendi ilgi duyduğu alandaki bilgilerini ve becerilerini geliştirmeyi amaçlar. Bu amaçla ders seçimlerinde doğru tercihlerde bulunulması öğrenciler için önemlidir.

Bu çalışmada ders seçimlerinde öğrencilere yardımcı olacak bir araç sunulmuştur. Öğrencilerin zorunlu derslerden aldıkları başarı puanları, dönem ortalama puanları ve cinsiyet bilgileri kullanılarak teknik seçmeli derslerden elde edecekleri başarı puanları tahmin edilmektedir.

Uygulamada YSA, ANFIS ve karar ağaçları yöntemleri kullanılmıştır. BSM429 dersinin verileri kullanılarak yöntemlerin karşılaştırmaları yapılmış ve YSA yönteminin diğer yöntemlere göre daha iyi performans gösterdiği görülmüştür. 10 yıllık bir veri kümesi içerisinde sadece 5 öğrenci 2 birimlik bir sapma ile tahmin edilmiştir. Gini kuralı kullanılarak oluşturulan karar ağacının doğru tahmin etme oranı %67, twoing kuralı kullanılarak oluşturulan karar ağacının doğru tahmin etme oranı %65, ANFIS yöntemi ile yapılan tahminin başarı oranı %59, YSA yöntemi ile yapılan tahminlerin başarı oranı %64 olarak bulunmuştur. Bu oranlardanda

görüldüğü gibi gini kuralıyla elde edilen karar ağacı daha fazla sayıda doğru tahminde bulunmuştur. Fakat tahmin çıktılarında çok sayıda sapmalar mevcuttur ve sapma miktarları fazladır. 4 birimlik, 3 birimlik sapmalarla cevap verdiği görülmüştür. Sapma miktarının çok fazla olması performansı çok fazla yanıltmaktadır. Doğru tahmin etme sayısı ise bu olumsuzluğu azaltacak derecede diğer yöntemlerden fazla değildir. Bu nedenle diğer derslerin tahmininde kullanılacak yöntem olarak YSA tercih edilmiştir. BSM421 dersi için YSA yöntemi uygulandığında elde edilen doğru tahmin etme oranı %63, BSM425 için %64, BSM435 için %65'dir. Tahmin sonucunda elde edilen değerler aralık içine alınarak değerlendirildiğinde BSM421 için doğruluk oranının %87, BSM425 için %91 ve BSM435 için %89 olduğu görülmüştür.

Uygulamada önceden ilişkili derslerin tespit edilmesi ve başarı notlarının diğer parametreler ile beraber değerlendirilmesi sonucunda yeterli veri sağlanması koşulu ile öğrencilerin seçmeli derslerden alacakları başarı puanlarının tahmininin mümkün olabileceği görülmüştür. Elde edilen tahmin modelinin bir bilgi sistemi ile entegrasyonu ile öğrenci ve öğretilere faydalı olabilecek bir bilişim alt yapısı sunulabilecektir.

7.2. Öneriler

Sakarya Üniversitesi bilgi sistemi olan SABİS sistemine entegre olabilecek bir yazılım geliştirilerek öğrenci ve danışmanlara bir web arayüzü sunulabilir. Bu arayüzün içerisinde öğrencinin transkript bilgileri ve tahmin edilmek istenen dersler (Teknik seçmeli dersler) olmalıdır. Transkript bilgileri öğrenci bilgi sistemi üzerinden elde edilip bir web servisi aracılığıyla ekrana getirilebilir. Tahmin edilecek olan derslerde ekranda listelenebilir. Bu listede yer alan derslerden öğrencinin elde edebileceği başarı puanı bu çalışmada gerçekleştirilmiş olan tahmin modeliyle tahmin edilerek öğrenci veya danışmana bir sonuç bilgisi gösterilebilir. Tahmin modelinin girdi parametresi olarak transkriptten alınacak olan ders başarı notları ve dönem ortalaması alınacak, model çalıştırdıktan sonra çıktı parametreside ekranda gösterilecek bir sonuç haline dönüştürülecektir. Örnek bir ekran görüntüsü Şekil 7.1'de gösterilmektedir.

