

**İFLAS OLASILIĞININ GENELLEŞTİRİLMİŞ DOĞRUSAL
MODELLER VE BİRLEŞİK AKTÜERYAL YAPAY SİNİR AĞLARI
YÖNTEMLERİ İLE DEĞERLENDİRİLMESİ: TAMAMLAYICI
SAĞLIK SİGORTASI UYGULAMASI**

**EVALUATION OF RUIN PROBABILITY USING GENERALIZED
LINEAR MODEL AND COMBINED ACTUARIAL NEURAL
NETWORKS METHODS: A COMPLEMENTARY HEALTH
INSURANCE APPLICATION**

NERMİN ÖDÜL OKUNAKOL

DR. ÖĞR. ÜYESİ BAŞAK BULUT KARAGEYİK

Tez Danışmanı

Hacettepe Üniversitesi

Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin

Aktüerya Bilimleri Anabilim Dalı için Öngördüğü

YÜKSEK LİSANS TEZİ olarak hazırlanmıştır.

ÖZET

İFLAS OLASILIĞININ GENELLEŞTİRİLMİŞ DOĞRUSAL MODELLER VE BİRLEŞİK AKTÜERYAL YAPAY SİNİR AĞLARI YÖNTEMLERİ İLE DEĞERLENDİRİLMESİ: TAMAMLAYICI SAĞLIK SİGORTASI UYGULAMASI

NERMİN ÖDÜL OKUNAKOL

Yüksek Lisans, Aktüerya Bilimleri Bölümü

Tez Danışmanı: Dr. Öğr. Üyesi Başak BULUT KARAGEYİK

Ocak 2024, 86 sayfa

Fiyatlandırma, sigorta sektöründe, özellikle de risk değişkenlerinin kapsamlı bir şekilde dikkate alındığı sağlık sigortasında çok önemli bir rol oynamaktadır. Fiyatlandırmanın önemi, sağlıkla ilgili potansiyel olaylarla ilişkili mali yükü doğru bir şekilde değerlendirme ve tahsis etme kabiliyetinde yatmaktadır. Sağlık sigortası fiyatlandırması yaş, önceden var olan koşullar, yaşam tarzı seçimleri ve bölgesel sağlık hizmeti maliyetleri gibi çeşitli risk faktörlerini dikkate alır. Sigorta şirketleri, bu değişkenleri fiyatlandırma modeline dahil ederek, bir bireyin veya grubun sağlık hizmeti ihtiyaçlarını karşılamamanın beklenen maliyetini yansıtan primler belirleyebilir. Ayrıca, doğru fiyatlandırma, sigortacıların riski etkin bir şekilde yönetmesine olanak tanıyarak primlerin beklenen talepleri karşılamaya yeterli olmasını ve aynı zamanda kârlılığını sürdürmesini sağlar. Sağlık sigortası poliçelerinin uygun şekilde fiyatlandırılmaması, daha yüksek sağlık riskine sahip bireylerin sigortaya kaydolma olasılığının daha yüksek olduğu ters seçime yol açabilir, bu da hasar maliyetlerinin artmasına ve sigortacı için mali kayıplara neden olur. Bu nedenle, risk değişkenlerine dayalı fiyatlandırma, sağlık sigortası piyasasının istikrarını ve sürdürülebilirliğini korumak için gereklidir ve sonuçta hem sigortacılara hem de poliçe sahiplerine fayda sağlar. Gelişmiş risk değerlendirmesi ve prim hesaplaması için sağlık sigortası alanında Genelleştirilmiş Doğrusal Modeller ve

Makine Öğrenimi metodolojilerinin kullanılmasını amaçlanmıştır. Genellikle GLM'lere dayanan geleneksel aktüeryal yaklaşımlar, çeşitli risk faktörleri ve sigorta talepleri arasındaki ilişkiyi modellemek için sağlam bir temel oluşturmaktadır. Bunun yanı sıra, gelişmiş makine öğrenimi tekniklerinin ortaya çıkmasıyla birlikte, sağlık sigortası ortamında tahmin doğruluğunu artırma ve karmaşık modelleri yakalama fırsatı da artmıştır. Bu kapsamda, aktüerlerin sıklıkla sigorta ürünlerinin fiyatlamasında kullandığı GLM'e ek olarak farklı bir yaklaşımla GLM çıktıları makine öğrenmesi algoritmalarından yapay sinir ağlarına yerleştirilerek modelin tahmininin güçlendirilmesi hedeflenmektedir.

Tez çalışmasına öncelikle GLM'lerin aktüerya bilimi ve sağlık sigortası alanındaki uygulamalarını vurgulayarak, kapsamlı bir GLM incelemesi ile başlanmaktadır. Daha sonra, GLM'lerin tahmin yeteneklerini tamamlamak ve genişletmek için makine öğrenimi algoritmalarından yapay sinir ağlarının entegrasyonu incelenerek birleşik aktüeryal yapay sinir ağları açıklanmıştır. Bu kapsamda, model ve algoritma birleşiminin, modellerin doğrusal olmayan ilişkileri, değişkenler arasındaki etkileşimleri nasıl etkili bir şekilde ele alabileceğini ve verilerdeki karmaşık kalıpları nasıl yakalayabileceği araştırılmaktadır. Önerilen yaklaşıma ilişkin analiz Türkiye'de faaliyet gösteren bir sigorta şirketine ait gerçek tamamlayıcı sağlık sigortası verisi kullanılarak gerçekleştirilmiştir. Çalışmada, öncelikle veri yapısına bakılmış ve veriyi modellemeye uygun hale getirmek için bazı düzenlemeler yapılmıştır. Hasar sayısına üst sınır belirlenmesi ve bazı sürekli değişkenlerin kategorize edilmesi bu düzenlemelere örnek olarak verilebilmektedir.

Klasik risk modeli varsayımı altında, hem frekans hem de şiddet için Birleşik Aktüeryal Yapay Sinir ağları modeli oluşturulmuş ve Genelleştirilmiş Linear Model için anlamlı açıklayıcı değişkenler belirlenerek en uygun modeller saptanmıştır. Elde edilen GLM çıktıları yapay sinir ağları için girdi nöronları olarak, YSA'nın katmanlı yapısına uyum sağlaması sağlanmıştır. Böylece kapalı kutu model olarak adlandırılan yapay sinir ağlarının hem açıklanabilirliği arttırılmış hem de tahmin gücü yükseltilmiştir. Bu işlem frekans ve şiddet modelleri için ayrı ayrı yapılmıştır. Çalışmada ayrıca, tahmin doğruluğu, model yorumlanabilirliği ve dinamik sağlık verilerine uyarlanabilirliği açısından GLM-Makine Öğrenimi birleşiminin geleneksel GLM'lere karşı performansı da değerlendirilmiştir. Makine öğrenimi tekniklerinin yerleşik GLM metodolojileriyle bütünleştirilmesinin etkisinin ölçülmesinde, klasik GLM ile yapılan modelleme ve birleşik aktüeryal yapay sinir ağlarına bağlı tahmin edilen primler hasar tutarını

karşılımları yönünden iflas olasılıkları hesaplanarak karşılaştırılmıştır. Çalışmanın uygulama aşamasında ayrıca, makine öğreniminin aktüeryal iş akışlarına dâhil etmenin etkilerinin, hem teknik yönleri hem de sigorta sektöründeki pratik uygulanabilirliği göz önünde bulundurarak analiz gerçekleştirilmiştir.

Çalışma, risk değerlendirme doğruluğunu artırmak ve sağlık sigortası sektöründe daha bilinçli karar verme süreçlerini teşvik etmek için gelişmiş modelleme yaklaşımlarının aksine birleşik aktüeryal yapay sinir ağlarının uygulanabilirliğini araştırmayı amaçlamaktadır.

Anahtar Kelimeler: Genelleştirilmiş Lineer Modeller, Derin öğrenme, Yapay Sinir Ağları, Bileşik Aktüeryal yapay sinir ağları, İflas Olasılığı

ABSTRACT

EVALUATION OF RUIN PROBABILITY USING GENERALIZED LINEAR MODEL AND COMBINED ACTUARIAL NEURAL NETWORKS METHODS: A COMPLEMENTARY HEALTH INSURANCE APPLICATION

NERMİN ÖDÜL OKUNAKOL

Master of Science, Department of Actuarial Sciences

Supervisor: Assist Prof. Dr. Başak BULUT KARAGEYİK

January 2024, 86 pages

In the insurance business, pricing is essential, especially for health insurance, where risk factors are taken into account significantly. Pricing is important because it can fairly estimate and distribute the cost of future health-related incidents. The cost of health insurance is determined by taking into consideration a number of risk variables, including age, lifestyle choices, pre-existing diseases, and local healthcare expenses. Insurance firms can set prices that represent the anticipated cost of meeting a person's or a group's healthcare needs by including these factors into the pricing model. Furthermore, accurate pricing helps insurers control risk by guaranteeing that premiums are high enough to pay for expected claims without compromising revenue. Inadequate pricing of health insurance policies may result in adverse selection, which increases the likelihood of enrollment for people with greater health risks and raises claims expenses and insurer losses. In light of this, risk-based pricing is crucial to preserving the health insurance market's sustainability and stability, which eventually benefits both insurers and policyholders. This study aims to use Generalized Linear Models (GLM) and Machine Learning (ML) methodologies in the field of health insurance for advanced risk

assessment and premium calculation. Traditional actuarial approaches, typically based on GLMs, provide a robust foundation for modeling the relationship between various risk factors and insurance claims. Additionally, with the emergence of advanced machine learning techniques, there are increased opportunities to enhance prediction accuracy and capture complex models in the health insurance environment. In this context, the objective is to strengthen the model's prediction by integrating GLM outputs into artificial neural networks, a different approach in addition to the frequently used GLM in insurance product pricing by actuaries.

The thesis begins with a comprehensive examination of GLMs, emphasizing their applications in actuarial science and the field of health insurance. Subsequently, it explores the integration of artificial neural networks from machine learning algorithms to complement and expand the predictive capabilities of GLMs, explaining the Combined Actuarial Neural Networks (CANN). This study investigates how the combination of models and algorithms can effectively address nonlinear relationships, interactions between variables, and capture complex patterns in data. The proposed approach's analysis is conducted using real complementary health insurance data from a Turkish insurance company. Initially, the data structure is examined, and adjustments are made to fit the modeling, such as setting upper limits for claim numbers and categorizing some continuous variables.

Under the assumption of a classical risk model, a CANN model for both frequency and severity is developed, and the most suitable models are determined with significant explanatory variables for GLM. The obtained GLM outputs are aligned to the layered structure of artificial neural networks as input neurons. This process has increased the explainability and prediction power of the so-called black box models of artificial neural networks, implemented separately for frequency and severity models. The study also evaluates the performance of the GLM-Machine Learning combination against traditional GLMs in terms of prediction accuracy, model interpretability, and adaptability to dynamic health data. The impact of integrating machine learning techniques with established GLM methodologies is measured by examining the effect of modeling results using classic GLM on the profitability of the insurance company and comparing bankruptcy probabilities according to both approaches. Additionally, in the application phase of the study, the effects of incorporating machine learning into actuarial workflows

are analyzed, considering both technical aspects and practical applicability in the insurance sector.

The study aims to investigate the applicability of combined actuarial neural networks as opposed to advanced modeling approaches to enhance risk assessment accuracy and encourage more informed decision-making processes in the health insurance sector.

Keywords: Generalized Linear Models, Deep Learning, Artificial Neural Networks, Combined Actuarial Neural Networks, Ruin Probability

TEŐEKKÜR

Tez alıřmamın her ařamasında benimle heyecanımı paylařan, deęerli bilgi ve deneyimiyle beni ynlendiren Dr. ęr. yesi Bařak BULUT KARAGEYİK'e

Deęerli eleřtiri ve katkıları iin tez savunma jri yelerim Prof. Dr. Sevtap SELUK KESTEL, Prof. Dr. Duru KARASOY'a

Bana tm kalbiyle inanan biricik ailem annem Filiz SARI ve babam Okay SARI'a

Hayatımdaki en byk destekim, yol arkadařım, sevgili eřim Alper OKUNAKOL'a

Beni mesleęimle tanıştıran zel insan amcam Okan SARI'a

Teőekkrlerimi sunarım.

İÇİNDEKİLER

ÖZET.....	i
ABSTRACT	iv
TEŞEKKÜR.....	vii
İÇİNDEKİLER.....	viii
ŞEKİLLER DİZİNİ.....	xi
ÇİZELGELER DİZİNİ	xiv
KISALTMALAR	xv
1.GİRİŞ	1
2. GENELLEŞTİRİLMİŞ DOĞRUSAL MODELLER.....	6
2.1. Genelleştirilmiş Doğrusal Modellerin Yapısı	7
2.1.1. Üstel Dağılım Ailesi.....	8
2.1.1.1. Poisson Dağılımı	8
2.1.1.2. Gamma Dağılımı	9
2.2. Varyans fonksiyonu.....	10
2.3. Ağırlık	10
2.4. Tweedie Dağılımı.....	10
2.5. Bağ Fonksiyonları (Link Functions)	11
2.5.1. Normal Bağ Fonksiyonlar	11
2.5.2. Log-Bağ fonksiyonu.....	11
2.5.3. Kanonik Bağ.....	11
2.6. En çok Olabilirlik Tahmini	12
2.6.1. Tahmin Edicilerin Özellikleri.....	12
2.6.1.1. Tekrarlanan Örnekleme	12
2.6.1.2. Yansızlık.....	12
2.6.1.3. Tutarlılık.....	12
2.6.1.4. Ortalama Hata Karesi	13
2.6.2. En çok Olabilirlik Tahmin Edicisi	13

2.6.2.1. Değişmezlik	14
2.6.2.2. Asimptotik Yansızlık	14
2.6.2.3. Minimum Varyanslılık.....	14
2.7. Fisher Bilgi	15
2.8. Ofset (Dengeleme Terimi)	15
2.9. Genelleştirilmiş Doğrusal Modellerin Aktüeryal Uygulamaları: Model Seçimi.	16
3. DERİN ÖĞRENME	17
3.1. Derin Öğrenme ve Temsil Öğrenimi	18
3.2. Yapay Sinir Ağları	18
3.2.1. İleri Beslemeli Ağlar.....	18
3.3. Derin İleri Beslemeli Sinir Ağları.....	20
3.3.1. Birleştirilmiş Aktüeryal Yapay Sinir Ağları.....	20
4. İFLAS TEORİSİ.....	21
4.1. Risk Süreci.....	21
4.2. İflas Olasılığı.....	21
4.3. İflas Olasılığının Değerlendirilmesi.....	22
5. UYGULAMA	23
5.1. Veri Seti ve Tanımlayıcı İstatistikler	23
5.1.1. Hasar Frekansına İlişkin Bazı İstatistikler	24
5.1.2. Hasar Şiddetine İlişkin Bazı İstatistikler	30
5.2. GLM Uygulaması	35
5.2.1. GLM Frekans Model Kurulumu ve Değerlendirilmesi	35
5.2.2. Poisson Log Bağlantı Fonksiyonu ve Seçim Kriterleri	35
5.2.3. GLM Frekans Yaklaşımı ile Değişken Seçimi ve Model Optimizasyonu....	36
5.2.4. Genelleştirilmiş Lineer Modellerle Hasar Sıklığının Analizi	36
5.2.5. Hasar Sıklığı Modellerinin Performans Karşılaştırması.....	37
5.2.6. Olabilirlik Oranı Testi ile GLM3 Model Değişkenlerinin Hasar Sıklığı Üzerindeki Etkisinin Değerlendirilmesi	38
5.2.6.1. GLM Frekans Modeli Karşılaştırmalı Analizi.....	39
5.2.6.2. GLM Frekans Tahminlerinin Logaritmik Ölçekte Hasar Frekansı Tahminlerinin Karşılaştırmalı Analizleri.....	42
5.2.7 GLM Şiddet Model Kurulumu ve Değerlendirilmesi.....	44

5.2.8. Gamma Log Bağlantı Fonksiyonu ve Seçim Kriterleri.....	45
5.2.9. GLM Şiddet Yaklaşımı ile Değişken Seçimi ve Model Optimizasyonu	47
5.2.10. GLM4 Modeli ile ANOVA Sonuçlarının İncelenmesi	48
5.2.11. LRT ile GLM4 Model Değişkenlerinin Hasar Tutar Güncel Sıklığı Üzerindeki Etkisinin Değerlendirilmesi.....	48
5.2.10.1 GLM Şiddet Modeli Karşılaştırmalı Analizi.....	49
5.2.10.1 GLM Şiddet Tahminlerinin Logaritmik Ölçekte Karşılaştırılması	50
5.3. Birleşik Aktüeryal Sinir Ağları	54
5.3.1. Sıralanmamış (Nominal) Kategorik Özellik Bileşenleri	55
5.3.2. Sürekli Özellik Bileşenleri	56
5.3.3. İkili Özellik Bileşenleri	56
5.3.4. Veri Ön İşleme ve Özellik Mühendisliği	57
5.3.5. CANN Hasar Frekans Modellenmesi.....	58
5.3.6. CANN Eğitim Sürecinde Hasar Frekans Değerlerinin Değişimi.....	59
5.3.7. CANN Modelinin Çeşitli Demografik Özelliklere Göre Frekans Tahmin Performansı	61
5.3.7. CANN Hasar Şiddetinin Modellenmesi.....	64
5.3.8. CANN Eğitim Sürecinde Hasar Şiddetinin Değerlerinin Değişimi	65
5.3.9. CANN Eğitim Sürecinde Hasar Değerlerinin Performansı.....	66
5.4. CANN ve GLM Modellerinin Karşılaştırılması.....	68
5.4.1. CANN ve GLM Performans Ölçütlerinin Karşılaştırılması.....	68
5.4.2. İflas Durumunun (Hasar-Prim Yeterliliği) Değerlendirilmesi	70
6. SONUÇ VE ÖNERİLER	72
KAYNAKÇA	74
EKLER	76
Ek 1 - Genelleştirilmiş Lineer Modellerle Hasar Sıklığı Tablosu.....	76
Ek 2 - Genelleştirilmiş Lineer Modellerle Hasar Güncel Tutar Sıklığı Tablosu	83
ÖZGEÇMİŞ	86

ŞEKİLLER DİZİNİ

Şekil 5.1. Hasar frekans bilgileri	25
Şekil 5.2. Yeni iş ve yenileme değişkenlerine bağlı hasar sıklıkları	26
Şekil 5.3. İl grup değişkenine hasar sıklıkları.....	27
Şekil 5.4. Coğrafi bölge değişkenine bağlı hasar sıklıkları	27
Şekil 5.5. Yaş değişkenine bağlı hasar sıklıkları	28
Şekil 5.6. Yaş grup değişkenlerine bağlı hasar sıklıkları.....	28
Şekil 5.7. Cinsiyet değişkenine bağlı hasar sıklıkları	29
Şekil 5.9. Meslek grup değişkenine bağlı hasar sıklıkları	30
Şekil 5.10. Yenileme grup değişkenine bağlı hasar şiddeti bilgisi	30
Şekil 5.11. İl grup değişkenine bağlı hasar şiddeti bilgisi	31
Şekil 5.12. Bölge grup değişkenine bağlı hasar şiddeti bilgisi	31
Şekil 5.13. Yaş değişkenine bağlı hasar şiddeti bilgisi.....	32
Şekil 5.14. Yaş grubu değişkenine bağlı hasar şiddeti bilgisi	32
Şekil 5.15. Cinsiyet değişkenine bağlı hasar şiddeti bilgisi.....	33
Şekil 5.17. Medeni durum değişkenine bağlı hasar şiddeti bilgisi	34
Şekil 5.18 Türkiye'deki il ve bölge bazındaki hasar veri değişimi	34
Şekil 5.19 Gözlenen frekans poisson modeli.....	35
Şekil 5.20 Poliçe yenileme durumu ve cinsiyete göre glm modelleri ile hasar frekansı tahmini	39
Şekil 5.21 Yaş gruplarına göre hasar frekansının gözlemlenmesi.....	40
Şekil 5.22 Medeni duruma göre hasar frekansının gözlemlenmesi	41
Şekil 5.23 Türkiye'deki illerin hasar frekans gözlemlenmesi	41
Şekil 5.24. GLM2 ve GLM1 modellerinin karşılaştırmalı değerlendirilmesi	42
Şekil 5.25. GLM1 ve GLM3 modellerinin karşılaştırmalı değerlendirilmesi	43
Şekil 5.26. GLM2 ve GLM3 modellerinin karşılaştırmalı değerlendirilmesi	43
Şekil 5.27 Gamma dağılımının sigorta hasar tutarları üzerindeki uygulanabilirliğinin istatistiksel değerlendirilmesi	45
Şekil 5.28 Yenileme gruplarına göre hasar şiddetinin gözlemlenmesi.....	49
Şekil 5.29 Yaş gruplarına göre hasar şiddetinin gözlemlenmesi	49
Şekil 5.30 Türkiye'deki il gruplarına göre hasar şiddetinin gözlemlenmesi	50

Şekil 5.31. GLM1 ve GLM2 Modellerinin hasar tutarı güncel şiddeti tahminlerinin logaritmik ölçekte karşılaştırılması	50
Şekil 5.33. GLM3 modellerinin hasar tutar güncel şiddet tahminlerinin logaritmik ölçekte karşılaştırmalı değerlendirmesi	51
Şekil 5.34. GLM3 ve GLM4 modellerinin hasar tutar güncel şiddet tahminlerinin logaritmik ölçekte karşılaştırmalı değerlendirmesi	52
Şekil 5.35. GLM ile sigorta primi tahmin modelinin gerçekleştirme başarısı ve potansiyel iflas riski analizi	53
Şekil 5.36. CANN hasar frekans modelinin akış diyagramı	58
Şekil 5.37. CANN eğitim sürecinde glm tüm değişkenler hasar frekans değerlerinin değişimi-1	59
Şekil 5.38. CANN eğitim sürecinde GLM'deki tüm değişkenlerin hasar frekans değerlerinin değişimi-2	60
Şekil 5.40. CANN modelinin cinsiyet bazında frekans tahmin performansını göstermektedir.	61
Şekil 5.40. CANN modelinin cinsiyet bazında frekans tahmin performansı	61
Şekil 5.41. CANN modelinin yaşgrup bazında frekans tahmin performansını gösterilmektedir.	61
Şekil 5.41. CANN modelinin yaşgrup bazında frekans tahmin performansı	61
Şekil 5.42. CANN modelinin medeni durum bazında frekans tahmin performansı	62
Şekil 5.43'te CANN modelinin çeşitli demografik özelliklere göre frekans tahmin performansı gösterilmektedir.	62
Şekil 5.43. CANN modelinin çeşitli demografik özelliklere göre frekans tahmin performansı	62
Şekil 5.44'te CANN şiddet modelinin akış diyagramı gösterilmiştir.	63
Şekil 5.44. CANN şiddet modelinin akış diyagramı	63
Şekil 5.45. CANN eğitim sürecinde hasar şiddet değerlerinin değişim analizi-1	65
Şekil 5.46. CANN eğitim sürecinde hasar şiddet değerlerinin değişim-2 analizini göstermektedir.	66
Şekil 5.46. CANN eğitim sürecinde hasar şiddet değerlerinin değişim analizi-2	66
Şekil 5.47. CANN eğitim sürecinde yen iş/yenileme bazında şiddet değerlerinin değişimi analizi	66

Şekil 5.48. CANN eğitim sürecinde yasgrup bazında şiddet değerlerinin değişim analizi	67
Şekil 5.49. CANN eğitim sürecinde ilGrup bazında şiddet değerlerinin değişim analizi	67
Şekil 5.50. CANN ile sigorta primi tahmin modeline ilişkin prim karşılaştırması	68
Şekil 5.51. CANN ve GLM karşılaştırma grafiği	70

ÇİZELGELER DİZİNİ

Çizelge 2.1. Üstel dağılım ailesi ve parametreleri [7].....	9
Çizelge 2.2. Bağ fonksiyonları [7]	11
Çizelge 3.1. Aktivasyon fonksiyonları [18]	19
Çizelge 5.2. Hasar adedi ve kazanılmış adet bilgisi	25
Çizelge 5.3. İl grup değişken bilgisi.....	26
Çizelge 5.4. GLM ile değişken seçimi model tablosu.....	36
Çizelge 5.5. Hasar sıklığı modelinin performans karşılaştırma tablosu.....	37
Çizelge 5.6. GLM3 modeli sonuçları tablosu	37
Çizelge 5.7. LTR testi sonuçları tablosu	39
Çizelge 5.8. GLM şiddet değişken model tablosu	47
Çizelge 5.9. GLM4 modeli tablosu	48
Çizelge 5.10. LTR testi sonuçları tablosu	48
Çizelge 5.11. CANN ve GLM performans metrikleri.....	69

KISALTMALAR

Simgeler

θ (Theta)	Tahmin ediciler veya parametresi
μ (Mu)	Kanonik parametre veya gerçek değerli konum parametresi
β (Beta)	Regresyon katsayılarını veya parametresi
U	Sigorta şirketinin risk sürecinde kullanılan bir değişken
σ (Sigma)	Standart sapmayı veya dağılımın genişliğini ifade eder
ln	Logaritma fonksiyonu

Kısaltmalar

GLM	Genelleştirilmiş Lineer Model
YSA	Yapay Sinir Ağı
CANN	Birleştirilmiş Aktüeryal Yapay Sinir Ağları
ML	Makine Öğrenimi
SAV	İsviçre Aktüerler Birliği
TSB	Türkiye Sigorta Birliği
AIC	Akaike Bilgi Kriteri
LRT	Likelihood Ratio Testi - Olabilirlik Oran Testi
NAV	Net Aktif Değeri
ANOVA	Yapılan Varyans Analizi
TSS	Tamamlayıcı Sağlık Sigortası
MAPE	Ortalama Yüzdesel Mutlak Hata
MAE	Ortalama Mutlak Hata
RMSE	Kök Ortalama Kare Hata
MSE	Ortalama Kare Hata

1.GİRİŞ

Sağlık sigortası, bireyleri, aile üyelerini veya grup poliçeleriyle şirket vb. personelleri, tıbbi masrafların ve bazı sağlık hizmetlerinin maliyetlerine karşı koruyan bir sigorta türüdür. Genellikle sağlık sigortası, poliçesi, sahibinin hastalık veya yaralanma durumunda tedavi masraflarını kısmen veya tamamen karşılar. Sağlık sigortasının temel amacı, beklenmedik yüksek tıbbi masrafların mali yükünü azaltmaktır. Sağlık sigortaları 1 yıllık kısa dönemli hayat-dışı poliçeleridir. Türkiye'deki sağlık sigortası sektöründe ciddi bir büyüme olduğu ve bu büyümenin önemli bir kısmının tamamlayıcı sağlık sigortasından kaynaklandığı gözlemlenmiştir. Tamamlayıcı sağlık sigortasının pazar payı, artan farkındalık ve talep ile önemli ölçüde artmıştır. Kamu sağlık sigortalarında kapsam dışı veya kısmen karşılanan sağlık harcamalarının maliyetlerini azaltmak amacıyla *Tamamlayıcı Sağlık Sigortası(TSS)* ile bireylere sağlık güvencesi sunmaktadır. Türkiye Sigorta Birliği (TSB) tarafından yayımlanan 2022 yılı sektör raporundaki sonuçlara göre tamamlayıcı sağlık branşı için toplam prim üretimi bir önceki yıl 2,616,075,833 TL iken 2022 yılında 7,808,624,560 TL'ye ulaşmıştır. Sektör 2022 yılı için tamamlayıcı sağlık sigortası prim üretiminde ise 2021 yılına göre %198,5'lik büyük bir artış göstermiştir. 2022 yılında TSS sigortalı sayısı 3 milyonu aşmış ve yılsonunda 2021 yılına göre %39 oranında artış göstermiştir [1].

