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

SAMSUN İLİNİN METEOROLOJİK DEĞİŞKENLERİNİN BETA REGRESYON VE YAPAY SİNİR AĞLARI İLE KARŞILAŞTIRMALI ANALİZİ

Yüksek Lisans Tezi

Abdallah K. A. BALAHA

Danışman

Prof. Dr. Pelin KASAP

SAMSUN
2025

TEZ KABUL VE ONAYI

Abdallah K.A. BALAHA tarafından, **Prof. Dr. Pelin KASAP** danışmanlığında hazırlanan **"SAMSUN İLİNİN METEOROLOJİK DEĞİŞKENLERİNİN BETA REGRESYON VE YAPAY SİNİR AĞLARI İLE KARŞILAŞTIRMALI ANALİZİ"** başlıklı bu çalışma, jürimiz tarafından 18.7.2025 tarihinde yapılan sınav sonucunda oy birliği ile başarılı bulunarak Yüksek Lisans Tezi olarak kabul edilmiştir.

	Unvanı Adı Soyadı **Üniversitesi** **Ana Bilim/Ana Sanat Dalı**	**Sonuç**
Başkan	Prof. Dr. Erol TERZİ Ondokuz Mayıs Üniversitesi İstatistik Ana Bilim Dalı	☒ Kabul ☐ Ret
Üye	Prof. Dr. Pelin KASAP (Danışman) Ondokuz Mayıs Üniversitesi İstatistik Ana Bilim Dalı	☒ Kabul ☐ Ret
Üye	Dr. Öğr. Üyesi Erdinç YÜCESOY Ordu Üniversitesi Matematik Ana Bilim Dalı	☒ Kabul ☐ Ret

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

Prof. Dr. Faik Ahmet SESLİ
Enstitü Müdürü

BİLİMSEL ETİĞE UYGUNLUK BEYANI

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

Etik Kurul Gerekli mi?

Evet ☐ (Gerekli ise ekler kısmına ekleyiniz)

Hayır ☒

18 /07 / 2025
Abdallah K. A. BALAHA

TEZ ÇALIŞMASI ÖZGÜNLÜK RAPORU BEYANI

Tez Başlığı: SAMSUN İLİNİN METEOROLOJİK DEĞİŞKENLERİNİN BETA REGRESYON VE YAPAY SİNİR AĞLARI İLE KARŞILAŞTIRMALI ANALİZİ

Yukarıda başlığı belirtilen tez çalışması için şahsım tarafından 18/08/2025 tarihinde intihal tespit programından alınmış olan özgünlük raporu sonucunda;

Benzerlik oranı : % 9

Tek kaynak oranı : % 3 çıkmıştır.

18 /08 / 2025
Prof. Dr. Pelin KASAP

ÖZET

BETA REGRESYON VE YAPAY SİNİR AĞLARI İLE SAMSUN İLİNİN METEOROLOJİK DEĞİŞKENLERİNİN KARŞILAŞTIRMALI ANALİZİ

Abdallah K. A. BALAHA
Ondokuz Mayıs Üniversitesi
Lisansüstü Eğitim Enstitüsü
İSTATİSTİK ANA BİLİM DALI
Yüksek Lisans, Temmuz/2025
Danışman: Prof. Dr. Pelin KASAP

Bu tez, Ocak 2008-Ocak 2025 tarihleri arasında Samsun Meteoroloji Genel Müdürlüğü'nden temin edilen meteorolojik veriler kullanılarak Samsun ili nem seviyelerinin modellenmesinde beta regresyon, kuadratik beta regresyon ve ileri beslemeli sinir ağları (FNN) yöntemlerinin karşılaştırmalı analizini sunmaktadır. Bağımlı değişken olan nem (%) ile ortalama sıcaklık, maksimum sıcaklık, minimum sıcaklık, ortalama işba sıcaklık, ortalama atmosfer basıncı, maksimum atmosfer basıncı, minimum atmosfer basıncı, denizin halı, güneşleme süresi, maksimum yatay görüş mesafesi, minimum yatay görüş mesafesi, ortalama bulut kapalılığı, toplam yağış manuel, rüzgâr yönü ve rüzgâr hızı arasındaki ilişkiler incelenmiştir. Nem değerlerinin sınırlı aralığı (0-100%) olması ve doğrusal olmayan ilişkileri nedeniyle geleneksel regresyon tekniklerinin yetersiz kaldığı tespit edilmiştir. Bu sınırlamaları aşmak için beta regresyon ve kuadratik beta regresyon modelleri uygulanmış ve sırasıyla geriye doğru seçim yöntemiyle 13 ve 20 anlamlı değişken belirlenmiştir. Adaptif moment tahmini optimizasyon algoritması ile 70 epoch boyunca eğitilen bir FNN mimarisi ile ikili çapraz entropi (BCE) kaybına dayalı ileriye doğru seçimle 12 etkin değişken seçilmiştir. MSE, MAE, BCE, AIC, R^2_{psuedo} performans metrikleri ve artık grafikleri üzerinden yapılan değerlendirmelerde her yöntemin avantaj ve sınırlılıkları ortaya konmuştur. FNN modeli, eğitim verisinde 0.9731 R^2_{psuedo} değeri ve 0.5753 BCE kaybı ile üstün performans sergileyerek karmaşık doğrusal olmayan ilişkileri yakalama yeteneğini kanıtlamıştır. FNN test verisinde ise beta regresyon modelleri ile benzer genelleme başarısı göstermiştir. Ancak, kuadratik beta regresyon da, beta regresyonu geride bırakmıştır. Artık analizleri tüm modellerde düşük yanlılık ve varyansı doğrularken, FNN için artık grafiğinin sıfıra daha yakın olduğu dikkat çekmiştir. Çalışma, meteorolojik modellemede doğrusal olmayan makine öğrenimi tekniklerinin etkinliğini vurgularken, yorumlanabilirlik (beta regresyon) ve tahmin gücü (FNN) arasındaki dengeye dikkat çekmektedir. Bulgular, tarım, enerji yönetimi ve iklim direnci planlaması gibi alanlarda nem tahmin doğruluğunu artırmaya yönelik değerli bilgiler sunmaktadır. Sonuçlar, gerçek dünya meteorolojik sorunlarını ele alırken istatistiksel doğruluk ve hesaplama karmaşıklığını dengeleyen bağlam odaklı model seçimini önermektedir.

Anahtar Sözcükler: Meteorolojik değişkenler, Beta regresyon, Kuadratik beta regresyon, İleri beslemeli sinir ağları, İkili çapraz-entropi, Adaptif moment tahmini algoritması

ABSTRACT

COMPARATIVE ANALYSIS OF METEOROLOGICAL VARIABLES OF SAMSUN PROVINCE WITH BETA REGRESSION AND ARTIFICIAL NEURAL NETWORKS

Abdallah K. A. BALAHA
Ondokuz Mayıs University
Institute of Graduate Studies
Department of Statistics
Master, July/2025
Supervisor: Prof. Dr. Pelin KASAP

This thesis presents a comparative analysis of beta regression, quadratic beta regression, and Feed-Forward Neural Networks (FNN) for modeling humidity levels in Samsun, Turkey, using meteorological data obtained from Samsun General Directorate of Meteorology between January 2008 and January 2025. The relationships between the dependent variable humidity (%) and mean temperature, maximum temperature, minimum temperature, mean dewpoint temperature, mean atmospheric pressure, maximum atmospheric pressure, minimum atmospheric pressure, sea carpet, sunshine duration, maximum horizontal visibility, minimum horizontal visibility, mean cloud cover, total precipitation, wind direction, and wind speed were investigated. Traditional regression techniques were deemed inadequate due to humidity's bounded nature (0–100%) and nonlinear relationships with predictors. Beta egression and quadratic beta regression models were employed to address these limitations, with backward selection identifying 13 and 20 significant variables, respectively. An FNN architecture, optimized using the adaptive moment estimation algorithm (ADAM) over 70 epochs, utilized forward selection to determine 12 impactful variables based on binary cross-entropy (BCE) loss. Diagnostic tools, including residual plots and performance metrics (MSE, MAE, BCE, AIC, R^2_{pseudo}), revealed advantages and limitations of each approach. The FNN achieved superior on the train data with a R^2_{pseudo} of 0.9731 and BCE loss of 0.5753, demonstrating its ability to capture complex nonlinear interactions. On the test data, FNN demonstrated similar generalization success to beta regression models. However, quadratic beta regression also outperformed beta regression. Residual analysis confirmed minimal bias and variance across all models, though FNN residuals where more closer to zero. The study highlights the effectiveness of nonlinear machine learning techniques in meteorological modeling, highlighting the trade-off between interpretability (beta regression) and predictive power (FNN). These findings provide valuable insights for improving humidity prediction accuracy, with applications in agriculture, energy management, and climate resilience planning. The results suggest context-driven model selection that balances statistical accuracy and computational complexity when addressing real-world meteorological problems.

Keywords: Meteorological variables, Beta regression, Quadratic beta regression, Feed-forward neural Networks, Binary cross-entropy, Adaptive moment estimation algorithm

ÖN SÖZ VE TEŞEKKÜR

Burada, danışmanım Prof. Dr. Pelin KASAP'a içten teşekkürlerimi sunmak istiyorum. "Bilginizi ve kaynaklarınızı paylaştığınız için teşekkür ederim."

Ayrıca, değerlendirmeleri için jüri üyelerine de teşekkür etmek istiyorum.

Abdallah K. A. BALAHA

İÇİNDEKİLER

TEZ KABUL VE ONAYI ..... **i**
BİLİMSEL ETİĞE UYGUNLUK BEYANI ..... **ii**
TEZ ÇALIŞMASI ÖZGÜNLÜK RAPORU BEYANI ..... **ii**
ÖZET ..... **iii**
ABSTRACT ..... **iv**
ÖNSÖZ VE TEŞEKKÜR ..... **v**
İÇİNDEKİLER ..... **vi**
KISALTMALAR ..... **vii**
ŞEKİLLER DİZİNİ ..... **viii**
TABLOLAR DİZİNİ ..... **ix**
1. GİRİŞ **1**
2. MATERYAL VE YÖNTEM **5**
2.1. Beta Regresyonu 5
2.2. İleri Beslemeli Yapay Sinir Ağları 9
2.3. Doğruluk Metrikleri ve Doğrulama Araçları 15
3. UYGULAMA **19**
3.1. Veri Kapsamı 19
3.2. Veri İşleme 21
3.3. Beta Regresyon Analizi 24
3.4. Kuadratik Beta Regresyon Analizi 29
3.5. İleri Beslemeli Sinir Ağı Analizi 34
3.6. Yöntemlerin Karşılaştırılması 40
4. SONUÇLAR **44**
KAYNAKLAR **45**
ÖZ GEÇMİŞ **47**

KISALTMALAR

MLE	: En Çok Olabilirlik Tahmini
LL	: Log-Olabilirlik Fonksiyonu
AIC	: Akaike Bilgi Kriteri
DF	: Serbestlik Derecesi
FNN	: İleri Beslemeli Sinir Ağı
ReLU	: Düzeltici Lineer Birim Fonksiyonu
ELU	: Üssel Lineer Birim Fonksiyonu
MSE	: Hata Kareler Ortalaması
MAE	: Ortalama Mutlak Hata
BCE	: İkili Çapraz Entropi
ADAM	: Adaptif Moment Tahmin Algoritması
CART	: Sınıflandırma ve Regresyon Ağaçları
PMM	: Tahmine Dayalı Ortalama Eşleştirme
RF	: Rastgele Ormanlar

ŞEKİLLER DİZİNİ

Şekil 2.1. Dört katmandan oluşan temel FNN. 9
Şekil 2.2. Aktivasyon fonksiyonlarının grafikleri. 11
Şekil 2.3. FNN ağırlıklarının başlatılması. 12
Şekil 2.4. Her bir ağırlığın ilgili hatasını geri yayma. 13
Şekil 2.5. Bir geri yayılım iterasyonundan sonra yeni ağırlıklar. 14
Şekil 2.6. Artıklar için bir grafik örneği. 18
Şekil 3.1. Nem verilerinin dağılımı. 20
Şekil 3.2. Nem verilerinin kutu grafiği. 20
Şekil 3.3. Nem ile ilgili parametreler arasındaki saçılım grafikleri. 24
Şekil 3.4. Beta regresyon modeli için artık grafiği. 28
Şekil 3.5. Kuadratik beta regresyon modeli için artık grafiği. 33
Şekil 3.6. FNN ağı. 35
Şekil 3.7. Farklı kayıp kriterleri ile model performansının karşılaştırılması. 36
Şekil 3.8. 50 epoch boyunca BCE kaybının yakınsaması. 36
Şekil 3.9. Nem verisinde FNN modeli için eğitim algoritmasına ait akış şeması. 37
Şekil 3.10. FNN modeline ait artık grafiği. 38
Şekil 3.11. FNN'nin 70 epoch boyunca eğitim ve test kayıplarının yakınsaması. 39
Şekil 3.12. FNN'nin 70 epoch boyunca eğitim ve test R^2_{pseudo} değerlerinin yakınsaması. 39

TABLOLAR DİZİNİ

Tablo 2.1. En iyi bilinen bazı bağ fonksiyonları. 8
Tablo 2.2. Yaygın olarak kullanılan aktivasyon fonksiyonları. 10
Tablo 2.3. ADAM optimizasyon algoritması adımları ve parametreleri. 15
Tablo 3.1. Nem verilerinin tanımlayıcı istatistikleri. 19
Tablo 3.2. Meteorolojik verilerdeki değişkenlere göre eksik değer sayısı..... 21
Tablo 3.3. Beta regresyon ve FNN modelleri için değişken bazında uygulanan eksik veri doldurma yöntemleri. 22
Tablo 3.4. Doldurma işlemi yapıldıktan sonra gözlem verilerinin tanımlayıcı istatistikleri.. 23
Tablo 3.5. Çeşitli beta regresyon modelleri için AIC ve R^2_{pseudo} değerleri. 25
Tablo 3.6. Cloglog fonksiyonuna bağlı beta regresyon katsayıları. 25
Tablo 3.7. Model seçimi sonrası cloglog fonksiyonuna bağlı beta regresyon katsayıları..... 27
Tablo 3.8. ϕ katsayıları (Hassasiyet modeli, kimlik bağ ile). 28
Tablo 3.9. Beta model performansının eğitim ve test verileri üzerindeki değerlendirmesi. .. 28
Tablo 3.10. Çeşitli kuadratik beta regresyon modelleri için AIC ve R^2_{pseudo} değerleri. 31
Tablo 3.11. Model seçimi sonrası cloglog fonksiyonuna bağlı kuadratik beta regresyon katsayıları. 32
Tablo 3.12. ϕ katsayısıları (kimlik bağlı hassasiyet modeli). 33
Tablo 3.13. Kuadratik beta regresyon model performansının eğitim ve test verileri üzerindeki değerlendirmesi..... 34
Tablo 3.14. FNN model performansının eğitim ve test verileri üzerindeki değerlendirmesi. 40
Tablo 3.15. Eğitim veri setinde beta regresyon, kuadratik beta regresyon ve FNN modellerinin performans karşılaştırması. 41
Tablo 3.16. Test veri setinde beta regresyon, kuadratik beta regresyon ve FNN modellerinin performans karşılaştırması..... 42
Tablo 3.17. Ortalama nem tahmin performansının beta regresyon, kuadratik beta regresyon ve FNN modelleri için karşılaştırılması..... 43

1. GİRİŞ

Gerçek dünyadaki birçok uygulamada, yanıt değişken belirli bir aralıkla sınırlı olabilmektedir. Olasılıklar veya oranlar gibi değişkenler genellikle 0 ile 1 arasında değer alır ve bu tür sınırlı yapılar klasik regresyon modellerinin temel varsayımlarını ihlal edebilir. Örneğin, bir seçimde belirli bir adayın oy oranını tahmin etmek ya da bir otelde dolu odaların yüzdesini modellemek gibi senaryolarda, yanıt değişkeninin doğal sınırları göz önünde bulundurulmalıdır. Klasik doğrusal regresyon, normallik ve homojen varyans varsayımlarına dayandığından, yanıt değişken belirli bir aralıkta olduğunda uygun tahminler üretmekte yetersiz kalabilmektedir. İstatistikçiler, bu tür değişkenleri modellemek için daha karmaşık regresyon teknikleri geliştirmişlerdir. Bu modeller, genellikle yanıt değişkendeki sınırlamaları dikkate almakta ve tahminlerin geçerliliğini artırmaktadır. Bu bağlamda, lojistik regresyon, probit regresyon gibi modeller önemli bir yere sahiptir. Lojistik regresyon, bir hastalığın olup olmaması, bir ürünün satın alınıp alınmaması gibi yanıt değişkenin iki olası sonucu olduğu durumlarda kullanılmaktadır. Lojistik regresyon modelinde, yanıt değişken için tahmin edilen olasılık, lojistik fonksiyon kullanılarak 0 ve 1 arasında sınırlandırılmaktadır. Bu, tahmin edilen değerlerin mantıklı ve geçerli olmasını sağlamaktadır (Hosmer vd., 2013). Probit regresyon da, yanıt değişkenin olasılığı, standart normal dağılımın kümülatif dağılım fonksiyonu (CDF) ile sınırlandırılmaktadır. Probit ve lojistik modeller arasındaki ana fark, olasılıkların modellenmesinde kullanılan fonksiyonların farklı olmasıdır (Cameron ve Trivedi, 2005).