Not Tahmini

Transkript Bilgileri

1. Yarıyıl -				2. Yarıyıl -			
Dersler	Başarı Notu	AKTS Kredisi		Dersler	Başarı Notu	AKTS Kredisi	
BİLİŞİM TEKNOLOJİLERİ	BB	4		WEB TEKNOLOJİLERİ	AA	4	
MATEMATİK I	DC	6		OLASILIK VE İSTATİSTİK	CC	4	
LINEER CEBİR	CC	3		FİZİK I	CC	6	
FİZİK I	DD	6		MATEMATİK II	DC	6	
INTRODUCTION TO COMPUTER ENGINEERING	BA	2		ALGORITHMS AND PROGRAMMING II	CC	5	
TÜRK DİLİ	BA	4		İNGİLİZCE	MU	4	
ALGORİTMALAR VE PROGRAMLAMA I	AA	6					
Toplam AKTS Kredisi			Ortalama	Toplam AKTS Kredisi			Ortalama
Dönem Sonu :	31	78	2,52	Dönem Sonu :	25	55	2,2
Genel:	31	78	2,52	Genel:	56	133	2,38
3. Yarıyıl -				4. Yarıyıl -			
Dersler	Başarı Notu	AKTS Kredisi		Dersler	Başarı Notu	AKTS Kredisi	
VERİ YAPILARI	BA	5		AYRILIK İŞLEMSEL YAPILAR	DD	6	
NESNEYE DAYALI PROGRAMLAMA	AA	6		COMPUTER ORGANİZASYON	DD	6	
MANTIK DEVRELERİ	BB	5		NESNEYE DAYALI ANALİZ VE TASARIM	AA	6	
SAYISAL ANALİZ	BA	5		PROGRAMLAMA DİLLERİNİN PRENSİPLERİ	DD	5	
DEFERANSİYEL DENKLEMLER	DC	6		ELEKTRONİK DEVRELER VE LABORATUVARI	DD	7	
ATATÜRK İLKELERİ VE İNKILAP TARİHİ	CC	4					
ELECTRIC CIRCUIT FUNDAMENTALS	DD	4					
Toplam AKTS Kredisi			Ortalama	Toplam AKTS Kredisi			Ortalama
Dönem Sonu :	35	95	2,71	Dönem Sonu :	30	45	1,6
Genel:	91	228	2,51	Genel:	121	276	2,28
5. Yarıyıl -				6. Yarıyıl -			
Dersler	Başarı Notu	AKTS Kredisi		Dersler	Başarı Notu	AKTS Kredisi	
VERTİKAL YÖNETİM SİSTEMLERİ	BA	7		YAZILIM MÜHENDİSLİĞİ	DD	5	
İŞLETİM SİSTEMLERİ	CC	6		WEB PROGRAMLAMA	AA	6	
BİLİŞİM HUKUKU(Fakülte/Sosyal Seç.)	DC	2		TOPLAM KALİTE YÖNETİMİ(Fakülte/Sosyal Seç.)	DC	2	
DATA COMMUNICATION	CC	5		BİLGİSAYAR AĞLARI	DD	6	
BİÇİMSEL DİLLER VE SOYUT MAKİNELER	BB	5		SİSTEM PROGRAMMING	CC	6	
İŞARETLER VE SİSTEMLER	GR	5		MİKROİŞLEMİCİLİK SİSTEMLERİ VE LABORATUVAR	CC	7	
Toplam AKTS Kredisi			Ortalama	Toplam AKTS Kredisi			Ortalama
Dönem Sonu :	30	64,5	2,15	Dönem Sonu :	32	64	2
Genel:	151	340,5	2,25	Genel:	103	404,5	2,21

7. Yarıyıl Teknik Seçmeli Dersler

Dersler	AKTS Kredisi	Tahmin Et
BİLGİSAYAR GRAFİĞİ(Bölüm Seç.)	5	<input type="button" value="Tahmin Et"/>
VERİ MADENLİĞİ(Bölüm Seç.)	5	<input type="button" value="Tahmin Et"/>
İNTERNET MÜHENDİSLİĞİ(Bölüm Seç.)	5	<input type="button" value="Tahmin Et"/>
YAPAY ZEKA(Bölüm Seç.)	5	<input type="button" value="Tahmin Et"/>

İNTERNET MÜHENDİSLİĞİ dersi için yapılan tahmin sonucunda **DC** ile **DD** arasında bir not alabileceğiniz tahmin edilmektedir.