Ülkemizde artan sağlık enflasyonu, sağlık hizmetlerine erişebilirlikteki kaygılar ve Covid-19 salgını nedenleri gibi bazı etkenlere bağlı olarak özel sağlık sigortalarına olan talepte artış gözlenmiştir. Bu durum birçok sigorta şirketinin özellikle TSS ürününe ilişkin portföy hacminin büyümesine sebep olmuştur. Genişleyen TSS portföyleri beraberinde hasar-prim oranlarının yükselmesine neden olmuştur. Yüksek hasar-prim oranları ise sigorta şirketinin karlılığını tehdit eden bir durum haline gelmiştir. Bu nedenle, aktüerler için sağlık sigortaları fiyat modellerinin oluşturulması ve adil primin belirlenmesi konusu gündeme gelmiştir.

Sağlık sigortası priminin hesaplanması, kapsam için uygun maliyeti belirlemek üzere çeşitli faktörlerin titiz bir değerlendirmesini içerir. Sigortacılar sigortalı bireyin yaşı, tıbbi geçmişi, yaşam tarzı ve konumu gibi temel unsurları dikkate alır. Buna ek olarak, seçilen teminat seviyesi, muafiyet gibi faktörler de prim hesaplamasında önemli bir rol oynar. Genellikle temel sağlık sigortası planının ötesinde teminat sağlayan tamamlayıcı sağlık sigortası sağlık sigortasına benzer bir yapı sunmakla birlikte daha kapsamlı bir

değerlendirme gerektirir. Tamamlayıcı sağlık sigortasının primi, sigortalı birey tarafından seçilen belirli ek faydalar ve teminat seçeneklerinden etkilenir. Bunlar arasında standart sağlık sigortası planları tarafından tipik olarak kapsanmayan diğer hizmetlerin yanı sıra, tamamlayıcı sağlık sigortasının geçerli olduğu bölge, hastane, bölüm ve doktor kısıtlamalarının yanı sıra, sigorta kullanımında getirilen sayı sınırlaması yer alabilir. Karmaşıklık, tamamlayıcı planlarda mevcut olan özelleştirmeden kaynaklanmaktadır. Bireyler teminatı kendi sağlık ihtiyaçlarını karşılayacak şekilde uyarlayabilirler. Sonuç olarak, sigortacılar tamamlayıcı sağlık sigortası primini belirlerken bu özel hizmetlerle ilişkili maliyetleri de hesaba katmalıdır. Bu incelikli yaklaşım, temel sağlık sigortası kapsamının ötesinde kapsamlı teminat arayan bireyler arasındaki sağlık hizmeti ihtiyaçlarının ve tercihlerinin çeşitliliğini yansıtmaktadır. Bu dengenin kurulması, bireylerin kapsamlı teminatları almasını sağlarken sigorta sisteminin mali açıdan yaşayabilirliğini de korumaktadır. Aktüerler, riski analiz etmek ve sigortalı için satın alınabilirliği sigortacı için sürdürülebilir iş uygulamaları ile dengeleyen primleri belirlemek için aktüeryal modeller ve istatistiksel veriler kullanırlar.

GLM kullanılarak sağlık sigortası primlerinin hesaplanması, uygun teminat maliyetini belirlemek için istatistiksel modellemeden yararlanan güçlü bir yaklaşımı temsil eder. GLM, sigortacıların çeşitli faktörleri eş zamanlı olarak analiz etmesine ve birleştirmesine olanak tanıyarak riskin daha kapsamlı ve doğru bir şekilde değerlendirilmesini sağlamaktadır. Bu bağlamda GLM, yaş, cinsiyet, tıbbi geçmiş, yaşam tarzı ve diğer ilgili faktörler gibi çeşitli değişkenlerin entegrasyonuna izin verir. Model, her bir değişkene genel risk profili üzerindeki etkisine göre uygun ağırlıklar atayarak sigortacıların primleri bireyselleştirilmiş risk faktörlerine göre uyarlamasına olanak tanır. Ayrıca GLM, doğrusal olmayan ilişkileri barındırarak farklı değişkenler arasındaki karmaşık etkileşimlerin daha gerçekçi bir temsilini sağlar. GLM'nin esnekliği, sigortacıların gelişen sağlık hizmetleri ortamına ve değişen demografik özelliklere uyum sağlamaları için onu güçlü bir araç haline getirmektedir. Sonuç olarak, GLM, sağlık sigortası primlerinin hesaplanmasında, hassasiyeti, adaleti ve fiyatlandırma yapılarını sigortalı bireylerin çeşitli ihtiyaçlarını karşılayacak şekilde ayarlama becerisini artırır. GLM, 1972 yılında Nelder ve Wedderburn [2] tarafından tanıtıldığından beri tahmine dayalı modellerde çok sayıda araştırmaya temel oluşturmuştur. 1996 yılında Haberman ve Renshaw [3], 2008 yılında De Jong ve Heller [4], 2010 yılında Frees [5] ve 2010 yılında Ohlsson ve Johansson [6] tarafından yapılan çalışmalarda hem hayat hem de hayat dışı sigortalarda

aktüeryal problemlerin çözümüne yönelik çok sayıda GLM uygulaması sunulmuştur. Ohlsson ve Johansson [6] çalışmasında bu değişkenler, çok sayıda düzeyi olan kategorik faktörler olarak yorumlanmıştır.

Ancak, istatistiksel modellerde kategorik faktörlerin dâhil edilmesi, bu düzeylerin bazıları için seyrek verilerin mevcut olduğu durumlarda güvenilir olmayan sonuçlara yol açabilmektedir. Sürekli değişkenleri modele dâhil etmek amacıyla genelleştirilmiş toplamsal modeller kullanılabilir. Bir başka yaklaşım, 2019 yılındaki Denuit [7]'in çalışmasındaki gibi, sürekli değişkenlere bir sınıflandırma yöntemi uygulanarak kategorik değişkenlere dönüştürebileceğidir.

Diğer sigortalardan farklı olarak sağlık sigortaları aynı tarife içinde birden çok sigorta planı da içerebilir. Bunlar sağlık hizmeti kapsamı ve faydalanan hastane sayısına göre değişkenlik gösterebilmektedir. Kapsamlı teminat sağlamak ve finansal sürdürülebilirliği sürdürmek arasında doğru dengeyi kurmak, bu çeşitli faktörlerin incelikli bir şekilde anlaşılmasını ve modellenmesini gerektirir. Sağlık hizmetleri ortamının sürekli evrimi ve bireysel sağlık profillerinin dinamik yapısı, sağlık sigortası primlerinin hesaplanmasını, uyarlanabilirlik ve hassasiyetin artması nedeniyle zorlaştırmaktadır.

Sağlık sigortası primlerinin hesaplanması makine öğrenimi algoritmalarının entegrasyonu ile dönüştürücü bir değişim geçirmiştir. Makine öğrenimi, risk faktörlerini daha doğru bir şekilde tahmin etmek ve değerlendirmek için geniş veri kümelerinden ve karmaşık modellerden yararlanarak sürece yeni bir karmaşıklık düzeyi getirmiştir.

Makine öğrenimi modelleri, yalnızca geleneksel demografik bilgilere dayanmak yerine, bireysel sağlık geçmişleri, yaşam tarzı seçimleri ve hatta gerçek zamanlı sağlık izleme verileri de dâhil olmak üzere çok sayıda değişkeni analiz edebilmektedir. Bu algoritmalar sürekli olarak öğrenilebilir ve yeniden adapte olabilmesi nedeniyle, yeni verilerle karşılaştıkça tahminlerini zaman içinde iyileştirebilmektedir. Makine öğreniminin tahmin gücü, sigortacıların primleri ayrıntılı bir düzeyde özelleştirmesine ve maliyetleri bireysel risk profillerine göre ayarlamasına olanak tanımaktadır. Bu, yalnızca fiyatlandırmanın adilliğini artırmakla kalmaz, aynı zamanda bireylerin benzersiz sağlık ihtiyaçlarıyla yakından uyumlu bir teminat almalarını da sağlamaktadır. Makine öğreniminin dinamik yapısı, gerçek zamanlı ayarlamalara olanak tanıyarak sigortacıların gelişen sağlık trendlerine ve bireysel sağlık durumlarına daha hızlı uyum sağlamalarını sağlar. Özünde, makine öğreniminin sağlık sigortası primlerinin hesaplanmasına dâhil edilmesi, sağlık

hizmetlerinin sürekli gelişen ortamında doğruluğu, adaleti ve uyarlanabilirliği optimize eden çığır açan bir yaklaşımı temsil etmektedir.

Aktüerya bilimlerinde makine öğrenimi yöntemlerinin uygulanması ile ilgili olarak birçok çalışma bulunmaktadır. 2020 yılı Gabrielli vd. [8] ve 2021 yılında Blier-Wong, Yapay Sinir Ağları (YSA) araştırılması ve teorilerin incelenmesi üzerine çalışmışlardır. 2003 yılında Dugas vd. [9], YSA modelinin birinci avantajının üstün istatistiksel performansa sahip olması, ikinci avantajının ise karmaşık etkileşimlerinin otomatik modellenmesi olduğunu belirtmiştir. İlk avantaj, literatürde fikir birliği olmadığı için bilimsel olarak henüz belirlenmemiştir (2003 yılı Dugas vd. [9], 2002 yılı Mano ve Rasa [10], 2021 yılı Blier-Wong vd. [11]). İkinci avantaj ise daha az belirsizdir. Çünkü GLM ile modelleme yapılırken etkileşimler manuel olarak modellenir. Bu da GLM'in karmaşık ilişkileri keşfetme becerisiyle sınırlı bir süreç olmasına neden olmaktadır. Böylece YSA potansiyel olarak bu alanda hem zaman hem de performans açısından GLM'e göre iki kat iyileştirme sağlamaktadır.

Belirtilen avantajlara rağmen, genel olarak makine öğrenimi modellerinin ve YSA modelinin uygulanması sigorta şirketlerinde henüz yeterince yer bulamamıştır. Ana dezavantajı YSA modellerinin yorum eksikliğidir. YSA modelleri yalnızca girdi ve çıktılardan analiz edilebildikleri için genellikle kara kutu modelleri olarak adlandırılır.

Makine öğrenmesi modelleri, modelin sonuca nasıl vardığını açıklamadan basitçe bir tahmin üretmektedir. Sigorta şirketlerinin karşılaşılabileceği riskleri belirleyebilmesi ve müşterinin nihai primine neyin katkıda bulunduğunu açıklayabilmesi önem arz etmektedir. GLM 'in aksine eğitim sürecinin stokastik davranışı göz önüne alındığında, iki YSA modeli benzer istatistiksel performansa sahip olabilir ancak aynı veri seti için farklı tahminler verebilir.

Yapay sinir ağlarının avantaj ve dezavantajlarına ilişkin tartışmada, 2019 yılında Wüthrich ve Merz [12], her iki yaklaşımın da en iyisini elde etme çabasıyla klasik GLM'yi YSA ile birleştiren yeni bir model önerdiler. Bu model, Birleşik Aktüeryal Sinir Ağı (CANN) olarak adlandırılmıştır. Pratikte bir GLM modelinin, başlangıç noktasından çok fazla sapmadan GLM modelinde bir gelişme sağlamak amacıyla bir atlama bağlantısı (skip-connection) kullanılarak bir YSA'na yerleştirilebileceğini önermektedir. Bu aynı zamanda yapay sinir ağlarının GLM'de dikkate alınmayan özelliklerinin arasındaki etkileşimleri keşfetmesine de olanak tanımaktadır. Wüthrich ve Merz [12], modellerin

GLM model parametresinin YSA'da eğitilebilir olup olmayacağına karar verebileceği CANN modelinin çeşitli yapılarını da önermiştir. Schelldorfer ve Wüthrich [13] tarafından İsviçre Aktüerler Birliği (SAV) için yayınlanan bir çalışmada hasar frekansı modellemesi ve Gabrielli vd. [8] hasar rezervi modeli hakkındaki makalesinde gerçekleştirilen Birleşik Aktüeryal Yapay Sinir ağlarını sunmuşlardır. Schelldorfer ve Wüthrich [13]. CANN, yaklaşımının eksik etkileşimlerinin sistematik olarak tanımlanmasını ve dolayısıyla GLM modelinden gelişmiş bir istatistiksel performans sağladığını belirtmiştir. Ayrıca Gabrielli vd. [8], önerilen modelin diğer bazı faydalarına, yani otomatik bir parametre seçim süreci ve gradyan iniş algoritmasının hızlı yakınsama, tahmin doğruluğunu incelemek için önyükleme yöntemlerine olanak sağladığına dikkat çekmiştir. Laporta [14] sağlık sigortası verisi ile CANN modeli için frekans-şiddet yaklaşımını uygulamıştır.

CANN yaklaşımı, GLM modelini temel aldığı için "en iyi model belirsizliğini" azaltırken, YSA'ların karmaşık değişken etkileşimlerini otomatik olarak tanımlama avantajını da sağlamaktadır. Ancak, YSA'nın yorumlanabilir olmaması durumu Birleşik Aktüeryal Sinir Ağı modelinde devam etmektedir. Aktüerler için kara kutu modellerinin uygulanabilirliğini değerlendirmek kilit bir konudur. Bu modelleri daha iyi anlamak için aktüeryal bağlamda modelin nasıl açıklanacağı da büyük önem arz etmektedir. Değişkenlerin etkisi ve değişkenlerin etkileşimleri, açıklanması gereken iki ana alandır. Literatürde, kara kutu modellerinin yorumlanmasını sağlayan çok sayıda yöntem ve araç bulunmaktadır. (2018 yılı Guidotti vd. [15] ve 2020 yılı Molnar [16] çalışmaları bulunmaktadır.

Bu tez çalışmasının birinci bölümünde literatür taramasına yer verilmiştir. İkinci bölümde genelleştirilmiş doğrusal modeller detaylı olarak anlatılmıştır. Tezin üçüncü bölümünde makine öğrenmesi algoritmalarından yapay sinir ağlarının teorisi aktüeryal kullanım alanlarından ve CANN modellerden bahsedilmiştir. Tezin dördüncü bölümü, GLM ve CANN modelini içeren bir analize dayanmaktadır. Analiz de Türkiye'de faaliyet gösteren özel bir sigorta şirketinin tamamlayıcı sağlık sigortası verisi kullanılmıştır. İlgili veri için GLM hasar frekansı modeli ve hasar şiddeti modeli kurulmuştur. GLM modelinden elde edilen katsayılar, Yapay sinir ağlarına yerleştirilerek bileşik aktüeryal CANN modeliyle prim tahmini yapılmıştır. GLM' den elde edilen primlerle ve CANN ile tahmin edilen primler karşılaştırılarak şirketin iflas durumu değerlendirilmiştir. Tezin beşinci ve son bölümünde çalışmanın çıktıları sunulmuş ve sonuçlar yorumlanmıştır.

2. GENELLEŐTİRİLMİŐ DOĐRUSAL MODELLER

GenelleŐtirilmiŐ Dođrusal Modeller (GLM), sađlık sigortası da dâhil olmak üzere çeŐitli alanlarda yaygın olarak kullanılan istatistiksel model sınıfıdır. GLM'ler, deđiŐkenler arasındaki iliŐkileri modellemek ve tahminler yapmak için esnek bir çerçeve sađladığı için aktüeryal hesaplamalarda sıklıkla tercih edilir. GLM, aktüeryal bilimlerde hasar tahminleri ve risk deđerlendirmeleri için kritik bir öneme sahiptir. GLM, tamamlayıcı sađlık sigortası uygulamalarında ise özellikle hasar verilerinin analizinde etkin bir rol oynamaktadır.

Sađlık sigortası alanında GLM'ler, fiyatlandırma, sigortalama ve risk deđerlendirmesi gibi birkaç kilit alana uygulanabilir. Aktüerler, sađlık sigortası poliçeleri için fiyatlandırma modelleri geliŐtirmek üzere GLM'leri kullanabilir. Bu modellerde, poliçe sahipleri için beklenen hasar maliyetlerini tahmin etmek için yaş, cinsiyet, tıbbi geçmiŐ ve diđer risk faktörleri dikkate alınmaktadır. Amaç, sigortalıların bireysel risklerini veya benzer risklere sahip olunan gruplarla iliŐkili risklerini dođru bir Őekilde yansıtan primi belirlemektir.

Hasar tutarı, sigortalının karŐılaŐtığı mali zararın miktarını ifade etmektedir. GLM, bu tutarların tahmini için bir dizi istatistiksel dađılım sunar. Hasar tutarının modellenmesinde Gamma veya ters Gauss dađılımları gibi sürekli dađılımlar tercih edilir. Bu dađılımlar, hasarın büyüklüğünü ve varyansını daha iyi yansıtmaları ve buna bađlı olarak sigorta Őirketlerinin hasar ödeme miktarlarında daha dođru tahmin edebilir olmaları nedeniyle tercih edilmektedir.

Hasar frekansı ise sigortalı olay başına ortalama hasar miktarını temsil etmektedir. Poisson veya negatif Binom dađılımları, nadir ve rastgele hasar olaylarını modellemek için kullanılır. Bu, hasar tutarlarının tahmininde, frekans ve yoğunluğun ayrı ayrı ele alınmasını gerektirir.

GLM, bađımsız deđerŐkenlerle bađımlı deđerŐkenler arasındaki iliŐkiyi dođrusal bir bađlamda ele alır. Bu, aktüerlerin, çeŐitli risk faktörlerini (örneğin, sigortalının yaŐı, sađlık durumu, geçmiŐ hasar geçmiŐi vb.) hasar tutarı ve frekansı ile iliŐkilendirmesine olanak tanımaktadır. Bu modelleme, poliçe fiyatlandırması ve risk yönetimi stratejilerinin geliŐtirilmesinde kritik bir rol oynamaktadır. Aktüeryal analizlerde GLM'nin kullanımı, daha dođru risk deđerlendirmeleri ve maliyetin önem taŐıdığı sigorta politikalarının geliŐtirilmesine katkı sađlamaktadır.

2.1. Genelleştirilmiş Doğrusal Modellerin Yapısı

Genelleştirilmiş doğrusal model, regresyon modellemesine veya denetimli öğrenme metoduna bir örnek oluşturabilmektedir [7]. Regresyon modellemesi temel olarak bağımsız (açıklayıcı) değişkenler yardımıyla bir yanıt değişkeninin bazı özelliklerini açıklamayı amaçlamaktadır. Böylece, değişkenler arasındaki ilişkiyi modelleme imkânı elde edilmektedir.

Geleneksel doğrusal regresyonda, yanıt değişkeninin (bağımlı değişkeni) normal olarak dağıldığı varsayılırken GLM’de, binom, Poisson ve gama dağılımları dâhil olmak üzere üstel dağılım ailesine mensup yanıt değişkenleri olabilir.

GLM’ler üç bileşen içerir: rastgele bir bileşen, sistematik bir bileşen ve ikisini birbirine bağlayan bir bağ fonksiyonu

Rastgele bileşen, Genelleştirilmiş doğrusal modellerde, rastgele bileşen bağımsız ve üstel dağılım ailesine ait rastgele değişkenlerden oluşmaktadır, aşağıdaki eşitlikte gösterilmiştir [17].

$$\exp\left(\frac{y\theta - a(\theta)}{\phi/(nv)}\right) c(y, \phi/(nv))$$

Sistematik bileşen, Sistematik bileşen, yanıt değişkenini açıklayıcı değişkenlerle doğrusal bir ilişki üzerinden bağlamaktadır. Doğrusallık, katsayıların (β_j) doğrusal kombinasyonu anlamındadır. Sistematik bileşen ve bağımlı değişken arasındaki ilişkiyi ifade eden regresyon denklemi aşağıdaki eşitlikte gösterilmiştir.

$$\text{score}_i = x_i^T \beta = \beta_0 + \sum_{j=1}^p \beta_j x_{ij}, \quad i = 1, 2, \dots, n.$$

Tasarım Matrisi, istatistikte genellikle tasarım matrisi olarak adlandırılan, tüm x_i ’lerin satır bazında bağlanmasıyla elde edilen X matrisidir.

Tasarım matrisi X, bağımsız değişkenlerin ve sabit terimin düzenlenmiş formunu içerir ve bu eşitlik aşağıdaki şekilde gösterilmiştir.

$$X = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}$$

Regresyon modeli bir sabit içerdiğinde, tasarım matrisi X , 1 değerlerinden oluşan bir sütun içerir. Tüm n yanıtlar için s_1, s_n skorlarını toplayan s vektörü aşağıdaki şekilde elde edilir. Skorlar vektörü s , regresyon katsayıları $X\beta$ ile ilişkilendirilmiş yanıt değişkenlerinin tahmini değerlerini temsil eder ve aşağıdaki eşitlik ile formüle edilmiştir.

$$s = (s_1, \dots, s_n)^T = X\beta.$$

Bağ fonksiyonu,

GLM'de bağ fonksiyonu g , yanıt değişkeninin beklenen μ_i değeri ile doğrusal tahmin ediciyi ilişkilendirir: $g(\mu_i) = \text{score}_i$. Bu sürekli, tersi alınabilir dönüşüm, yanıt değişkeninin ortalama değerini açıklayıcı değişkenlerle bağlar ve $g(\mu_i) = \text{score}_i \Leftrightarrow \mu_i = g^{-1} = (\text{score}_i)$ şeklinde ifade edilmektedir.

2.1.1. Üstel Dağılım Ailesi

Üstel Dağılım Ailesi, geliştirilmiş bir modelin olasılık yoğunluk fonksiyonunu temsil eder, θ konum ve ϕ ölçek parametreleri ile v ağırlığı ve $a(\theta)$, $b(\theta)$ fonksiyonları modelin şekli ve normalleştirmesini belirler. Aktüeryal uygulamalarda, v hacim ölçüsünü ifade etmektedir. θ ve ϕ , dağılımın konumunu ve ölçeğini, μ ortalama ve σ^2 varyansını tanımlamaktadır.

$$\left\{ \frac{p_Y(y)}{f_Y(y)} \right\} = \exp\left(\frac{y\theta - a(\theta)}{\phi/v} \right) c(y, \phi/v), \quad y \in S \quad (2.1)$$

Üstel aileye ait geliştirilmiş form, aşağıdaki şekilde elde edilir.

$$\exp\left(\frac{y\theta - a(\theta)}{b(\phi, v)} \right) c(y, \phi, v)$$

2.1.1.1. Poisson Dağılımı

Bir Poisson dağılımlı yanıt değişkeni Y , $S = \{0, 1, 2, \dots\}$ kümesinden değerler alır ve olasılık fonksiyonuna sahiptir. Olasılık fonksiyonu aşağıdaki gibidir.

$$p_Y(y) = \exp(-\lambda) \frac{\lambda^y}{y!}, \quad y = 0, 1, 2, \dots \quad (2.2)$$

$Y \sim \text{Poisson}(\lambda)$ ise,

$$E[Y] = \lambda \quad \text{ve} \quad \text{Var}[Y] = \lambda' \text{dir.} \quad (2.3)$$

2.1.1.2. Gamma Dağılımı

Gamma, yüksek bir tepe noktasına ve uzun kuyruğa sahip sağa çarpık bir dağılımdır. Bu nedenle, Gamma dağılımı sigorta şirketi tarafından ödenen hasarları modellemek için sıklıkla kullanılmaktadır.

Y , $\alpha > 0$ ve $\tau > 0$ parametreleri ile Gamma dağılımına sahip olmak üzere ve $Y \sim \text{Gama}(\alpha, \tau)$ olarak gösterilir. Bu durumda yoğunluk fonksiyonu aşağıdaki biçimde gösterilir. Olasılık yoğunluk fonksiyonu aşağıdaki gibi yazılabilir.

$$f_Y(y) = \frac{y^{\alpha-1} \tau^\alpha \exp(-\tau y)}{\Gamma(\alpha)}, \quad y > 0$$

$Y \sim \text{Gama}(\alpha, \tau)$ için ortalama ve varyans değerleri verilmiştir.

$$E[Y] = \frac{\alpha}{\tau}$$

ve

$$\text{Var}[Y] = \frac{\alpha}{\tau^2} = \frac{1}{\alpha} E[Y]^2$$

Üstel dağılım ailesi ve parametreleri Çizelge 2.1'de gösterilmektedir.

Çizelge 2.1. Üstel dağılım ailesi ve parametreleri [7]

Dağılım	θ	$a(\theta)$	ϕ	$\mu = E[Y]$	$\text{Var}[Y]$
Ber(q)	$\ln \frac{q}{1-q}$	$\ln(1 + \exp(\theta))$	1	q	$\mu(1 - \mu)$
Bin(m, q)	$\ln \frac{q}{1-q}$	$m \ln(1 + \exp(\theta))$	1	$m q$	$\mu \left(1 - \frac{\mu}{m}\right)$
Geo(q)	$\ln(1 - q)$	$-\ln(1 - \exp(\theta))$	1	$\frac{1 - q}{q}$	$\mu(1 + \mu)$
Poi(μ)	\ln	$\exp(\theta)$	1	μ	μ
NB(μ, σ^2)	μ	$\frac{\theta^2}{2\sigma^2}$	σ^2	μ	ϕ
Exp(μ)	$-\frac{1}{\mu}$	$-\ln(-\theta)$	1	μ	μ^2
Gam(μ, α)	$-\frac{1}{\mu}$	$-\ln(-\theta)$	$\frac{1}{\alpha}$	μ	$\phi \mu^2$
Gau(μ, α)	$\frac{1}{2\mu^2}$	$-\sqrt{-2\theta}$	$\frac{1}{\alpha}$	μ	$\phi \mu^3$

2.2. Varyans fonksiyonu

Varyans fonksiyonu $V(\mu)$, üstel dağılım ailesinde ortalama μ ile varyans arasındaki ilişkiyi tanımlar; θ üzerinden μ 'nin değişimiyle ve fonksiyonu $a(\cdot)$ ile ilişkilidir. Yanıt değişkeninin varyansı $\text{Var}[Y] = \frac{\phi}{v} V(\mu)$ formülüyle kanonik parametre üzerinden hesaplanmaktadır. (\cdot) varyans fonksiyonu aşağıdaki gibi tanımlanır.

$$V(\mu) = \frac{d^2}{d\theta^2} a(\theta) = \frac{d}{d\theta} \mu(\theta).$$

Yanıt değişkeninin varyansını, kanonik parametre üzerinden hesaplanması aşağıdaki eşitlikle gösterilmiştir.

$$\text{Var}[Y] = \frac{\phi}{v} V(\mu).$$

2.3. Ağırlık

Aktüeryal analizlerde, yanıtlar genellikle hasara oranlı toplam hasar veya hasar-prime oran olarak ifade edilir. Toplam ağırlık ve ağırlıklandırılmış ortalama, sırasıyla toplam v ve ağırlıklı ortalama ile hesaplanmaktadır.