Oran türündeki sürekli veriler için ise beta regresyon daha uygun bir modeldir. Ferrari ve Cribari-Neto (2004) tarafından önerilen bu yöntem, bağımlı değişkenin (0,1) aralığında sürekli dağıldığı ve beta dağılımına sahip olduğu varsayımına dayanır. Beta dağılımı, iki parametre aracılığıyla farklı şekiller alabilen esnek bir dağılımdır. Cribari-Neto ve Zeileis (2010) tarafından geliştirilen betareg R-paketi, oransal verilerin çarpıklığını ve değişen varyans gibi özelliklerini dikkate alan tanı ve tahmin araçları sunmaktadır. Bu model, hem teorik hem de uygulamalı çalışmalarda geniş yer bulmuştur. Örneğin, Cribari-Neto (2023), Brezilya'da COVID-19 ölüm oranını analiz etmek için beta regresyonu kullanarak, kentleşme ve gelirin ölüm oranlarının ortalamasını belirlemede önemli olduğunu, hekim yoğunluğunun ise hassasiyet

parametresini etkilediğini göstermiştir. Benzer şekilde, Ünlü ve Aktaş (2017), Türkiye illeri için refah endekslerini modellemek amacıyla beta regresyonu kullanmışlardır.

Kuadratik beta regresyon modeli ise, bağımsız değişkenlerin ikinci dereceden terimlerini ekleyerek doğrusal olmayan ilişkileri daha iyi temsil etmeyi amaçlamaktadır. Bu genişletilmiş model, özellikle karmaşık yapılı oransal verilerde öngörü gücünü artırmaktadır.

Diğer taraftan, doğrusal olmayan ilişkileri tanımlamak için makine öğrenme algoritmaları da giderek daha fazla önem kazanmıştır. Geleneksel lineer regresyon modelleri yalnızca değişkenler arasındaki lineer ilişkileri öğrenmekle sınırlı iken, bir makine öğrenme algoritması olan ileri beslemeli sinir ağları (FNN), verilerdeki karmaşık ve lineer olmayan desenleri kavramada avantajlı bir yöntemdir. Lau ve arkadaşları (2019) tarafından yapılan çalışmaya göre, lineer regresyon modelleri yalnızca düz çizgiler veya düzlemlerle uyum sağlayabilirken, sinir ağları sınıflandırma veya kümeleme gibi lineer olmayan şekillere sahip sorunları ele alabilmektedir. Li ve Wang (2019) tarafından yapılan çalışmaya göre, sinir ağları, herhangi bir sürekli türevlenebilir fonksiyonu yaklaşık olarak modelleyebilmekte, bu da herhangi bir sorunun düzgün bir eğri olarak ifade edilebileceği anlamına gelmektedir. Baker ve Richards (1999)'a göre, sinir ağları, veriler için lineer, lojistik veya polinomial olmak üzere en iyi modele dinamik olarak uyum sağlayabilmektedir. Schmidt-Hieber (2020) tarafından ReLU aktivasyon fonksiyonu ile derin sinir ağları kullanılarak parametrik olmayan regresyonun gerçekleştirilmesine dair bir çalışma yapılmıştır. FNN'ler, ReLU veya sigmoid gibi aktivasyon fonksiyonları aracılığıyla giriş özelliklerinin katmanlı dönüşümlerinden yararlanarak konut fiyat tahmini ve borsa eğilim analizi gibi durağan desen tanıma görevlerinde üstün performans sergilemektedir. Hornik (1991), FNN'lerin yeterli sayıda gizli birim sağlandığında herhangi bir sürekli fonksiyonu yaklaşık olarak ifade edebileceğini kanıtlayarak bu ağların teorik temelini ortaya koymuştur. Xu ve Zhang (2021), FNN'lerin istatistiksel desen tanıma alanındaki yararlarını özellikle konut fiyatlarının tahmininde göstererek, bu modellerin emlak verilerindeki karmaşık ve doğrusal olmayan ilişkileri başarıyla işleyebildiğini ortaya koymuştur. Abiy ve arkadaşları (2022) ise FNN'leri iklim modellemesinde kullanarak Florida körfezindeki tuzluluk seviyelerini tahmin etmede başarılı bir şekilde uygulamış ve bu ağların çevresel verilerdeki karmaşık yapılarla başa çıkma yeteneklerini göstermişlerdir.

Nem modellemesi, meteorolojik değişkenlerin karmaşık etkilerini anlamaya yönelik önemli bir adımdır. Bu nedenle, birçok araştırmacı nem tahmininde daha hassas ve güvenilir modeller geliştirmeye yönelik çeşitli teknikler kullanmışlardır. Rezaei Melal vd. (2024), kapalı bir süs bitkisi serasındaki nem oranını tahmin etmek üzere yedi makine öğrenmesi algoritmasının heterojen topluluk öğrenmesinin kümelenmesine dayalı bir metamodel geliştirmişlerdir. İran'daki bu süs bitkisi serasından elde edilen 600.000'in üzerinde sensör kaydıyla eğitilen model, $R^2 = 0.965$ değeriyle önceki yaklaşımları geride bırakarak topluluk tekniklerinin tekil modellere kıyasla üstünlüğünü ortaya koymuştur. Jung vd. (2022), domates seralarında buharlaşma-terleme ve bağıl nemin tahmini için "Convolutional Neural Network - Long Short-Term Memory (CNN-LSTM)" mimarisine dayalı derin öğrenme modeli geliştirmiştir. Sensör ve kontrolör verileriyle eğitilen model, geleneksel yöntemlere kıyasla üstün performans sergileyerek nem yönetimi ve sulama stratejilerinde duyarlılık sağlamıştır.

Bu tezde, Samsun Meteoroloji müdürlüğünden alınan Ocak 2008-Ocak 2025 tarihleri arasındaki meteorolojik veriler kullanılarak, Samsun ilindeki nem yüzdesi değişkeninin modellenmesi amaçlanmıştır. Bu amaçla, beta regresyon, kuadratik beta regresyon ve ileri beslemeli sinir ağları (FNN) yöntemleri arasında karşılaştırmalı bir analiz yapılmıştır. Bağımlı değişken olarak nem oranı (%) ele alınmış; ortalama sıcaklık, maksimum sıcaklık, minimum sıcaklık, ortalama işba sıcaklık, ortalama atmosfer basıncı, maksimum atmosfer basıncı, minimum atmosfer basıncı, denizin halı, güneşleme süresi, maksimum yatay görüş mesafesi, minimum yatay görüş mesafesi, ortalama bulut kapalılığı, toplam yağış manuel, rüzgâr yönü ve rüzgâr hızı olmak üzere toplam 15 açıklayıcı değişken ile ilişkileri analiz edilmiştir. Nem değişkeninin %(0–100) sınırlı aralıkta yer alması ve açıklayıcı değişkenlerle doğrusal olmayan ilişkiler göstermesi nedeniyle klasik regresyon yönteminin kısıtlı kaldığı değerlendirilmiş ve bu sınırlamaları aşmak amacıyla beta regresyon ile kuadratik beta regresyon modelleri tercih edilmiştir. Bu modellerde değişken seçimi, geriye doğru eleme yöntemiyle gerçekleştirilmiş ve modelin anlamlı değişkenleri belirlenmiştir. FNN model, adaptif moment tahmini (ADAM) optimizasyon algoritması ile eğitilmiş ve değişken seçimi, ikili çapraz entropi (BCE) kayıp fonksiyonu temelinde ileriye doğru seçim yöntemiyle yapılmıştır. Tüm modeller; ortalama mutlak hata (MAE), hata kareler ortalaması (MSE), Akaike bilgi kriteri (AIC), ikili çapraz entropi (BCE) gibi

istatistiksel değerlendirme ölçütleri ve artık grafikleri aracılığıyla yöntemsel olarak karşılaştırılmıştır.

Bu tez çalışması dört bölümden oluşmaktadır. İkinci bölüm, temel kavramlar ve teorik altyapıyı içermektedir. Üçüncü bölüm, veri setinin kapsamı, veri işleme, uygulama ve karşılaştırma analizlerine ayrılmıştır. Dördüncü bölümde ise sonuçlar yer almaktadır.

2. MATERYAL VE YÖNTEM

Bu bölümde, beta regresyon ve ileri beslemeli sinir ağları konularının teorik altyapıları ve modellerin değerlendirilmesi için kullanılan ölçütler ve doğrulama araçları verilmektedir.

2.1. Beta Regresyonu

Bir değişkenin üstel dağılım ailelerine uygun olması için, aşağıdaki denklemde belirtilen fonksiyonel formda ifade edilebilir olması gerekmektedir:

$$f(y|\theta) = e^{\eta(\theta)t(y) - C(\theta))}h(y) \tag{2.1}$$

Yukarıdaki denklemdeki $\eta(\theta)$ ve $C(\theta)$ fonksiyonları, dağılımın parametrelerine bağlı iken, $t(y)$ ve $h(y)$ fonksiyonları değişkene bağlıdır. Üstel dağılım ailelerine uygun bir değişkenin olabilirlik fonksiyonu aşağıdaki şekilde ifade edilmektedir:

$$L(\theta|y) = e^{\eta(\theta)\sum_{i=1}^{n} t(y_i) - nC(\theta))} \prod_{i=1}^{n} h(y_i) \tag{2.2}$$

Beta dağılımı, üstel dağılım ailesinin bir üyesidir. (0,1) aralığında olan ve pozitif olası değerlere sahip rastgele değişkenlerin davranışını modellemek için kullanılan çok yönlü bir dağılımdır. Oranlar ve yüzdeler beta dağılımı ile modellenmektedir. Beta dağılımı, beta fonksiyonu kavramını içermektedir. α ve β pozitif reel sayılar olsun. $B(\alpha, \beta)$ olarak tanımlanan beta fonksiyonu aşağıdaki şekilde tanımlanmaktadır:

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx \tag{2.3}$$

Gamma ve beta fonksiyonları arasındaki ilişki ise

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \tag{2.4}$$

şeklinde ifade edilir. Burada

$$\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx \tag{2.5}$$

Gamma fonksiyonunu göstermektedir.

Tanım 2.1.1. Bir X rastgele değişkeni aşağıdaki beta olasılık yoğunluk fonksiyonuna sahipse,

$$f(x) = \begin{cases} \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & 0 < x < 1 \\ 0 & , \quad d.d. \end{cases} \tag{2.6}$$

o zaman X bir beta rastgele değişkeni olmaktadır ve $X \sim B(\alpha,\beta)$ olarak gösterilmektedir. X rastgele değişkeni $B(\alpha,\beta)$ dağılımına sahipse, o zaman X rastgele değişkeninin beklenen değeri ve varyansı aşağıdaki şekilde bulunmaktadır:

$$E(X) = \frac{\alpha}{\alpha+\beta} \tag{2.7}$$

$$Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{2.8}$$

Ferrari ve Cribari-Neto'nun (2004) çalışmasında tanıtılan beta regresyon modelleri sınıfı, $(0,1)$ açık standart birim aralığında değerler alan y sürekli değişkenini modellemek için kullanışlıdır. Eğer yanıt değişken (a,b) aralığında $(a < b)$ sayısal değerler alıyorsa, bu durumda y'yi doğrudan modellemek yerine $\frac{y-a}{b-a}$ modellemesi yapılır.

α ve β parametreli beta dağılımının olasılık yoğunluk fonksiyonu (2.6) denkleminde verilmiştir. Ancak beta regresyon, genellikle bağımlı değişkenin ortalamasını (μ) modellemek için daha kullanışlıdır. Ayrıca modeli ϕ hassasiyet parametresi içerecek şekilde tanımlamak da yaygın olarak kullanılır. Bu amaçla, Ferrari ve Cribari-Neto (2004), bir ϕ hassasiyet parametresi ile birlikte bağımlı değişkenin ortalaması μ için bir regresyon yapısı elde etmek amacıyla $\mu = \frac{\alpha}{\alpha+\beta}$ ve $\phi = \alpha + \beta$ şeklinde farklı bir parametrelendirme önermişlerdir. Yeni parametrelendirme ile Y'nin olasılık yoğunluk fonksiyonu aşağıdaki şekilde ifade edilmektedir:

$$f(y;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma\big((1-\mu)\phi\big)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} \qquad \textbf{(2.9)}$$

Burada, $0 < \{\mu, y\} < 1$ ve $\phi > 0$ dir ve beta dağılımının beklenen değeri $E(y) = \mu$ ve varyansı $Var(y) = \frac{\mu(1-\mu)}{\phi+1}$ olmaktadır. Sabit μ için, ϕ arttıkça y'nin varyansı azalmaktadır. ϕ^{-1} yayılım parametresi olduğu için ϕ'ye hassasiyet parametresi denilmektedir.

Beta dağılımının üstel dağılım ailesi içinde yer almasının uygunluğu, (2.1) eşitliğindeki fonksiyonel formun ifade edilmesiyle aşağıdaki gibi yazılmaktadır:

$$f(y;\mu,\phi) = e^{\mu\phi\ln(y) - \ln\left(\frac{\Gamma(\mu\phi)}{\Gamma(\phi)}\right)} \frac{(1-y)^{(1-\mu)\phi}}{y(1-y)\Gamma\big((1-\mu)\phi\big)} \qquad \textbf{(2.10)}$$

Beta regresyonunun genel fikri, bağ fonksiyonu (g) kullanılarak $(0,1)$ aralığındaki değerleri reel sayıların sürekli aralığına çevrilmesiyle açıklanmaktadır. Bu alanda, verilerimizin beta dağılımına uyduğu varsayılmaktadır ve bu varsayıma dayanarak regresyon gerçekleştirilmektedir. $(0,1)$ aralığında katsayıları doğru yorumlamak için, verileri bağ fonksiyonunun tersi ile yeniden dönüştürmek gerekmektedir. Regresyon reel sayı alanında gerçekleştirilip ardından sınırlı alana uyarlanarak, model tarafından üretilen tüm tahminlerin (0,1) aralığında kalması sağlanmaktadır.

Öncelikle, her bir veri noktası için yanıt değişken y'nin ortalaması μ olan bir beta dağılımı ile temsil edildiği varsayılarak regresyon modeli kurulmaktadır. Daha sonra, μ aşağıdaki şekilde tanımlanmaktadır:

$$g(\mu_i) = \sum_{j=1}^{m} \beta_j x_{ij} \qquad i = 1,2,\dots,n \qquad \textbf{(2.11)}$$

Burada, $\beta = (\beta_1, \beta_2, \dots, \beta_m)^\top$ bilinmeyen regresyon parametrelerinin bir vektörüdür, $x = (x_{i1}, \dots, x_{im})$ vektörü bağımsız değişkenleri; g ise standart birim aralığında $(0,1)$ monoton ve çift türevlenebilir bir bağ fonksiyonudur. En iyi bilinen bağ fonksiyonlarından bazıları Tablo 2.1.'de verilmiştir (Dünder ve Cengiz, 2020).

Tablo 2.1. En iyi bilinen bazı bağ fonksiyonları.

Bağ Fonksiyonu	Formül
Log	$\eta = \ln \mu$
Logit	$\eta = \ln\left(\frac{\mu}{1-\mu}\right)$
Loglog	$\eta = -\ln(-\ln(\mu))$
Cloglog	$\eta = \ln(-\ln(1-\mu))$
Probit	$\Phi^{-1}(\mu)$
Cauchit	$\eta = \tan(\pi(\mu - 0.5))$

$\Phi^{-1}(\cdot)$, standart normal dağılımın kümülatif dağılım fonksiyonunun tersidir.

Daha sonra, Cribari-Neto ve Zeileis (2010) tarafından önerilen MLE yöntemi kullanılarak log-olabilirlik fonksiyonu maksimize edilerek regresyon parametreleri tahmin edilmektedir. Beta regresyon analizi için elde edilen log-olabilirlik fonksiyonu aşağıdaki şekilde ifade edilmektedir:

$$LL(\mu, \phi) = \sum_{i=1}^{n} LL(\mu_i, \phi) \qquad \textbf{(2.12)}$$

Burada

$$LL(\mu_i, \phi) = LL\left(g^{-1}\left(\sum_{j=1}^{m} \beta_j X_{ij}\right), \phi\right) =$$

$$\ln\big(\Gamma(\phi)\big) - \ln\big(\Gamma(\mu_i\phi)\big) - \ln\Big(\Gamma\big((1-\mu_i)\phi\big)\Big) + (\mu_i\phi - 1)\ln(y_i) + \big((1-\mu_i)\phi - 1\big)\ln(1-y_i)$$

şeklindedir. **(2.13)**

Denklem 2.11'e göre, kuadratik beta regresyon, doğrusal değişkenlerin kareleri gibi uygun dönüşümler eklenerek kolayca uygulanabilmektedir; bu genişletme

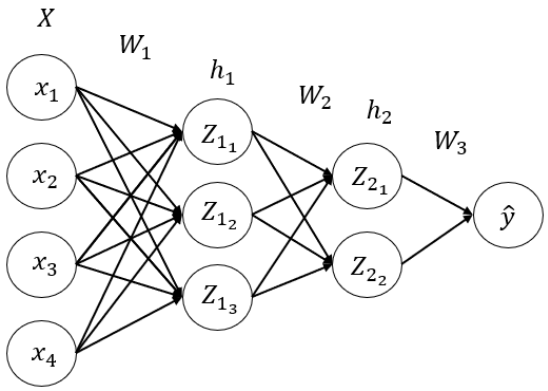
değişkenler arasındaki eğrisel ilişkileri daha iyi yakalamaya olanak sağlar ve modelin tahmin performansını artırabilmektedir.