Şekil 7.1. SABİS ile entegre not tahmin ekran görüntüsü

KAYNAKLAR

- [1] A. A. KARDAN, H. SADEGHI, S. S. GHIDARY, and M. R. F. SANI, "Prediction of student course selection in online higher education institutes using neural network," *Comput. Educ.*, vol. 65, pp. 1–11, 2013.
- [2] O. TAYLAN and B. KARAGÖZOĞLU, "An adaptive neuro-fuzzy model for prediction of student's academic performance," *Comput. Ind. Eng.*, vol. 57, no. 3, pp. 732–741, Oct. 2009.
- [3] V. O. OLADOKUN, A. T. ADEBANJO, and O. E. C. OWABA, "Predicting Students' Academic Performance using Artificial Neural Network: A Case Study of an Engineering Course .," vol. 9, no. 1, pp. 72–79, 2008.
- [4] R. D. IBRAHIM Z, "Predicting Students' Academic Performance: Comparing Artificial Neural Network, Decision Tree and Linear Regression, 2007," *21st Annual SAS Malaysia Forum, 5th September 2007, Shangri-La Hotel, Kuala Lumpur, 2007*.
- [5] T. D. GEDEON and S. TURNER, "Explaining student grades predicted by a neural network," *Proc. 1993 Int. Conf. Neural Networks (IJCNN-93-Nagoya, Japan)*, vol. 1, 1993.
- [6] D. ŞENGÜR and A. TEKİN, "Öğrencilerin Mezuniyet Notlarının Veri Madenciliği Metotları İle Tahmini," *Int. J. INFORMATICS Technol.*, vol. 6, 2013.
- [7] N. GÜNER and E. ÇOMAK, "Mühendislik Öğrencilerinin Matematik I Derslerindeki Başarısının Destek Vektör Makineleri Kullanılarak Tahmin Edilmesi. (Turkish)," *Predict. Perform. First Year Eng. Students Calc. by Using Support Vector Mach.*, vol. 17, no. 2, pp. 87–96, 2011.
- [8] A. S. BOZKIR, E. SEZER, and B. GÖK, "Öğrenci Seçme Sınavında (Öss) Öğrenci Başarımını Etkileyen Faktörlerin Veri Madenciliği Yöntemleriyle Tespiti," 2009.
- [9] I. LYKOURANTZOU, I. GIANNOUKOS, V. NIKOLOPOULOS, G. MPARDIS, and V. LOUMOS, "Dropout prediction in e-learning courses through the combination of machine learning techniques," *Comput. Educ.*, vol. 53, pp. 950–965, 2009.