Toplam ağırlık formülü aşağıdaki eşitlikte gösterilmektedir.

$$v. = \sum_{i=1}^n v_i \quad (2.4)$$

Ağırlıklandırılmış ortalama eşitlik (2.4)'te değiştirilmiş formu olarak aşağıdaki biçimde elde edilmiştir.

$$\frac{1}{v.} \sum_{i=1}^n v_i Y_i$$

2.4. Tweedie Dağılımı

Tweedie modelleri, üstel dağılım ailesinde güç parametresi ξ kullanılarak tanımlanan bir varyans fonksiyonuna sahiptir ve bu fonksiyon dağılımın şeklini belirler. Normal, Poisson, Gama ve Ters Gaussian dağılımları, belirli ξ değerleri için Tweedie modeline örnek oluşturmaktadır.

$$V(\mu) = \mu^\xi$$

Normal dağılımı için $\{\xi = 0\}$, Poisson dağılımı için $\{\xi = 1\}$, Gamma dağılımı için $\{\xi = 2\}$, ve Inverse Gaussian $\{\xi = 3\}$ değerlerini alır.

2.5. Bağ Fonksiyonları (Link Functions)

GLM, bağımsız değişkenlerle yanıt değişkeninin beklenen değerini ilişkilendiren bağ fonksiyonu aracılığıyla modellenmektedir. Bu fonksiyon, beklenen değerleri doğrusal tahmin ediciye dönüştürür ve ters çevrilebilir. Bu dönüşüm, ortalama değerleri tahmin etmek için açıklayıcı değişkenlerin doğrusal bir fonksiyonudur ve düzeltilmiş formülüyle doğru ortalama tahminlerine ulaşmaktadır.

2.5.1. Normal Bağ Fonksiyonlar

Bağ fonksiyonunun ana amacı, regresyon katsayılarını kısıtlamadan doğrusal tahmincileri yanıtın aralığıyla uyumlu hale getirmektir ki bu da modelin tahminlerini daha doğru ve anlamlı yapar.

2.5.2. Log-Bağ fonksiyonu

Bağ fonksiyonu, birçok sigorta uygulamasında negatif olmayıp, genellikle bir hasar sıklığı veya hasar şiddetini temsil eder ve bu nedenle $\mu_i > 0$ olur. Bir log-link fonksiyonu bu kısıtı kaldırır çünkü bu durumun her zaman yerine getirildiğini garanti eder.

2.5.3. Kanonik Bağ

Bağlantı fonksiyonu, skoru kanonik parametreyle aynı değer yaptığında buna kanonik bağ fonksiyonu denir. Kanonik bağ aşağıdaki özdeşliği doğrular.

$$\theta_i = \text{score}_i$$

$g(\mu_i) = \text{skor}_i$ ve $\mu_i = a'(\theta_i)$ olduğunda, kanonik bağlantı şu şekilde tanımlanır:

$$\theta_i = \text{score}_i \Leftrightarrow g(\alpha'(\theta)_i) = \theta_i \Leftrightarrow g^{-1} = a'$$

Çizelge 2.2'de Bağ fonksiyonları açıklanmıştır.

Çizelge 2.2. Bağ fonksiyonları [7]

Bağ Fonksiyon	$g(\mu)$	Kanonik Bağ
Identity (Kimlik)	μ	Normal
Log(Logaritma)	$\ln \mu$	Poisson
Power(Güç)	μ^ξ	Gamma($\xi = -1$)
		Inverse Gaussian($\xi = -2$)
Logit(Lojit)	$\ln\left(\frac{\mu}{1-\mu}\right)$	Binomial

Böylece, kümülatif fonksiyon $a(\cdot)$, her üstel dağılım ailesi dağılımının kendi özel kanonik bağına sahip olacak şekilde kanonik bağ fonksiyonunu belirler.

Kanonik parametre, GLM'e dayalı analizi büyük ölçüde basitleştirir, ancak diğer bağ fonksiyonları da kullanılabilir. Bu tam olarak GLM yaklaşımının güçlü yönüdür.

2.6. En çok Olabilirlik Tahmini

Aktüerler, prim hesaplamalarında yanıt değişkenlerinin bilinmeyen bir parametre olan θ 'ye bağlı olduğunu düşünürler ve veriye dayalı olarak bu θ için bir tahmin yaparlar. Bu tahmin, θ 'nın gerçek değerine olabildiğince yakın olmalıdır ve örneklem değişkenliği nedeniyle dalgalanabilir. Seçilen tahmin değeri, parametrenin nokta tahmini olarak bilinir. Bu nokta tahmini, aşağıdaki fonksiyon ile ifade edilmektedir:

$$\hat{\theta} = \hat{\theta}(Y_1, Y_2, \dots, Y_n).$$

Burada, tahmin edicinin değerleri, veri setinin değişkenliğine bağlı olarak değişir.

2.6.1. Tahmin Edicilerin Özellikleri

2.6.1.1. Tekrarlanan Örnekleme

Genel olarak, birden fazla θ tahmin edici vardır ve aktüerler bu tahmin edicilerin değerlerini inceleyerek karşılaştırırlar. Tahmin edicilerin özellikleri, tekrarlayan örnekleme yoluyla elde edilir. Belirli bir popülasyondan alınan örneklerin tekrarlarından elde edilen tahminlerin ortalaması, bu tahminlerin matematiksel beklentisini verir. $\text{Var}[\hat{\theta}]$, bu tahminlerin genelindeki kararlılığı ölçmektedir.

$$E[\hat{\theta}] = \lim_{k \rightarrow \infty} \frac{1}{k} \sum_{j=1}^k \hat{\theta}_j$$

2.6.1.2. Yansızlık

θ parametresi için bir tahmin edici $\hat{\theta}$ aşağıdaki koşulun sağlanması durumunda yansızdır:

$$E[\hat{\theta}] = \theta.$$

Yansızlık, sonsuz sayıda örnek üzerinde θ_k 'nin $k = 1, 2, \dots$ bağımsız örneklerden elde edildiği $\theta_k - \theta$ tahmin hatalarının ortalama olarak birbirini götürmesi anlamına gelir.

2.6.1.3. Tutarlılık

Tutarlılık, örneklem büyüklüğü arttıkça, bir tahmin edicinin gerçek parametre değerine yaklaşmasını ifade eden bir asimptotik özelliktir. θ 'nın bir tahmin edicisi için, örneklem sayısı sonsuza giderken gerçek θ değerine yaklaşıyorsa ve tahmin edicinin varyansı sıfıra düşüyorsa tutarlı olduğunu söyleyebiliriz. Yani, n örneklem sayısı arttıkça, tahmin edicinin

gerçek değer etrafında yoğunlaşma olasılığı artar. Bu, büyük sayılar yasasının bir sonucudur.

Tahmin edici için tutarlılık: $\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| < \epsilon) = 1$ hepsi için $\epsilon > 0$.

Tahmincinin varyansı: $\lim_{n \rightarrow \infty} \text{Var}[\hat{\theta}_n] = 0$

Ortalama varyans: $\text{Var}[Y] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[Y_i] = \frac{\text{Var}[Y]}{n} \rightarrow 0, n \rightarrow \infty$.

2.6.1.4. Ortalama Hata Karesi

$\hat{\theta}$ 'nin doğruluğu geleneksel olarak ortalama kare hatasıyla ölçülür. Ortalama kare hatası ve bunun tahmin ediciler için önemi gösterilmiştir.

$$\text{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2].$$

Bu, hipotetik tekrarlanan örnekleme durumunda sonsuz sayıda örneklemin ortalama karesel tahmini hatası $\hat{\theta} - \theta$ 'ye karşılık gelir. Yansız bir tahmin edicinin varyansı, onun hassasiyetini gösterir.

$$\hat{\theta} \text{ yansız} \Rightarrow \text{MSE}(\hat{\theta}) = \text{Var}[\hat{\theta}].$$

ve

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + (E[\hat{\theta}] - \theta)^2.$$

2.6.2. En çok Olabilirlik Tahmin Edicisi

En yüksek olasılık tahmini, gözlemlenen değerleri en olası kılan parametre değeridir. Olasılık fonksiyonu $L(\theta)$, bu gözlemlerin birleşik olasılığını temsil eder ve bağımsız gözlemler için çarpım olarak ifade edilir. En yüksek olasılık tahmin edicisi, bu fonksiyonu maksimize eden parametre değeridir. Olasılık fonksiyonunun logaritması alındığında, çözüm için gerekli olan türev sıfıra eşitlenir. Bu işlem, parametrenin tahmini değerini bulmak için kullanılır ve bu değer, modelin parametre uzayında benzersiz bir maksimuma sahiptir.

Olasılık fonksiyonu:
$$L(\theta) = \prod_{i=1}^n \exp\left(\frac{y_i \theta - a(\theta)}{\phi}\right) c(y_i, \phi).$$

$$L(\theta) = \ln \mathcal{L}(\theta)$$

Log-olasılık fonksiyonu:
$$L(\theta) = \sum_{i=1}^n \left(\frac{y_i \theta - a(\theta)}{\phi} + \ln c(y_i, \phi) \right)$$

$$= \frac{n(\bar{y} - a(\theta))}{\phi} + \sum_{i=1}^n \ln c(y_i, \phi)$$

Örnekleme ortalaması:
$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$\frac{d}{d\theta} L(\theta) = 0.$$

$$0 = \frac{d}{d\theta} L(\theta)$$

En yüksek olasılık tahmin

edici için türev:
$$= \sum_{i=1}^n \frac{d}{d\theta} \frac{n(\bar{y} \theta - a(\theta))}{\phi}$$

$$= \frac{n(\bar{y} - a'(\theta))}{\phi}$$

2.6.2.1. Değişmezlik

Bire bir fonksiyon h kullanılarak elde edilen en çok olabilirlik tahmini $\widehat{h(\theta)} = h(\hat{\theta})$ şeklindedir. Bu yöntem, üstel dağılım ailesi için geçerli olup, ortalama tahmini $\hat{\mu} = \bar{y}$ olarak verilmektedir. Hasarsızlık durumları için, Poisson ortalaması λ 'nın en çok olabilirlik tahmini olarak kullanılmaktadır.

2.6.2.2. Asimptotik Yansızlık

Örnekleme büyüdükçe beklenen değer $E[\hat{\theta}]$ θ 'ya yaklaşır. Yine, bu büyük sayılar yasasının bir sonucudur, çünkü \bar{y} , 1 olasılıkla μ 'ye yaklaşır.

2.6.2.3. Minimum Varyanslılık

Tüm tahmin ediciler sınıfında, büyük örneklem için θ minimum varyansa sahiptir ve bu nedenle mümkün olan en doğru tahmin edicidir. Aktüerler genellikle büyük miktarda

veriyle uğraştığı için bu tahmin prosedürünü sigorta çalışmalarında yürütmek cazip hale gelmektedir.

2.7. Fisher Bilgi

Fisher bilgisini elde etmek için, Fisher bilgi matrisi $I(\beta)$ 'nin (j, k) ögesini yazılmalıdır.

$$\begin{aligned} I_{jk}(\beta) &= E \left[\frac{\partial^2}{\partial \beta_j \partial \beta_k} L(\beta) \right] \\ &= E \left[\sum_{i=1}^n \frac{(Y_i - \mu_i) x_{ij}}{\text{Var}[Y_i]} \frac{\partial \mu_i}{\partial s_i} \sum_{l=1}^n \frac{(Y_l - \mu_l) x_{lk}}{\text{Var}[Y_l]} \frac{\partial \mu_l}{\partial s_l} \right] \\ &= \sum_{i=1}^n \sum_{l=1}^n \text{Cov}[Y_i, Y_l] \frac{x_{ij} x_{lk}}{\text{Var}[Y_i] \text{Var}[Y_l]} \frac{\partial \mu_i}{\partial s_i} \frac{\partial \mu_l}{\partial s_l} \end{aligned}$$

Yanıt değişkeni bağımsız olduğundan kovaryans terimleri kaybolur, yani $i \neq l$ için $\text{Cov}[Y_i, Y_l] = 0$. Bu değer, sadece $i = l$ 'ye karşılık gelen terimler kalır.

$$I_{jk}(\beta) = \sum_{i=1}^n \frac{x_{ij} x_{ik}}{\text{Var}[Y_i]} \left(\frac{\partial \mu_i}{\partial s_i} \right)^2$$

Yardımcı ağırlığı tanımlarsak;

$$\tilde{v}_i = \frac{1}{\text{Var}[Y_i]} \left(\frac{\partial \mu_i}{\partial s_i} \right)^2 = \frac{v_i}{\phi V(\mu_i)} \left(\frac{\partial \mu_i}{\partial s_i} \right)^2$$

$\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_n$ girişleriyle $n \times n$ diyagonal W matrisi. Matris formunda,

$$I(\beta) = X^T \tilde{W} X = \sum_{i=1}^n \tilde{v}_i x_i x_i^T \text{ olarak elde edilir.}$$

2.8. Ofset (Dengeleme Terimi)

Ofsetler, aktüeryal modellerde bireysel sabit katkılar olarak işlev görür ve regresyon analizinde sabit bir katsayı ile çarpılıp modele eklenir. Bu, modeldeki bireysel özellikleri yansıtır.

$$\text{score}_i = \text{offset}_i + x_i^{T\beta}$$

2.9. Genelleştirilmiş Doğrusal Modellerin Aktüeryal Uygulamaları: Model Seçimi

GLM'lerin aktüeryal uygulamaları özellikle model seçimi üzerine odaklanmaktadır. Model seçimi, üstel dağılım ailesinin seçimi, bağ fonksiyonunun seçimi ve özellik seçimi ile mühendisliği gibi çeşitli faktörleri içerir.

Üstel Dağılım Ailesi Üyesi Seçimi: Ağır kuyruklu veriler için üstel dağılım ailesinden bir model seçmek, logaritmik dönüşümlerle mümkündür. Bu dönüşümler, verilerin daha normal bir dağılıma sahip olmasını sağlar ancak bias düzeltmeleri gerekebilir.

Bağ Fonksiyonunun Seçimi: Modelin ortalama değeri ile doğrusal tahminci arasındaki bağı kurmak için genellikle kanonik bağ fonksiyonu kullanılır. Pratik gereksinimler, farklı bağ fonksiyonlarına ihtiyaç duyabilir. Örneğin, gama modeli için log-link bağlantısı daha uygun olabilir.

Özellik Seçimi ve Mühendisliği: Modelin performansını artırmak için en uygun özelliklerin seçilmesi ve gereksiz özelliklerin atılması önemlidir. Bu süreç, özelliklerin dönüştürülmesi veya yeni özelliklerin türetilmesini içerebilir.

Kategorik Özellik Bileşenleri ve Kukla Kodlama: Kategorik değişkenlerin, modelleme için uygun hale getirilmesi gerekmektedir. Kukla kodlama, bu değişkenleri bir Öklid uzayına yerleştirmek için yaygın bir yöntemdir.

İkili Özellik Bileşenleri: İkili özellikler, sadece iki değer alabilen Bernoulli değişkenleridir ve kukla kodlama ile kolayca işlenebilir.

Sürekli Özellik Bileşenleri: Sürekli değişkenler, bazen kategorik sınıflara ayrılarak veya matematiksel dönüşümlerle işlenerek modelleme için uygun hale getirilir.

Modeldeki etkileşimlerin nasıl eklenmesi gerektiği ve özellik mühendisliğinin önemi vurgulanmaktadır. Etkileşimler ve fonksiyonel form seçimi, modelin doğruluğunu ve genelleme yeteneğini artırabilir.

Aktüeryal problemleri çözmek için GLM içindeki uygun bir model seçimi, belirli faktörleri ve seçenekleri dikkate almayı gerektirmektedir [18].

3. DERİN ÖĞRENME

Teknolojinin hızla ilerlediği günümüz dünyasında, yapay zekâ ve özellikle derin öğrenme ve yapay sinir ağları birçok sektörde devrim yaratmaktadır. Derin öğrenme, insan beyninin işleyişini taklit eden algoritmalar serisi kullanarak karmaşık veri setlerinden öğrenme yeteneğine sahip bir yapay zekâ biçimidir. Derin öğrenme, yapay sinir ağlarının, karmaşık ve çok katmanlı yapıları kullanarak verileri öğrenme ve analiz etme yeteneği olarak tanımlanabilir. Bu yaklaşım, büyük veri setlerinde karmaşık ilişkileri keşfetmek, tahminlerde bulunmak ve desenleri tanımak için son derece etkilidir.

Yapay sinir ağları ise, insan beynindeki nöronların ağ benzeri yapılarını taklit eden ve veriler arasındaki karmaşık ilişkileri modelleyebilen yapay zekâ modelleridir.

Sigortacılık sektörü, derin öğrenme ve yapay sinir ağlarının sunduğu bu yenilikçi yaklaşımlardan önemli ölçüde yararlanmaktadır. Bu teknolojiler, risk değerlendirme, hasar tahmini, dolandırıcılık tespiti ve müşteri hizmetleri gibi alanlarda kullanılarak sigortacılık işlemlerini daha verimli, doğru ve hızlı hale getirmektedir.

Sağlık sigortası alanında ise, derin öğrenme ve yapay sinir ağları, özellikle hasta verilerinin analizinde ve risk değerlendirmelerinde kritik roller oynamaktadır. Bu teknolojiler, hastaların sağlık geçmişi, yaşam tarzı ve genetik bilgilerini analiz ederek, sağlık risklerini daha doğru bir şekilde tahmin etmekte ve böylece bireyler için daha uygun ve maliyet odaklı sigorta paketleri sunmaktadır.

Ayrıca, sağlık hizmetlerinin kullanımı ve maliyetleri üzerinde daha derin analizler yaparak, sigorta şirketlerinin maliyet yönetimini optimize etmelerine yardımcı olmaktadır.

"Combined Actuarial Neural Network" (CANN), aktüerya alanında kullanılan bir yapay zekâ uygulamasıdır. Aktüerya, riski ölçme, yönetme sanatı ve bilimi olarak tanımlanabilir. Bu alanda, özellikle sigortacılık ve finans sektörlerinde, riskleri ve gelecekteki mali yükümlülükleri tahmin etmek için çeşitli matematiksel ve istatistiksel yöntemler kullanılır.

CANN, derin öğrenme ve yapay sinir ağları teknolojilerini aktüeryal hesaplamalarla birleştirir. Bu birleşim sayesinde, daha karmaşık risk yapılarını modelleyebilir ve daha doğru tahminler yapabilir.

Özellikle, büyük veri setlerini işleyebilme ve karmaşık desenleri algılama yetenekleri sayesinde, geleneksel aktüeryal yöntemlerden daha etkili sonuçlar elde edilebilmektedir. CANN uygulamaları, sigorta primlerinin belirlenmesinden, talep tahminlerine, risk

değerlendirmelerinden, dolandırıcılık tespitine kadar geniş bir yelpazede kullanılabilir. Bu teknolojinin kullanımı, aktüeryal analizlerin doğruluğunu artırırken, aynı zamanda daha hızlı ve verimli işlem yapılmasına olanak tanır.

Aktüeryal alanın yanı sıra, finans, sağlık hizmetleri ve diğer birçok sektörde de CANN'ın potansiyeli giderek daha fazla fark edilmektedir.

Bu sayede, daha karmaşık ekonomik ve sosyal sistemlerin analizi için yeni kapılar açılmaktadır

3.1. Derin Öğrenme ve Temsil Öğrenimi

GLM'ler, x değişkenleri verildiğinde yanıt Y 'nin beklenen değerini modelleyen bir regresyon fonksiyonu kullanılmaktadır.

Derin öğrenme, bu yapıyı katmanlar kullanarak otomatik özellik mühendisliği ile genişletmektedir; her katman, veriye yeni bir temsil katmanı ekler. GLM'lerde, derinlemesine işlenmiş özellikler kanonik bağ fonksiyonu ve regresyon parametreleri ile birleştirilerek beklenen yanıtı tahmin eder.

Derin öğrenme, bu işlemi optimize ederek test verilerinde genelleştirmiştir.

$$x \mapsto \mu(x) = E_{\theta(x)}[Y] = g^{-1} \langle \beta, x \rangle. \quad (3.1)$$

$$\mu(x) = E_{\theta(x)}[Y] = g^{-1} \langle \beta, z^{(d:1)}(x) \rangle, \quad (3.2)$$

3.2. Yapay Sinir Ağları

Yapay Sinir Ağları (Artificial Neural Networks - ANN), insan beyninin bilgi işleme şeklini taklit etmeyi amaçlayan hesaplama sistemleridir. Bu sistemler, genellikle çok sayıda işlem birimi olan düğümlerden (neuronlar) oluşur.

Bu düğümler birbiriyle çeşitli ağırlıklı bağlantılar aracılığıyla iletişim kurar.

3.2.1. İleri Beslemeli Ağlar

İleri beslemeli sinir ağları, düşük ve derin olmak üzere iki ana tipe ayrılmaktadır. Regresyon fonksiyonlarını modellemek için kullanılırlar. Sinir ağları, sürekli fonksiyonları yaklaşım için esnek bir çerçeve sunar ve çok yönlü bir araç kutusuna benzetilmektedir.

Çizelge 3.1 Aktivasyon fonksiyonları ve türevlerini gösterilmektedir.

Çizelge 3.1. Aktivasyon fonksiyonları [18]

	Aktivasyon fonksiyonu	Türev
Sigmoid (lojistik) aktivasyon	$\phi(x) = (1 + e^{-x})^{-1}$	$\phi' = \phi(1 - \phi)$
Hiperbolik tanjant aktivasyon	$\phi(x) = \tanh(x)$	$\phi' = 1 - \phi^2$
Üstel aktivasyon	$\phi(x) = \exp(x)$	$\phi' = \phi$
Adım fonksiyonu aktivasyon	$\phi(x) = 1_{(x \geq 0)}$	
Düzeltilmiş doğrusal birim (ReLU) aktivasyon	$\phi(x) = x1_{(x \geq 0)}$	

İleri Beslemeli Ağların Yapısı:

$z^{(m)} = (z_1^{(m)}, \dots, z_{q_m}^{(m)})$ katmanlarının her bir $z_j^{(m)}$, $1 \leq j \leq q_m$ nöronunun olduğu regresyon fonksiyonlarıdır. q_m $1 \leq m \leq d$, bir GLM yapısına sahiptir. İlk bileşen $z_1^{(m)} = 1$ 'dir.

Aktivasyon Fonksiyonu:

Bir sinir ağı mimarisinin ilk önemli bileşeni (hiperparametre), $\phi : \mathbb{R} \rightarrow \mathbb{R}$ aktivasyon fonksiyonunun seçimidir. Lineer olmayan regresyon fonksiyonlarına yaklaşmak istediğimiz için, bu aktivasyon fonksiyonları doğrusal (linear) olmalıdır.

Bu, türevleri içeren model kalibrasyonu algoritmalarında bir avantaj oluşturur. Hiperbolik teğet aktivasyon fonksiyonu aşağıdaki eşitliği sağlar.

$$x \mapsto \phi(x) = \frac{e^x}{1 + e^x} = (1 + e^{-x})^{-1} \in (0,1).$$

Sinir ağı katmanları:

Bir sinir ağı (birkaç) sinir ağı katmanından oluşur. $q_{m-1}, q_m \in \mathbb{N}$ hiper parametrelerini ve $\phi : \mathbb{R} \rightarrow \mathbb{R}$ aktivasyon fonksiyonunu seçilir. Bir sinir ağı katmanı, aşağıdaki forma sahip bir eşlemedir.

$$z^{(m)}: \mathbb{R}^{q_{m-1}} \rightarrow \mathbb{R}^{q_m} \quad z \mapsto z^{(m)}(z) = \left(z_1^{(m)}(z), \dots, z_{q_m}^{(m)}(z) \right)'$$

İleri beslemeli sinir ağı mimarisi:

İleri beslemeli bir sinir ağı mimarisi, birkaç sinir ağı katmanının bir bileşimidir. İlk katman, tam olarak $X \subset \mathbb{R}^q$ özellik uzayımız olan girdi katmanıdır ve bu bölümün geri kalanı için

$q_0 = q$ olarak ayarlanır. Sinir ağı mimarisinin derinliğini gösteren bir hiperparametre $d \in \mathbb{N}$ seçilir. $1 \leq m \leq d$ için m 'inci gizli katmandaki aktivasyonlar şu şekilde elde edilir:

$$x \in \mathcal{X} \mapsto z^{(m:1)}(x) \stackrel{\text{def}}{=} (z^{(m)} \circ \dots \circ z^{(1)})(x) \in \mathbb{R}^{q_m}.$$

3.3. Derin İleri Beslemeli Sinir Ağları

Derin sinir ağları, regresyon fonksiyonu özellik bileşenleri arasında etkileşimleri var ise model kalibrasyonunda daha verimlidir.

$d \geq 2$ derinliğe sahip ileri beslemeli sinir ağları aşağıdaki şekilde verilir:

$$x \in \mathcal{X} \mapsto \mu(x) = \log \lambda(x) = \beta_0 + \sum_{j=1}^{q_d} \beta_j z_j^{(d:1)}(x) = \langle \beta, z^{(d:1)}(x) \rangle$$

3.3.1. Birleştirilmiş Aktüeryal Yapay Sinir Ağları

CANN, GLM'leri ve sinir ağlarını birleştirerek aktüeryal modellemeyi geliştiren bir yöntemdir. Bu model, GLM ile modellenen doğrusal bir bileşen ve sinir ağı ile modellenen doğrusal olmayan bir bileşen içerir. Bu bileşim, her iki yöntemin avantajlarını kullanarak daha esnek ve etkili tahminler sağlar [19]. CANN, sigorta oranları, fiyatlandırma ve hasar sayımında geleneksel yöntemlere üstünlük göstermiştir.