2.2. İleri Beslemeli Yapay Sinir Ağları

İleri beslemeli yapay sinir ağları (FNN'ler), insan beyninin işlevlerini öğrenerek yeni bilgiler oluşturmak suretiyle taklit etmektedir. Zurada (1992) tarafından yapılan çalışmaya göre, biyolojik sinir ağlarını taklit eden sentetik ağlar olup, yapı ve yetenekler açısından önemli ölçüde farklılık göstermektedir. Zhang vd. (1998) tarafından yapılan çalışmaya göre, FNN'lerin tahmindeki uygulamaları incelenmiş ve bunların uyum sağlama ve lineer olmayan yeteneklerine dikkat çekilmiştir. Bu çalışmada, FNN'lerin potansiyeli vurgulanırken, yapısal optimizasyon ve deneysel doğrulama konusundaki zorlukları da göz önünde bulundurulmaktadır. FNN'lerin temel unsurları ve tanımları aşağıdaki bölümlerde tartışılacaktır.

2.2.1. Mimari yapısı

Çok katmanlı FNN, giriş, gizli ve çıkış katmanlarından oluşmakta ve birbirine bağlı nöronlar içermektedir (bkz. Şekil 2.1). Her bir katmandaki nöron sayısı, ağın yapısını tanımlamaktadır. Aynı katmandaki nöronlar doğrudan bağlantılı değildir, bunun yerine ağırlıklarla bağlanmaktadır.



Şekil 2.1. Dört katmandan oluşan temel FNN.

2.2.2. Aktivasyon fonksiyonu (σ)

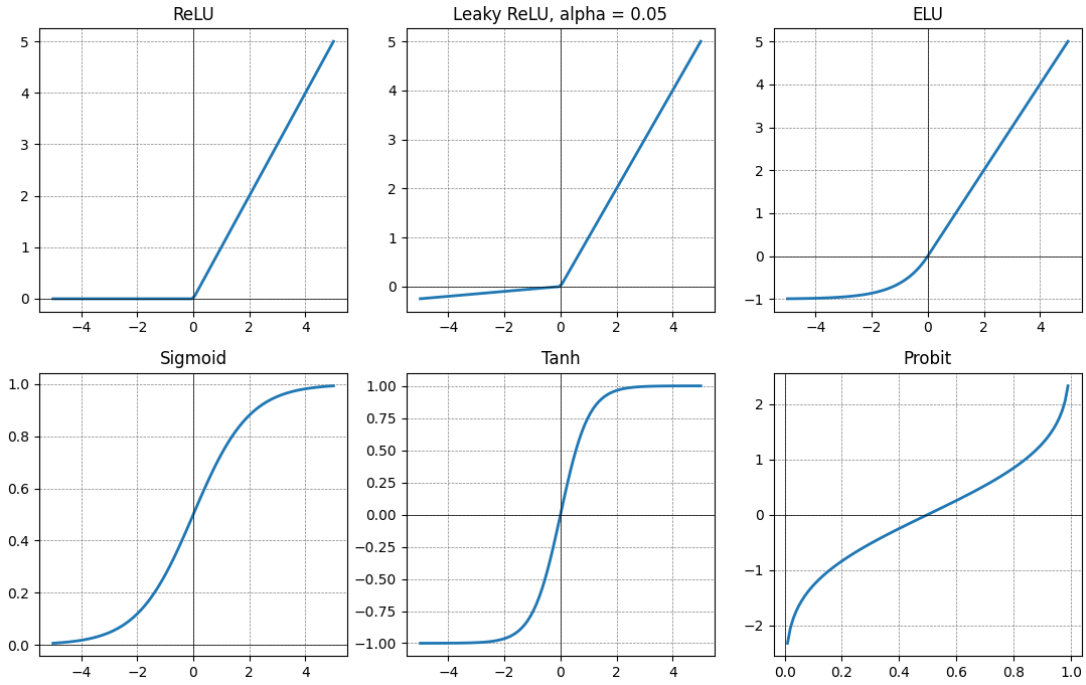
Bu fonksiyon, girişler ve çıkışlar arasındaki haritalamaları belirlemekte ve ağın performansını etkilemektedir. Uygun bir aktivasyon fonksiyonunun seçilmesi, örneğin lineer veya polarize, ağın davranışını etkilemektedir. Lineer olmayan aktivasyonlarda, istenilen çıktı değerlerine ulaşmak için eğim parametresinin belirlenmesi kritik öneme sahiptir. Eğim parametresi, aktivasyon fonksiyonunun herhangi bir noktada ne kadar dik olduğunu belirlemektedir. Eğer eğim dik ise, fonksiyonun girdideki değişimlere güçlü bir şekilde tepki verdiği anlamına gelmektedir. Buna karşılık, daha yumuşak bir eğim daha kademeli bir tepkiyi göstermektedir.

Tablo 2.1'de belirtilen cloglog, probit ve log-log gibi fonksiyonlar, Gomes vd. (2011) tarafından açıklandığı üzere, sinir ağlarının performansını artırmak için kullanılır.

Tablo 2.2. Yaygın olarak kullanılan aktivasyon fonksiyonları.

Aktivasyon Fonksiyonu	**Formül**
ReLU	$max(0, x)$
Leaky ReLU	$max(\alpha x, x), \alpha \in \mathbb{R}$
ELU	$\begin{cases} x, & x \geq 0 \\ \alpha(e^x - 1), & x < 0 \end{cases}, \alpha \in \mathbb{R}$
Sigmoid	$\frac{1}{1 + e^{-x}}$
Tanh	$\tanh x$

Tablo 2.2, literatürde yaygın olarak kullanılan aktivasyon fonksiyonlarının bir listesini sunmaktadır ve bu fonksiyonların ilgili grafikler Şekil 2.2'de gösterilmektedir (Dubey vd., 2022).



Şekil 2.2. Aktivasyon fonksiyonlarının grafikleri.

2.2.3. Öğrenme algoritması

Sinir ağlarındaki ağırlıkları belirlemek için çeşitli algoritmalar kullanılmaktadır ve bunlar arasında en yaygın kullanılanlardan biri, Hecht-Nielsen (1989) tarafından tanıtılan geri yayılım öğrenme algoritmasıdır. Bu algoritma, bir FNN'nin gerçek ve tahmin edilen çıktıları arasındaki farklara dayanarak ağırlıkları güncellemektedir. Geri yayılım algoritmasındaki öğrenme parametresi, istenilen çıktılara ulaşmada önemli bir rol oynar ve algoritmanın yürütülmesi sırasında dinamik olarak ayarlanabilmektedir.

Şekil 2.1'de açıklanan FNN'de geri dönüldüğünde, geri yayılım algoritması aşağıdaki ardışık adımlarla gerçekleştirilmektedir (Hecht-Nielsen, 1989):

1. İleri Besleme İşlemi: Aşağıdaki formül kullanılarak giriş değişkenleri üzerinde ileri besleme işlemi gerçekleştirilir:

$$\hat{y} = \sigma\left(W_3 * \sigma\left(W_2 * \sigma(W_1 * X)\right)\right) \quad \textbf{(2.14)}$$

2. Hata Karşılaştırması: Çıktı, uygun bir hata fonksiyonuna ε (örneğin, BCE) dayalı olarak istenen çıktı ile karşılaştırılır.

3. Hata Hesaplama: Her bir ağırlıkla ilişkili hata hesaplanır. Bunu başarmanın bir yolu, hata fonksiyonu ε'nin her bir ağırlığa göre gradyanlarının hesaplandığı gradyan iniş algoritmasının kullanılmasıdır:

$$\nabla\varepsilon = \left(\frac{\partial\varepsilon}{\partial w_{1,1}^{(1)}}, \frac{\partial\varepsilon}{\partial w_{1,2}^{(1)}}, \ldots, \frac{\partial\varepsilon}{\partial w_{2,1}^{(3)}} \right) \tag{2.15}$$

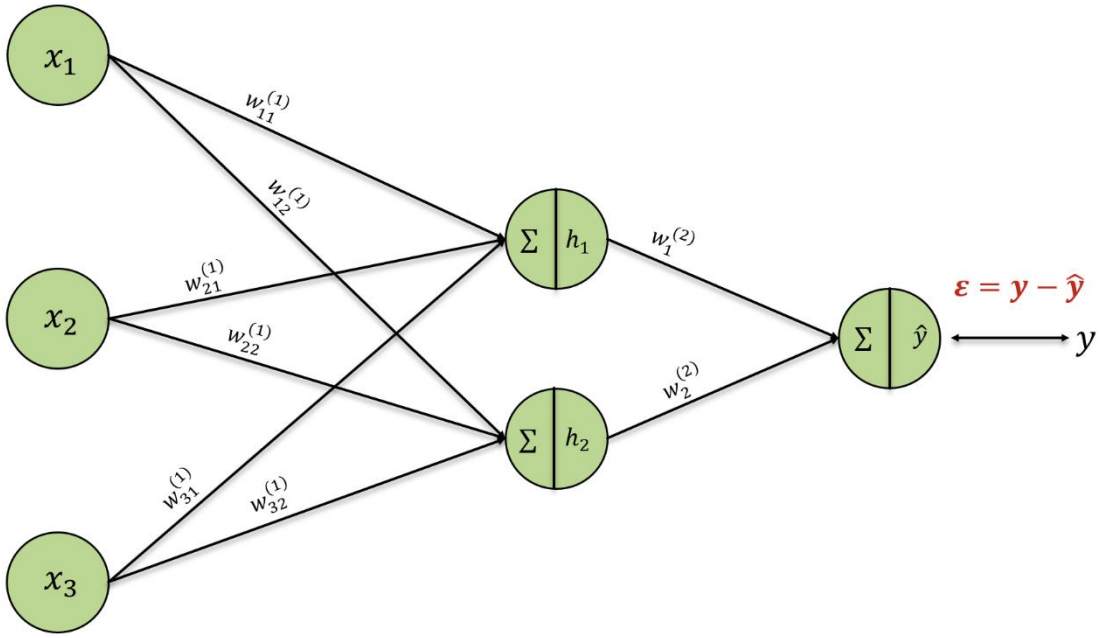
4. Ağırlık Güncelleme: Hatayı ağırlıklara dağıtmak ve ağırlıkları aşağıdaki formülü kullanarak güncellemek için geri besleme işlemi gerçekleştirilir:

$$w_{ij}^{(k)} \leftarrow w_{ij}^{(k)} - \alpha \frac{\partial\varepsilon}{\partial w_{ij}^{(k)}} \tag{2.16}$$

Burada α öğrenme hızı sabiti ve k katman indeks numarasıdır.

5. Tatmin edici bir model elde edilene kadar 1-4 adımları tekrarlanır.

Bir sonraki örnekte, hatayı ağ boyunca geri yayarak ağırlıkların güncellenmesi süreci açıklanmaktadır. Bu süreçte, bazı görseller sunulmaktadır ve ağın nasıl eğitildiği adım adım gösterilmektedir:



Şekil 2.3. FNN ağırlıklarının başlatılması.

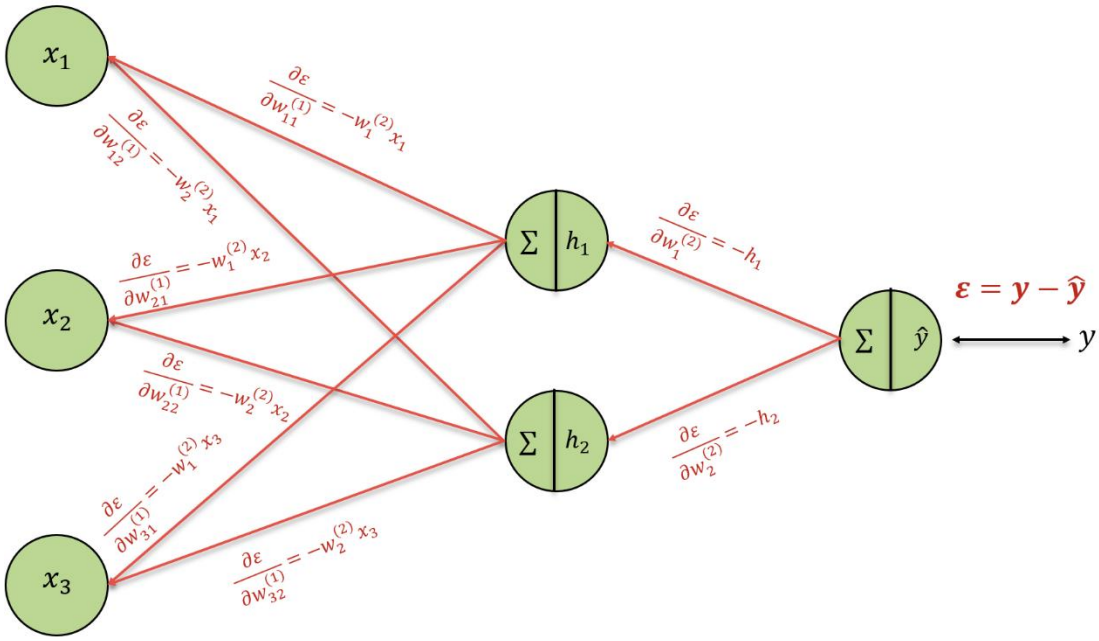
Şekil 2.3'de gösterildiği gibi, üst simge $(*)$ katman numarasını belirtmektedir. Bu örnekte, birden fazla girdi, bir gizli katman ve bir çıktı bulunmaktadır. Ağırlıklar

rastgele değerlerle başlatıldıktan sonra, ileri besleme yayılımı aşağıdaki matris çarpımıyla gerçekleştirilmektedir:

$$\hat{y} = \begin{pmatrix} w_1^{(2)} & w_2^{(2)} \end{pmatrix}_{1\times 2} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}_{2\times 1}$$

$$= \begin{pmatrix} w_1^{(2)} & w_2^{(2)} \end{pmatrix}_{1\times 2} \left(\begin{pmatrix} w_{11}^{(1)} & w_{21}^{(1)} & w_{31}^{(1)} \\ w_{12}^{(1)} & w_{22}^{(1)} & w_{32}^{(1)} \end{pmatrix}_{2\times 3} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}_{3\times 1} \right)_{2\times 1} \quad \textbf{(2.17)}$$

Ardından, kayıp fonksiyonu hesaplandıktan sonra hata ağ boyunca geri yayılmaktadır. Daha önce belirtildiği gibi, hatanın geri yayılmasının bir yolu, Şekil 2.4'de açıklandığı gibi, her bir ağırlık ile ilgili olarak kayıp fonksiyonu ε'nin gradyanının hesaplanmasıdır.



Şekil 2.4. Her bir ağırlığın ilgili hatasını geri yayma.

Her türev, zincir kuralı yardımıyla kolayca hesaplanmaktadır. Örneğin, kayıp fonksiyonu ε'nin $w_{11}^{(1)}$'e göre türevi, aşağıdaki şekilde hesaplanmaktadır:

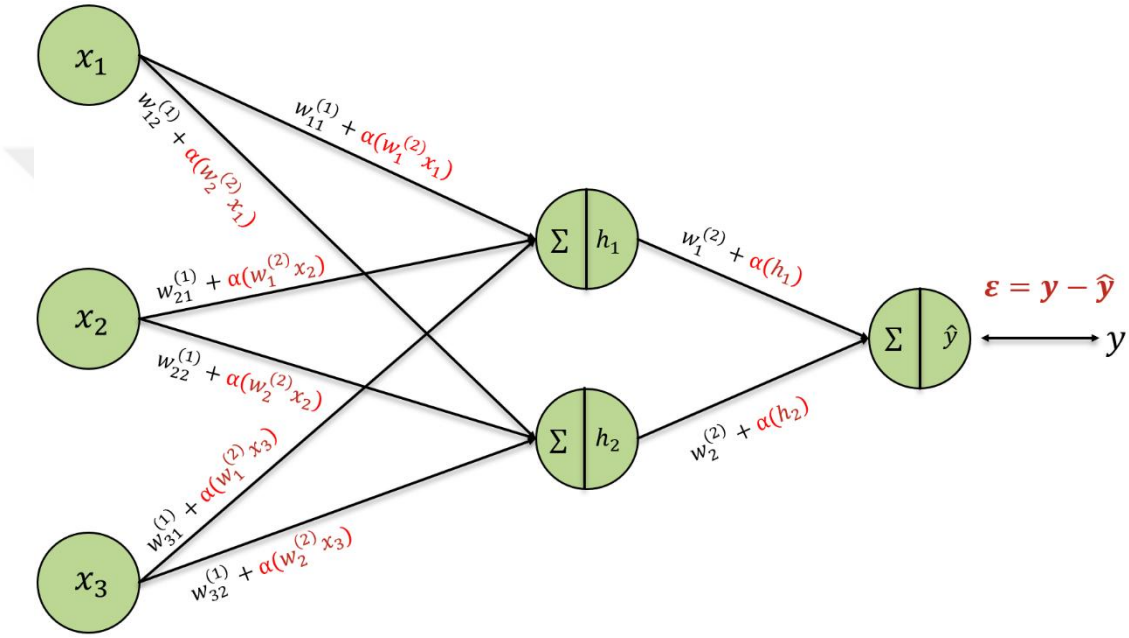
$$\frac{\partial \varepsilon}{\partial w_{11}^{(1)}} = \frac{\partial \varepsilon}{\partial h_1} \frac{\partial h_1}{\partial w_{11}^{(1)}} = \frac{\partial (y - \hat{y})}{\partial h_1} \frac{\partial \left(w_{11}^{(1)} x_1 + w_{21}^{(1)} x_2 + w_{31}^{(1)} x_3 \right)}{\partial w_{11}^{(1)}} = \left(\frac{\partial y}{\partial h_1} - \frac{\partial \hat{y}}{\partial h_1} \right) x_1$$

$$= \left(0 - \frac{\partial \left(w_1^{(2)} h_1 + w_2^{(2)} h_2 \right)}{\partial h_1} \right) x_1 = -w_1^{(2)} x_1 \quad \textbf{(2.18)}$$

Veya genel bir formül olarak:

$$\frac{\partial \varepsilon}{\partial w_{ij}^{(1)}} = -w_j^{(2)} x_i \quad ; \quad \frac{\partial \varepsilon}{\partial w_i^{(2)}} = -h_i \tag{2.19}$$

Bu gradyanları ağırlıkları güncellemek için kullanmanın bir yolu, gradyan inişi yöntemiyle, gradyanların (veya bunların bir kısmının) ilgili ağırlıklardan çıkarılmasını içermektedir. Bu işlem, belirli bir iterasyon sayısı kadar veya yakınsama sağlanana kadar tekrarlanmaktadır. Güncellenmiş ağırlıklar Şekil 2.5'de gösterilmektedir.