- [10] W. W. GUO, "Incorporating statistical and neural network approaches for student course satisfaction analysis and prediction," *Expert Syst. Appl.*, vol. 37, pp. 3358–3365, 2010.
- [11] S. HERZOG, "Estimating Student Retention and Degree-Completion Time: Decision Trees and Neural Networks Vis-à-Vis Regression," *New Dir. Institutional Res.*, no. 131, pp. 17–33, 2006.
- [12] Y. ZHANG, S. OUSSENA, T. CLARK, and H. KIM, *Use Data Mining To Improve Student Retention In Higher Education – A Case Study*. 2007.
- [13] J. N. BIN HAMID, "Modelling Academic Achievement with Respect to Self-Readiness of Public Universities.," Universiti Utara Malaysia.
- [14] S. SAVAŞ, N. TOPALOĞLU, and M. YILMAZ, "Veri madenciliği ve Türkiye'deki uygulama örnekleri," *İstanbul Ticaret Üniversitesi Fen Bilim. Derg.*, no. 21, pp. 1–23, 2012.
- [15] A. OĞUZLAR, "Veri Ön İşleme," *Erciyes Üniversitesi İktisadi ve İdari Bilim. Fakültesi Derg.*, no. 21, pp. 67–76, 2004.
- [16] W. G. BAXT, "Use of an Artificial Neural Network for Data Analysis in Clinical Decision-Making: The Diagnosis of Acute Coronary Occlusion," *Neural Computation*, vol. 2, pp. 480–489, 1990.
- [17] W. G. BAXT, "Use of an artificial neural network for the diagnosis of myocardial infarction.," *Ann. Intern. Med.*, vol. 115, pp. 843–8, 1991.
- [18] H. B. BURKE, "Artificial neural networks for cancer research: Outcome prediction," *Semin. Surg. Oncol.*, vol. 10, no. 1, pp. 73–79, Jan. 1994.
- [19] H. B. BURKE, P. H. GOODMAN, D. B. ROSEN, D. E. HENSON, J. N. WEINSTEIN, F. E. HARRELL, J. R. MARKS, D. P. WINCHESTER, and D. G. BOSTWICK, "Artificial neural networks improve the accuracy of cancer survival prediction.," *Cancer*, vol. 79, pp. 857–862, 1997.
- [20] I. GUYON, "Applications Of Neural Networks To Character Recognition," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 05, no. 01n02, pp. 353–382, Jun. 1991.
- [21] S. KNERR, L. PERSONNAZ, and G. DREYFUS, "Handwritten digit recognition by neural networks with single-layer training.," *IEEE Trans. Neural Netw.*, vol. 3, pp. 962–968, 1992.
- [22] G. L. MARTIN and J. A. PITTMAN, "Recognizing Hand-Printed Letters and Digits Using Backpropagation Learning," *Neural Comput.*, vol. 3, no. 2, pp. 258–267, Jun. 1991.

- [23] H. BOURLARD and N. MORGAN, "Continuous speech recognition by connectionist statistical methods," *IEEE Trans. Neural Netw.*, vol. 4, no. 6, pp. 893–909, Jan. 1993.
- [24] R. P. LIPPMANN, "Review of Neural Networks for Speech Recognition," *Neural Comput.*, vol. 1, no. 1, pp. 1–38, Mar. 1989.
- [25] E. I. ALTMAN, G. MARCO, and F. VARETTO, "Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience)," *J. Bank. Financ.*, vol. 18, no. 3, pp. 505–529, May 1994.
- [26] R. C. LACHER, P. K. COATS, S. C. SHARMA, and L. F. FANT, "A neural network for classifying the financial health of a firm," *Eur. J. Oper. Res.*, vol. 85, no. 1, pp. 53–65, Aug. 1995.
- [27] M. LESHNO and Y. SPECTOR, "Neural network prediction analysis: The bankruptcy case," *Neurocomputing*, vol. 10, no. 2, pp. 125–147, Mar. 1996.
- [28] K. Y. TAM and M. Y. KIANG, "Managerial Applications of Neural Networks: The Case of Bank Failure Predictions," *Manage. Sci.*, vol. 38, no. 7, pp. 926–947, Jul. 1992.
- [29] R. L. WILSON and R. SHARDA, "Bankruptcy prediction using neural networks," *Decision Support Systems*, vol. 11, pp. 545–557, 1994.
- [30] G. ZHANG, M. Y. HU, B. E. PATUWO, and D. C. INDRO, "Artificial neural networks in bankruptcy prediction: general framework and cross-validation analysis," *Eur. J. Oper. Res.*, vol. 116, pp. 16–32, 1999.
- [31] G. P. ZHANG, "Neural networks for classification: a survey," *IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev.)*, vol. 30, 2000.
- [32] KRAIPEERAPUN PAWALAI, "Neural network classification based on quantification of uncertainty," Murdoch University, 2009.
- [33] K. KAYAER and T. YILDIRIM, "Medical Diagnosis on Pima Indian Diabetes Using General Regression Neural Networks," pp. 78–81.
- [34] Y. SUN, F. KARRAY, and S. AL-SHARHAN, "Hybrid soft computing techniques for heterogeneous data classification," *2002 IEEE World Congr. Comput. Intell. 2002 IEEE Int. Conf. Fuzzy Syst. FUZZ-IEEE'02. Proc. (Cat. No.02CH37291)*, vol. 2, 2002.
- [35] P. JEATRAKUL and K. W. WONG, "Comparing the performance of different neural networks for binary classification problems," *2009 Eighth Int. Symp. Nat. Lang. Process.*, 2009.
- [36] E. ÖZTEMEL, *Yapay Sinir Ağları*, 2nd ed. Papatya Yayıncılık, 2006.