4. İFLAS TEORİSİ

İflas olasılığı, bir işletmenin finansal yükümlülüklerini yerine getirememesi riskini, varlık-yükümlülük dengesi üzerinden değerlendirilir. Bu analiz, işletmenin finansal durumunu ve geleceğini belirlemek için varlık ve yükümlülüklerin yönetimini inceler. Varlıklar, işletmenin değerleri; yükümlülükler ise borçları ifade etmektedir. Varlıkların yükümlülüklerden az olması iflas riskini artırmaktadır. İflas olasılığının hesaplanması, bu dengenin formülle ifadesini gerektirmektedir.

Aktif/Pasif Oranı (Asset to Liability Ratio): Bu oran, işletmenin varlıklarının toplam yükümlülüklerine oranını hesaplar.

$$\text{Aktif / Pasif Oranı} = \frac{\text{Toplam Varlıklar}}{\text{Toplam Yükümlülükler}}$$

Bu oran 1'den büyükse, işletme genellikle karlı kabul edilir çünkü varlıkları yükümlülüklerini karşılamada yeterli olacaktır. Oran 1'den küçükse, potansiyel iflas riski artmaktadır.

Net Aktif Değeri (Net Asset Value): işletmenin finansal karlılığını varlıkların toplam değerinden yükümlülüklerin çıkarılmasıyla hesaplar. Pozitif NAV finansal sağlık, negatif NAV iflas riski işaretidir. Sigorta şirketlerinde, primler ve hasar ödemeleri NAV üzerindeki temel etkilerdir; sermaye negatifse iflas gerçekleşmektedir.

4.1. Risk Süreci

Sigorta şirketinin risk süreci prim ve hasar süreçleri dikkate alınarak aşağıdaki eşitlik yardımıyla elde edilmektedir.

$$U(t) = u + ct - S(t), \quad t \geq 0, \quad (4.1)$$

Bu eşitlikte, u başlangıç sermayesi, c birim zamandaki prim oranını ve $S(t)$ toplam hasar tutarını göstermektedir [20].

4.2. İflas Olasılığı

İflas teorisi, bir sigorta şirketinin rezervlerinin yetersiz kalma riskini inceler. İflas olasılığı, sermayenin negatif olma şansını, başlangıç sermayesi ve prim-hasar döngüsü üzerinden değerlendirilir. Bu olasılık $\Psi(u) = Pr\{U(t) < 0, t > 0\}$ ile ifade edilmektedir, $U(t)$ belirli bir zamandaki sermayeyi temsil etmektedir [20]. Bu çalışmada GLM ve CANN

Modellerinin karşılaştırılmasında her iki yöntemle göre tahmin Edilen Primlerin ödenen hasarları karşılamadaki yeterliliği incelenmiştir. Analiz, teorik olarak iflas olasılığı hesabı yerine öneri modellerin gerçek veri üzerindeki sonuçları açısından inceleme yapılmıştır. Bu şekilde iki yöntemle göre tahmin edilen primler, gerçek primler ve hasar ödemeleri açısından iflas olasılığı bir tür performans ölçüsü olarak kullanılmaktadır.

4.3. İflas Olasılığının Değerlendirilmesi

Genel anlamda uygulanan veri seti için GLM ile CANN modelinin, karşılaştırılmasında her iki yöntemle göre tahmin edilen primlerin, gerçek tazminatlar ile karşılaştırılmaları yapılmıştır. Teorik olarak iflas olasılığı hesabı yerine, kullanılan verideki sonuçlar açısından bir tür performans ölçüsü olarak iflas olasılığı kullanılmıştır.

5. UYGULAMA

5.1. Veri Seti ve Tanımlayıcı İstatistikler

Tez çalışmasının uygulama aşamasında, Türkiye’de aktif olarak faaliyet göstermekte olan özel bir sigorta şirketinin 2019, 2020, 2021, 2022 yılları ile 2023 yılının ilk 9 aylık dönemine ilişkinin tamamlayıcı sağlık sigortası poliçeleri verisi kullanılmıştır. Veri 44,318 adet gözlem ve 8 değişken içermektedir. Veri kümesinde bağımsız değişkenler sürekli ve kategoriktir. Bunlar poliçe sahibinin yaşı, cinsiyeti, vücut kitle endeksi, il, coğrafi bölge, meslek, medeni durum, yeni iş/yenileme bilgilerini içermektedir. Her bir poliçe başına poliçe numarası, hasar sayısı, hasar tutarı ve poliçe primi bilgisi bulunmaktadır. Ayrıca hasar tutarları değerleri aylık hesaplanan *Ayakta Tedavi Enflasyonu* (TÜİK harcama gruplarına göre Hasar Tutar) [21] ile güncellenerek yeni bir sütun yaratılmış ve şiddet modelinde bağımlı değişken olarak bu veri kümesi kullanılmıştır. Veri seti ile ilgili analizlerde R stüdyo bilgisayar programları kullanılmıştır [22]. Veri içerisinde kullanılan değişkenler ve açıklamaları Çizelge 5.1’de gösterilmektedir.

Çizelge 5.1. Değişkenler ve açıklamaları

Değişken Adı	Değişken Açıklaması
PoliceNo	Poliçe numarası.
Yenileme	Yenileme ve Yeni İş olmak üzere iki kategorilidir.
Yaş	Sürekli bir değişkendir.
YasGrup	Yasların 10 kategorik değişkene gruplanmasıdır.
Cinsiyet	Kadın ve Erkek olmak üzere iki kategoride açıklanmıştır.
VKI	Vücut kitle endeksini değişkenidir. Sürekli Değişkendir.16 ile 40 arasında değerler alır.
Bölge	Türkiye'nin 7 coğrafi bölgesini içeren kategorik bir değişkendir. Marmara, Akdeniz, İç Anadolu, Karadeniz, Güney Doğu Anadolu, Doğu Anadolu, Ege bölgesi olmak üzere.
İl	Türkiye'nin 81 ilidir. Kategorik değişkendir.
ilGrup	81’ il 6 gruba kategorileştirilmiştir. Çizelge 5.3’te yer verilmiştir.
VKIGrup	VKI 5 kategoriye ayrılmış kategorik değişkenidir.
MedeniDurum	Evli, Bekâr, Boşanmış ve Dul olarak 4 kategorilidir.
Prim	Toplam Prim değişkenidir.
KazAdet	Kazanılmış prim âdetidir.
HasarTutar	Gerçekleşen hasar tutarlarıdır.
Meslek	9 kategorili ayrılmış meslek gruplarıdır. Mavi Yaka, Beyaz Yaka, Öğretmen, Emekli, Memur, Çalışmayan, Öğrenci, Bebek, Diğer
Medeni Durum	Evli, Bekâr, Boşanmış ve Dul olarak 4 kategorilidir.
HasarTutarGuncel	Gerçekleşen hasar tutarlarının enflasyonla güncellenmiş hali.
HasarAdet	Hastaneye gidiş âdeti. Minimum gidiş sayısı 0, maksimum 10’dur.

Çizelge 5.1’de açıklanan değişkenlerdeki verileri kategorize etmek ve değişken bazında veri yetersizliği gibi durumların önüne geçmek amacıyla yaş değişkeni yaş grubu, vücut kitle endeksi değeri, vücut kitle endeksi grubu, il bilgisi ise il grubu değişkeni olarak gruplandırılmıştır.

Kazanılan adet hesaplanmasında çeyreklik dönem dikkate alınmıştır. Bunun nedeni sağlık sigortalarının mevsimsellik durumlarından etkilenmesidir. Kış aylarında daha sık hastaneye gidişlerin olması bu duruma örnek olarak verilebilir.

Veriye ilişkin incelemeler ve analiz öncesinde yapılan tüm düzeltmeler aşağıda açıklanmaktadır.

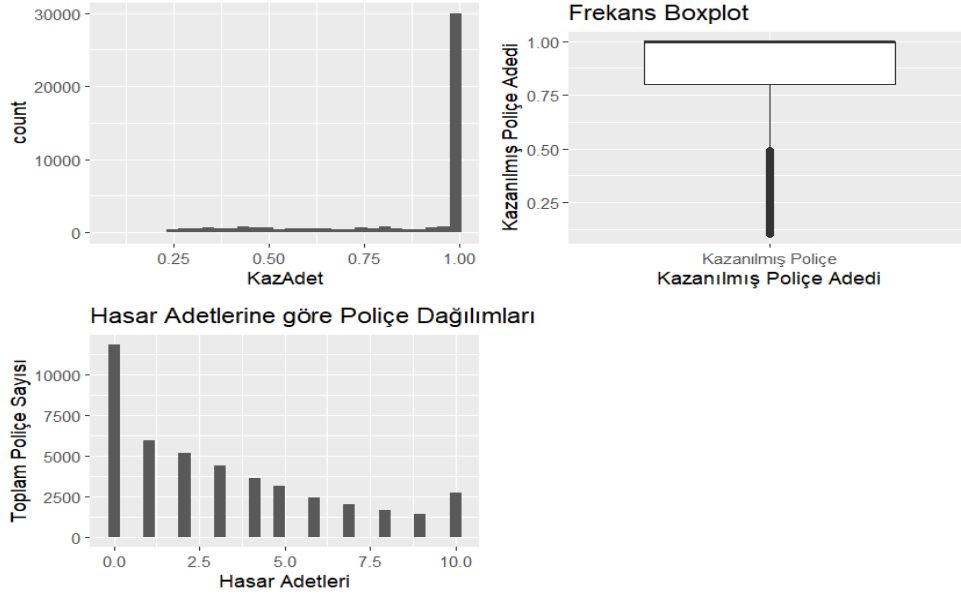
- Portföye ilişkin her bir poliçe birbirinden bağımsızdır. Aynı aile üyeleri için de farklı poliçe numaraları bulunmaktadır. Bu bağlamda çalışma bağımsızlık varsayımıyla ilerletilmiştir. Veri tabanından bireyler arasındaki ilişkisellik tespit etmeye uygun bir poliçe üretim sistemi bulunmamaktadır.
- Bu tamamlayıcı sağlık sigortası 10 kullanım adedi ile sınırlandırılmış bir üründür. Bu nedenle, hasar frekansı maksimum 10 olacaktır. İlgili poliçe 10’dan fazla kez kullanılmış ise kullanım sayısı 10 olarak kabul edilmiştir. Bununla ilgili veride düzeltme yapılmıştır.
- Sigortalı hastaneye gittiğinde hastalığına dair bir tanı kodu oluşturulmaktadır. Tüm muayene, test ve işlemler aynı tanı koduna yazılmaktadır. Böylece birden fazla işlem olmasına rağmen aynı tanı koduna yazıldığı için hasar âdeti 1 olarak yazılmaktadır.
- Veride yenileme numarası sütunu bulunmaktadır. Her poliçe numarasının benzersiz olması amacıyla poliçe numaralarının yanına yenileme numarası getirilerek yeni bir sütun oluşturulmuştur.
- Veri setinde hasarın gerçekleştiği il çoğunlukla İstanbul ve Ankara İleride gözlenmekte birlikte 81 ilden düzenlenmiş poliçeler bulunmaktadır.
- Vücut kitle endeksi 0-6 yaş aralığında kullanılmamaktadır. 0-6 yaş için yüzdeler dilim tablosunda bakılmıştır. Bu sebeple vücut kitle endeksi 0-6 yaş aralığındaki poliçeler ayrı bir grup olarak değerlendirilmiştir.

5.1.1. Hasar Frekansına İlişkin Bazı İstatistikler

Kazanılmış adet, bir poliçenin süresini yıllık birim halinde ölçülmesidir. Poliçenin bir yıl içerisinde kazanılma oranını verir. Tamamlanmış muhasebe yılına karşılık exposure değeri 1’dir. Örneğin ocak ayında kesilen bir poliçenin ilk çeyrek sonunda exposure değeri 0.25

olacaktır. Çoğunlukla veride sonlanmış poliçeler yer aldığı için kazanılmış adet grafikte çoğunlukla 1 olarak gösterilmektedir.

Şekil 5.1'te veriye ilişkin hasar adetlerini göstermektedir. Bu değişken, poliçe süresince sigortalının hastaneye gidiş âdetini temsil etmektedir.



Şekil 5.1. Hasar frekans bilgileri

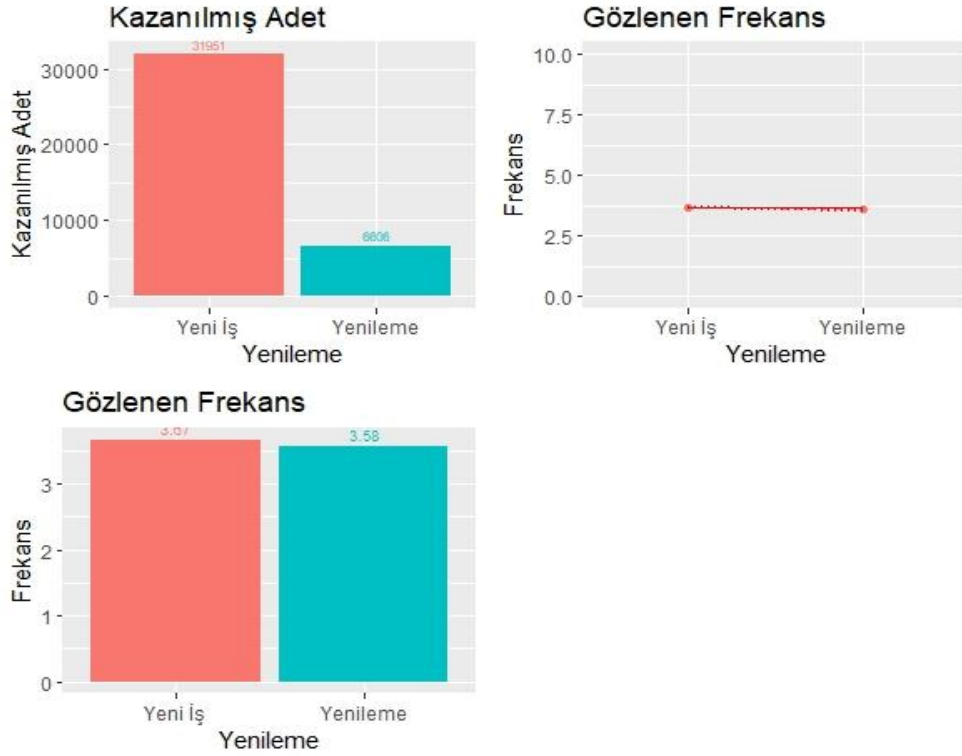
Portföye ilişkin ortalama hasar frekansı 3.65 olarak hesaplanmıştır. Poliçelerdeki hasar adetlerinin sayısına bağlı olarak toplam kazanılmış adet il tablosu Çizelge 5.2'de gösterilmektedir.

Çizelge 5.2. Hasar adedi ve kazanılmış adet bilgisi

H.Adet	N	K.Adet
0	11877	8804
1	5949	4779
2	5160	4489
3	4418	4043
4	3604	3409
5	3114	2990
6	2440	2376
7	2035	1995
8	1635	1613
9	1387	1374
10	2699	2685

Yenileme ve yeni iş olarak yazılan poliçe frekansında çok fazla bir fark olmakla birlikte yeni iş frekansının daha fazla olduğu gözlenmektedir. Yeni işlerden kazanılan poliçe adedi yenilemeye göre daha fazladır.

Yeni iş ve yenileme değişkenlerine bağlı olarak hasar sıklık bilgileri Şekil 5.2’de gösterilmektedir.



Şekil 5.2. Yeni iş ve yenileme değişkenlerine bağlı hasar sıklıkları

İl değişkeninin gruplandırılmasıyla elde edilen (*İl Grup Değişkeni*) şirketin bağlı bulunduğu ikili sigorta firmasının anlaşmalı özel hastane sayılarına bağlı olarak oluşturduğu gruplama sistemine göre oluşturulmuştur.

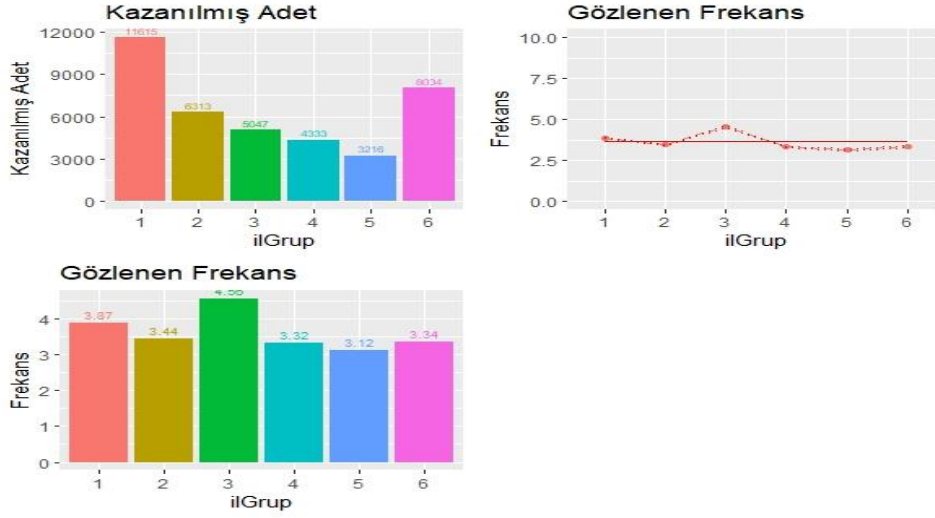
Grup numarası ve gruba ilişkin il bilgileri Çizelge 5.3’de gösterilmektedir.

Çizelge 5.3. İl grup değişken bilgisi

İ.G.N	İl Bilgileri	Gözlenen Frekans
1	İstanbul	3.87
2	Ankara	3.44
3	Kocaeli, Ordu, Samsun, Sivas	4.56
4	Bursa, Eskişehir, Konya, Sakarya	3.32
5	Antalya, Diyarbakır, Hatay, İzmir, Tekirdağ, Van	3.12
6	Diğer iller,	3.34

İl grup değişken bilgisi göz önünde bulundurulduğunda, 3’üncü il grubunun 4.56 ile en yüksek hasar frekansına sahip olduğu gözlenmektedir.

İl grup değişkenine bağlı olarak hasar sıklık bilgileri Şekil 5.3’de gösterilmektedir.



Şekil 5.3. İl grup değişkenine hasar sıklıkları

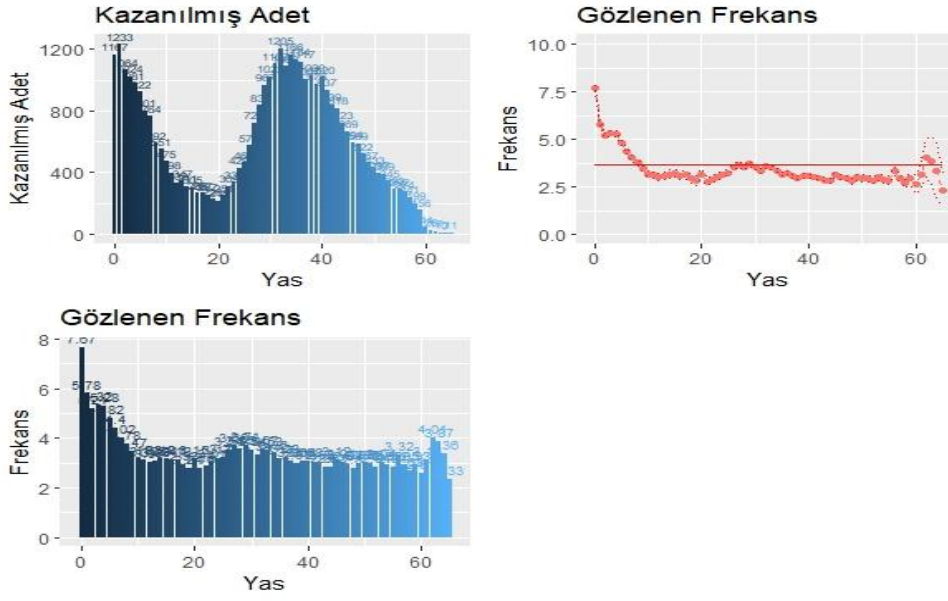
Poliselerin düzenlenmiş olduğu coğrafi bölgeler incelendiğinde, yedi bölgeye göre yapılan sınıflandırmada, en fazla kazanılmış âdetin yoğunluklu olarak Marmara bölgesinden geldiği gözlemlenmektedir. Ancak kullanım frekansı en yüksek bölge Karadeniz bölgesi olarak belirlenmiştir. Coğrafi Bölge değişkenine bağlı olarak hasar sıklık bilgileri Şekil 5.4’te gösterilmektedir.



Şekil 5.4. Coğrafi bölge değişkenine bağlı hasar sıklıkları

Poliseler sahiplerinin yaşları dikkate alınarak yapılan incelemede, bebeklerde frekansın 7.67 olduğu görülmüştür. Gözlenen frekans sayısının, 60 yaşından sonra arttığı da belirlenmiştir.

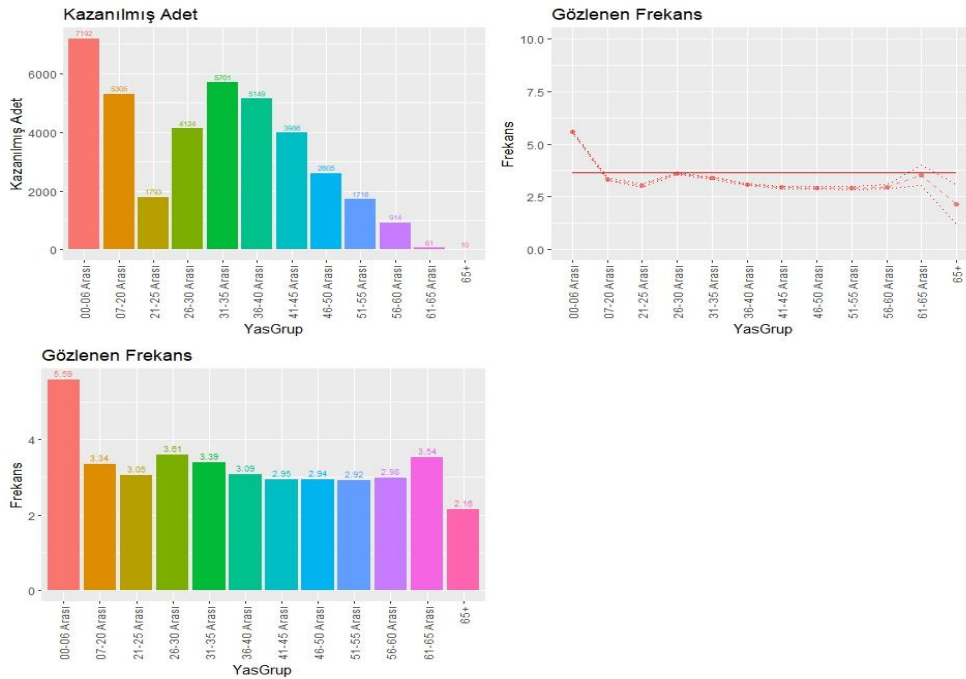
Yaş değişkenine bağlı olarak hasar sıklık bilgileri Şekil 5.6'da gösterilmektedir.



Şekil 5.5. Yaş değişkenine bağlı hasar sıklıkları

Yaş bilgilerinden elde edilen yaş gruplarında ise, en yüksek frekansın 0-6 yaş grubunda olduğu ve bu oranın 5.59 olduğu gözlemlenmiştir.

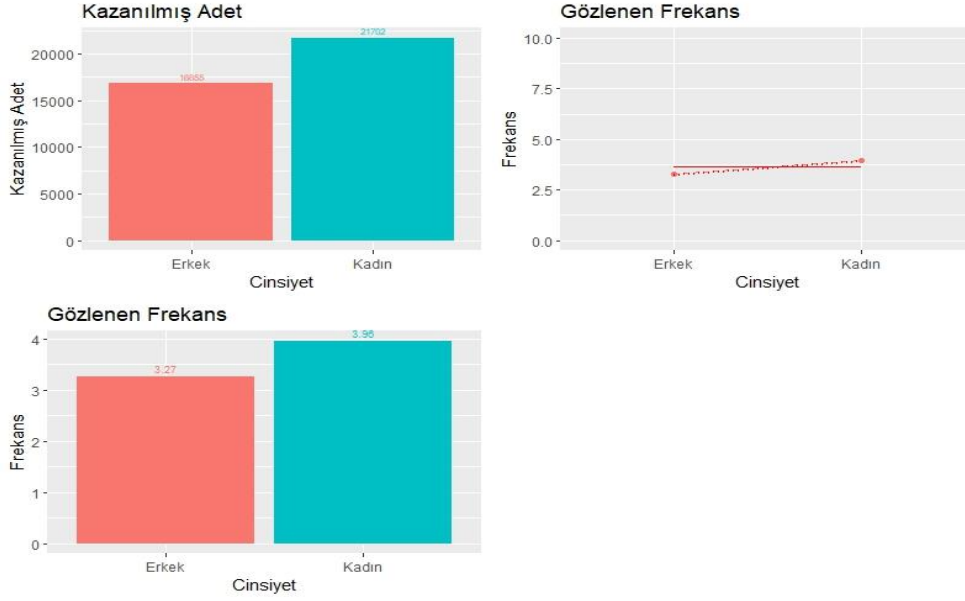
Yaş grubu değişkenine bağlı olarak hasar sıklık bilgileri Şekil 5.6'da gösterilmektedir.