Şekil 2.5. Bir geri yayılım iterasyonundan sonra yeni ağırlıklar.

Her ağırlık ayarlaması, hatayı azaltan yönde ağırlıkları hareket ettirmektedir. Bu yinelemeli süreç, kayıp fonksiyonu bir minimuma ulaşana kadar devam etmektedir ve bu, ağın verilen giriş verileri için optimal ağırlıkları öğrendiğini göstermektedir.

2.2.4. Adaptif Moment Tahmini

Adaptif moment tahmini (ADAM), 2014 yılında Kingma ve Ba tarafından tanıtılan gradyan inişi tabanlı algoritmalardan biridir. Bu algoritma, stokastik amaç fonksiyonlarının birinci dereceden gradyan tabanlı optimizasyonuna odaklanmakta ve daha düşük dereceli momentlerin adaptif tahminlerini kullanmaktadır. Momentum kullanmak, yerel minimumlardan kaçınmak için faydalı görülmektedir.

Algoritmanın uygulanması kolaydır, hesaplama açısından verimlidir, az bellek gereksinimi vardır, gradyanların diyagonal yeniden ölçeklenmesine karşı değişmez ve

veri ve/veya parametreler açısından büyük olan problemler için uygundur (Kingma ve Ba, 2014).

Tablo 2.3. ADAM optimizasyon algoritması adımları ve parametreleri.

ADAM Algoritması
Girdi: α: Adım boyutu $\beta_1, \beta_2 \in [0, 1)$: Moment tahminleri için üstel azalma oranları $f(\theta)$: Parametreleri θ olan stokastik amaç fonksiyonu θ_0: Başlangıç parametre vektörü $m_0 \leftarrow 0$ (1. moment vektörünü başlat) $v_0 \leftarrow 0$ (2. moment vektörünü başlat) t ← 0 (Zaman adımını başlat)
Çıktı: θ_t (Sonuç parametreleri)
1 $\boldsymbol{\theta_t}$ **Yakınsamadıkça:**
2 $t \leftarrow t + 1$
3 $g_t \leftarrow \nabla_\theta f_t(\theta_{t-1})$ (t zaman adımında stokastik amaç fonksiyonu ile ilgili gradyanları hesapla)
4 $m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$ (Yanlı ilk moment tahminini güncelle)
5 $v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$ (Yanlı ikinci moment tahminini güncelle)
6 $\widehat{m}_t \leftarrow \frac{m_t}{1-\beta_1^t}$ (Yanlılık düzeltilmiş ilk moment tahminini hesapla)
7 $\widehat{v}_t \leftarrow \frac{v_t}{1-\beta_2^t}$ (Yanlılık düzeltilmiş ikinci moment tahminini hesapla)
8 $\theta_t \leftarrow \theta_{t-1} - \alpha \cdot \frac{\widehat{m}_t}{\sqrt{\widehat{v}_t}+\varepsilon}$ (Parametreleri güncelle)
9 Bitir

2.3. Doğruluk Metrikleri ve Doğrulama Araçları

Burada, modellerin değerlendirilmesi için temel doğruluk metrikleri ve doğrulama araçları ele alınmaktadır. Bu metrikler, MSE, MAE, BCE, AIC ve R^2_{pseudo} gibi, modelin performansı, uyumu ve karmaşıklığı hakkında fikir sunmaktadır. Ayrıca, modelleri doğrulamak ve performanslarını incelemek için tahminlere ait hataların görselleştirilmesi amacıyla artık grafikleri ele alınmaktadır.

2.3.1. Hata Kareler Ortalaması (MSE)

Hata kareler ortalaması (MSE), tahmin edilen $\hat{Y}$ değerleri ile gerçek Y değerleri arasındaki ortalama kare farkını nicelleştiren istatistiksel bir metriktir. Bir tahmincinin veya modelin kalitesini değerlendirmek için yaygın olarak kullanılmaktadır (Bickel ve Doksum, 2015).

MSE formülü aşağıdaki şekildedir:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \widehat{Y}_i\right)^2 \quad \textbf{(2.20)}$$

2.3.2. Ortalama Mutlak Hata (MAE)

Ortalama Mutlak Hata (MAE), tahmin edilen değerler ile gerçek değerler arasındaki ortalama mutlak farkı ölçmektedir. MSE'deki kare hata farkının aksine, MAE farkların karesini almaz, bu da onu daha anlaşılır kılar (Robeson ve Willmott, 2023).

MAE aşağıdaki şekilde hesaplanmaktadır:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|Y_i - \widehat{Y}_i\right| \quad \textbf{(2.21)}$$

MAE, aykırı değerlere karşı dayanıklıdır ve model performansının doğrudan bir değerlendirmesini sağlamaktadır.

2.3.3. İkili Çapraz Entropi (BCE)

BCE (Binary Cross Entropy veya log loss olarak da bilinir), ikili sınıflandırma görevlerinde yaygın olarak kullanılan bir kayıp fonksiyonudur. Tahmin edilen olasılıklar ile gerçek ikili etiketler arasındaki farklılığı değerlendirmektedir (Hurtik vd., 2022). Her bir veri noktası için, BCE, gerçek etiket y'nin (pozitif sınıf için 1, negatif sınıf için 0) tahmin edilen olasılık temelinde log olabilirlik hesaplamaktadır.

BCE'nin formülü aşağıdaki şekildedir:

$$BCE = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(\widehat{y}_i) + (1 - y_i)\log(1 - \widehat{y}_i) \quad \textbf{(2.22)}$$

BCE, tahmin edilen olasılığın gerçek etiketten uzaklaştığı durumlarda yanlış tahminler için daha yüksek kayıp üretmektedir.

Örneğin, aşağıdaki tabloda 4 adet gerçek ve tahmin edilen etiket değeri verildiği varsayılmaktadır.

Gerçek	1	0	0	1
Tahmin Edilen	0.8	0.2	0.6	0.9

Önceki tanıma göre, BCE kayıp değeri şu şekilde kolayca hesaplanacaktır: $BCE = -\frac{1}{4}(1 * \log 0.8 + (1-1) * \log(1-0.8) + 0 * \log 0.2 + (1-0) * \log(1-0.2) + 0 * \log 0.6 + (1-0) * \log(1-0.6) + 1 * \log 0.9 + (1-1) * \log(1-0.9)) \cong 0.367$

2.3.4. Akaike Bilgi Kriteri (AIC)

AIC (Akaike Information Criterion), genellikle regresyon gibi modellerin seçiminde kullanılan bir istatistiksel modelin performansını değerlendiren bir kriterdir (Akaike, 1973). Log-olabilirlik ve model parametrelerine dayanır, uyum ve karmaşıklığı ölçmektedir.

AIC'nin formülü aşağıdaki şekildedir (Bozdogan, 1987):

$$AIC(k) = -2LL + 2k \quad \textbf{(2.23)}$$

Burada, LL modelin log-olabilirlik değerini gösterir ve k parametrelerin sayısını belirtmektedir. Daha düşük bir AIC, daha iyi uyum ve daha az karmaşıklık anlamına gelir, bu da aşırı uyumu kontrol etmeye yardımcı olmaktadır.

2.3.5. Pseudo R-Kare

Ferrari ve Cribari-Neto (2004) tarafından rapor edilen R^2_{pseudo}, gözlenen yanıt değişkeni y ile modelin tahmin ettiği ortalama değer $\hat{\mu}$ üzerinde uygulanan bağ fonksiyonları $g(y)$ ve $g(\hat{\mu})$ arasındaki korelasyonun karesidir:

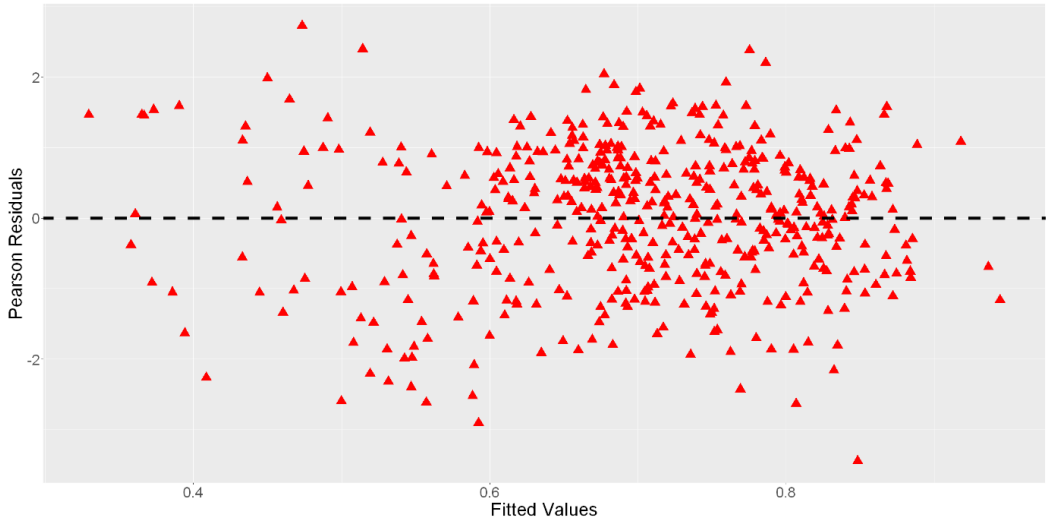
$$R^2_{pseudo} = cor(g(y), g(\hat{\mu}))^2 \quad \textbf{(2.24)}$$

Tahmin edilen değerler ile gerçek değerler arasındaki korelasyonun büyüklüğü ne kadar yüksek olursa, korelasyonun pozitif veya negatif olmasına bakılmaksızın, R^2_{pseudo} o kadar büyük olmaktadır.

2.3.6. Artık Grafiği

Regresyon modellerinin geçerliliğini değerlendirmek için artıkların grafiklerinin önemli araçlar olduğu düşünülmektedir. Model varsayımlarıyla ilgili potansiyel sorunlar belirlenmekte ve aykırı değerler tespit edilmektedir. Bu grafikler, özellikle normal olmayan hata terimlerine sahip genelleştirilmiş lineer modeller için oldukça faydalı bulunmaktadır (Beyer, 2002).

Artık grafiğinde, dikey eksende artık değerler, yatay eksende ise bağımsız değişken gösterilmektedir.



Şekil 2.6. Artıklar için bir grafik örneği.

3. UYGULAMA

3.1. Veri Kapsamı

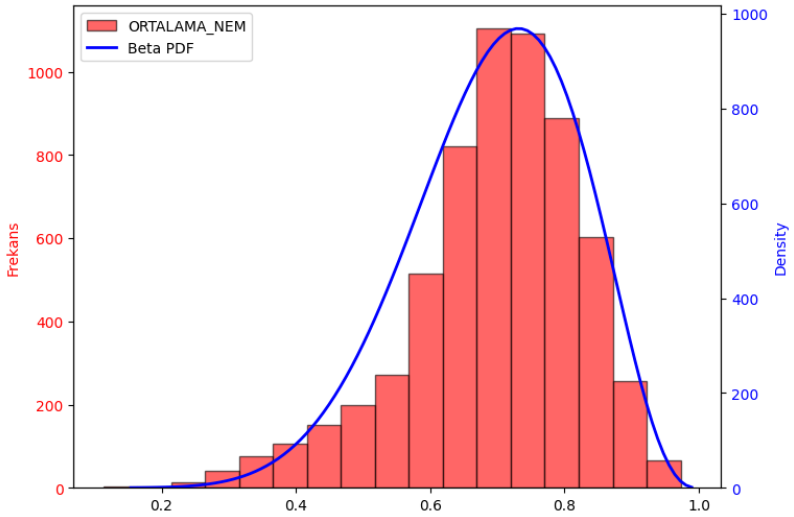
Bu çalışmada, 01 Ocak 2008 ve 01 Ocak 2025 tarihleri arasında, Samsun ilindeki Meteoroloji Genel Müdürlüğü'nden günlük hava durumu kayıtları olmak üzere 6210 veri alınmıştır. Bu kayıtlar, 15 bağımsız değişken üzerinde yapılandırılmıştır. Bu değişkenler, Ortalama sıcaklık (C°), maksimum sıcaklık (C°), minimum sıcaklık (C°), ortalama işba sıcaklık (C°), ortalama atmosfer basıncı (hPa), maksimum atmosfer basıncı (hPa), minimum atmosfer basıncı (hPa), denizin halı (m), güneşleme süresi (saat), maksimum yatay görüş mesafesi (km), minimum yatay görüş mesafesi (km), ortalama bulut kapalılığı (8Okta), toplam yağış manuel (mm), rüzgâr yönü (°) ve rüzgâr hızı (m/s) değişkenleridir. Bağımlı değişken olarak ortalama nem (%) belirlenmiştir.

Meteorolojik çalışmalarda nem değişkeninin modellenmesi, tarım, enerji ve altyapı gibi çeşitli sektörler için önemli olan nem seviyelerinin doğru bir şekilde tahmin edilmesine olanak sağlamaktadır. Bu tür modeller, yağış tahminleri, buharlaşma oranlarının anlaşılması ve küf oluşumu veya sıcaklık stresi gibi nemle ilgili sorunların öngörülmesine yardımcı olmaktadır.

Tablo 3.1, veri setindeki nem verilerinin dağılımının tanımlayıcı istatistiklerini sunmaktadır:

Tablo 3.1. Nem verilerinin tanımlayıcı istatistikleri.

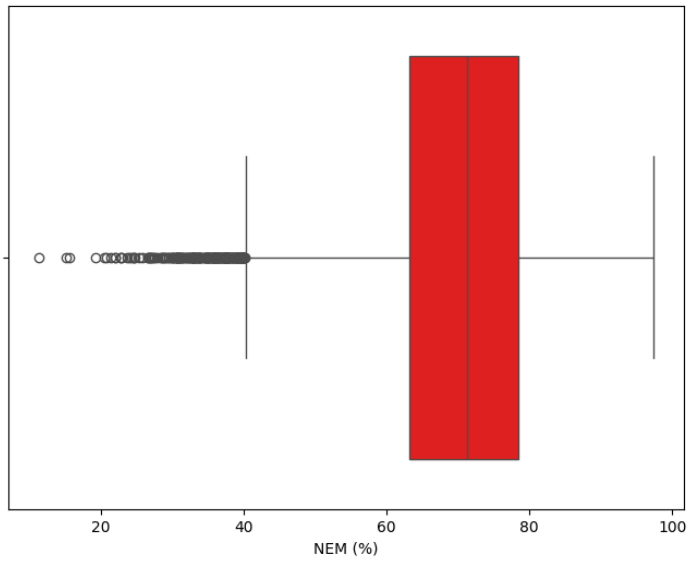
Gözlem Sayısı	6210
Ortalama	69.59
Std. Sapma	12.77
Min	11.3
%25	63.2
%50	**71.2**
%75	78.5
Maks	97.3



Şekil 3.1. Nem verilerinin dağılımı.

Verilerin, parametreleri $\alpha = 9.4567$ ve $\beta = 4.0739$ olan bir beta dağılımına uygun olduğu görülmektedir. Bu tanımlayıcı istatistikler ve dağılım grafiği ile, bu bilgilerin bir beta regresyon modeli oluşturmak için kullanılması amaçlanmaktadır.

Veri setinin çarpıklığı yaklaşık olarak -0.4257'dir ve bu, şekil 3.1 de kolayca görülebildiği gibi hafif bir negatif çarpıklık olduğunu göstermektedir. Basıklık ise yaklaşık olarak -0.1076 olup, dağılımın normal dağılıma benzer şekilde orta seviyede bir zirveye sahip olduğunu ve kuyruklarının normal dağılıma yakın olduğunu göstermektedir.



Şekil 3.2. Nem verilerinin kutu grafiği.

Nem verilerine ait yukarıdaki kutu grafiğinde, kutu grafiğinin alt bıyığının altında bazı aykırı değerler görülmektedir. Bu aykırı değerler, verilerin büyük çoğunluğundan anlamlı derecede daha düşük olan nem değerlerini temsil etmektedir.

3.2. Veri İşleme

Verilere göre, bazı değişkenlerde tutarsızlıklar ve eksik değerler tespit edilmiştir. Bu eksikliklerin giderilmesi için uygun doldurma yöntemleri uygulanmıştır.

Tablo 3.2. Meteorolojik verilerdeki değişkenlere göre eksik değer sayısı.