- [37] H. AL NAGEIM, R. NAGAR, and P. J. G. LISBOA, “Comparison of neural network and binary logistic regression methods in conceptual design of tall steel buildings,” *Constr. Innov. Information, Process. Manag.*, vol. 7, no. 3, pp. 240–253, 2007.
- [38] J. A. BULLINARIA, “Radial Basis Function Networks: Introduction,” 2004.
- [39] G. KUMAŞOĞLU and B. BOLAT, “Yapay Sinir Ağlarıyla Müzikal Tür Tanıma,” 2011.
- [40] Ö. KIŞI and S. AFŞAR, “Yapay Sinir Ağı ve Bulanık-Yapay Sinir Ağı Yöntemleri Kullanılarak Tava Buharlaşıma Tahmini,” *Tarım Bilim. Araştırma Derg.*, vol. 3, no. 1, pp. 45–51, 2010.
- [41] F. HEIMES and B. VAN HEUVELN, “The normalized radial basis function neural network,” in *SMC’98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No.98CH36218)*, vol. 2, pp. 1609–1614.
- [42] Q.-J. ZHANG, K. C. GUPTA, and V. K. DEVABHAKTUNI, “Artificial neural networks for RF and microwave design — from theory to practice,” *IEEE Trans. Microw. Theory Tech.*, vol. 51, pp. 1339–1350, 2003.
- [43] TAKAGI TOMOHIRO and SUGENO MICHIO, “Fuzzy Identification of Systems and Its Applications to Modeling and Control,” 1985.
- [44] R. BABUSKA and H. VERBRUGGEN, “Neuro-fuzzy methods for nonlinear system identification,” *Annu. Rev. Control*, vol. 27 I, pp. 73–85, 2003.
- [45] B. LAZZERINI, “Sugeno-Type Fuzzy Inference,” 2009. .
- [46] A. KAUR and A. KAUR, “Comparison of Mamdani-Type and Sugeno-Type Fuzzy Inference Systems for Air Conditioning System,” *Int. J. Soft Comput. Eng.*, vol. 2, no. 2, 2012.
- [47] M. A. ÇAKIROĞLU, E. ERENOĞLU, S. KASAP, and Y. EKİZ, “Çelik Lif Katkılı Betonların Tahribatsız Deney Yöntemleriyle Elde Edilen Basınç Dayanımının ANFIS Metoduyla Tahmini,” vol. 3, no. 3, pp. 14–22, 2011.
- [48] D. G.S.V.P. RAJU, V. MARY SUMALATHA, K. V. RAMANI, and K. V. LAKSHMI, “Solving Uncertain Problems using ANFIS,” *Int. J. Comput. Appl.*, vol. 29, no. 11, pp. 14–21, Sep. 2011.
- [49] M. ÖZTÜRK, M. HANÇER, M. S. ULUTAŞ, M. CAN, M. TARTIK, A. GÜNEL, U. ŞAHİN, A. D. ŞAHİN, AND D. E. AKYÜZ, “Bulanik Mantik Hesaplamalarına Dayalı Binalarda Isi Kayıp-Kazanç Yaklaşımı,” 2010.