Şekil 5.6. Yaş grup değişkenlerine bağlı hasar sıklıkları

Cinsiyet değişkeni kadın ve erkek olarak sınıflandırılmış olup cinsiyet değişkenine bağlı olarak yapılan hasar sıklık bilgileri Şekil 5.7'de gösterilmektedir. Cinsiyet değişkenine

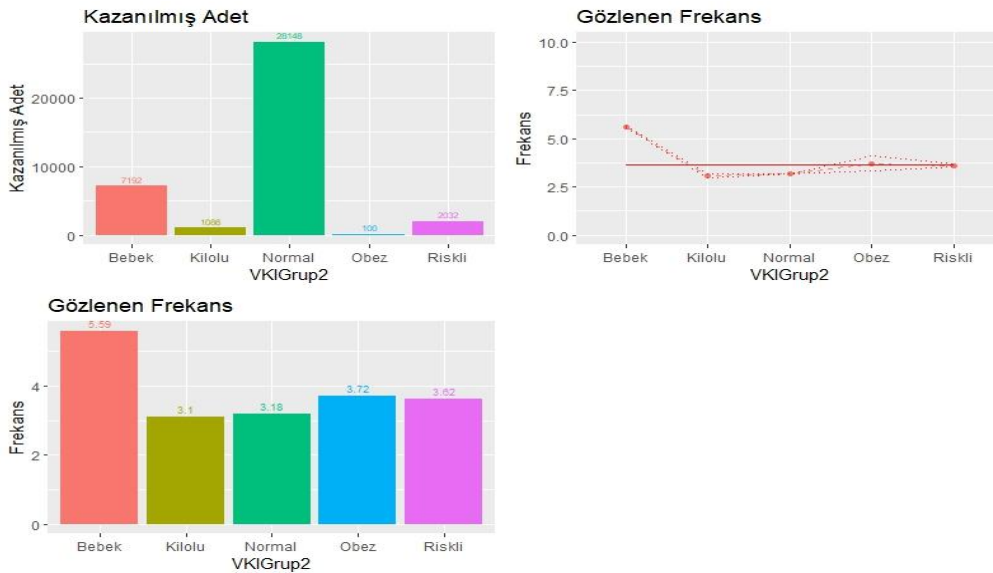
bağlı olarak incelemede, kadınların hastaneye gidiş sayısının erkeklerden daha fazla olduğu gözlemlenmiştir.



Şekil 5.7. Cinsiyet değişkenine bağlı hasar sıklıkları

Vücut kitle endeksine göre yapılan incelemede, en çok bebeklerin daha sonra da obez ve riskli olarak değerlendirilmiş olan ve sürprimin uygulanmakta olduğu grubun frekansının en yüksek olduğu gözlemlenmiştir.

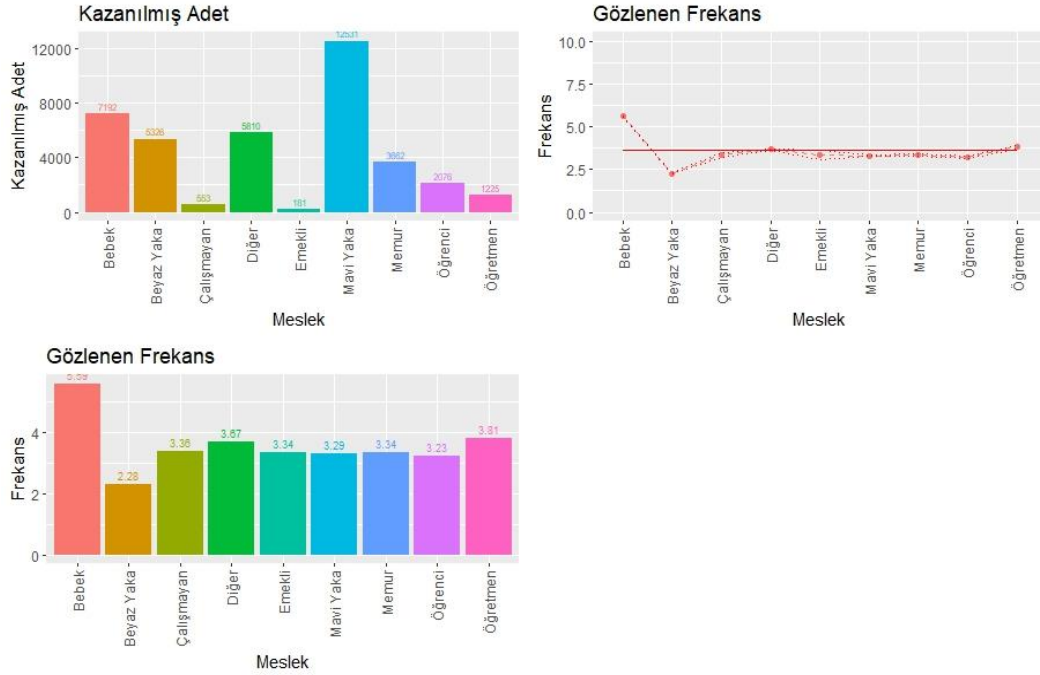
Vücut kitle endeksine bağlı olarak hasar sıklık bilgileri Şekil 5.8'de gösterilmektedir.



Şekil 5.8. Vücut kitle endeksi değişkenine bağlı hasar sıklıkları

Meslek grupları için yapılan incelemede, en çok kazanılan poliçe âdetinin mavi yakalı meslek grubundan geldiği gözlemlenmiştir. Hasar frekansı yüksek meslek grubunun ise

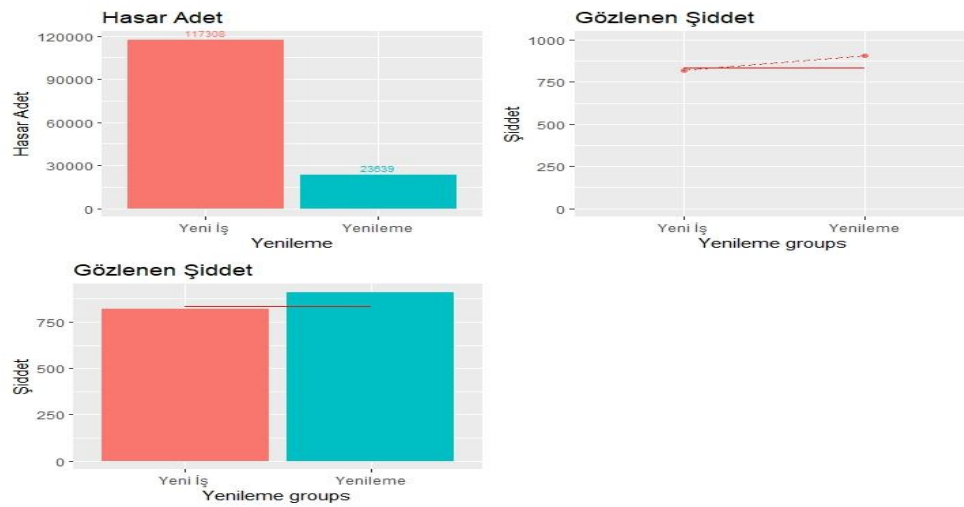
öğretmen olduğu belirlenmiştir. Meslek grubu değişkenine bağlı olarak hasar sıklık bilgileri Şekil 5.9’da gösterilmektedir.



Şekil 5.9. Meslek grup değişkenine bağlı hasar sıklıkları

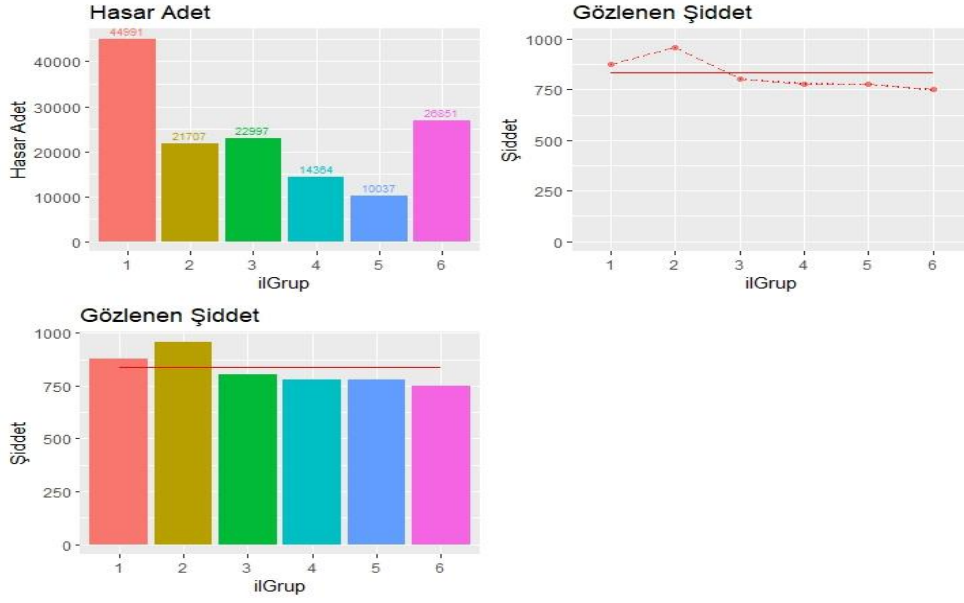
5.1.2. Hasar Şiddetine İlişkin Bazı İstatistikler

Portföy ortalama hasar şiddeti 835 TL olarak hesaplanmıştır. Hasar şiddeti açısından, yeni işlerde hasar âdeti daha yüksek olmakla birlikte, yenilemelerde gözlenen şiddetin daha fazla olduğu belirlenmiştir. Yenileme değişkenine bağlı olarak hasar şiddet bilgileri Şekil 5.10’da gösterilmektedir.



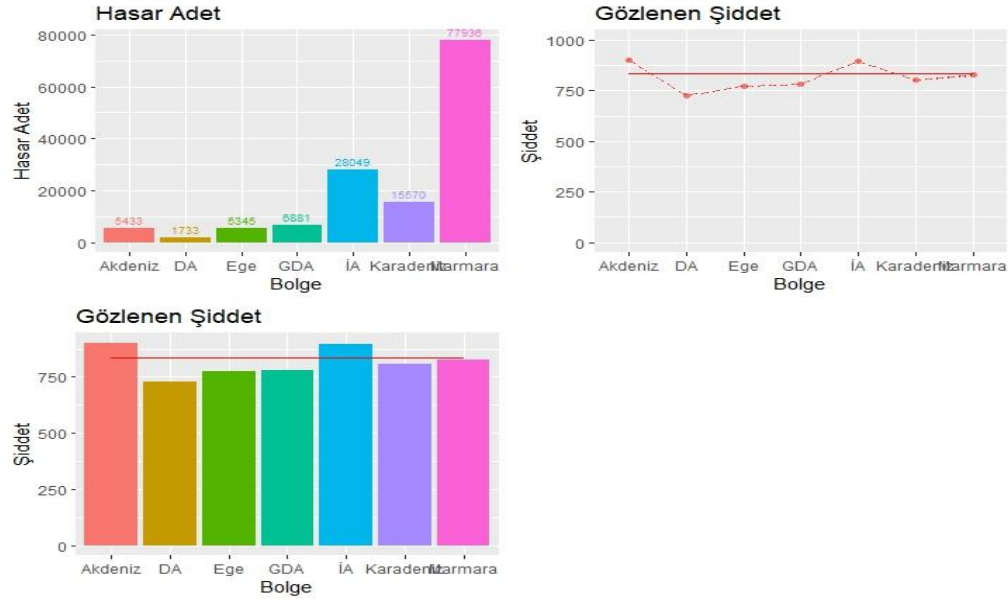
Şekil 5.10. Yenileme grup değişkenine bağlı hasar şiddeti bilgisi

İl grup değişkenine bağlı olarak hasar şiddet bilgileri Şekil 5.11’de gösterilmektedir.



Şekil 5.11. İl grup değişkenine bağlı hasar şiddeti bilgisi

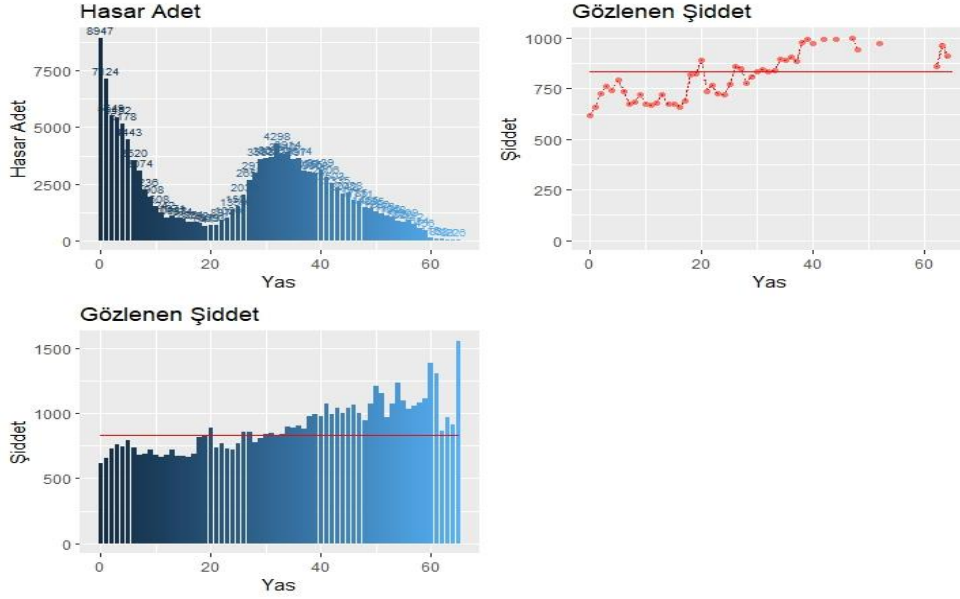
Bölge bazında hasar şiddetinin incelenmesinde, hasar Sayısının en fazla olduğu bölge Marmara Bölgesi olarak belirlenmiş olmasına rağmen, İç Anadolu ve Akdeniz bölgelerinde hasar şiddetinin en yüksek olduğu gözlenmiştir. Bölge değişkenine bağlı olarak hasar şiddet bilgileri Şekil 5.12’de gösterilmektedir.



Şekil 5.12. Bölge grup değişkenine bağlı hasar şiddeti bilgisi

Yaş değişkenine bağlı yapılan incelemede, 0-1 yaş arasında en yüksek hasar adetlerinin görüldüğü, hasar şiddetinde ise 60 yaş ve sonrası için en yüksek değerlerin gözlemlendiği

belirlenmiştir. Yaş değişkenine bağlı olarak hasar şiddet bilgileri Şekil 5.13’de gösterilmektedir.



Şekil 5.13. Yaş değişkenine bağlı hasar şiddeti bilgisi

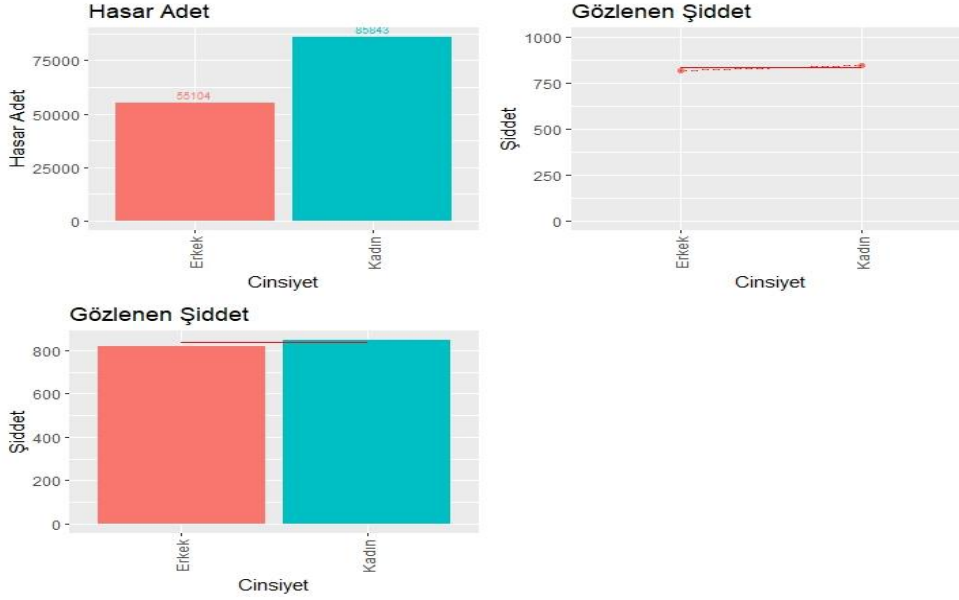
Yaş grup değişkenine bağlı yapılan incelemede, 0-6 yaş grubunda en yüksek hasar adetlerinin görüldüğü, hasar şiddetinde ise 65+ yaş grubu için en yüksek değerlerin gözlemlendiği belirlenmiştir.

Yaş grubu değişkenine bağlı olarak hasar şiddet bilgileri Şekil 5.14’de gösterilmektedir.



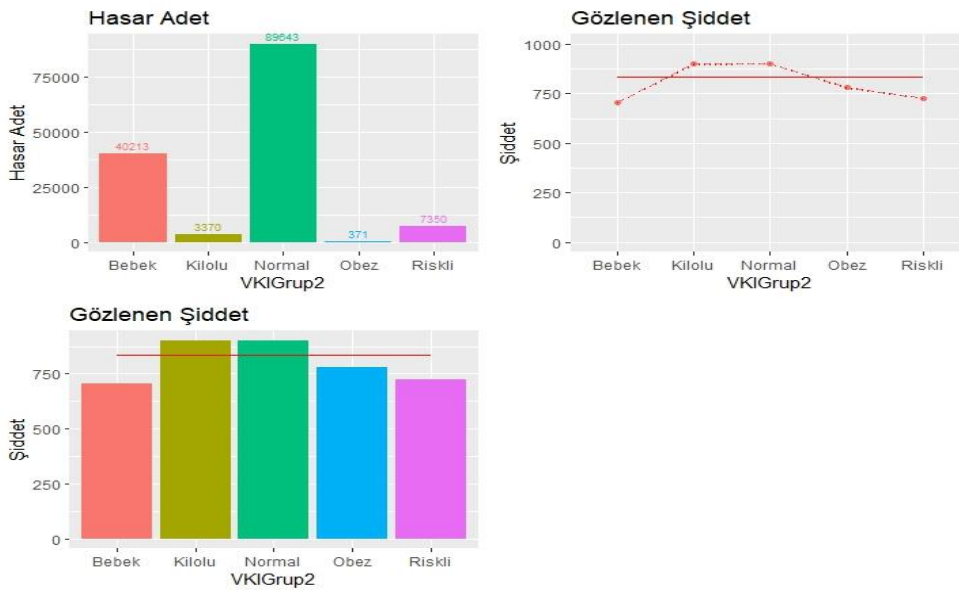
Şekil 5.14. Yaş grubu değişkenine bağlı hasar şiddeti bilgisi

Cinsiyet deęişkenine baęlı yapılan incelemede, kadın grubunda hasar sayısının ve hasar şiddetinin daha yüksek olduęu belirlenmiştir. Cinsiyet deęişkenine baęlı olarak hasar şiddet bilgileri Şekil 5.15’de gösterilmektedir.



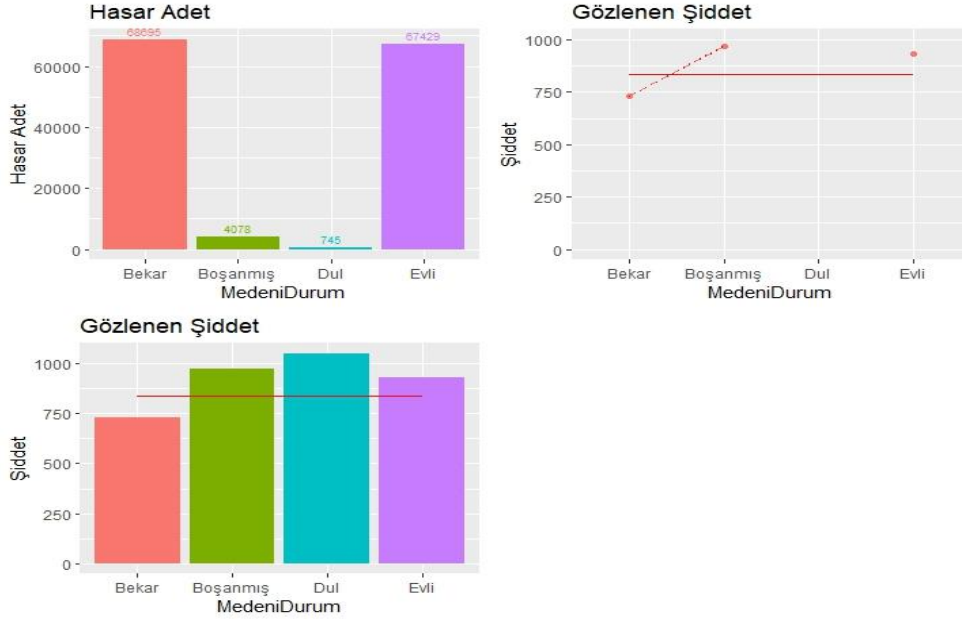
Şekil 5.15. Cinsiyet deęişkenine baęlı hasar şiddeti bilgisi

Vücut kitle endeksi deęişkenine baęlı yapılan incelemede, normal vücut kitle endeksine sahip risk grubunda en yüksek hasar adetlerinin görüldüğü, hasar şiddetinde ise kilolu olarak nitelendirilen vücut kitle endeksi 18-25 aralığının üstünde olan grup için en yüksek değerlerin gözleendięi belirlenmiştir. Yaş grubu deęişkenine baęlı olarak hasar şiddet bilgileri Şekil 5.17’de gösterilmektedir.



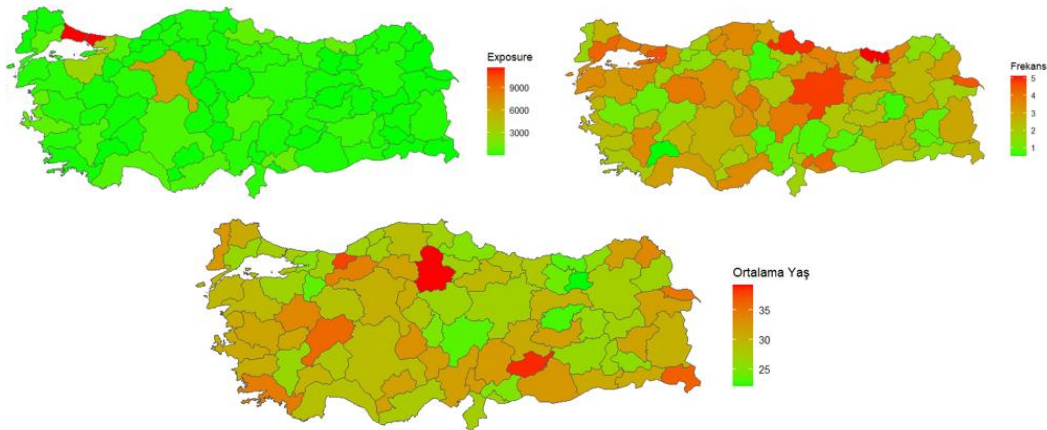
Şekil 5.16. Vücut kitle deęişkenine baęlı hasar şiddeti bilgisi

Medeni durum deęişkenine baęlı yapılan incelemede, hasar adedinin en yüksek çocuk grubunda gözleendięi, en yüksek hasar şiddetine sahip grubun ise medeni hali dul olanlar olduęu belirlenmiştir. Medeni durum deęişkenine baęlı olarak hasar şiddet bilgileri Şekil 5.17’de gösterilmektedir.



Şekil 5.17. Medeni durum deęişkenine baęlı hasar şiddeti bilgisi

İncelemede ayrıca, Türkiye’deki il ve bölge bazındaki hasar veri deęişimini ölçmek, ekonomik, sosyal ve demografik gelişmelerin analizdeki önemini görselleştirmek adına ortalama yaş, hasar sayısı ve hasar şiddet bilgileri Şekil 5.18’de sunulmuştur.



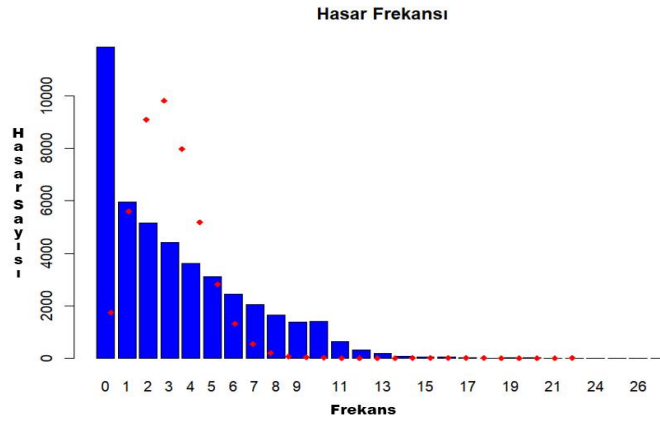
Şekil 5.18 Türkiye’deki il ve bölge bazındaki hasar veri deęişimi

5.2. GLM Uygulaması

5.2.1. GLM Frekans Model Kurulumu ve Değerlendirilmesi

Sigorta sektöründe hasar sayılarının modellenmesi ve frekans tahmini önemli bir kriter olup, genellikle sayım verisi için Poisson modeli kullanılmaktadır. Bu model, olayların bağımsız ve belirli bir zaman aralığında sabit bir oranla gerçekleştiği varsayımına dayanır. Sigorta sektöründe, hasarlar genellikle nadir ve rastgele olaylar olarak kabul edilir, bu nedenle hasar sayısı için Poisson modeli uygun bir seçim olacaktır.

Poisson modeli, veri setindeki hasar sayılarının dağılımını ve bu dağılımın bağımsız değişkenler (örneğin yaş, cinsiyet, coğrafi konum) ile ilişkisinin anlaşılmasına yardımcı olmaktadır. Bu, hasar riskini daha iyi tahmin etmek ve primleri daha doğru belirlemek için kritik öneme sahiptir. Şekil 5.19'da hasar frekansına ait poisson modeli gösterilmektedir.



Şekil 5.19 Gözlenen frekans poisson modeli

- Düşük hasar sayıları (0, 1, 2 gibi) için gözlemlenen frekanslar beklenen Poisson frekanslarına yakın görünmektedir.
- Ancak, hasar sayısı arttıkça, gözlemlenen ve beklenen frekanslar arasındaki uyum azalmakta ve bazı yüksek hasar sayılarında gözlemlenen frekanslar beklenenden daha düşük gözlenmektedir.
- Genel olarak, hasar sayısının az olduğu başlangıç kısımlarında Poisson dağılımına daha iyi bir uyum gözükmektedir.

5.2.2. Poisson Log Bağlantı Fonksiyonu ve Seçim Kriterleri

Poisson modelinde genel tercih log bağlantı fonksiyonu kullanılmasıdır. Bu tercihin nedeni, bağımsız değişkenlerin hasar sayısı üzerinde çarpan etkisi olması ve hasar sayısının sıfır ve pozitif tam sayılarla sınırlı olmasıdır.

Log bağlantı, modelin bu özelliklere uyum sağlamasına yardımcı olur ve hasar sayısının negatif olmasını engeller.

Tahmin edilen hasar sayısı, primleri belirlemede, risk yönetiminde ve finansal rezervlerin ayarlanmasında önemli bir rol oynar. Doğru tahminler, şirketin karlılığını ve finansal sağlığını doğrudan etkileyebilir.