Değişken	**Eksik Değer Sayısı**
Minimum Sıcaklık (C°)	1
Denizin halı (m)	31
Güneşleme süresi (saat)	324
Maksimum yatay görüş mesafesi (km)	29
Minimum yatay görüş mesafesi (km)	29
Ortalama bulut kapalılığı (8Okta)	29
Toplam yağış manuel (mm)	3481
Rüzgâr yönü (°)	333
Rüzgâr hızı (m/s)	333

Bu süreçte **mice (Multivariate Imputation by Chained Equations)** paketi kullanılmış olup, farklı modeller (Beta regresyon ve FNN) için ayrı ayrı doldurma işlemleri, en iyi performansın elde edilmesi amacıyla gerçekleştirilmiştir. Aşağıdaki tabloda, her bir değişken için kullanılan doldurma yöntemi ve hangi modelde uygulandığı gösterilmektedir:

- ✓ PMM (Predictive Mean Matching) yöntemi sürekli değişkenler için tercih edilmiştir.
- ✓ RF (Random Forest) yöntemi, dağılımı daha karmaşık olan ve eksiklik oranı yüksek değişkenlerde kullanılmıştır.
- ✓ CART (Classification and Regression Trees) yöntemi ise özellikle bimodal veya dönemsel değişim içeren değişkenlerde uygulanmıştır.

- ✓ Toplam yağış manuel değişkeni, yağışsız günlerin eksiklik olarak değerlendirilmemesi nedeniyle doğrudan sıfır ile doldurulmuştur.

- ✓ Yatay görüş idx adlı yeni bir değişken eklenmiş olup, ölçüm cihazının değişim tarihini göstermektedir; bu değişkende, 0 değeri cihaz değişikliği öncesini, 1 değeri ise sonrası dönemi belirtmektedir.

- ✓ Rüzgâr yönü açısal bir değişken olduğu için, doğrudan ortalamasını almak anlamsız ve yanıltıcı sonuçlara yol açabilir. Bu nedenle, yön bilgisini trigonometrik olarak temsil eden "Sin rüzgâr yönü" ve "Cos rüzgâr yönü" değişkenleri türetilmiş ve analizlerde bu bileşenler kullanılmıştır. Bu yaklaşım, açısal verinin regresyon modellere daha uygun biçimde entegre edilmesini sağlamaktadır.

Tablo 3.3. Beta regresyon ve FNN modelleri için değişken bazında uygulanan eksik veri doldurma yöntemleri.

Değişken Adı	**Beta Regresyon**	**FNN**
Minimum Sıcaklık	PMM	PMM
Denizin halı	PMM	PMM
Güneşleme süresi	RF	RF
Maksimum yatay görüş mesafesi	CART	PMM
Minimum yatay görüş mesafesi	CART	PMM
Ortalama bulut kapalılığı	CART	PMM
Toplam yağış manuel	0	0
Sin rüzgâr yönü	PMM	PMM
Cos rüzgâr yönü	PMM	PMM
Rüzgâr hızı	RF	PMM

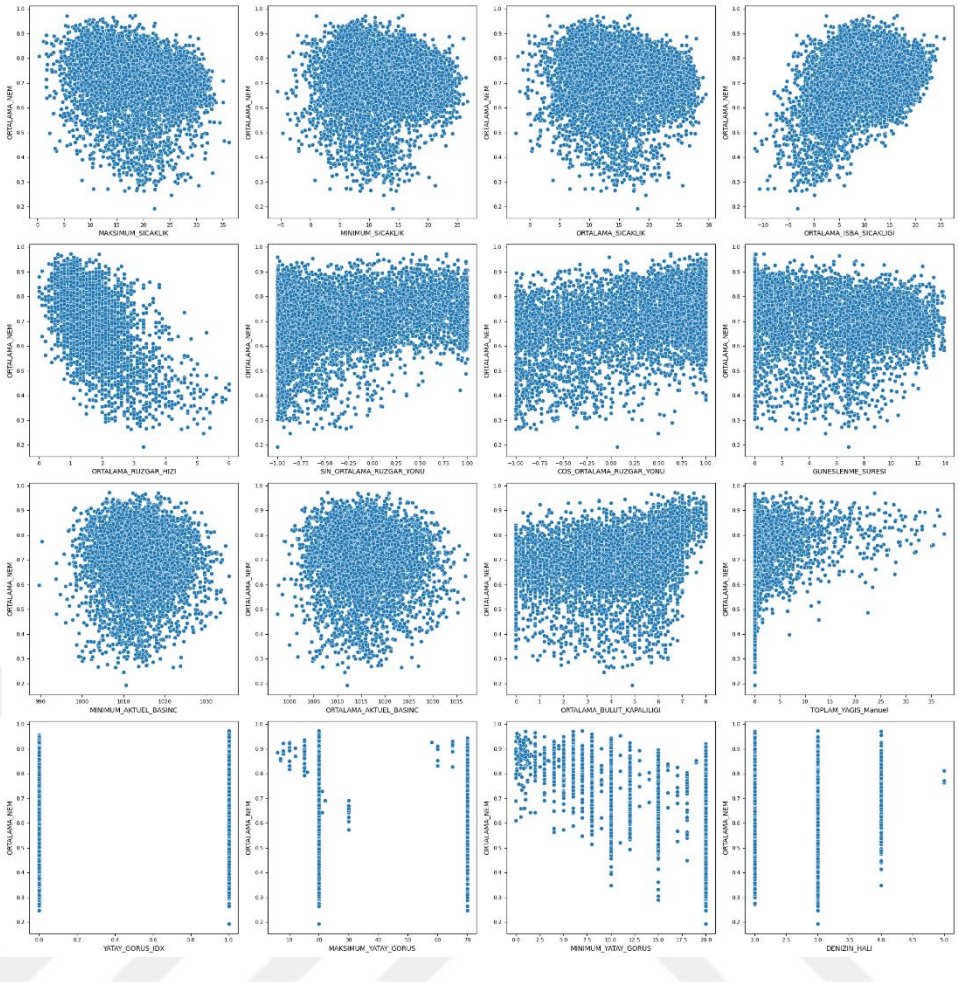
Doldurma işleminden sonra, her iki analiz seti için aykırı gözlemler Mahalanobis uzaklığı temel alınarak tanımlanmış ve çıkarılmıştır. Bu işlemde, Ki-kare dağılımı kullanılarak %99.9 güven düzeyine göre eşik değeri belirlenmiş ve bu değerin üzerindeki uzaklığa sahip gözlemler temizlenmiştir.

Tablo 3.4, aykırı değerler çıkarıldığında 5990 meteorolojik gözlemin özetini sunmakta olup, parametreler arasında belirgin bir varyasyon olduğunu göstermektedir. Sıcaklık ölçümleri geniş aralıklara sahip olup, standart sapmaların 6.44–7.04 civarında seyretmesi dikkat çekicidir; buna karşın, atmosfer basıncı değerleri nispeten sabit

kalmaktadır. Öte yandan, bulut örtüsü ve güneşlenme süresi daha geniş bir değişkenlik sergilerken, denizin hali ve yatay görüş indekslerinde gözlenen uniformluk ile tezat oluşturmaktadır.

Tablo 3.4. Doldurma işlemi yapıldıktan sonra gözlem verilerinin tanımlayıcı istatistikleri.

	Sayı	Ort.	Std	Min	25%	50%	75%	Maks
Maksimum Sıcaklık	5990	19.57	7.04	0.30	13.80	19.80	26.00	36.10
Minimum Sıcaklık	5990	12.64	6.44	-5.60	7.40	12.30	18.30	26.70
Ortalama Sıcaklık	5990	15.90	6.72	-2.30	10.30	15.70	22.08	29.50
Ortalama İşba Sıcaklığı	5990	10.04	6.95	-11.70	4.60	9.90	16.10	25.70
Maksimum Aktüel Basınç	5990	1017.73	6.15	1000.8	1013.20	1017.10	1021.70	1040.10
Minimum Aktüel Basınç	5990	1013.42	6.32	989.60	1009.10	1013.00	1017.50	1035.40
Ortalama Aktüel Basınç	5990	1015.50	6.17	997.40	1011.00	1014.90	1019.40	1037.20
Denizin Hali	5990	2.53	0.58	2.00	2.00	2.00	3.00	5.00
Güneşlenme Suresi	5990	5.45	4.29	0.00	0.90	5.70	9.20	13.90
Yatay Görüş idx	5990	0.66	0.47	0.00	0.00	1.00	1.00	1.00
Maksimum Yatay Görüş	5990	28.16	18.53	6.00	20.00	20.00	20.00	70.00
Minimum Yatay Görüş	5990	16.32	5.46	0.00	10.00	20.00	20.00	20.00
Ortalama Bulut Kapalılığı	5990	4.04	2.18	0.00	2.30	4.30	6.00	8.00
Toplam Yağış Manuel	5990	2.09	6.42	0.00	0.00	0.00	1.00	204.60
Sin Ortalama Rüzgâr Yönü	5990	-0.25	0.63	-1.00	-0.83	-0.39	0.22	1.00
Cos Ortalama Rüzgâr Yönü	5990	0.28	0.68	-1.00	-0.31	0.52	0.91	1.00
Ortalama Rüzgâr Hızı	5990	1.69	0.76	0.00	1.20	1.60	2.00	6.00
Ortalama Nem	5990	0.70	0.12	0.19	0.64	0.71	0.79	0.97



Şekil 3.3. Nem ile ilgili parametreler arasındaki saçılım grafikleri.

Şekil 3.3'da belirtildiği gibi, nem ile diğer parametreler arasındaki ilişki lineer görünmemektedir ve değerler, belirli bölgelerde yoğunlaşmıştır. Bu nedenle, normal veya beta regresyonu gibi geleneksel regresyon yöntemleri, bu senaryo için doğru modeller sunamamaktadır.

Veri seti, %80 eğitim ve %20 test olacak rastgele bir şekilde ikiye ayrılmış olup, tüm model kurma ve seçim işlemleri yalnızca eğitim seti üzerinde gerçekleştirilmiştir.

3.3. Beta Regresyon Analizi

Bu bölümde, tüm hesaplamalar R sürüm 4.3.3'deki 'betareg' paketi kullanılarak yapılmıştır. 'betareg()' fonksiyonunun argümanları şu şekildedir:

betareg(formula, data, subset, na.action, weights, offset, link = c("logit", "probit", "cloglog", "cauchit", "loglog"), link.phi = NULL, type = c("ML", "BC", "BR"), control = betareg.control(...), model = TRUE, y = TRUE, x = FALSE, ...).

Nem değişkeni ile beta regresyon modellemesi uygulanmıştır. Dört farklı bağ fonksiyonu kullanılmış ve Tablo 3.5'te bunların AIC ve R^2_{pseudo} değerleri sunulmuştur. Sonuçlar, cloglog bağ fonksiyonunun daha üstün bir uyum sağladığını göstermektedir. Optimal bir bağ fonksiyonunun seçimi, modelin uyumunu önemli ölçüde iyileştirir. Bu nedenle cloglog bağ ile elde edilen sonuçlar tercih edilmiştir.

Tablo 3.5. Çeşitli beta regresyon modelleri için AIC ve R^2_{pseudo} değerleri.

Bağ Fonksiyon	**DF**	**AIC**	$\boldsymbol{R^2_{pseudo}}$
Probit	19	-19494.578	0.9211
Loglog	19	-17858.664	0.8796
Cloglog	19	**-20802.688**	**0.9455**
Cauchit	19	-17018.374	0.6816

Tablo 3.6, bağımsız değişkenlerin anlamlılığını gösteren standart hatalarla birlikte beta regresyon katsayılarını sunmaktadır. İstatistiksel anlamlılık kodları, regresyon analizinde değişkenlerin etkisinin gücünü ifade eder. Buna göre, pr-değeri 0.001'den küçük olanlar (***) çok yüksek düzeyde anlamlı, 0.01'den küçük olanlar (**) yüksek düzeyde anlamlı, 0.05'ten küçük olanlar (*) anlamlı, 0.1 veya daha büyük olan değişkenler ise istatistiksel olarak anlamlı değildir ve bu nedenle sembolsüz olarak gösterilir.

Tablo 3.6. Cloglog fonksiyonuna bağlı beta regresyon katsayıları.

Değişken	**Tahmin**	**Std. Hata**	***z* değer**	**Pr(> |*z*|)**
Sabit (Intercept)	2.50	0.27	9.27	< 2e-16***
Maksimum Sıcaklık	-7.83e-04	1.12e-03	-0.699	0.48452
Minimum Sıcaklık	7.19e-03	1.60e-03	4.498	6.87e-06***
Ortalama Sıcaklık	-0.113	2.42e-03	-46.505	< 2e-16***

Ortalama İşba Sıcaklığı	0.11	7.14e-04	154.002	< 2e-16***
Maksimum Aktüel Basınç	6.08e-04	1.36e-03	0.448	0.65415
Minimum Aktüel Basınç	-2.71e-03	1.31e-03	-2.065	0.03897*
Ortalama Aktüel Basınç	5.52e-04	2.43e-03	0.227	0.82076
Denizin Hali	-6.42e-03	2.43e-03	-2.641	0.00827**
Güneşlenme Süresi	-2.13e-03	4.17e-04	-5.096	3.48e-07***
Yatay Görüş Idx	1.04e-03	3.21e-03	0.323	0.74663
Maksimum Yatay Görüş	-1.65e-04	8.33e-05	-1.978	0.0479*
Minimum Yatay Görüş	-5.16e-03	2.84e-04	-18.189	< 2e-16***
Ortalama Bulut Kapalılığı	-2.03e-04	9.43e-04	-0.215	0.82959
Toplam Yağış Manuel	3.74e-03	2.92e-04	12.816	< 2e-16***
Sin Ortalama Rüzgâr Yönü	6.58e-03	2.20e-03	2.995	0.00274**
Cos Ortalama Rüzgar Yönü	5.13e-03	2.04e-03	2.512	0.01202*
Ortalama Rüzgâr Hızı	-1.03e-02	2.22e-03	-4.651	3.31e-06***

Ortalama sıcaklık, ortalama nem üzerinde güçlü negatif bir etki göstermekte olup, yüksek sıcaklıkların nem seviyelerini belirgin şekilde azalttığı ifade edilmektedir. Buna karşın, ortalama işba sıcaklığı, nemi anlamlı derecede artıran güçlü bir pozitif etki sergilemektedir. Atmosfer basıncı açısından, yalnızca minimum aktüel basınç anlamlı şekilde negatif etki göstermekte, maksimum ve ortalama basınç değerleri istatistiksel olarak anlamlı bulunmamıştır. Güneşlenme süresi de nem üzerinde negatif etki yapmaktadır. Ayrıca, yatay görüş ile ilgili ölçümler (maksimum yatay görüş, minimum yatay görüş) anlamlı derecede negatif etki sergilemekte olup, azalan görünürlüğün artan neme bağlı olduğu öne sürülmektedir. Son olarak, toplam yağış manuel değişkeni, daha fazla yağışın daha yüksek neme bağlı olduğunu gösteren anlamlı pozitif etki sergilemektedir.

Modele, ana katkıda bulunan değişkenlerin seçimi için geriye doğru seçim işlemi yapılmıştır ve tablo 3.7' de verilmiştir. R^2_{pseudo} değeri neredeyse hiç değişmemektedir,

ancak AIC değeri -20809.984'e kadar gerilemektedir. Log-olabilirlik 15 serbestlik derecesi ile 10420 olarak kaydedilmiştir.

Model seçimi sonucunda, 17 bağımsız değişkenden 13'ünün nem seviyeleri üzerinde önemli bir etkiye sahip olduğu görülmüştür. Dikkat çekici bir şekilde, bazı değişkenler negatif bir etki gösterirken, bazıları da pozitif bir etki sergilemektedir.

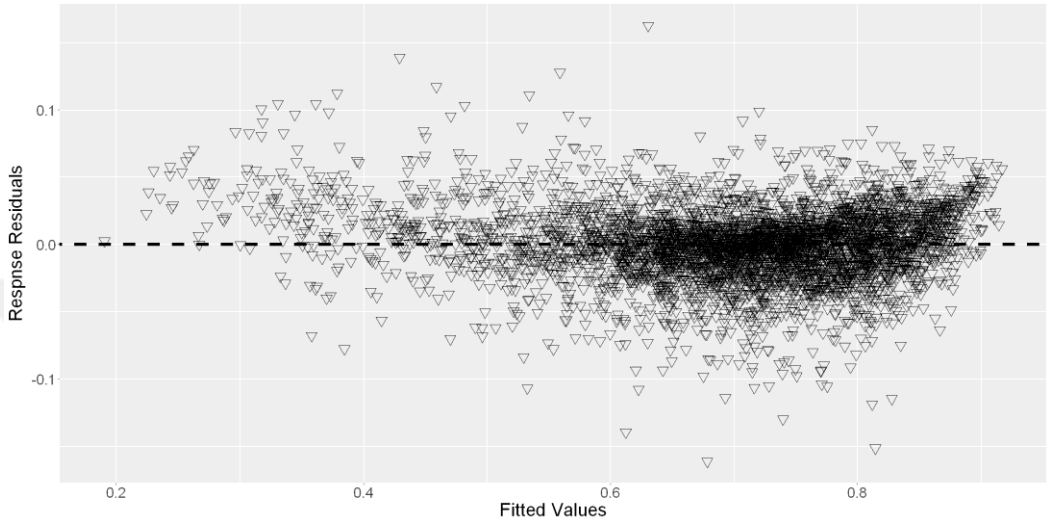
Tablo 3.7. Model seçimi sonrası cloglog fonksiyonuna bağlı beta regresyon katsayıları.