- [50] “Sugeno Fuzzy Models,” 2013. [Online]. Available: <http://www.bindichen.co.uk/post/AI/takagi-sugeno-fuzzy-model.html>. Erişim Tarihi:05.03.2014
- [51] J.-S. R. JANG, “Input selection for ANFIS learning,” *Proc. IEEE 5th Int. Fuzzy Syst.*, vol. 2, 1996.
- [52] E. MADUKO, “Development and testing of a neuro-fuzzy classification system for ios data in asthmatic children .”
- [53] Ö. DEMİREL, A. KAKİLLİ, and M. TEKTAŞ, “Anfis Ve Arma Modelleri İle Elektrik Enerjisi,” vol. 25, no. 3, pp. 601–610, 2010.
- [54] K. POLAT and S. GÜNEŞ, “An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease,” *Digital Signal Processing*, vol. 17. pp. 702–710, 2007.
- [55] M. A. YURDUSEV and M. FIRAT, “Adaptive neuro fuzzy inference system approach for municipal water consumption modeling: An application to Izmir, Turkey,” *J. Hydrol.*, vol. 365, pp. 225–234, 2009.
- [56] A. NOURELDIN, A. EL-SHAFIE, AND M. REDA TAHA, “Optimizing neuro-fuzzy modules for data fusion of vehicular navigation systems using temporal cross-validation,” *Eng. Appl. Artif. Intell.*, vol. 20, pp. 49–61, 2007.
- [57] I. GÜLER and E. D. UBEYLI, “Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients.,” *J. Neurosci. Methods*, vol. 148, pp. 113–121, 2005.
- [58] S.-P. LO, “The Application of an ANFIS and Grey System Method in Turning Tool-Failure Detection,” *The International Journal of Advanced Manufacturing Technology*, vol. 19. pp. 564–572, 2002.
- [59] M. A. BOYACIOGLU and D. AVCI, “An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: The case of the Istanbul stock exchange,” *Expert Syst. Appl.*, vol. 37, pp. 7908–7912, 2010.
- [60] U. ÇAYDAŞ, A. HASÇALIK, AND S. EKICI, “An adaptive neuro-fuzzy inference system (ANFIS) model for wire-EDM,” *Expert Syst. Appl.*, vol. 36, pp. 6135–6139, 2009.
- [61] J.-S. R. JANG and C.-T. S. C.-T. SUN, “Neuro-fuzzy modeling and control,” *Proc. IEEE*, vol. 83, 1995.
- [62] M. A. SHOOREHDELI, M. TESHNEHLAB, A. K. SEDIGH, and M. A. KHANESAR, “Identification using ANFIS with intelligent hybrid stable learning algorithm approaches and stability analysis of training methods,” *Appl. Soft Comput.*, vol. 9, no. 2, pp. 833–850, Mar. 2009.

- [63] M. KUMAR and D. P. GARG, "Intelligent Learning Of Fuzzy Logic Controllers Via Neural Network And," 2004, pp. 1–8.
- [64] F. M. F. MASCIOLI, G. M. VARAZI, and G. MARTINELLI, "Constructive algorithm for neuro-fuzzy networks," *Proc. 6th Int. Fuzzy Syst. Conf.*, vol. 1, 1997.
- [65] J.-S. R. JANG, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Trans. Syst. Man. Cybern.*, vol. 23, 1993.
- [66] L. C. YING and M. C. PAN, "Using adaptive network based fuzzy inference system to forecast regional electricity loads," *Energy Convers. Manag.*, vol. 49, pp. 205–211, 2008.
- [67] A. SENGUR, "Wavelet transform and adaptive neuro-fuzzy inference system for color texture classification," *Expert Syst. Appl.*, vol. 34, pp. 2120–2128, 2008.
- [68] E. D. ÜBEYLI, "Adaptive neuro-fuzzy inference system employing wavelet coefficients for detection of ophthalmic arterial disorders," *Expert Syst. Appl.*, vol. 34, no. 3, pp. 2201–2209, Apr. 2008.
- [69] M. BURAGOHAİN and C. MAHANTA, "A novel approach for ANFIS modelling based on full factorial design," *Appl. Soft Comput. J.*, vol. 8, pp. 609–625, 2008.
- [70] H. ÖZKAN, "K-Means Kümeleme Ve K-Nn Siniflandırma Algoritmalarının Öğrenci Notları Ve Hastalık Verilerine Uygulanması," 2013.
- [71] H. ARSLAN, "Sakarya üniversitesi Web Sitesi Erişim Kayıtlarının Web Madenciliği İle Analizi," *Fen Bilim. Enstitüsü Derg.*, 2008.
- [72] G. G. EMEL and Ç. TAŞKIN, "Veri Madenciliğinde Karar Ağaçları Ve Bir Satış Analizi Uygulanması," *Eskişehir Osmangazi Üniversitesi Sos. Bilim. Derg.*, 2005.
- [73] Y. MA, "Data Warehousing, OLAP, And Data Mining: An Integrated Strategy For Use At FAA."
- [74] C. BOUNSAYTHIP and E. R. RUNSALA, "Overview of Data Mining for Customer Behavior Modeling," 2001.
- [75] M. PAL, "Ensemble Learning with Decision Tree for Remote Sensing Classification," *Proc. World Acad. Sci. Eng. Technol. Vol 26 Parts 1 2 December 2007*, vol. 26, pp. 735–737, 2007.