5.2.3. GLM Frekans Yaklaşımı ile Değişken Seçimi ve Model Optimizasyonu

Çizelge 5.4'te hasar frekansını tahmin etmek için önerilen modelin değişken seçim süreci özetlenmiştir. Kullanılan değişkenlere bağlı olarak sırasıyla GLM (Model1) , GLM2 (Model 2) ve GLM3 (Model3) modelleri oluşturulmuştur.

Çizelge 5.4. GLM ile değişken seçimi model tablosu

	YasGrup	Cinsiyet	il	Yenileme	MedeniDurum	Meslek	VKIGrup
GLM1	✓	✓	✓	✓	✓	✓	✓
GLM2	✓	✓	✓	✓	✓		✓
GLM3	✓	✓	✓	✓	✓		

Model, bireylerin yaş grupları (YaşGrup), cinsiyeti (Cinsiyet), coğrafi konumları (il), poliçe yenileme durumları (Yenileme), medeni durumları (MedeniDurum) ve meslekleri (Meslek) gibi demografik ve bireysel özelliklerini içermektedir. Ayrıca, bireylerin sağlık durumunu yansıtan vücut kitle indeksi (VKIGrup) değişkeni de modele dâhil edilmiştir. Bu değişkenlerin seçimi, hasar frekansı üzerindeki potansiyel etkileri temel alınarak yapılmıştır ve sigorta şirketinin risk yönetimi ve fiyatlandırma stratejilerini geliştirmeye yönelik önemli bilgiler sağlamaktadır.

5.2.4. Genelleştirilmiş Lineer Modellerle Hasar Sıklığının Analizi

Model1, Model2 ve Model3 yaş grupları, cinsiyet ve coğrafi konum gibi demografik değişkenlerin yanı sıra, poliçe yenileme durumu ve medeni durum gibi değişkenleri de içermektedir. Değişkenlerin hasar sıklığı üzerindeki etkisini gösteren tahmin katsayılarını, standart hataları, z değerlerini ve p-değerlerine bağlı olarak ve bu değişkenlerin çoğunun hasar sıklığını anlamlı bir şekilde etkilediği elde edilmiştir. Model2, meslek değişkenini dışarıda bırakırken, Model3 hem meslek hem de vücut kitle indeksi (VKI) değişkenlerini dışarıda bırakarak daha basit bir model yapısını benimsemiştir. Analizde anlamlılık düzeyi 0.05 olarak seçilmiştir.

5.2.5. Hasar Sıklığı Modellerinin Performans Karşılaştırması

Hasar sıklığı modellerinin performans karşılaştırması Çizelge 5.5’te verilmektedir.

Çizelge 5.5. Hasar sıklığı modelinin performans karşılaştırma tablosu

Model	Par.	Ak. Değeri	Ö. İçi Değ.	Ö. Dışı Değ.	Ort. Frekans
GLM1-Hepsi	108	165942	243.7916	243.8797	3.6722
GLM2-Meslek yok	100	167177	247.3511	246.8888	3.6662
GLM3-Meslek, VKI yok	96	167222	247.4975	247.0158	3.6667

Çizelge 5.5’de GLM1 modeli tüm değişkenleri içerirken, GLM2 modeli meslek değişkenini ve GLM3 ise vücut kitle indeksi (VKI) değişkenini dışarıda bırakarak modellerin karmaşıklığını azaltmıştır. Modellerin performansları açısından akaike bilgi kriteri değerleri gözlemlenmiştir. GLM1, değişken açısından en kapsamlı model olup minimum Akaike Bilgi Kriteri değerine sahiptir. Çalışmanın bulguları, model seçiminde karmaşıklığın ve veriye uyumun dengelenmesinin önemini vurgular. Modeller arası performans karşılaştırması, sigorta şirketlerinin risk değerlendirme yaklaşımlarını iyileştirmelerine ve hasar sıklığını daha doğru tahmin edebilmelerine olanak tanımaktadır.

Modeller, Meslek ve VKI grupları değişkenlerinin belirli bir sınıfı için katsayı üretememiştir. Bu durum eksik veri veya doğrusal bağımlılık problemiyle açıklanabilir. Veri seti kontrol edildiğinde eksik veri yer almadığı gözlenmiştir. “Meslek -Öğretmen” ve “VKIGrup – Bebek” değişkenlerinde hem tekil olarak hem de birlikte dikkate alındığında doğrusal bağımlılık probleminin ortaya çıktığı belirlenmiştir. Bu nedenle öneri modelin bu değişkenleri içermeyecek şekilde seçilmesi önem taşımaktadır. GLM3 Modeli ile yapılan varyans analizi sonuçları Çizelge 5.6 da verilmiştir.

Çizelge 5.6. GLM3 modeli sonuçları tablosu

Terim	Ser. Derecesi	Sapma	K. Ser. Derecesi	Artık Hatalar
Sabit katsayı			35091	117453
YasGrupGLM	11	6815.6	35080	110637
CinsiyetGLM	1	1512.7	35079	109125
ilGLM	79	5445.9	35000	103679
YenilemeGLM	1	191.3	34999	103487
MedeniDurumGLM	3	132	34996	103355

Çizelge 5.6’deki modelde henüz herhangi bir değişken yokken, artık hatalar 117453’tür. Bu değer, modelin hiçbir açıklayıcı değişken içermediğindeki başlangıç noktasını temsil eder.

YasGrupGLM: yaş gruplarına ilişkin değişkenin eklendiği durumdaki artık varyans 6815,6 puan azalmıştır. Bu, yaş gruplarının hasar frekansı üzerinde önemli bir etkiye sahip olduğunu göstermektedir.

Değişken 11 farklı kategoriden oluşmakta ve modelin artık hatalarının 110637'ye düştüğü gözlenmiştir.

Cinsiyet değişkeni de modele eklendiği durumlardaki modelde, artık hatalar 1512,7 puan daha azalmıştır. Bu, cinsiyetin de hasar frekansı üzerinde anlamlı bir etkisi olduğunu göstermektedir ve modelin artık hatalar 109125'e düştüğü gözlenmiştir.

ilGLM: Coğrafi konumu temsil eden, bu değişken modele dâhil edildiğinde, artık hatalar 5445,9 puan azalmıştır. Bu, farklı illerdeki hasar frekansının farklılık gösterdiğini ve bu değişkenin de modele önemli bir katkıda bulunduğunu göstermektedir.

YenilemeGLM: Poliçe yenileme durumu değişkeni eklendiğinde, artık hatalar yalnızca 191,3 puan azalmıştır. Bu, poliçe yenilemenin hasar frekansı üzerinde daha düşük bir etkiye sahip olduğunu gösterir ancak hala model tarafından anlamlı bulunmuştur.

MedeniDurumGLM: Medeni durumu temsil eden bu değişkenin model eklenmesi, artık hatalar 132 puan daha azaltmıştır ve artık hatalar 103355'e düştüğü gözlenmiştir.

Her adımda, artık hatalar 'daki azalma, eklenen değişkenin modelin hasar frekansını açıklama yeteneğini artırdığını göstermektedir.

Df (serbestlik derecesi) değeri, değişkenin kategorilerinin (veya seviyelerinin) sayısını gösterir ve artık hatalardaki azalmanın istatistiksel anlamlılığını değerlendirmek için kullanılmaktadır.

Her bir değişken için artık hatalardaki azalma ne kadar büyükse, o değişkenin model için o kadar önemli olduğu sonucuna varılabilir.

Genel olarak, GLM3 modelindeki değişkenlerin tümü hasar frekansının belirlenmesinde önemli rol oynamakta ve modelin toplam varyansı açıklama kapasitesini iyileştirmektedir. Ancak, en büyük etkiyi yaratan değişkenlerin yaş grupları ve coğrafi konum olduğu görülmektedir.

5.2.6. Olabilirlik Oranı Testi ile GLM3 Model Değişkenlerinin Hasar Sıklığı Üzerindeki Etkisinin Değerlendirilmesi

Olabilirlik Oranı Testi (Likelihood Ratio Testi - LRT) testi sonuçları tablosu çizelge 5.7'de gösterilmektedir.

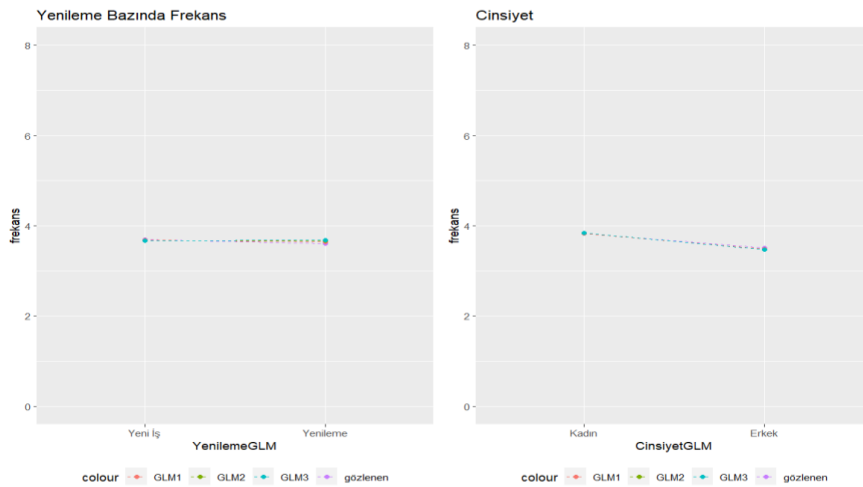
Çizelge 5.7. LTR testi sonuçları tablosu

Terim	S. Der.	Sapma	AIC	LTR	Pr(>Chi)
Boş		103355	183726		< 2.2e-16
YasGrupGLM	11	107857	188205	4501.4	< 2.2e-16
CinsiyetGLM	1	104544	184913	1189.1	< 2.2e-16
ilGLM	79	108927	189140	5572	< 2.2e-16
YenilemeGLM	1	103538	183907	182.6	< 2.2e-16
MedeniDurumGLM	3	103487	183852	132	< 2.2e-16

Çizelge 5.7'deki LTR, modelden bir terim çıkarıldığında model uyumunun ne kadar kötüleştiğini ölçen bir testtir ve bu test sonuçları, modeldeki değişkenlerin önemini istatistiksel olarak değerlendirmekte kullanılır. Modelden hiçbir değişken çıkarılmadan önce sapma değeri 103355'tir. Her bir değişken sırasıyla modelden çıkarıldığında, sapmadaki artış ve AIC değerlerindeki değişim, bu değişkenlerin modelde ne kadar önemli olduğunu gösterir. Örneğin, 'YasGrupGLM' değişkeni çıkarıldığında LTR'de 4501.4'lük bir artış gözlemlenmiş ve bu, yüksek derecede anlamlı bir p-değeri (< 2.2e-16) ile desteklenmiştir. Benzer şekilde, diğer değişkenlerin çıkarılması da modelin LTR değerinin önemli ölçüde artırmış ve bu da onların model için önemini vurgulamıştır. Bu bulgular, hasar sıklığı modellemesinde, yaş grubu, cinsiyet, coğrafi konum, poliçe yenileme durumu ve medeni durumun önemli değişkenler olduğunu ve bu değişkenlerin dikkate alınmasının modelin tahmin gücünü artıracaklarını göstermektedir.

5.2.6.1. GLM Frekans Modeli Karşılaştırmalı Analizi

Şekil 5.20'de poliçe yenileme durumu ve cinsiyetin hasar frekansı tahminlerine olan etkileri üç farklı GLM modeli için verilmektedir.

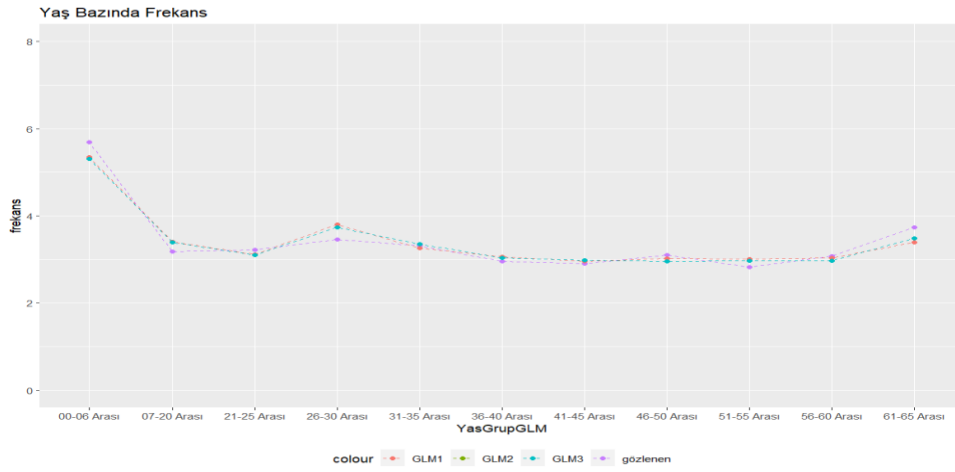


Şekil 5.20 Poliçe yenileme durumu ve cinsiyete göre glm modelleri ile hasar frekansı tahmini

Elde edilen grafikler, modellerin gözlemlenen verilere göre tutarlı tahminler yaptığını ortaya koymaktadır. Poliçe yenileme durumunun her iki cinsiyet için de hasar frekansı üzerinde belirleyici bir etkisi olmadığı, ancak cinsiyetin hasar frekansını etkileyen önemli bir faktör olduğu görülmektedir.

Erkeklerde gözlemlenen hasar frekansının kadınlara kıyasla daha düşük olduğu bulgusu, risk değerlendirme ve prim belirleme stratejilerinin cinsiyet özelliklerine göre düzenlenmesi gerektiğini vurgulamaktadır.

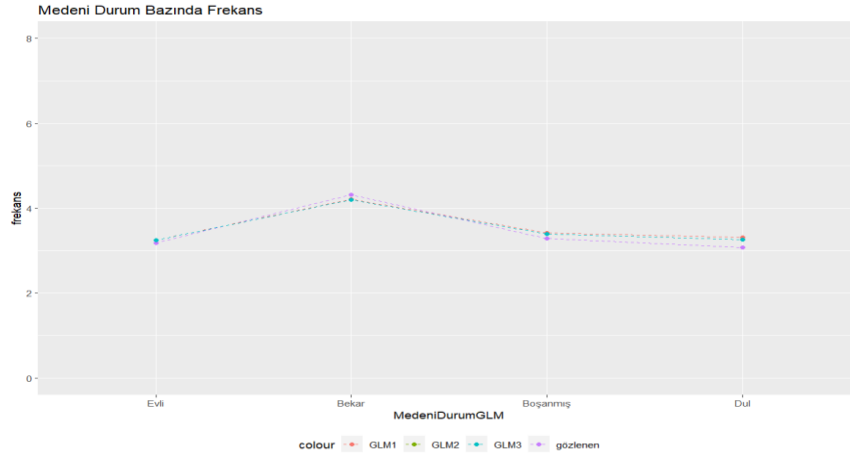
Bu analiz, sigorta şirketlerinin daha doğru risk tahminleri yapmalarına ve cinsiyet ile yenileme durumuna göre özelleştirilmiş poliçeler sunmalarına olanak tanıyacak değerli bilgiler sağlamaktadır. Şekil 5.21’de yaş gruplarına bağlı olarak GLM1, GLM2 ve GLM3’deki hasar frekansı değerleri verilmektedir.



Şekil 5.21 Yaş gruplarına göre hasar frekansının gözlemlenmesi

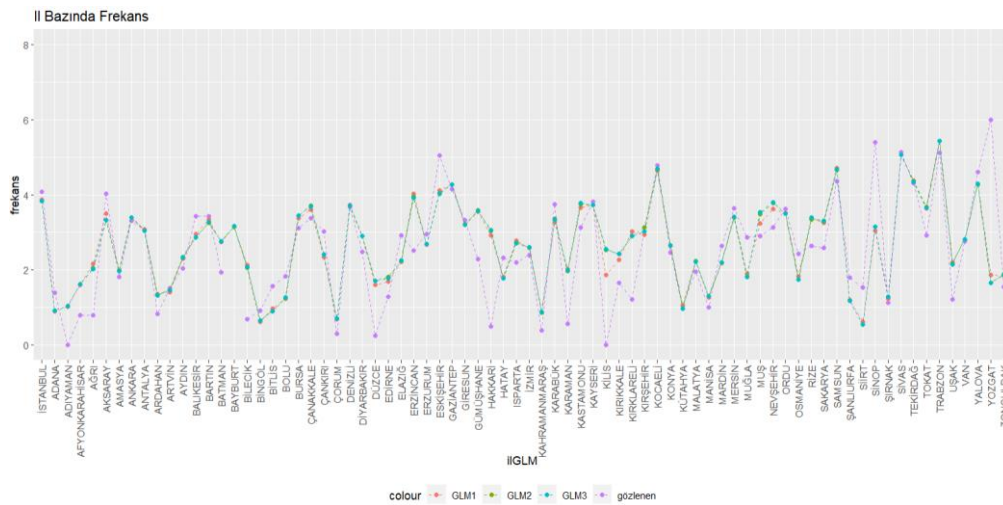
GLM1, GLM2 ve GLM3 modelleri, yaş gruplarına dayalı olarak ayrıştırılmış sağlık sigortası hasar verileri kullanılarak kurulmuştur. Gözlemlenen verilerle modellerin tahminleri karşılaştırılmış, her modelin yaş grupları bazında nasıl bir performans sergilediği incelenmiştir. Genç yaş gruplarında GLM1 modelinin, diğer modellere göre daha yüksek frekans tahminleri sunduğu; buna karşılık, orta yaş gruplarında tüm modellerin gözlemlenen verilerle benzer sonuçlar verdiği tespit edilmiştir. Bu sonuçlar, risk değerlendirme ve sigorta primlerinin belirlenmesinde yaş faktörünün önemini vurgulamakta ve farklı modelleme tekniklerinin etkinliğini ortaya koymaktadır. Yapılan bu analizle, aktüeryal çalışmalar için model seçimi ve parametre tahminleri konusunda önemli iç görüler elde edilmiştir. Bu çalışma, sağlık sigortası poliçe tasarımında ve risk yönetiminde yaş faktörünün dikkate alınmasının gerekliliğini ve bu alandaki modellerin

özelleştirilmesinin önemini ortaya koymaktadır. Şekil 5.22 Medeni durum değişkenine göre GLM1, GLM2 ve GLM3’deki hasar frekans sayıları gösterilmektedir.



Şekil 5.22 Medeni duruma göre hasar frekansının gözlemlenmesi

Evli, bekâr, boşanmış ve dul bireylerin hasar bildirim frekansları, Genelleştirilmiş Lineer GLM1, GLM2 ve GLM3 kullanılarak modellenmiş ve gözlemlenen verilerle karşılaştırılmıştır. Analiz sonuçları, evli bireylerin biraz daha yüksek frekansa sahip olduğunu, diğer medeni durum kategorilerinin ise benzer frekans değerleri sergilediğini ortaya koymaktadır. Bu bulgular, sigorta şirketlerinin medeni durumu bir risk faktörü olarak değerlendirmesinin potansiyel önemini göstermektedir. Ayrıca, kullanılan istatistiksel modellerin, çeşitli medeni durum kategorileri arasındaki frekans tahminlerini başarılı bir şekilde yapabildiği ve bu modellerin risk yönetimi ve prim belirleme süreçlerinde etkili araçlar olduğunu göstermiştir. Şekil 5.23 Türkiye'deki il bazında GLM1, GLM2 ve GLM3 modellerindeki hasar frekans değerlerini göstermektedir.



Şekil 5.23 Türkiye'deki illerin hasar frekans gözlemlenmesi

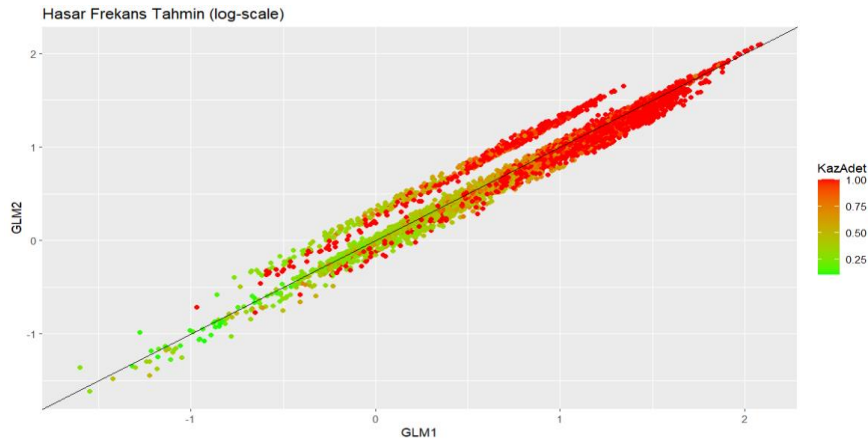
GLM1, GLM2 ve GLM3 modelleri, farklı coğrafi ve sosyo-ekonomik özelliklere sahip illere uyarlanarak, hasar frekanslarının nasıl farklılık gösterdiğini incelemiştir. Gözlemlenen verilerle yapılan karşılaştırmalar, modellerin performansını değerlendirmekte ve hasar frekansları üzerinde etkili olabilecek yerel faktörler hakkında önemli iç görüler sağlamaktadır.

Bu tür analizler, sigorta şirketlerine risk değerlendirmesi ve prim belirleme süreçlerinde yardımcı olabilecek veri odaklı yaklaşımlar sunmaktadır. Ayrıca, elde edilen sonuçlar, politika yapıcılar ve yerel yönetimler için sağlık ve güvenlik alanlarında iyileştirme yapma konusunda rehberlik edebilir.

Sigorta sektöründe veri analitiği ve modellemenin önemini vurgulamakta ve Türkiye'nin farklı illerindeki hasar frekanslarını anlama yolunda katkıda bulunmaktadır.

5.2.6.2. GLM Frekans Tahminlerinin Logaritmik Ölçekte Hasar Frekansı Tahminlerinin Karşılaştırmalı Analizleri

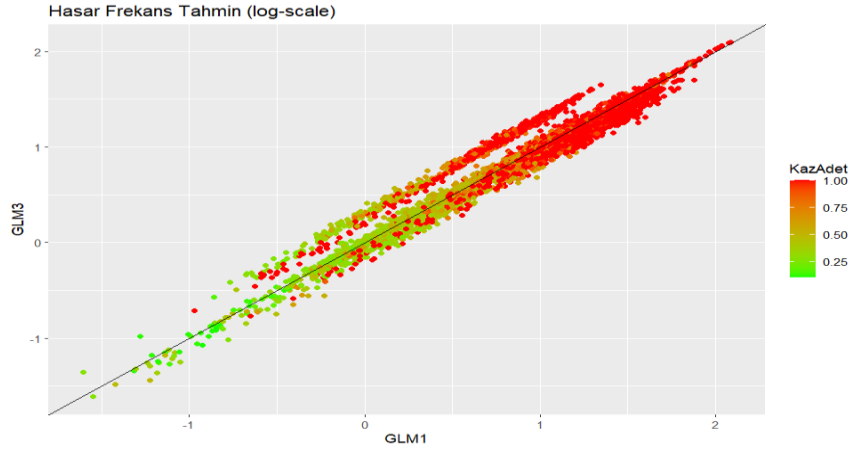
Bu bölümde hasar frekansları yönünde karşılaştırma yapılmaktadır. Şekil 5.24Te GLM2 ve GLM1 modellerinin karşılaştırmalı grafiği göstermektedir.



Şekil 5.24. GLM2 ve GLM1 modellerinin karşılaştırmalı değerlendirilmesi

Sigorta hasar frekansı tahminlerinde iki farklı GLM1 ve GLM2 performansını logaritmik ölçekte değerlendirmektedir. Analiz, modellerin her biri için yapılan tahminlerin nasıl bir ilişki gösterdiğini ve sigorta poliçesi başına düşen ortalama hasar sayısı (KazAdet) değişkeni üzerindeki etkilerini incelemektedir. Renk gradyanı kullanarak, hasar frekansı tahminlerinin yoğunluk dağılımı görselleştirilmiş ve modellerin tahmin gücü karşılaştırılmıştır. Bulgular, modellerin genel olarak uyumlu tahminler sunduğunu, ancak belli durumlarda birbirinden farklılaşabileceğini göstermektedir. Bu tür karşılaştırmalar,

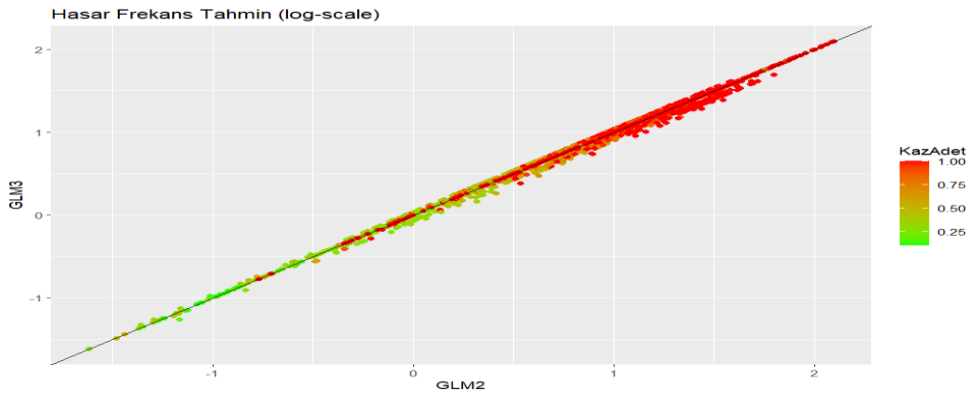
sigorta şirketlerine ve politika yapıcılara, risk değerlendirmesi ve prim belirleme süreçlerinde kritik bilgiler sağlayabilir. Tez, bu analiz tekniklerinin sigorta sektöründeki uygulamaları hakkında derinlemesine bir bakış sunmaktadır ve bu modellerin gelecekteki tahmin ve fiyatlandırma stratejilerine katkıda bulunabilecek potansiyeli değerlendirilmiştir. Şekil 5.25. GLM1 ve GLM3 modellerinin karşılaştırmasını göstermektedir.



Şekil 5.25. GLM1 ve GLM3 modellerinin karşılaştırmalı değerlendirilmesi

Modeller arası tahminlerin uyumlu olup olmadığı ve sigorta poliçesi başına düşen ortalama hasar sayısının modellerin tahmin kabiliyeti üzerindeki etkisi incelenmiştir. Nokta yoğunluğunun ve renk dağılımının analizi, her iki modelin de genel olarak güçlü bir ilişki içinde olduğunu ve benzer tahminler yaptığını göstermiştir. Ancak, renk skalası üzerindeki gradyanlar, kazaların frekansında bazı farklılıkları da ortaya koymaktadır. Bu bulgular, risk değerlendirmesi ve prim fiyatlandırma stratejileri için önemli olup, sigorta şirketlerine modeller arası seçim yapma ve poliçe tasarımında kararlar alırken yol gösterici olmuştur.