Değişken	**Tahmin**	**Std. Hata**	***z* değer**	**Pr(> |z|)**
Sabit (Intercept)	2.46	2.61e-01	9.441	< 2e-16 ***
Ortalama Rüzgâr Hızı	-1.01e-02	2.19e-03	-4.592	4.39e-06 ***
Minimum Yatay Görüş	-5.16e-03	2.79e-04	-18.483	< 2e-16 ***
Ortalama İşba Sıcaklığı	0.11	6.97e-04	158.11	< 2e-16 ***
Ortalama Sıcaklık	-0.114	1.29e-03	-88.471	< 2e-16 ***
Toplam Yağış Manuel	3.74e-03	2.91e-04	12.865	< 2e-16 ***
Güneşlenme Süresi	-2.10e-03	3.70e-04	-5.67	1.43e-08 ***
Minimum Aktüel Basınç	-2.37e-03	4.79e-04	-4.954	7.28e-07 ***
Minimum Sıcaklık	7.63e-03	1.31e-03	5.824	5.73e-09 ***
Sin Ortalama Rüzgâr Yönü	6.83e-03	2.17e-03	3.151	0.00163 **
Maksimum Yatay Görüş	-1.79e-04	6.74e-05	-2.652	0.00799 **
Denizin Halı	-6.51e-03	2.40e-03	-2.716	0.00661 **
Cos Ortalama Rüzgâr Yönü	5.13e-03	1.98e-03	2.588	0.00965 **
Ortalama Aktüel Basınç	8.58e-04	5.44e-04	1.577	0.11488

ϕ hassasiyet parametresi, lineer regresyon çerçevesinde değişiklik göstermektedir ve tablo 3.8’ de verilmiştir.

Tablo 3.8. ϕ katsayıları (Hassasiyet modeli, kimlik bağ ile).

	Tahmin	**Std. Hata**	***z* değer**	**Pr(> |z|)**
$\boldsymbol{\phi}$	248.57	5.07	49.02	< 2e-16 ***



Şekil 3.4. Beta regresyon modeli için artık grafiği.

Yukarıdaki şekilde gösterilen artık grafiği, modelimizin verilere uygunluğunu değerlendirmek için bir teşhis aracıdır. Artık grafiği, sıfır etrafında rastgele dağılım gösteren bir nokta deseni ortaya koymakla birlikte, tahmin edilen değerlerin (fitted values) 0.65 ile 0.85 arasında olduğu durumda bir kümelenme deseni olduğu görülmektedir. Özellikle, artıkların sınırlı aralığı (-0.1612 ila 0.1625) aşırı aykırı değerlerin yokluğuna işaret etmektedir.

Bu model, test verileri üzerinde değerlendirilmiş ve elde edilen sonuçlar aşağıda belirtilmiştir.

Tablo 3.9. Beta model performansının eğitim ve test verileri üzerindeki değerlendirmesi.

Ölçüt	**Eğitim Veri Seti**	**Test Veri Seti**
$\boldsymbol{R^2_{pseudo}}$	0.9455	0.9467
MSE	0.000736	0.000771

MAE	0.019844	0.020284
BCE	0.576413	0.580655

Bu sonuçlardan, modelin eğitim verisine yüksek uyum sağlamakla kalmayıp, yeni veriye genelleme yapma kapasitesinin de etkili olduğu; eğitim ve test veri seti arasındaki performans metriklerindeki minimal farkların, modelin güvenilirliğini yansıttığı sonucuna varılmaktadır.

3.4. Kuadratik Beta Regresyon Analizi

Kuadratik regresyon, lineer regresyona kuadratik bir terim eklenmesiyle lineer olmayan ilişkileri modellemeye olanak tanımaktadır. Bu yaklaşım veriye bir parabol uydurarak, veri setlerindeki eğrisel eğilimleri yakalamak için kullanışlıdır. Bu bölümde, beta regresyona kuadratik bir terim ekleyerek genişletilen kuadratik beta regresyonu ele alınmıştır. Bu, $(0,1)$ aralığındaki daha karmaşık ilişkileri modellemeye ve tahmin-değişken ilişkilerindeki eğriliği incelemeye olanak tanımaktadır.

Beta regresyonun çoklu kuadratik regresyon versiyonunu oluşturmak için aşağıdaki formülasyonlar izlenmektedir:

$y_1, \dots, y_n$ birbirinden bağımsız rastgele bir örneklem olsun. Her bir $i = 1, \dots, n$ için y_i, ortalaması μ_i ve bilinmeyen hassasiyet parametresi ϕ olan ve denklem (2.9)'da verilen yoğunluk fonksiyonuna sahip beta dağılımından gelsin. Beta regresyonunun tanımına göre genel form aşağıdaki şekilde ifade edilmektedir:

$$\begin{cases} \eta_1 = g(\mu_1) = \sum_{j=1}^{m} \beta_j x_{1j} + e_1, \\ \eta_2 = g(\mu_2) = \sum_{j=1}^{m} \beta_j x_{2j} + e_2 \\ \vdots \\ \eta_n = g(\mu_n) = \sum_{j=1}^{m} \beta_j x_{nj} + e_2, \end{cases} \quad \textbf{(3.1)}$$

veya

$$\eta_i = g(\mu_i) = \sum_{j=1}^{m} \beta_j x_{ij} + e_i \quad , \quad i = 1, \dots, n \tag{3.2}$$

Burada $g(\mu_i)$ lineer tahminci olup, $g(\cdot): (0,1) \mapsto \mathbb{R}$ bir bağ fonksiyonudur. Bu matris notasyonunda aşağıdaki şekilde yazılmaktadır:

$$\boldsymbol{\eta} = \boldsymbol{X} \cdot \boldsymbol{\beta} + \boldsymbol{e} \tag{3.3}$$

Burada,

$$\begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} \tag{3.4}$$

şeklindedir. Çoklu kuadratik beta regresyon modelinin genel formu, esasında normal çoklu kuadratik regresyon modelininkiyle aynıdır. Tek fark, tahminci değişkenlerin alanını gerçek sayılar kümesine uydurmak için bir bağ fonksiyonun kullanılmasıdır. Formül aşağıdaki şekildedir:

$$\eta_i = g(\mu_i) = \beta_0 + \sum_{j=1}^{m} \beta_j x_{ij} + \sum_{j=1}^{m} \sum_{k=1}^{m} \beta_{ik} x_{ij} x_{ik} \tag{3.5}$$

Burada, β_0 sabittir ve β_j, β_{jj} ve β_{jk} ($j \neq k$ olduğunda) sırasıyla lineer, saf kuadratik ve etkileşim katsayılarıdır.

Regresyon parametrelerinin tahmini, MLE yöntemi kullanılarak log-olabilirlik fonksiyonunun maksimize edilmesiyle elde edilmektedir. Log-olabilirlik fonksiyonu aşağıdaki şekildedir:

$$LL(\beta, \phi) = \sum_{i=1}^{n} LL(\mu_i, \phi) \tag{3.6}$$

Burada

$$LL(\mu_i, \phi) = LL\left(g^{-1}\left(\beta_0 + \sum_{j=1}^{m} \beta_j x_{ij} + \sum_{j=1}^{m} \sum_{k=1}^{m} \beta_{ik} x_{ij} x_{ik} \right), \phi \right) \tag{3.7}$$

Aynı süreç izlenerek, veriler modele yüklenmektedir ve nem, yanıt değişken olarak tutulmaktadır. Tablo 3.10'da belirtilen dört farklı bağ fonksiyonu tekrar test edilmektedir. Kuadratik beta fonksiyonda da, yine cloglog bağ fonksiyonu üstün bir performans sergilemektedir.

Tablo 3.10. Çeşitli kuadratik beta regresyon modelleri için AIC ve R^2_{pseudo}değerleri.

Bağ Fonksiyon	**DF**	$\boldsymbol{R^2_{pseudo}}$	**AIC**
Probit	136	0.9561	-21945.607
Loglog	136	0.9436	-21505.213
Cloglog	136	**0.9615**	**-22176.399**
Cauchit	136	0.8053	-20128.675

Tablo 3.5'te sunulan bulgularla karşılaştırıldığında, R^2_{pseudo} değerleri daha yüksek olup, bu da modelin daha iyi uyum sağladığını gösterirken, AIC değerleri ise daha düşük olup, bu da modelin daha iyi olduğuna işaret etmektedir. Kuadratik değişkenlerin (toplamda 134 değişken) kullanılması nedeniyle, model özetinin tamamını incelemek pratik olmamaktadır. Bu sebeple, doğrudan geriye doğru seçim işlemi uygulanmıştır.

Bu süreçte, model geriye doğru seçim işlemi ile 114 adım geçirmiş ve nihai olarak en önemli 20 katkıda bulunan değişken belirlenmiştir. Sonuç olarak, R^2_{pseudo} değeri hafif bir azalma göstererek 0.9579'a, AIC değeri ise hafif bir artış göstererek -21958.567 yerleşmektedir. Log-olabilirlik değeri, 22 serbestlik derecesi ile 11000'dir.

Bu analiz, kalan 20 bağımsız değişkenin nem seviyeleri üzerinde önemli bir etkisi olduğunu ortaya koymaktadır. Dikkat çekici bir şekilde, bazıları negatif bir etki gösterirken, diğerleri pozitif bir etki göstermektedir.

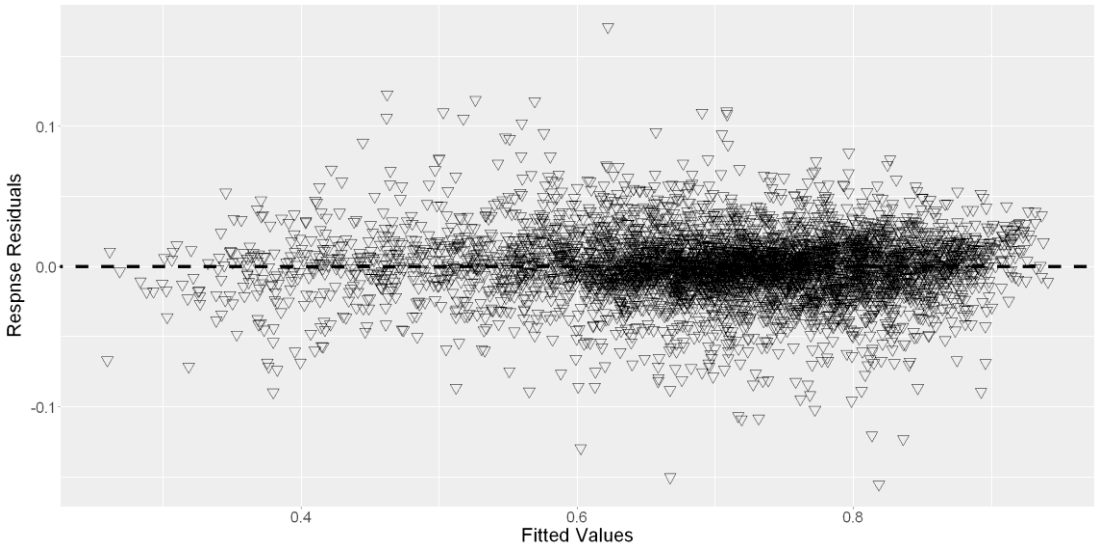
Tablo 3.11. Model seçimi sonrası cloglog fonksiyonuna bağlı kuadratik beta regresyon katsayıları.

Değişken	Tahmin	Std. Hata	*z* değer	$Pr(>\|z\|)$
Sabit (Intercept)	1.03	1.11e-02	93.243	< 2e-16 ***
Ortalama Sıcaklık	-0.181	3.80e-03	-47.599	< 2e-16 ***
Ortalama İşba Sıcaklığı	0.17	2.09e-03	81.466	< 2e-16 ***
Ortalama Sıcaklık^2	5.23e-03	3.09e-04	16.943	< 2e-16 ***
Ortalama İşba Sıcaklıgı^2	1.56e-03	1.10e-04	14.143	< 2e-16 ***
Toplam Yağış Manuel^2	-1.64e-04	2.47e-05	-6.648	2.98e-11 ***
Ortalama Sıcaklık* Minimum Sıcaklık	-3.16e-03	3.32e-04	-9.512	< 2e-16 ***
Ortalama İşba Sıcaklıgı* Minimum Sıcaklık	3.98e-03	3.17e-04	12.57	< 2e-16 ***
Minimum Sıcaklık* Ortalama Aktüel Basınç	-3.13e-04	6.40e-05	-4.896	9.80e-07 ***
Minimum Sıcaklık* Maksimum Aktüel Basınç	3.29e-04	6.41e-05	5.131	2.88e-07 ***
Ortalama Sıcaklık * Ortalama İşba Sıcaklığı	-7.66e-03	3.15e-04	-24.333	< 2e-16 ***
Ortalama Sıcaklık* Minimum Yatay Görüş	8.05e-04	1.06e-04	7.611	2.72e-14 ***
Ortalama Sıcaklık * Ortalama Rüzgâr Hızı	-6.10e-04	1.16e-04	-5.279	1.30e-07 ***
Ortalama İşba Sıcaklığı* Minimum Yatay Görüş	-7.01e-04	1.07e-04	-6.57	5.03e-11 ***
Maksimum Aktüel Basınç* Güneşlenme Süresi	-2.20e-06	4.05e-07	-5.443	5.25e-08 ***
Maksimum Aktüel Basınç* Minimum Yatay Görüş	-6.75e-06	6.83e-07	-9.884	< 2e-16 ***
Maksimum Aktüel Basınç* Toplam Yağış Manuel	9.19e-06	6.66e-07	13.813	< 2e-16 ***

Güneşlenme Süresi * Yatay Görüş Idx	2.51e-03	3.25e-04	7.736	1.03e-14 ***
Toplam Yağış Manuel*Yatay Görüş Idx	-3.32e-03	4.64e-04	-7.146	8.90e-13 ***
Minimum Yatay Görüş*Cos Ortalama Rüzgâr Yönü	-1.46e-03	2.45e-04	-5.974	2.32e-09 ***
Cos Ortalama Rüzgâr Yönü* Denizin Halı	1.15e-02	1.70e-03	6.794	1.09e-11 ***

Tablo 3.12. ϕ katsayısıları (kimlik bağlı hassasiyet modeli).

	Tahmin	**Std. Hata**	***z* değer**	**Pr(> |*z*|)**
$\boldsymbol{\phi}$	317.306	6.475	49.01	<2e-16 ***



Şekil 3.5. Kuadratik beta regresyon modeli için artık grafiği.

Artık grafiği, sıfır etrafında rastgele bir dağılım gösteriyor gibi görünmektedir. Ancak, tahmin edilen değerler (fitted values) 0.60 ile 0.85 arasında olduğunda burada da kümelenme deseni ortaya çıkmıştır. Özellikle, artıkların sınırlı aralığı (-0.155 ila 0.171), bu aralığın lineer beta regresyon modelinden daha dar olduğunu ve aykırı değerlerin yokluğunu da işaret etmektedir.

Bu model, eğitim ve test veri setleri üzerinde değerlendirilmiş olup elde edilen performans metrikleri Tablo 3.13'de yer almaktadır.

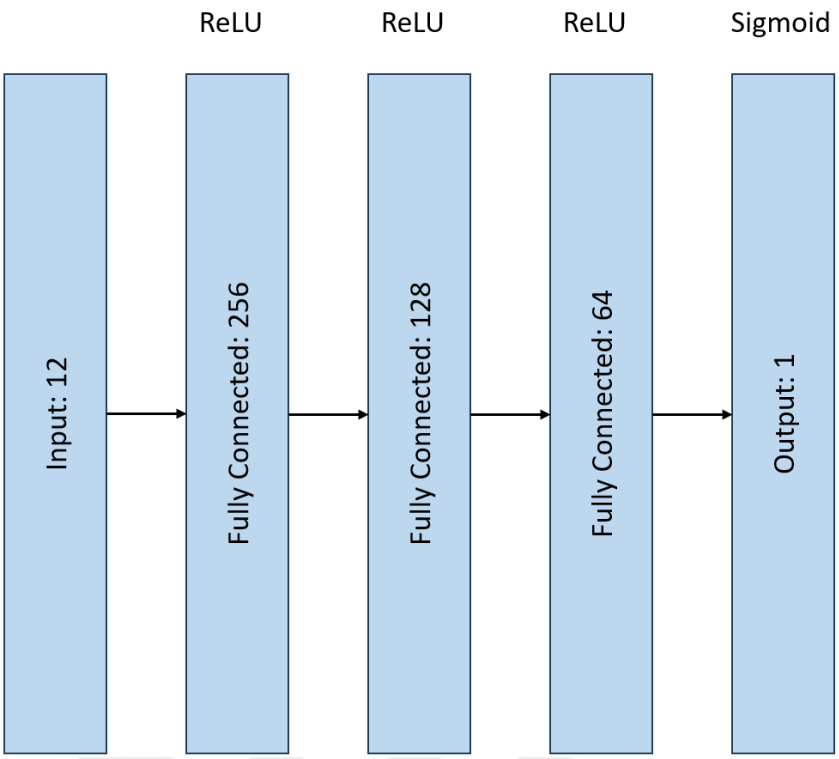
Tablo 3.13. Kuadratik beta regresyon model performansının eğitim ve test verileri üzerindeki değerlendirmesi.