- [76] M. ZORMAN, V. PODGORELEC, P. KOKOL, M. PETERSON, M. SPROGAR, AND M. OJSTERSEK, "Finding the right decision tree's induction strategy for a hard real world problem," *Int. J. Med. Inform.*, vol. 63, pp. 109–121, 2001.
- [77] S. L. SALZBERG, "C4.5: Programs for Machine Learning," 1994, pp. 235–240.
- [78] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, vol. 5. 1984, pp. 95–96.
- [79] J. Mingers, "An Empirical Comparison of Pruning Methods for Decision Tree Induction," vol. 243, pp. 227–243, 1989.
- [80] A. S. ALBAYRAK and Ş. KOLTAN YILMAZ, "Veri Madenciliği: Karar Ağacı Algoritmaları Ve İmkb Verileri Üzerine Bir Uygulama," *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilim. Fakültesi Derg.*, vol. 14, pp. 31–52, 2009.
- [81] Y. ÖZKAN, *Veri madenciliği yöntemleri*. Papatya, 2008.
- [82] A. FIĞLALI and O. ENGİN, "Akış Tipi Çizelgeleme Problemlerinin Genetik Algoritma Yardımı ile Çözümünde Uygun Çaprazlama Operatörünün Belirlenmesi," *Doğuş Üniversitesi Derg.*, 2002.
- [83] G. G. EMEL and Ç. TAŞKIN, "GENETİK ALGORİTMALAR ve UYGULAMA ALANLAR," *Uludağ Üniversitesi İktisadi ve İdari Bilim. Fakültesi Dergis*, vol. XXI, no. 1, pp. 129–152, 2002.
- [84] Ü. TAŞ, "Fizyolojik sistemlerin yapay zeka teknikleri kullanılarak modellenmesi ve kontrolü için eğitim amaçlı bir simülasyon tasarımı," Marmara Üniversitesi, 2001.
- [85] A. F. G. SEBER AND A. J. LEE, *Linear regression analysis*. 2003.
- [86] F. GÜLTEKİN, "REGRESYON ANALİZİ," Balıkesir.
- [87] A. A. NAJAH, A. EL-SHAFIE, O. A. KARIM, AND O. JAAFAR, "Water quality prediction model utilizing integrated wavelet-ANFIS model with cross-validation," *Neural Comput. Appl.*, vol. 21, no. 5, pp. 833–841, Nov. 2010.
- [88] N. R. DRAPER AND H. SMITH, *Applied Regression Analysis (Wiley Series in Probability and Statistics)*. 1998, p. 736.
- [89] B. EFRON, "Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation," *J. Am. Stat. Assoc.*, vol. 78, p. 316, 1983.