Şekil 5.26. GLM2 ve GLM3 modellerinin karşılaştırmasını göstermektedir.



Şekil 5.26. GLM2 ve GLM3 modellerinin karşılaştırmalı değerlendirilmesi

İki farklı GLM2 ve GLM3 arasında hasar frekansı tahminlerinin uyumunu ve tutarlılığını logaritmik ölçekte incelenmiştir. Her bir modelin tahminleri, poliçe başına düşen ortalama hasar sayısına göre değerlendirilmiş ve renk skalası kullanılarak görselleştirilmiştir. Analiz, modellerin benzer sonuçlar ürettiğini ve hasar frekansının değerlendirilmesinde birbirleriyle güçlü bir korelasyona sahip olduğunu ortaya koymuştur. Bu sonuçlar, risk değerlendirme ve aktüeryal fiyatlandırma için önemli olup, sigorta şirketlerine ve analistlere model seçimi ve poliçe tasarımında kritik bilgiler sağlamaktadır.

GLM Frekans Model PSUEDOCODE Açıklaması

GLM frekans modelinin geliştirilmesi süreci, kapsamlı bir veri işleme ve analiz sürecini içerir. İlk olarak veri yüklenmiştir.

Veri ön işleme aşamasında, veri setindeki değişkenlerin tipleri düzenlenmiş ve faktör düzeyleri tanımlanmıştır. Bu adım, veri setinin modelleme için uygun hale getirilmesini sağlamıştır.

Tanımlayıcı istatistikler belirlemek için veri setinin yapısını ve özet istatistiklerini sunarak, veri setinin genel bir görünümünü sağlamıştır.

GLM model kurulumu, glm fonksiyonu kullanılarak gerçekleştirilmiştir. Bu model, Poisson regresyonunu kullanarak, frekansın bağımsız değişkenlere bağlı olarak nasıl değiştiğini modellemek amacıyla kurulmuştur. Modelin kurulumu sırasında, bağımsız değişkenler olarak demografik ve coğrafi faktörler gibi çeşitli değişkenler dikkate alınmıştır.

Model tahminleri ile test veri seti üzerinde yapılmış eğitim veri seti için model tahminlerini sağlamıştır. Bu tahminler, modelin performansını değerlendirmek için kullanılmıştır.

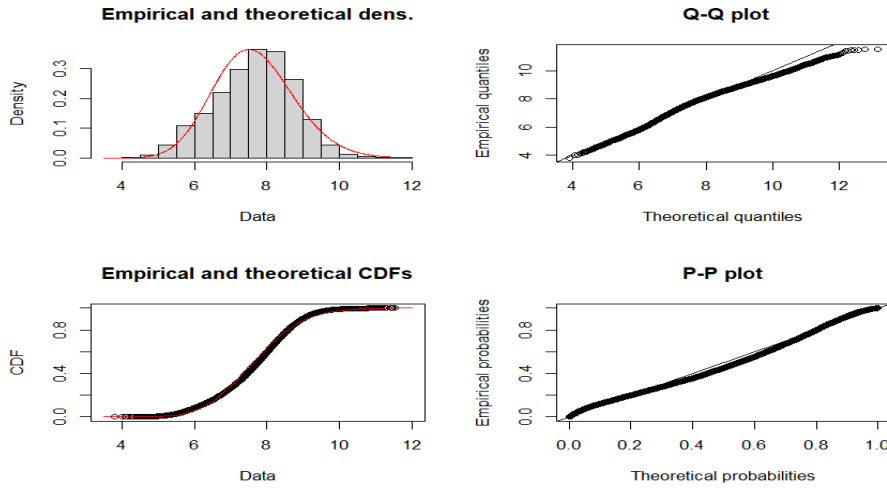
Modelin performansı, Poisson Deviance fonksiyonu ile ölçülmüştür. Bu fonksiyon ile modelin gerçek değerlere ne kadar yakın tahminler yaptığını gösteren bir sapma skoru hesaplanmıştır. Son olarak, modelin tahmin sonuçları ve gerçek veriler, grafiksel olarak görselleştirilmiştir.

5.2.7 GLM Şiddet Model Kurulumu ve Değerlendirilmesi

Gamma dağılımı ve log bağlantı fonksiyonu, genellikle sürekli ve pozitif değerler alan veri setleri için kullanılır ve sigorta sektöründe hasar tutarlarının modellenmesinde sıkça tercih edilir. Hasar tutarları genellikle asimetric dağılım gösterir; bu, çok sayıda düşük tutarlı hasar ve az sayıda çok yüksek tutarlı hasarın olduğu anlamına gelir. Gamma dağılımı, bu

tür asimetrik ve pozitif sürekli veriler için uygun bir seçimidir çünkü gamma dağılımı, yalnızca pozitif değerler alır ve bu da hasar tutarlarının doğasına uygundur çünkü hasar tutarları negatif olamaz.

Şekil 5.27’de Hasar tutarlarının gamma dağılımına ilişkin yoğunluk, QQ, PP ve kümülatif dağılım fonksiyonu gösterilmektedir.



Şekil 5.27 Gamma dağılımının sigorta hasar tutarları üzerindeki uygulanabilirliğinin istatistiksel değerlendirilmesi

Hasar tutarlarının Gamma dağılımı ile modellemesinin istatistiksel olarak ne derece uygun olduğunu dört farklı uygunluk testi kullanarak değerlendirmektedir. Deneyimsel ve teorik yoğunluk fonksiyonlarının karşılaştırılması, veri setinin Gamma dağılımına genel bir uyum gösterdiğini ortaya koymuştur.

Q-Q ve P-P grafikleri, modelin merkezi eğilimdeki verileri iyi yakaladığını göstermiştir. Kümülatif dağılım fonksiyonları arasındaki uyum, modelin genel eğilimi doğru bir şekilde yakaladığını göstermektedir.

5.2.8. Gamma Log Bağlantı Fonksiyonu ve Seçim Kriterleri

Gama dağılımı ve log bağlantı fonksiyonu ve seçim kriterlerine ilişkin temel açıklamalar şu şekildedir;

Esneklik:

Dağılımın şekli ve ölçeği üzerinde kontrol sağlayarak, geniş bir yelpazede veri seti davranışlarına uyum sağlayabilir. Bu esneklik, çeşitli büyüklükteki hasarlar için dağılımın şeklini etkili bir şekilde modellemeye olanak tanımaktadır.

Çarpıklık:

Hasar tutarları için gamma dağılımı, sağa çarpık veri setlerini modellemek için uygun olan parametreler içermektedir.

Log Bağlantı Fonksiyonu:

Log bağlantı, bağımsız değişkenlerin hasar tutarları üzerindeki çarpan etkisini modellemek için kullanılır. Bu, hasar tutarlarındaki yüzdelik değişimlerin bağımsız değişkenlerdeki birim değişimlerle ilişkilendirilmesini sağlar ve model tahminlerinin pozitifliğini garanti etmektedir.

Seçim Kriterleri:

Model seçiminde, Akaike Bilgi Kriteri (AIC) gibi istatistiksel ölçütler, veriye en iyi uyan ve en az karmaşıklığa sahip modeli bulmak için kullanılır.

Gamma dağılımı ve log bağlantı, genellikle bu tür veriler için düşük AIC değerleri üretir, dolayısıyla modelin veriye iyi uyduğunu ve aşırı uyum sorunlarını (overfitting) minimize ettiğini göstermektedir.

GLM Şiddet Model PSUEDOCODE Açıklaması

GLM Şiddet Modelinizin geliştirilmesi sürecinde coğrafi veri analizi, veri manipülasyonu ve görselleştirme için temel oluşturmuştur. Veri yükleme adımında, veri setinin yapısı incelenmiştir.

Veri ön işleme sürecinde, gereksiz veriler filtrelenmiş, değişken tipleri düzeltilmiş ve kategorik değişkenler için faktör düzeyleri tanımlanmıştır.

Modelleme öncesinde, benzersiz veri grupları belirlenmiş ve veri seti bu gruplara göre bölünmüştür. Grupların dağılımı kontrol edilmiştir.

GLM Şiddet Modeli kurulumunda, glm fonksiyonu kullanılarak Gamma dağılımı temelli bir regresyon modeli oluşturulmuştur.

Bu modelde, hasar tutarının çeşitli demografik ve coğrafi faktörlere bağlı olarak nasıl değiştiği incelenmiştir.

Model kurulumu sırasında, farklı kombinasyonlarda bağımsız değişkenler kullanılarak çeşitli modeller oluşturulmuş ve bu modellerin varyans değişkenlikleri belirlenmiştir.

Modelin tahmin performansı eğitim ve test veri setleri üzerinde değerlendirilmiştir. Bu tahminler, modelin performansını değerlendirmek için kritik öneme sahiptir. Son olarak, modelin performansını görselleştirmek için çeşitli grafikler oluşturulmuştur.

5.2.9. GLM Şiddet Yaklaşımı ile Değişken Seçimi ve Model Optimizasyonu

Çizelge 5.8'deki GLM yaklaşımını kullanarak oluşturduğumuz dört farklı istatistiksel modelin değişken seçim sonuçları karşılaştırmalı olarak sunulmaktadır.

Çizelge 5.8. GLM şiddet değişken model tablosu

	YasGrup3	İlGrup	Yenileme	Cinsiyet	MedeniDurum	Meslek	VKIGrup2
Model1	✓	✓	✓	✓	✓	✓	✓
Model2	✓	✓	✓	✓	✓	✓	
Model3	✓	✓	✓			✓	
Model4	✓	✓	✓				

Her bir model, belirli kriterlere göre seçilmiş değişkenlerle inşa edilmiştir. Model 1, en kapsamlı model olarak tüm değişkenleri içermekte ve temel referans modelimiz olarak kullanılmıştır. Model 2 ve Model 3, sırasıyla 'Meslek' ve 'VKIGrup2' değişkenlerini dışlayarak oluşturulan alternatif modellerdir. Model 4 ise en minimaliz model olarak, sadece temel demografik değişkenleri içeren bir modeldir. Modellerin performans karşılaştırması, Akaike Bilgi Kriteri (AIC) gibi istatistiksel ölçümler kullanılarak yapılmıştır.

Yaş grupları, coğrafi bölgeler, yenileme durumu, meslek grupları ve diğer sosyal demografik faktörlerin hasar şiddetine etkisi dört ayrı model üzerinden değerlendirilmiştir. Model 1 tüm değişkenleri içerirken, Model 2'de vücut kitle indeksi "VKI", Model 3'te "VKI," cinsiyet ve medeni durum ve Model 4'te meslek, cinsiyet, medeni durum ve "VKI" değişkenleri çıkarılarak farklı kombinasyonların etkileri analiz edilmiştir.

Her modelin istatistiksel anlamlılığı, z-değerleri ve p-değerleri ile değerlendirilmiştir. Modeller arasındaki farklılıklar Akaike Bilgi Kriteri (AIC) ile karşılaştırılmıştır. Bu analizler, hasar şiddetini etkileyen önemli faktörlerin belirlenmesine ve daha etkili risk yönetimi stratejilerinin geliştirilmesine katkı sağlamaktadır.

Analizler sonucunda, yaş gruplarının ve coğrafi grupların hasar şiddeti üzerinde anlamlı etkileri olduğu, bazı meslek gruplarının ve yenileme durumunun da önemli olduğu

gözlemlenmiştir. Özellikle, “İlGrup” ve “yasgrup3” değişkenleri tüm modellerde istatistiksel olarak anlamlı çıkmıştır.

5.2.10. GLM4 Modeli ile ANOVA Sonuçlarının İncelenmesi

Çizelge 5.9’deki ANOVA tablosu, bu terimlerin modeldeki toplam varyans içinde ne kadarını açıkladığını ve kalan varyansın büyüklüğünü göstermektedir.

Çizelge 5.9. GLM4 modeli tablosu

Terim	Ser. Derecesi	Sapma	K. Ser. Derecesi	Kalan Sapma
Sabit Değerler			25484	13615
yasgrup3GLM	4	638.20	25480	12977
ilGrupGLM	5	171.81	25475	12805
YenilemeGLM	1	49.08	25474	12756

Modelin hiçbir terim içermediği durumu ile karşılaştırıldığında, yaş gruplarının toplam varyans üzerinde en büyük etkiye sahip olduğu gözlemlenmiştir, bu da 638.20 sapma değeri ile ifade edilmektedir.

Coğrafi grupların etkisi 171.81 sapma değeri ile daha az belirgin, ancak yine de anlamlıdır. Yenileme durumu ise modeldeki varyansın küçük bir kısmını açıklamakta ve 49.08 sapma değeri ile en az etkiye sahip görünmektedir.

Bu sonuçlar, demografik ve coğrafi faktörlerin hasar şiddetini anlamada önemli değişkenler olduğunu göstermektedir.

Kalan sapma değerleri modelin bağımlı değişkenin varyansını ne kadar iyi açıkladığına dair fikir vermektedir. Özellikle, yaş grupları ve coğrafi grupların hasar şiddeti üzerindeki etkilerinin sigorta primlerinin daha doğru hesaplanması ve risk yönetiminin iyileştirilmesi için kullanılabilir bilgiler sağladığı düşünülmektedir.

5.2.11. LRT ile GLM4 Model Değişkenlerinin Hasar Tutar Güncel Sıklığı Üzerindeki Etkisinin Değerlendirilmesi

Çizelge 5.10’deki GLM4 modeli kapsamında LTR kullanılarak değerlendirilmiştir.

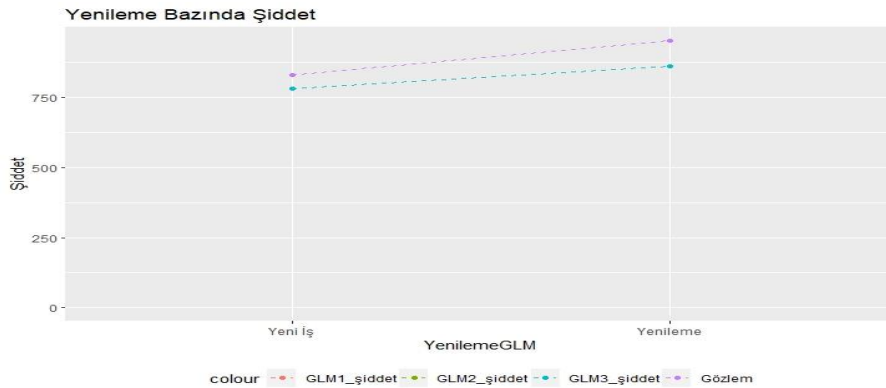
Çizelge 5.10. LTR testi sonuçları tablosu

Terim	Ser. Derecesi	Sapma	AIC	LRT	Pr(>Chi)
Bos		12756	385251		< 2.2e-16
yasgrup3GLM	4	13326	385803	560.01	< 2.2e-16
ilGrupGLM	5	12937	385419	178.04	< 2.2e-16
YenilemeGLM	1	12805	385297	48.26	< 2.2e-16

Modelde incelenen deęişkenler arasında yaşı grupları (yasgrup3GLM), il grupları (ilGrupGLM) ve yenileme durumu (YenilemeGLM) bulunmaktadır. Analiz sonuçlarına göre, istatistiksel olarak tüm deęişkenlerin hasar tutar şiddeti üzerinde anlamlı etkileri olduęu ortaya çıkmıştır. Modelde hiçbir terim olmadan (NULL model) elde edilen serbestlik derecesi ve AIC deęeri ile karşılaştırıldığında, eklenen her bir deęişkenin modelin uyumunu önemli ölçüde iyileştirdięi gözlemlenmiştir. Yaşı gruplarının eklenmesi modelde en büyük iyileşmeyi sağlamış ve 560.01 LRT deęeri ile en yüksek etkiyi göstermiştir. İl gruplarının etkisi 178.04 LRT deęeri ile anlamlı olmakla birlikte, yaşı gruplarına göre daha düşük bir etkiye sahiptir. Yenileme durumu ise 48.26 LRT deęeri ile modelin uyumuna en az katkıda bulunan deęişkendir. Bu bulgular, sigorta şirketlerinin hasar şiddetini tahmin ederken yaşı ve coęrafi faktörleri daha fazla dikkate almaları gerektiğini önermektedir. Ayrıca, AIC deęerlerinin her bir deęişken eklenmesi durumunda azaldığı görülerek, modelin deęişkenlerle birlikte tahmin yeteneğinin arttığı belirlenmiştir.

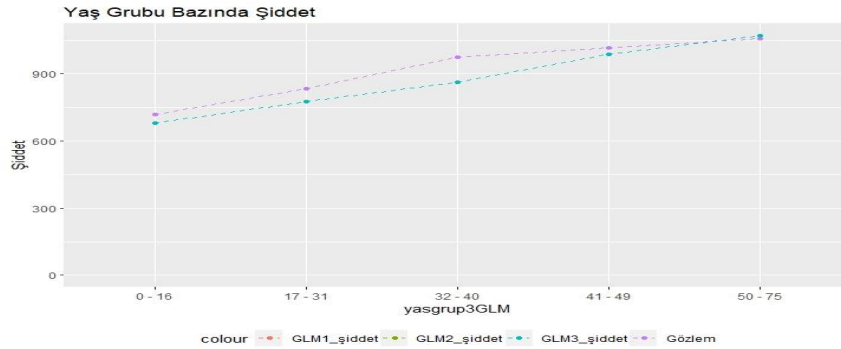
5.2.10.1 GLM Şiddet Modeli Karşılaştırmalı Analizi

Şekil 5.28’de yenileme gruplarına göre yenileme bazında şiddeti göstermektedir.



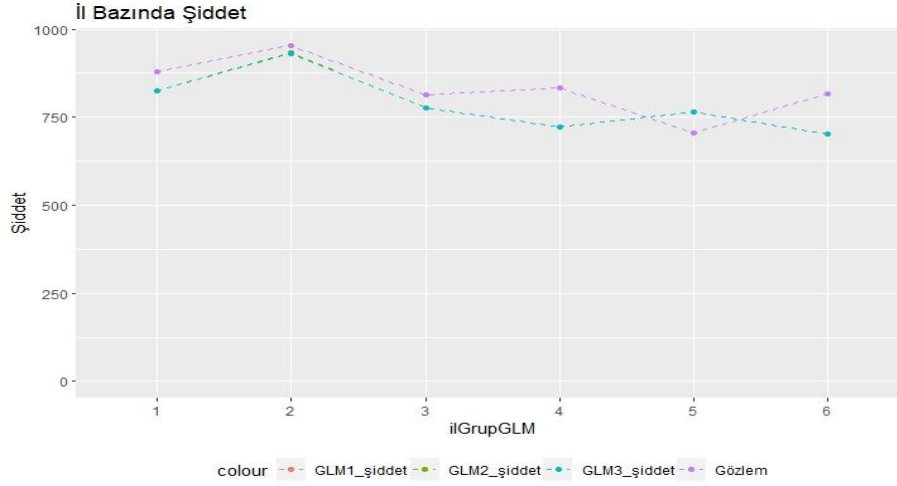
Şekil 5.28 Yenileme gruplarına göre hasar şiddetinin gözlemlenmesi

Şekil 5.29’da yaşı gruplarına göre hasar şiddetinin gözlenmesini göstermektedir.



Şekil 5.29 Yaşı gruplarına göre hasar şiddetinin gözlemlenmesi

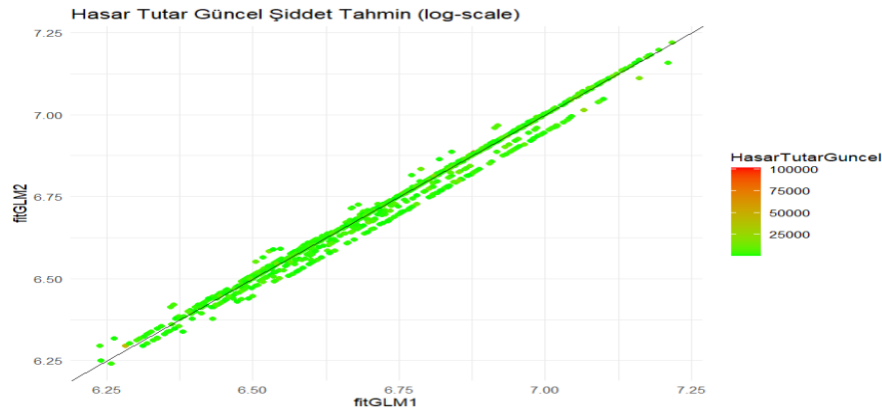
Şekil 5.30'de Türkiye'deki il gruplarına göre hasar şiddetinin gözlemlenmesi göstermektedir.



Şekil 5.30 Türkiye'deki il gruplarına göre hasar şiddetinin gözlemlenmesi

5.2.10.1 GLM Şiddet Tahminlerinin Logaritmik Ölçekte Karşılaştırılması

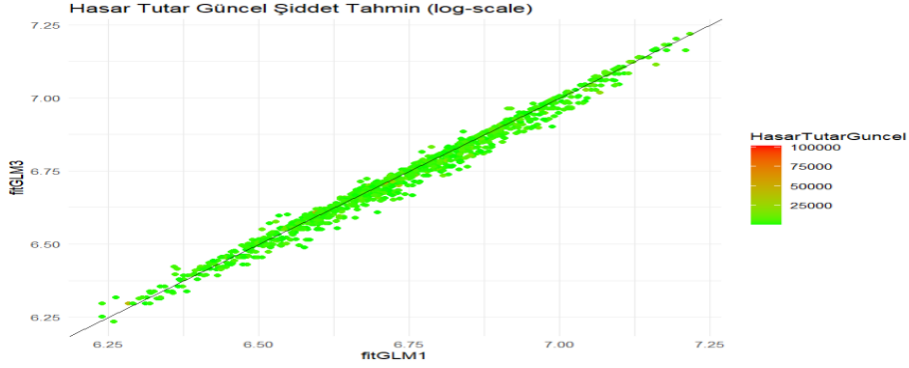
Şekil 5.31'de GLM1 ve GLM2 Modellerinin hasar tutarı güncel şiddeti tahminlerinin logaritmik ölçekte karşılaştırılması gösterilmektedir.



Şekil 5.31. GLM1 ve GLM2 Modellerinin hasar tutarı güncel şiddeti tahminlerinin logaritmik ölçekte karşılaştırılması

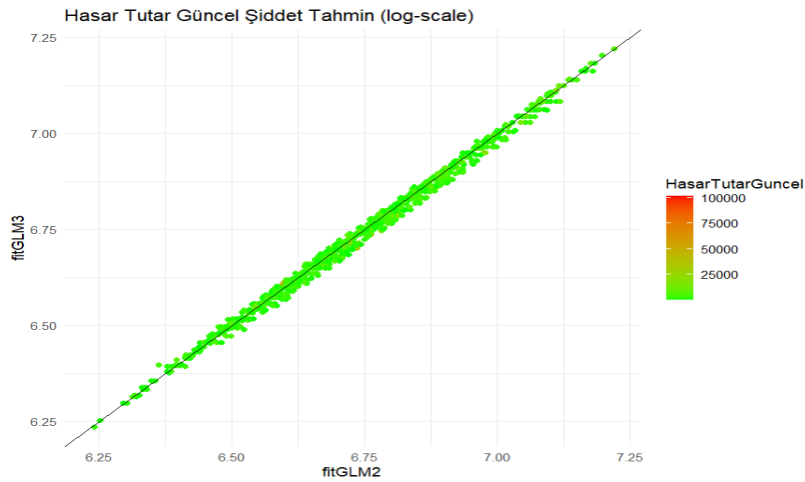
GLM1 ve GLM2 modellerinin tahminleri logaritmik ölçekte karşılaştırılmıştır. Grafik, hasar tutarı tahminlerinin her iki model arasındaki uyumunu ve örtüşme derecesini görselleştirmektedir. Renk gradyanı, her bir noktanın temsil ettiği hasar tutarının miktarını ifade ederken, kırmızıya doğru renkler yüksek hasar tutarlarını, yeşil renkler ise düşük hasar tutarlarını temsil etmektedir. Görseldeki noktaların büyük çoğunluğunun 1:1 çizgisinin etrafında sıkı bir şekilde gruplandığı görülmekte, bu da model tahminlerinin birbirleriyle yüksek oranda uyumlu olduğunu göstermektedir.

Bu uyum, sigorta şirketlerinin hasar şiddetini daha doğru tahmin etmelerine ve risk yönetim stratejilerini geliştirmelerine olanak tanıyan önemli bir göstergedir. Şekil 5.32. GLM1 ve GLM3 modelleri arasında hasar tutarı güncel şiddet tahminlerinin logaritmik ölçekte karşılaştırmalı analizini göstermektedir.



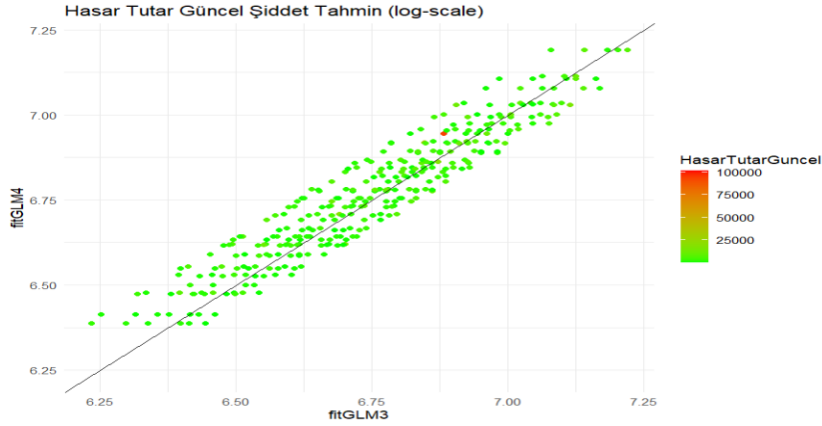
Şekil 5.32. GLM1 ve GLM3 modelleri arasında hasar tutarı güncel şiddet tahminlerinin logaritmik ölçekte karşılaştırmalı analizi

GLM1 ve GLM3 modellerinin hasar şiddeti tahminleri logaritmik ölçekte karşılaştırılmış ve her iki modelin tutarlılık seviyesi değerlendirilmiştir. Analiz sonuçları, noktaların 1:1 eğimli doğru etrafında yoğunlaşmasıyla, iki modelin tahminlerinin yüksek derecede uyum içinde olduğunu göstermektedir. Renk skalası, gözlemlenen hasar tutarlarının dağılımını yansıtmaktadır. Şekil 5.33'de GLM3 modellerinin hasar tutar güncel şiddet tahmini karşılaştırılmasını göstermektedir.