Ölçüt	Eğitim Veri Seti	Test Veri Seti
R^2_{pseudo}	0.95794	0.95758
MSE	0.000595	0.000629
MAE	0.017269	0.017622
BCE	0.576000	0.580278

Bu sonuçlardan, modelin hem eğitim verisine yüksek uyum sağladığı hem de yeni veriye genelleme yapma kapasitesinin etkili olduğu; eğitim ve test veri seti arasındaki performans metriklerindeki minimal farkların modelin stabilitesi ve güvenilirliğini yansıttığı sonucuna varılmıştır.

3.5. İleri Beslemeli Sinir Ağı Analizi

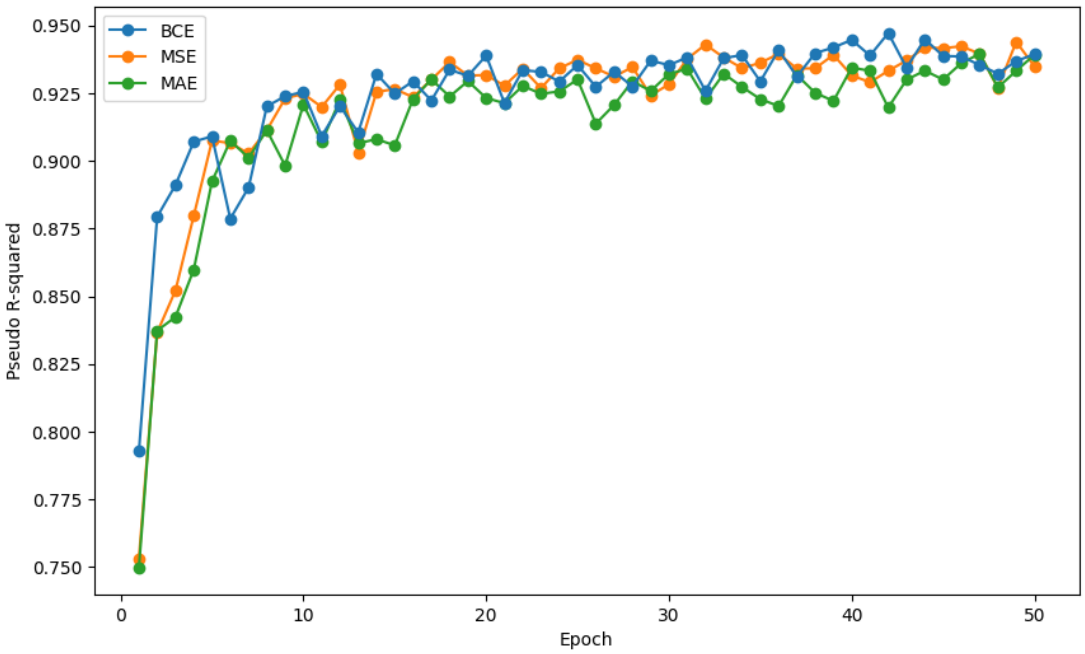
FNN, Şekil 3.6'da gösterildiği gibi beş katmandan oluşmaktadır. Çoklu deneyler sonucunda, aktivasyon fonksiyonlarının [ReLU, ReLU, ReLU, Sigmoid] kombinasyonunun model performansını en üst düzeye çıkardığı tespit edilmiştir. Ayrıca, 17 değişken arasından ileriye doğru seçim yöntemi uygulanmış ve analizler sonucunda modelin performansının, 12 değişken kullanıldığında en iyi seviyeye ulaştığı belirlenmiştir. Bu değişkenler, ortalama rüzgâr hızı, minimum yatay görüş, ortalama işba sıcaklığı, ortalama sıcaklık, toplam yağış manuel, minimum sıcaklık, maksimum yatay görüş, yatay görüş idx, maksimum sıcaklık, ortalama bulut kapalılığı, cos ortalama rüzgâr yönü, güneşlenme süresidir. Bu bölümde, tüm grafikler ve hesaplamalar Python'da PyTorch paketini kullanarak gerçekleştirilmiştir.



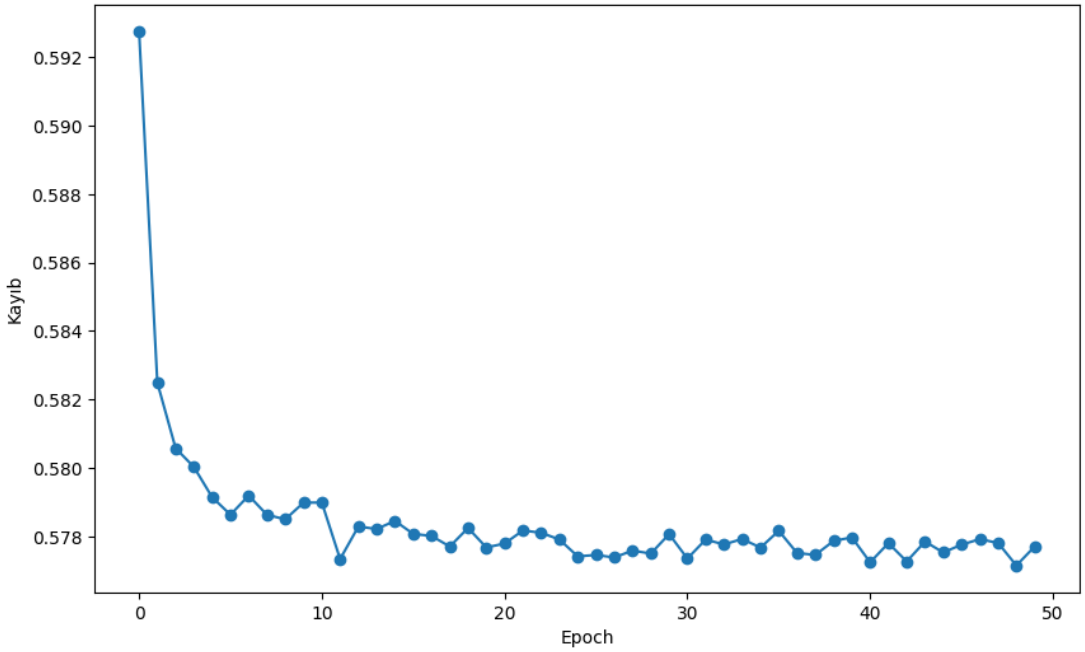
Şekil 3.6. FNN ağı.

Bu modelde, gizli katmanlarda ReLU kullanılması, gradyan kaybolması sorununu azaltarak hızlı ve etkili bir öğrenmeyi destekler. Böylece ağ, karmaşık ve doğrusal olmayan örüntüleri daha verimli bir şekilde öğrenebilir. Öte yandan, çıktı katmanında kullanılan sigmoid fonksiyonu, çıktıları [0, 1] aralığına sınırlar. Bu aralık, çıktının olasılıkları temsil ettiği durumlarda, örneğin ikili sınıflandırma veya olasılık temelli yorumlamalar gerektiren diğer uygulamalarda oldukça faydalıdır.

50 epoch boyunca üç farklı kayıp kriteri olan ikili çapraz entropi (BCE), ortalama mutlak hata (MAE) ve ortalama kare hata (MSE) kriterleri karşılaştırılmış ve şekil 3.7’de verilmiştir. Bulgular, diğer kayıp kriterlerine kıyasla BCE'nin modelin performansını artırdığını göstermektedir.



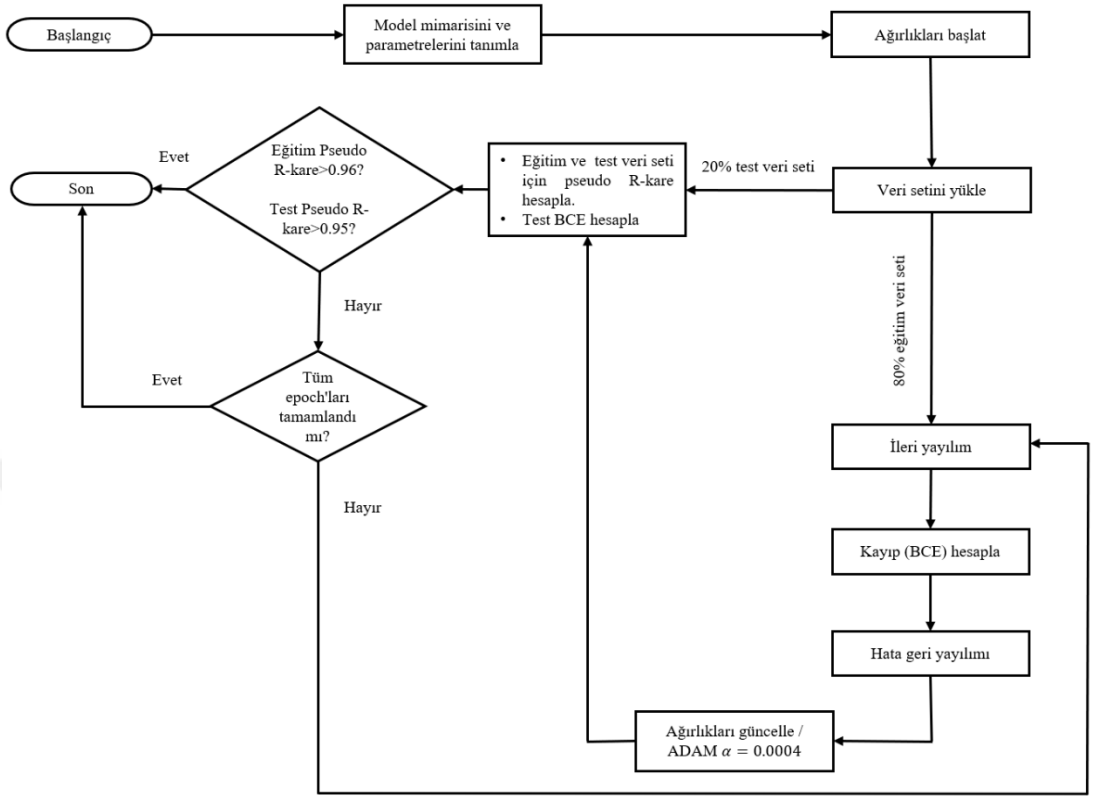
Şekil 3.7. Farklı kayıp kriterleri ile model performansının karşılaştırılması.



Şekil 3.8. 50 epoch boyunca BCE kaybının yakınsaması.

Ek olarak, optimizasyon algoritması olarak adaptif moment tahmin algoritması (ADAM) kullanılmıştır.

Aşağıdaki eğitim algoritması akış şeması, Python 3.10 sürümünde PyTorch modülü kullanılarak uygulanmıştır.



Şekil 3.9. Nem verisinde FNN modeli için eğitim algoritmasına ait akış şeması.

Bu modelin eğitim veri üzerinde R^2_{pseudo} değeri 0.9731'e ulaşırken, BCE kaybı değeri 0.5753 olarak kaydedilmiştir.

Hata terimlerinin $\varepsilon_i \sim \mathcal{B}(\mu_i, \phi)$ olduğu varsayımı altında, log-olabilirlik (2.12) formül esas alınarak hesaplanmıştır. Burada $\mu_i = \widehat{y_i} = FNN(x_i; \theta)$ modelin tahmin değeri; θ ise sinir ağına ait tüm ağırlıklar ve sabit terimlerinden oluşan parametre vektörünü ifade etmektedir.

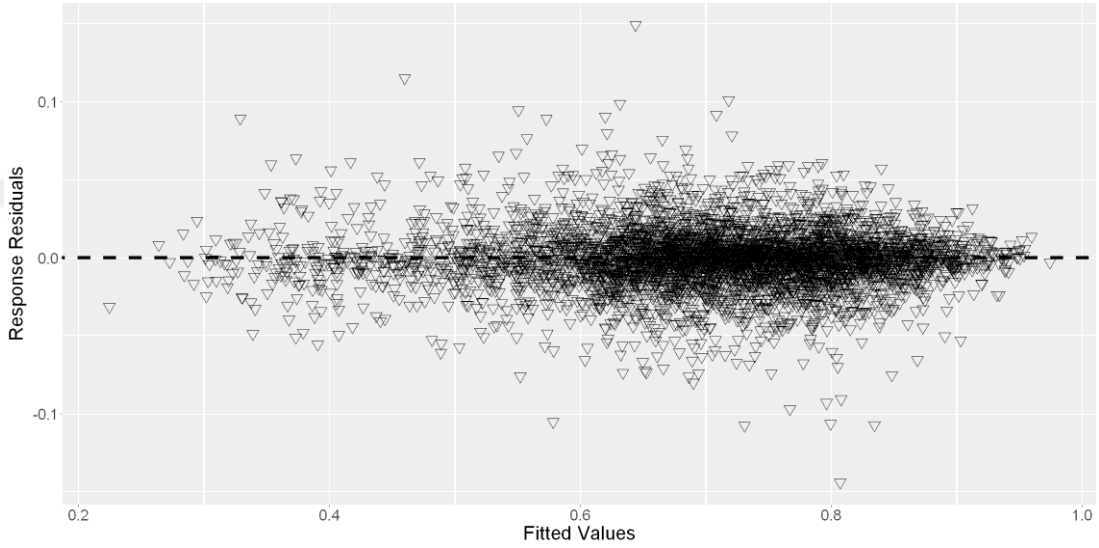
scipy.optimize.minimize_scalar ile yapılan optimizasyonda $\hat{\phi} = 498.479$ bulunmuş ve buna karşılık gelen maksimum log-olabilirlik değeri 12083.96'dir.

Toplam tahmin edilmesi gereken parametre sayısı, katman ağırlıkları ve sabit terimler dahil olmak üzere

$$k = 12 * 256 + 256 * 128 + 128 * 64 + 64 * 1 + 256 + 128 + 64 + 1 = 44545$$

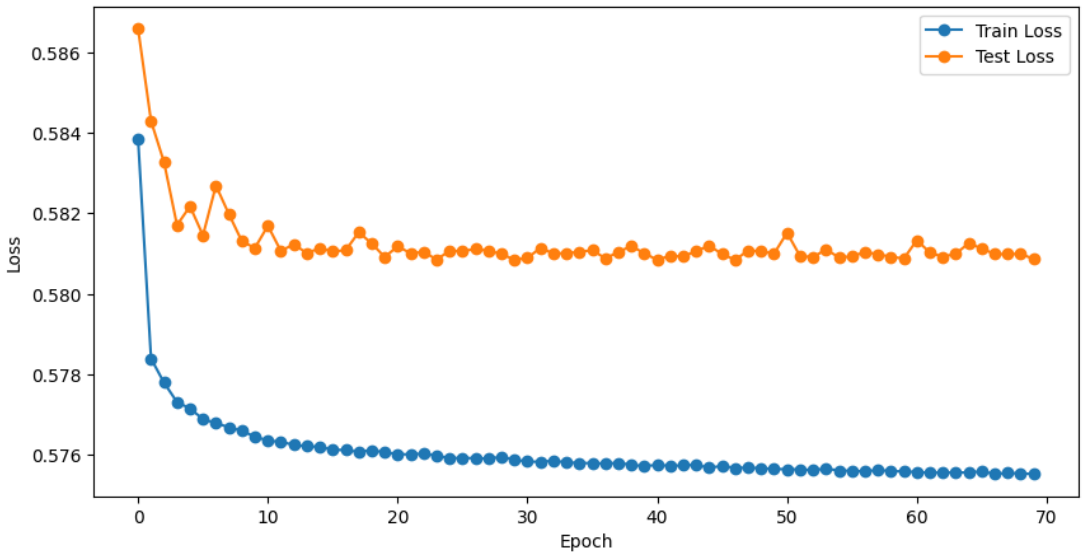
olarak hesaplanmıştır. Bu durumda AIC değeri aşağıdaki şekilde bulunmuştur:

$$AIC = -2LL + 2k = 20377.081$$



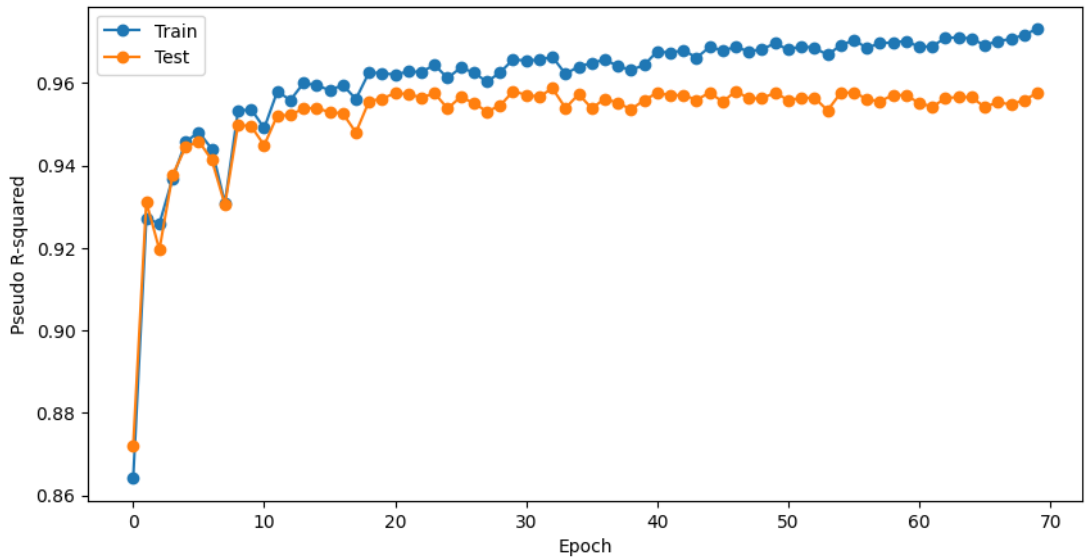
Şekil 3.10. FNN modeline ait artık grafiği.