- [90] P. BURMAN, E. CHOW, AND D. NOLAN, “A cross-validators method for dependent data,” 1994.
- [91] P. HALL, “Large Sample Optimality Of Least Squares Cross-Validation in Density Estimation ,” *Ann. Stat.*, vol. 11, 1983.
- [92] Ş. G. ÖĞÜDÜCÜ, “Veri Madenciliği Temel Sınıflandırma Yöntemleri.”
- [93] A. MOORE AND J. SCHNEIDER, “Cross Validation.” 2011.
- [94] R. GUTIERREZ-OSUNA, “Cross Validation.”
- [95] Ö. KELEŞOĞLU and A. FIRAT, “Tuğla Duvardaki ve Tesisattaki Isı Kaybının Yapay Sinir Ağları İle Belirlenmesi,” *Fırat Üniv. Fen ve Müh. Bil. Der.* , pp. 133–141, 2006.
- [96] M. AKTAS and H. İ. OKUMUS, “Doğrudan Moment Kontrollü Asenkron Motorun Stator Direncinin Yapay Sinir Ağı ile Kestirimi,” 2003, pp. 189–191.

EKLER

EK A. Betimleyici İstatistik Tabloları

BSM429:

Descriptive Statistics

	Mean	Std. Deviation	N
BSM429	3,94	1,767	192
Cinsiyet	1,29	,453	192
OD6	2,0349	,52890	192
BSM103	4,21	1,810	192
BSM101	3,72	1,577	192
BSM102	4,68	1,671	192
BSM210	3,59	1,602	192
MAT217	3,52	1,856	192
IST108	3,26	1,782	192

BSM421:

Descriptive Statistics

	Mean	Std. Deviation	N
BSM421	2,874	1,4827	183
Cinsiyet	1,251	,4350	183
OrtalamaDonem6	2,1051	,59292	183
BSM102	4,443	1,7402	183
BSM103	4,038	1,9082	183
BSM207	3,743	1,5600	183
BSM210	3,399	1,5225	183
MAT113	3,798	1,7313	183

BSM435:**Descriptive Statistics**

	Mean	Std. Deviation	N
BSM435	4,533	1,7948	137
BSM304	4,693	1,4929	137
BSM305	4,701	1,5921	137

BSM425:**Descriptive Statistics**

	Mean	Std. Deviation	N
BSM425	2,395	1,2317	81
BSM103	3,840	2,0091	81
BSM207	3,654	1,4677	81
BSM306	4,074	1,6642	81
MAT217	3,358	1,9125	81
Cinsiyet	1,198	,4006	81
OrtalamaDonem6	2,1757	,65815	81

EK B. Ağırlık Sensörü Sınıfı.

```

function [yaprakSayisi, hata] = AgacSec(egitimgirdi, egitimcikti)
hata = 0;
yaprakSayisi = 1;
leafs = 1:1:100;
MaxLeaf = numel(leafs);
err = zeros(MaxLeaf,1);
for i=1:MaxLeaf
    treeTwoing = treefit(egitimgirdi,
egitimcikti,'splitmin',i,'method','classification','splitcriterion','twoing','names',{'Cinsiy
et' 'OD6' 'BSM210' 'MAT217'});
    %N = numel(testgirdi);
    yfit = treeval(treeTwoing, egitimgirdi);
    fark = egitimcikti-yfit;
    foldAdet = numel(egitimgirdi);
    tmp = sum(fark.*fark)/foldAdet;
    err(i) = tmp;
    if tmp<hata
        hata=tmp;
        yaprakSayisi = i;
    end
end
plot(leafs,err);
xlabel('Min Leaf Size');
ylabel('error');
end

```

ÖZGEÇMİŞ

Sümeyye KAYNAK, 1989 yılında Sakarya'da doğdu. İlköğrenimini ve orta öğrenimini Sakarya'da tamamladı. 2007 yılında Sakarya Anadolu Lisesi'nden mezun oldu. 2012 yılında Sakarya Üniversitesi Bilgisayar Mühendisliği Bölümünü'nden bölüm 3.sü olarak mezun oldu. 2008 ve 2012 yılları arasında Sakarya Üniversitesi Uzaktan Eğitim Araştırma ve Uygulama Merkezinde öğrenci asistan olarak çalışmıştır. 2012 yılında Sakarya Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar ve Bilişim Mühendisliği'nde yüksek lisans eğitimine başlamıştır. 2012 yılının Ekim ayından itibaren Sakarya Üniversitesi Fen Bilimleri Enstitüsü'nde araştırma görevlisi olarak çalışmaktadır.