Yukarıda gösterilen artık grafiği, beta regresyon modellerine kıyasla üstün bir model uyumu sergilemektedir. Noktaların sıfır etrafında rastgele bir dağılım sergilediği ve sıfıra yakın bir yoğunlaşma olduğu gözlemlenmektedir. Tahmin edilen değerlerin (fitted values) 0.65 ile 0.9 arasında olduğu durumlarda bir kümelenme deseni fark edilmektedir. Ayrıca, artıkların sınırlı aralığında (response residuals) (-0.1439 ile 0.1491), beta regresyon modellerine göre daha az aykırı değer olduğu dikkate değerdir.



Şekil 3.11. FNN'nin 70 epoch boyunca eğitim ve test kayıplarının yakınsaması.

Şekil 3.11, FNN eğitim ve test kayıplarının 70 epoch boyunca istikrarlı şekilde azalarak kararlı yakınsama sağlandığını ve aşırı uyum olmadığını göstermektedir.



Şekil 3.12. FNN'nin 70 epoch boyunca eğitim ve test R^2_{psuedo} değerlerinin yakınsaması.

Şekil 3.12, hem eğitim hem test setleri için 70 epoch boyunca R^2_{psuedo} değerlerinin yakınsaması gösterilmiştir. Bu, modelin kararlılığını ortaya koymaktadır.

Bu model, eğitim ve test veri setleri üzerinde değerlendirilmiş olup elde edilen performans metrikleri Tablo 3.14'de yer almaktadır.

Tablo 3.14. FNN model performansının eğitim ve test verileri üzerindeki değerlendirmesi.

Ölçüt	Eğitim Veri Seti	Test Veri Seti
R^2_{pseudo}	0.9731	0.9576
MSE	0.000393	0.000651
MAE	0.013880	0.017558
BCE	0.575315	0.580867

3.6. Yöntemlerin Karşılaştırılması

Samsun meteoroloji müdürlüğünden alınan orijinal verilerde normallik varsayımından sapmalar gözlemlenmekte olup, nem ile 17 meteorolojik değişken arasındaki korelasyonun kümelenmiş ve lineer olmayan desenler sergilediği tespit edilmiştir. Bu bağlamda, nemin modellenmesinde beta dağılımının esnekliği değerlendirilmiştir.

Öncelikle, beta regresyon uygulanmakta olup; cloglog bağ fonksiyonunun tercih edilmesi neticesinde, 19 serbestlik derecesi ile modelin R^2_{pseudo} değeri 0.9455, AIC değeri ise -20802.689 olarak hesaplanmıştır. Nem üzerinde etki eden önemli değişkenleri belirlemek amacıyla gerçekleştirilen geriye doğru seçiminde, bağımsız değişkenlerden 13 tanesinin anlamlı etkiye sahip olduğu saptanmış; bu süreçte R^2_{pseudo} değeri sabit kalmış (0.9455) ve AIC değeri hafif düşüş göstererek -20809.984'e gerilemiştir.

Beta regresyon modeline ait artık grafikleri incelendiğinde, hata terimlerinin sıfır etrafında rastgele dağıldığı, artıkların (-0.1612 ila 0.1625) sınırlı aralıkta yer aldığı ve aşırı aykırı değerlerin bulunmadığı görülmüştür.

Ardından, 0 ile 1 arasında değişen karmaşık, lineer olmayan ilişkileri daha iyi yakalayabilmek amacıyla, ikinci dereceden terim içeren kuadratik beta regresyon uygulanmıştır. Bu yöntemde, dört farklı bağ fonksiyonu test edilmiş olup, cloglog fonksiyonunun en iyi sonuçları verdiği tespit edilmiştir. 134 değişkenin dahil edilmesi sonrasında, geriye doğru seçimi ile en önemli 20 anlamlı değişken belirlenmiş; son

modelin değeri 0.9579, AIC değeri -21958.567, log-olabilirlik değeri 11000 ve serbestlik derecesi 22 olarak hesaplanmıştır. Modelin artık grafikleri, hata terimlerinin sıfır etrafında rastgele dağıldığını ve aykırı değerlerin yokluğunu ortaya koymaktadır (artık aralığı: -0.1553 ila 0.1708).

FNN modeli; giriş katmanı, üç gizli katman ve çıkış katmanından oluşmakta olup, sırasıyla 12, 256, 128, 67 ve 1 nöron içermektedir. Her bir katmanda optimal aktivasyon fonksiyonları ReLU, ReLU, ReLU ve Sigmoid olarak belirlenmiştir. Model, BCE kayıp kriteri, 0.0004 öğrenme oranı ve ADAM optimizasyon algoritması kullanıldığında en iyi performansı sergilemiş; 70. epoch'tan sonra değeri 0.9731'e ulaşmış ve BCE kayıp değeri 0.5753 olarak hesaplanmıştır. Buna ek olarak, log-olabilirlik değeri 11996.788, bu değere bağlı olarak hesaplanan AIC ise 65096.42'dir. Ayrıca, artık grafikleri, hata terimlerinin önceki modellere göre daha dar bir aralıkta dağıldığını (artık aralığı: -0.1439 ila 0.1491) göstermektedir.

Tablo 3.15'de eğitim veri setinde, tablo 3.16'da ise test veri setinde beta regresyon, kuadratik beta regresyon ve FNN modellerinin performans karşılaştırmaları yapılmıştır.

Tablo 3.15. Eğitim veri setinde beta regresyon, kuadratik beta regresyon ve FNN modellerinin performans karşılaştırması.

	Beta Regresyon	**Kuadratik Beta Regresyon**	**FNN**
R^2_{pseudo}	0.94551	0.95794	0.97314
LL	10420	11000	12084
AIC	-20809.984	-21958.567	20377.081
MAE	0.019844	0.017269	0.013880
MSE	0.000736	0.000595	0.000393
BCE	0.57641	0.57600	0.57532

Artıklar Alt Sınırı	-0.161200	-0.155323	-0.143942
Artıklar Üst Sınırı	0.162481	0.170830	0.149121

Tablo 3.16. Test veri setinde beta regresyon, kuadratik beta regresyon ve FNN modellerinin performans karşılaştırması.

	Beta Regresyon	**Kuadratik Beta Regresyon**	**FNN**
R^2_{pseudo}	0.94672	0.95758	0.95756
MAE	0.020283	0.017622	0.017558
MSE	0.000771	0.000629	0.000651
BCE	0.580654	0.580278	0.580867

Bu çerçevede, karşılaştırılan modeller arasında, FNN modelinin özellikle eğitim veri setinde daha üstün bir performans sergilediği; ancak test veri setinde FNN ve kuadratik beta regresyon modellerinin sonuçlarının benzerlik arz ettiği gözlemlenmiştir. Tüm modellerde, artık dağılımlarının sıfır etrafında rastgele seyretmesi, modellerin büyük ölçüde yanlılık içermediğini ve hata terimlerinin öngörülebilir özellikler taşıdığını ortaya koymuştur.

Mevcut veri setinden rastgele olarak 20 gözlem seçilmiştir. Tablo 3.17'de, gerçek ortalama nem değerleri (y) ile beta regresyonu, kuadratik beta regresyonu ve FNN modelleri kullanılarak elde edilen tahmin ve hata değerleri sunulmuştur. Ayrıca, modellerin performansı MAE, MSE ve BCE metrikleriyle değerlendirilmiştir.

Tablo 3.17. Ortalama nem tahmin performansının beta regresyon, kuadratik beta regresyon ve FNN modelleri için karşılaştırılması.

y	$\widehat{y}_{beta}$	$\widehat{y}_{k.beta}$	$\widehat{y}_{FNN}$	$\lvert\varepsilon_{beta}\rvert$	$\lvert\varepsilon_{k.beta}\rvert$	$\lvert\varepsilon_{FNN}\rvert$
0.758	0.72186	0.72550	0.72233	0.03614	0.03250	0.03567
0.650	0.65049	0.63963	0.63221	0.00049	0.01037	0.01779
0.598	0.60754	0.59861	0.59945	0.00954	0.00061	0.00145
0.898	0.88319	0.89588	0.87500	0.01481	0.00212	0.02300
0.425	0.41078	0.43055	0.42972	0.01422	0.00555	0.00472
0.617	0.63856	0.63979	0.64493	0.02156	0.02279	0.02793
0.897	0.87353	0.87970	0.87672	0.02347	0.01730	0.02028
0.585	0.55513	0.56758	0.58365	0.02987	0.01742	0.00135
0.758	0.75925	0.74943	0.75781	0.00125	0.00857	0.00019
0.735	0.74524	0.74447	0.73188	0.01024	0.00947	0.00312
0.493	0.47438	0.48634	0.49772	0.01862	0.00666	0.00472
0.689	0.68690	0.68439	0.69010	0.00210	0.00461	0.00110
0.757	0.78210	0.72491	0.73973	0.02510	0.03209	0.01727
0.722	0.71386	0.70887	0.71932	0.00814	0.01313	0.00268
0.568	0.57478	0.57330	0.56674	0.00678	0.00530	0.00126
0.580	0.58628	0.56714	0.58004	0.00628	0.01286	0.00004
0.606	0.61778	0.58888	0.60403	0.01178	0.01712	0.00197
0.748	0.73088	0.72724	0.73011	0.01712	0.02076	0.01789
0.718	0.76539	0.75755	0.75365	0.04739	0.03955	0.03565
0.638	0.64794	0.64660	0.63163	0.00994	0.00860	0.00637
			Toplam	0.31483	0.28738	**0.22445**
			MAE	0.01574	0.01437	**0.01122**
			MSE	0.00039	0.00032	**0.00027**
			BCE	0.60106	0.60083	**0.60079**

4. SONUÇLAR

Bu çalışmada, nem ve diğer meteorolojik değişkenler arasındaki ilişkileri modellemek amacıyla beta regresyon, kuadratik beta regresyon ve FNN yöntemleri kullanılmıştır. Beta regresyon modeli, cloglog bağ fonksiyonu ile birlikte nemi modellemede en iyi uyumu sağlamış ve önemli bağımsız değişkenlerin etkilerini başarıyla yakalamıştır. Geriye doğru seçim işlemi ile, 17 değişkenden 13'ünün nem seviyeleri üzerinde anlamlı etkilere sahip olduğu gösterildi ve diğer modellerle karşılaştırıldı.

Kuadratik beta Regresyon, bağımlı ve bağımsız değişkenler arasındaki lineer olmayan ilişkileri yakalamak adına etkili bir yöntem olarak ortaya çıkmıştır. Cloglog bağ fonksiyonu bu modelde de en iyi sonuçları vermiştir, modelin uyumu ve tahmin gücü biraz daha da artmıştır. Geriye doğru seçim işlemi ile, 114 değişkenden 20'sinin nem seviyeleri üzerinde anlamlı etkilere sahip olduğu gösterildi ve diğer modellerle karşılaştırıldı. Artık grafiklerinde, artıkların rastgele dağıldığı ve aykırı değerlerin olmadığı gözlemlenmiştir, bu da modellerin güvenilirliğini artırmaktadır.

Son olarak, FNN modelinde, BCE değeri üzerinde ileriye doğru seçim işlemi ile en etkin 12 değişken seçilmiş; FNN modeli diğer yöntemlerle karşılaştırıldığında en yüksek R^2_{pseudo} genel ortalama değerini (0.9699) ve düşük BCE kaybını (0.5764) sağlamıştır. FNN modeli, karmaşık ilişkileri yakalamada etkili olmuş ve diğer tüm modellerden daha iyi bir uyum göstermiştir. Sonuç olarak, beta regresyonun mevcut veri setini modellemede kullanışlı bir araç olduğu, ancak sinir ağı tabanlı modeller kadar güçlü olmadığı söylenebilir.

Bu tez çalışmasında kullanılan ortalama nem tahmini yapan modeller, tarımsal takibi iyileştirebilir. Doğru nem öngörüleri, sera içi iklim kontrolü ile açık alandaki hastalık ve mantar risklerinin yönetiminde kritik rol oynayabilir. Öngörülen nem düzeylerine göre sulama ve havalandırma stratejileri optimize edilerek bitki gelişimi desteklenebilir, verim artar, su ve enerji tasarrufu sağlanmış olur. Böylece sürdürülebilir tarım uygulamaları için sağlam bir temel oluşturulmuş olur.

KAYNAKLAR

Abiy, A. Z., Wiederholt, R. P., Lagerwall, G. L., Melesse, A. M., & Davis, S. E. (2022). Multilayer feedforward artificial neural network model to forecast Florida Bay salinity with climate change. *Water, 14*(21), 3495.

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, 60 (2), 255-265.

Baker, B. D., & Richards, C. E. (1999). A comparison of conventional linear regression methods and neural networks for forecasting educational spending. *Economics of Education Review*, 18 (4), 405-415.

Beyer, W. H. CRC Standard Mathematical Tables, 31st ed. Boca Raton, FL: CRC Press, pp. 536 and 571, 2002.

Bickel, Peter J.; Doksum, Kjell A. (2015). Mathematical Statistics: Basic Ideas and Selected Topics. Vol. I (Second ed.). p. 20.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3), 345-370.

Cameron, A.C., & Trivedi, P.K. (2005). Microeconometrics: Methods and Applications. Cambridge University Press.

Cribari-Neto, F. (2023). A beta regression analysis of COVID-19 mortality in Brazil. *Infectious Disease Modelling*, 8, 309–317.

Cribari-Neto, F., & Zeileis, A. (2010). Beta regression in R. *Journal of Statistical Software*, 34 (2), 1-24.

Dubey, S. R., Singh, S. K., & Chaudhuri, B. B. (2022). Activation functions in deep learning: A comprehensive survey and benchmark. *Neurocomputing*. arXiv preprint arXiv:2109.14545.

Dünder, E., & Cengiz, M. A., (2020). Model Selection in Beta Regression Analysis Using Several Information Criteria and Heuristic Optimization. *Journal of New Theory*.

Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7), 799-815.

Gomes, G. S. da S., Ludermir, T. B., & Lima, L. M. M. R. (2011). Comparison of new activation functions in neural network for forecasting financial time series. *Neural Computing and Applications, 20*(3), 417–439.

Hecht-Nielsen. (1989). Theory of the backpropagation neural network. *In International 1989 Joint Conference on Neural Networks* (pp. 593-605). Washington, DC, USA.

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks, 4*(2), 251–257.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359–366.

Hosmer, D.W., Lemeshow, S., & Sturdivant, R.X. (2013). Applied Logistic Regression. John Wiley & Sons.

Hurtik, P., Tomasiello, S., Hula, J., & Hynar, D. (2022). Binary cross-entropy with dynamical clipping. *Neural Computing and Applications, 34*(14), 12029–12041

Jung, D.-H., Lee, T. S., Kim, K., & Park, S. H. (2022). A deep learning model to predict evapotranspiration and relative humidity for moisture control in tomato greenhouses. *Agronomy*, 12(9), 2169.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Lau, E. T., Sun, L., & Yang, Q. (2019). Modelling, prediction and classification of student academic performance using artificial neural networks. *SN Appl. Sci*., 1, 982.

Li, M., & Wang, J. (2019). An empirical comparison of multiple linear regression and artificial neural network for concrete dam deformation modelling. *Mathematical Problems in Engineering*, 2019, Article ID 7620948, 1-13.

Rezaei Melal, S., Aminian, M., & Shekarian, S. M. (2024). A machine learning method based on stacking heterogeneous ensemble learning for prediction of indoor humidity of greenhouse. *Journal of Agriculture and Food Research*, 16, 101107.

Robeson, S. M., & Willmott, C. J. (2023). Decomposition of the mean absolute error (MAE) into systematic and unsystematic components. *PLoS ONE*, 18(2), e0279774.

Schmidt-Hieber, J. (2020). Nonparametric regression using deep neural networks with ReLU activation function. *The Annals of Statistics*, 48 (4).

Unlu, H., & Aktas, S. (2017). Beta regression for the indicator values of well-being index for provinces in Turkey. *Journal of Engineering Technology and Applied Sciences*, 2 (2), 101-111.

Xu, X., & Zhang, Y. (2021). House price forecasting with neural networks. *Intelligent Systems with Applications, 12*, 200052.

Zhang, G. P., Patuwo, B. E., & Hu, Y. M. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14, 35–62.

Zurada, J. M. (1992). Introduction of artificial neural systems. St. Paul: West Publishing.

ÖZ GEÇMİŞ

Abdallah K. A. Balaha, Filistin uyrukludur. 2012–2013 yılları arasında Gazze'de bulunan Khalid Ben Al-Waleed Fen Lisesinde eğitim almıştır. 2013–2017 yılları arasında Al-Azhar Üniversitesi Matematik Bölümünde lisans eğitimi tamamlamış; aynı zamanda Bilgisayar Bilimi alanında yan dal yapmıştır. 2022 yılında Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı yüksek lisans programına başlamış, 2025 yılında mezun olmuştur. Yüksek lisans sürecinde özellikle İstatistik Teorisi alanında uzmanlaşmıştır.

İletişim Bilgileri

ORCID ID : 0000-0001-6277-2653.

Yayınlar:

Balaha, A. ve Uslu, V.R. (2024, February 3–5). Comparative analysis of meteorological variables of Samsun Province with beta regression and artificial neural network . 7th Hagia Sophia International Conference on Multidisciplinary Scientific Studies, Istanbul, Türkiye.