Şekil 5.33. GLM3 modellerinin hasar tutar güncel şiddet tahminlerinin logaritmik ölçekte karşılaştırmalı değerlendirmesi

GLM2 ve GLM3 modellerinin logaritmik ölçekte tahminleri, noktaların 1:1 doğrusu etrafında ne kadar yakın gruplandığına bakarak değerlendirilmiştir. Noktaların yoğunlaşması, modellerin tahminlerinin büyük ölçüde uyum içinde olduğunu ve benzer sonuçlar verdiğini göstermektedir. Renk gradyanı, hasar tutarlarının model tahminlerine nasıl yansıdığını ortaya koymakta olup, daha yüksek hasar tutarları kırmızıya, daha düşük tutarlar ise yeşile doğru renklendirilmiştir. Şekil 5.34’de GLM3 ve GLM4 modellerinin hasar tutar güncel şiddet tahminlerinin karşılaştırılmasını göstermektedir.



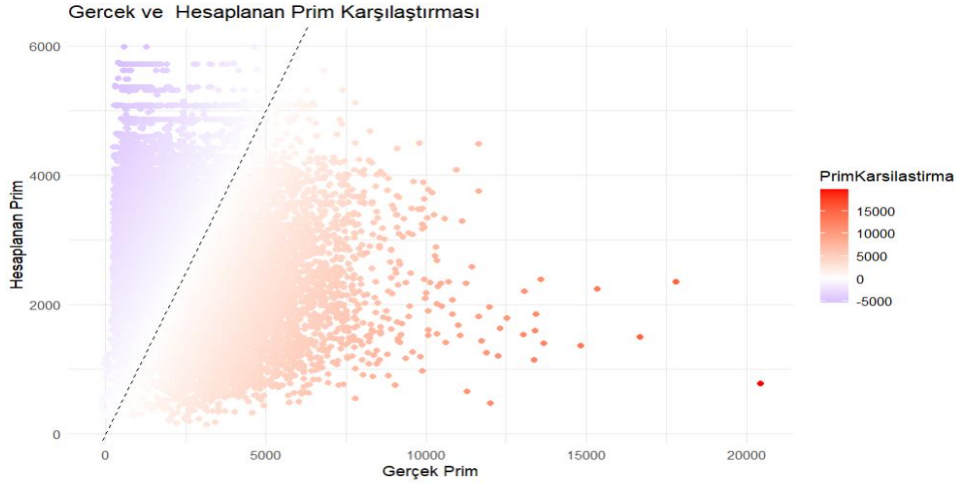
Şekil 5.34. GLM3 ve GLM4 modellerinin hasar tutar güncel şiddet tahminlerinin logaritmik ölçekte karşılaştırmalı değerlendirmesi

Hasar tutarlarının güncel değerlerinin tahmin edilmesi için GLM yöntemleri kullanılmıştır. Analiz, hasar tutarlarının büyüklüklerini ve ilişkili risk faktörlerini logaritmik ölçekte değerlendirmektedir. GLM3 ve GLM4 arasındaki ilişki, öngörü doğruluğunun ve modelin tutarlılığının göstergesi olarak kullanılmıştır.

Renk skalasındaki değişiklikler, farklı hasar tutarlarının frekansını ve dağılımını görselleştirmekte, bu da risk yönetimi ve fiyatlandırma stratejilerinin geliştirilmesi için değerli bilgiler sağlamaktadır.

Scatter plot üzerinde görülen eğimli çizgi, model tahminlerinin doğruluk derecesini yansıtmakta, bu da modelin performansını ve güvenilirliğini göstermektedir. Bu çalışmanın bulguları, hasar tahmin süreçlerinin optimizasyonuna katkıda bulunmayı amaçlamakta ve genelleştirilmiş lineer modellerin sigorta sektöründeki uygulamaları için bir çerçeve sunmaktadır.

Şekil 5.35’de GLM ile sigorta primi tahmin modelinin gerçekleştirme başarısı ve potansiyel iflas riski analizini göstermektedir.



Şekil 5.35. GLM ile sigorta primi tahmin modelinin gerçekleştirme başarısı ve potansiyel iflas riski analizi

Şekil 5.35’de prim tahminlerinin gerçekleşen ve prim değerleriyle karşılaştırılmasını göstermektedir. Gerçek primler ile tahmin edilen primler arasındaki ilişkiyi ortaya koyan bu görsel, bize modelin tahmin gücü ve prim yetersizliği problemi hakkında önemli bilgiler sunmaktadır.

Kesikli çizginin altında yoğunlaşan noktalar, genel olarak modelin tahmin ettiği primlerin gerçekte tahsil edilenden daha düşük olduğunu gösterir. Bu durum, iflas olasılığı açısından şirket için potansiyel bir riski işaret edebilir. Zira yetersiz prim tahminleri, beklenmedik hasar ödemeleri karşısında şirketin yeterli sermayeye sahip olmamasına yol açabilir. Diğer yandan, grafikte göze çarpan kırmızı renkli yüksek değerler, hesaplanan primlerin gerçek primlerden anlamlı derecede yüksek çıktığı durumları temsil eder. Bu, şirketin kar etme olasılığını artırabilir, çünkü tahsil edilen prim miktarı beklenen hasar ödemelerini aşar.

GLM kullanarak yapılan frekans tahminlerinde Poisson dağılımı ve şiddet tahminlerinde Gamma dağılımı tercih edilmiştir. Bu istatistiksel modeller, hasar oluşum sıklığını ve şiddetini modellemek için kullanılır. GLM-Frekans modeli, sigorta poliçesi başına beklenen hasar sayısını tahmin ederken, GLM-Şiddet modeli, bir hasarın maliyetini tahmin etmekte kullanılır. Hesaplanan yeni prim değerleri, bu iki modelin çıktılarının birleşimiyle elde edilmiştir ve şirketin uzun vadeli finansal sağlığı için kritik önem taşımaktadır.

Yüksek çıkan hesaplanan primler, şirketin rezervlerini güçlendirebilir ve büyük hasar olaylarına karşı daha dirençli hale getirebilir. Bu durum, şirketin iflas olasılığını azaltırken aynı zamanda yatırımcı güvenini ve piyasa itibarını artırabilir. Ancak, primlerin aşırı yüksek belirlenmesi müşteri sadakati ve pazar payı üzerinde olumsuz etkiler yaratabilir.

Özetle, grafikte sunulan veriler, şirketin prim stratejilerini ve risk yönetimini gözden geçirmesi için bir fırsat sunmaktadır.

Yapılan tahminlerin gerçek verilerle uyumlu olup olmadığına dair detaylı analizler, şirketin pazardaki konumunu güçlendirme ve finansal sürdürülebilirlik hedeflerine ulaşma yolunda önemli birer yol gösterici olmaktadır.

Bu bilgiler ışığında, şirketin prim politikalarını ve risk değerlendirme yöntemlerini optimize etmesi, sektördeki rekabetçi avantajını korumasına ve müşteri memnuniyetini artırmasına katkı sağlayacaktır.

5.3.Birleşik Aktüeryal Sinir Ağları

Bu bölümde, tamamlayıcı sigorta poliçesi veri setinin sinir ağları kullanılarak işlenmesi için gerekli ön işleme yöntemleri ele alınmaktadır. Ön işleme sürecinde, veri setindeki özellik bileşenlerinin ölçeği, modelin uyum sürecine önemli etkilerde bulunabilir.

Bu nedenle, verilerin doğru şekilde işlenmesi ve ölçeklendirilmesi büyük önem taşır. Veri setinde bulunan sıralanmamış kategorik ve sürekli özellikler ayrı ayrı ele alınmalıdır. Sıralı kategorik özellikler tam sayılarla, ikili kategorik özellikler ise 0 ve 1 ile kodlanmalıdır. Bu kodlama, modelin daha verimli öğrenmesini sağlar ve verilerin sinir ağı tarafından daha iyi anlaşılmasına yardımcı olur. Ayrıca, sinir ağlarında kullanılan aktivasyon fonksiyonunun seçimi de önemlidir. Eğer simetrik olmayan bir aktivasyon fonksiyonu tercih edilirse, ikili kategorik özellikler $\pm 1/2$ olarak ayarlanabilir. Bu ayarlama, en uygun şekilde sokma algoritmalarının daha basit ve etkili çalışmasını sağlayabilir. Bu tür bir ayar, modelin daha hızlı ve doğru bir şekilde eğitilmesine olanak tanır ve sonuçta daha doğru tahminler elde edilmesine katkıda bulunmaktadır.

Sinir ağlarına yönelik veri ön işleme, verilerin doğru bir şekilde kodlanması ve ölçeklendirilmesi üzerine yoğunlaşır.

Bu süreç, modelin veri setini daha iyi anlamasını ve veri setindeki karmaşık ilişkileri öğrenmesini sağlar. Bu yaklaşım, özellikle tamamlayıcı sigorta poliçesi veri seti gibi karmaşık ve çeşitli özelliklere sahip veri setlerinde oldukça önemlidir.

5.3.1. Sıralanmamış (Nominal) Kategorik Özellik Bileşenleri

Tamamlayıcı sağlık sigortası poliçeleri üzerinde gerçekleştirilen analizlerde kritik bir öneme sahip olan sıralanmamış kategorik özelliklerin işlenmesi süreçleri ele alınmaktadır. Genelleştirilmiş Doğrusal Modeller (GLM) kullanılarak frekans ve şiddet modelleri geliştirilirken katman, "MedeniDurumGLM", "YenilemeGLM", "CinsiyetGLM" ve "MeslekGLM" gibi gömülü kategorik değişkenlerdir. "YasGrupGLM" ve "ilGLM" gömüsüz kategorik değişkenleridir. Çünkü modelleme sürecinde 'sıcak kodlama (one-hot coding) veya 'sahte değişken' (dummy variable) dönüşümü yoluyla işlenmektedir. Gömülü kategorik değişkenler ise kategorik halini ile işlenmektedir.

Frekans modelinde "YasGrupGLM" ve "ilGLM" gömüsüz değişkenler referans kategori olarak seçilmişken, şiddet modelinde "yasgrup3GLM" ve "ilGrupGLM" gömüsüz değişkenler referans kategori olarak belirlenmiştir. Bu seçimler, söz konusu değişkenlerin modellerin tahmin performansları üzerindeki etkilerinin incelenmesi ve referans düzey olarak kabul edilen grupların diğer gruplarla karşılaştırılmasını sağlamak amacıyla yapılmıştır.

Veri setindeki saymaca kategorik değişkenlerin sayısallaştırılmasını ve ikili değişkenlere (dummy variables) dönüştürülmesini sağlamıştır. Bu süreç, "YenilemeGLM", "MedeniDurumGLM", "MeslekGLM" gibi gömülü kategorik değişkenler için belirli seviyeleri tam sayılara dönüştürmeyi, "YasGrupGLM" ve "ilGLM" gibi gömüsüz kategorik değişkenler için ise her bir kategorik seviyeyi ayrı bir ikili değişkene dönüştürmeyi içerir.

Bu dönüşüm, sinir ağlarının kategorik verileri daha etkin bir şekilde işlemesine ve modelin daha doğru tahminler yapmasına olanak tanır.

Modelleme amaçları doğrultusunda, frekans ve şiddet modelleri için en uygun prim hesaplaması gerçekleştirilmektedir. Bu hesaplamalar, Poisson ve Gamma dağılımları kullanılarak yürütülmekte ve sonuç olarak elde edilen en uygun prim değerleri, poliçe fiyatlandırma stratejilerinin belirlenmesinde önemli bir rol oynamaktadır.

Kategorik özelliklerin modelleme üzerindeki etkisi, bu prim değerlerinin doğruluğunu ve sigorta şirketinin risk yönetimi uygulamalarının etkinliğini doğrudan etkilemektedir. Ön işleme süreçlerinin doğru bir şekilde gerçekleştirilmesi, modelin genel başarısını artırmakta ve sigorta sektöründe karşılaşılan riskleri daha iyi anlamayı sağlamaktadır.

5.3.2. Sürekli Özellik Bileşenleri

Tamamlayıcı sigorta poliçesi veri setinde kullanılan sürekli özellik bileşenleri, modelleme süreçlerinde önemli bir rol oynar. Bu bileşenler, teorik olarak gelişmiş bir ağ yapısının seçilmesi halinde genellikle ön işleme gerektirmez. Ancak, pratikte bu özelliklerin benzer ölçeklere sahip olması ve eşit bir dağılıma sahip olmaları gerekmektedir. Bu gereklilik, ölçümleme algoritmalarının çoğunlukla gradyan iniş yöntemlerini kullanmasından kaynaklanır. Bu yöntemlerin doğru çalışabilmesi için, tüm bileşenlerin aynı ölçekte olması önemlidir.

Sürekli özelliklerin ölçeklendirilmesinde popüler bir yöntem olan MinMaxScaler, özellikleri $[-1, 1]$ aralığına getirir. Bu ölçeklendirme, modelin daha etkili öğrenmesine yardımcı olur. Ancak, veri setindeki aykırı değerlerin varlığı, MinMaxScaler'ın yoğunlaştırılmış özellik bileşenleri üretmesine yol açabilir. Bu durumda, özelliklerin daha eşit ve sağlam bir şekilde dönüştürülmesi için log-dönüşümü veya kantil (quantile) dönüşümü gibi alternatif yöntemler tercih edilebilir. Bu tür dönüşümler, modelin aykırı değerlerden daha az etkilenmesini sağlar ve daha kararlı bir öğrenme sürecine katkıda bulunmaktadır.

Bu sürekli özellik bileşenlerinin doğru şekilde işlenmesi, modelin genel performansını ve özellikle GLM-Frekans ve GLM-Şiddet modelleri kullanılarak en uygun PRIM değerinin hesaplanmasında kritik bir öneme sahiptir. Doğru ölçeklendirme ve dönüşüm tekniklerinin kullanılması, modelin veri setinin özelliklerini daha iyi anlamasını ve bu anlayışı sigorta poliçelerinin primlerini tahmin etmek için kullanmasını sağlamaktadır.

5.3.3. İkili Özellik Bileşenleri

İkili özellikler, modelleme süreçlerinde sıklıkla karşımıza çıkan ve özel dikkat gerektiren özellik türlerinden biridir. Örnek olarak cinsiyet, sigorta poliçelerinde sıklıkla kullanılan önemli bir ikili özelliktir.

Makine öğrenimi literatüründe, ikili özelliklerin daha yüksek boyutlu bir uzaya gömülmesi yaygın bir uygulama olmasına rağmen, bu yaklaşım her zaman mantıklı olmayabilir. Yüksek boyutlu bir uzaya gömme işlemi, modelin karmaşıklığını artırabilir ve özellikle küçük veri setlerinde aşırı uyuma (overfitting) yol açabilir. Bunun yerine, ikili kategorik özellik bileşenlerini $\pm 1/2$ olarak ayarlamak daha etkili bir yaklaşım olabilir. Bu yöntem, özelliklerin daha basit bir şekilde model içerisinde temsil edilmesini sağlar ve modelin eğitim sürecini hızlandırabilir.

Bu yaklaşım, özellikle sinir ağları gibi karmaşık modellerde önem kazanır. Aktivasyon fonksiyonlarının simetrik olmayan doğası göz önünde bulundurulduğunda, ikili özelliklerin $\pm 1/2$ olarak ayarlanması, en uygun şekilde sokma sürecini basitleştirebilir ve modelin daha hızlı ve etkili bir şekilde öğrenmesine yardımcı olabilir. Ayrıca, bu yöntem, modelin özellikleri ayrı ayrı değerlendirmesine ve daha doğru tahminler yapmasına olanak tanımaktadır.

5.3.4. Veri Ön İşleme ve Özellik Mühendisliği

Tezin bu bölümünde, tamamlayıcı sigorta poliçesi veri seti için uygulanan veri ön işleme ve özellik mühendisliği teknikleri özetlenmektedir. Bu teknikler, modelin eğitim sürecini iyileştirmek ve tahmin başarısını artırmak için hayati önem taşımaktadır.

Sürekli Özelliklerin Ölçeklenmesi: Sürekli özellikler, modelin daha iyi genelleme yapabilmesi için $[-1, 1]$ aralığına ölçeklenir. Bu ölçeklendirme, özelliklerin eşit derecede öneme sahip olmasını sağlar ve gradyan iniş gibi en uygun şekilde sokma algoritmalarının daha verimli çalışmasına yardımcı olur.

İkili Özelliklerin Ayarlanması: İkili özellikler $\pm 1/2$ değerine ayarlanır. Bu, sinir ağları için özellikle faydalıdır çünkü modelin bu tür özellikleri daha kolay öğrenmesini sağlar ve aktivasyon fonksiyonlarının simetrik olmayan yapısına uyum sağlar.

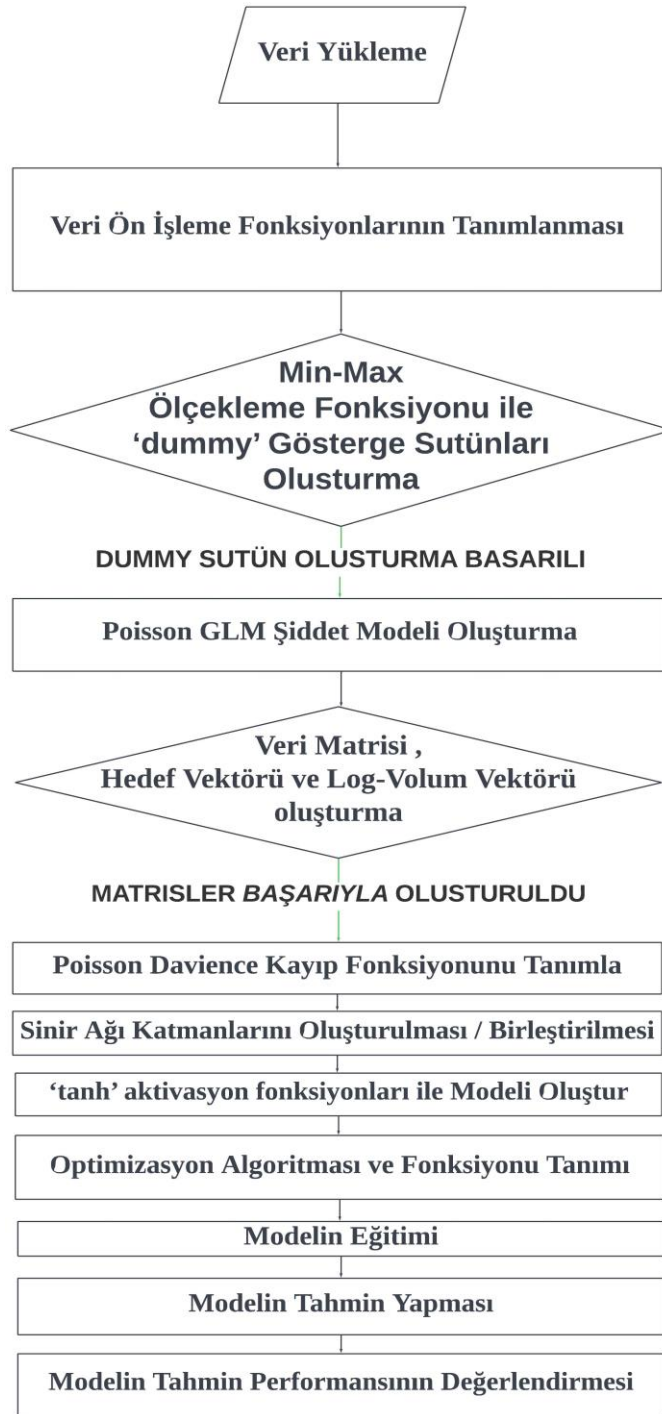
Kategorik Özelliklerin İşlenmesi: Kategorik özellikler, sayısal hale getirilir ve gerekli durumlarda ölçeklenir.

One-hot kodlama ve dummy kodlama teknikleri kullanıldığında özellikler doğrudan modelde kullanılabilir hale getirilir ve ayrıca ölçeklendirme yapılmaz. Gömme (embedding) işlemi uygulandığında ise, sayısallaştırma yapılmadan doğrudan gömme işlemi gerçekleştirilir. Sinir ağları ve diğer makine öğrenimi modelleri için veri ön işleme ve özellik mühendisliğinin önemini vurgular. Doğru uygulanan bu süreçler, modelin veri setindeki karmaşık ilişkileri daha iyi anlamasını ve bu anlayışı tahminlerde kullanmasını sağlar.

Ayrıca, bu süreçler, modelin eğitim süresini kısaltır ve genelleme yeteneğini artırır, böylece tahminlerin doğruluğunu ve güvenilirliğini iyileştirir.

Tamamlayıcı sigorta poliçesi veri seti üzerinde bu tekniklerin uygulanması, en uygun *prim* değerinin hesaplanmasında modelin başarısını önemli ölçüde artıracaktır.

Şekil 5.36'de CANN hasar frekans modelinin akış diyagramını göstermektedir.



Şekil 5.36. CANN hasar frekans modelinin akış diyagramı

5.3.5. CANN Hasar Frekans Modellenmesi

Modelimiz, veri setinin %80'ini içeren bir eğitim seti ve %20'sini içeren bir test seti kullanılarak geliştirilmiştir. Modelde, sürekli değişkenler için 'Min-Max' ölçekleme fonksiyonu kullanılmıştır. Ayrıca, kategorik değişkenler için dummy sütunlar

oluşturulmuştur. Bu yaklaşım, modelin hem sürekli hem de kategorik değişkenleri daha etkili bir şekilde işlemesini sağlar.

CANN-Frekans modelimiz, farklı yeteneklere sahip dört gizli katmandan oluşmaktadır. Bu katmanlarda sırasıyla 20, 15, 10 ve 5 nöron kullanılmıştır. Modelde 'tanh' aktivasyon fonksiyonları kullanılmıştır. 'tanh', non-linear ilişkileri yakalama konusunda etkili olup, modelin karmaşık veri yapılarını daha iyi anlamasını sağlamaktadır.

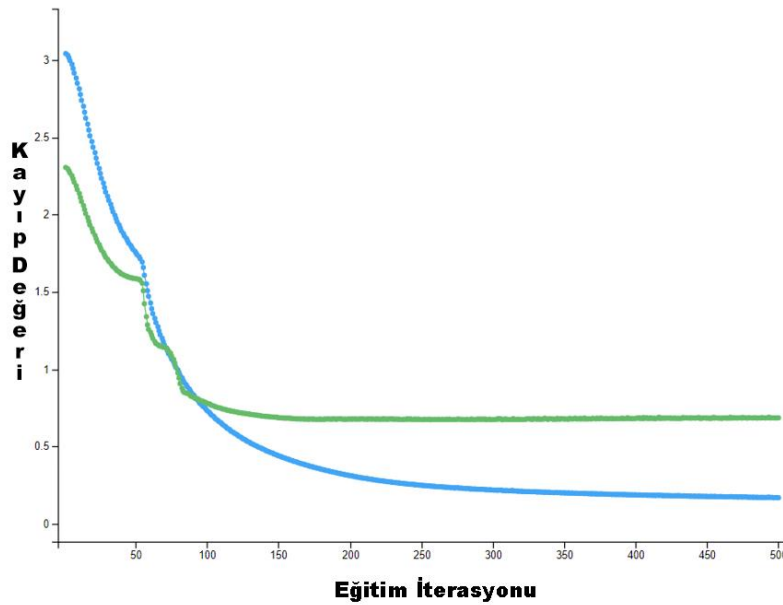
Model, veriler üzerindeki öğrenmeyi optimize etmek için 500 iterasyon (epoch) boyunca eğitilmiştir. Her bir iterasyonda 10,000 örneklemlik bir grup (batch) kullanılmıştır. Bu süreç, modelin veri setinin genel yapısını daha iyi öğrenmesine imkân tanımıştır. Eğitim sürecinde, modelin genelleştirme yeteneğini ve aşırı uyuma karşı direncini test etmek için %20 oranında bir doğrulama seti kullanılmıştır.

Modelde, Poisson dağılımı temelli bir GLM kullanılmıştır. Modelin derlenmesinde 'nadam' optimizasyon algoritması kullanılmıştır. Ayrıca, modelin hedef değişken üzerindeki hataları minimize etmek için 'poisson' kayıp fonksiyonu tercih edilmiştir.

Model performansının değerlendirilmesinde Poisson Deviance metriği kullanılmıştır.

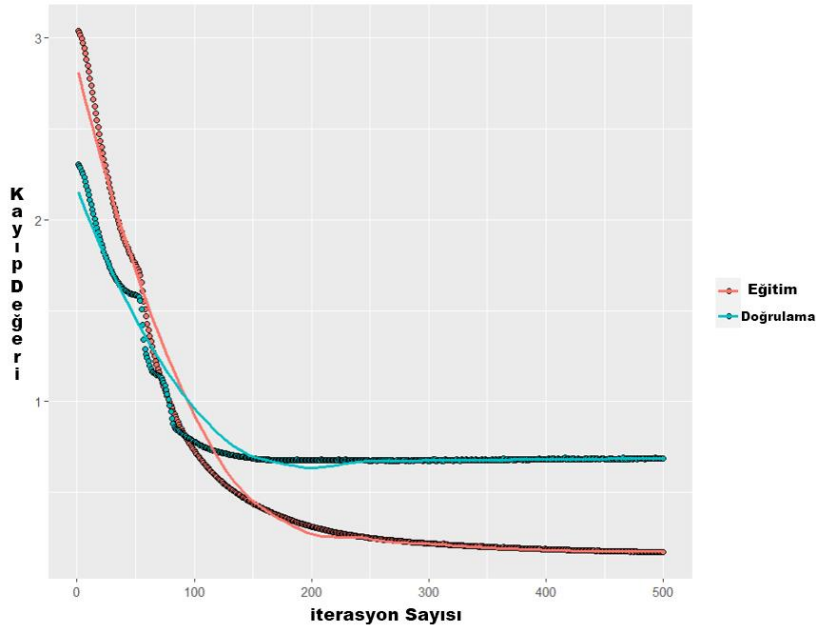
5.3.6. CANN Eğitim Sürecinde Hasar Frekans Değerlerinin Değişimi

Şekil 5.37'de CANN eğitim sürecinde GLM'deki tüm değişkenleri hasar frekans değerlerinin değişimini göstermektedir.



Şekil 5.37. CANN eğitim sürecinde glm tüm değişkenler hasar frekans değerlerinin değişimi-1

Şekil 5.38. CANN eğitim sürecinde GLM'deki tüm değişkenlerin hasar frekans değerlerinin değişimini göstermektedir.



Şekil 5.38. CANN eğitim sürecinde GLM'deki tüm değişkenlerin hasar frekans değerlerinin değişimi-2

Şekil 5.37 ve Şekil 5.38'deki CANN modelinin eğitim süreci boyunca, hasar frekansı tahminlerinde gözlemlenen kayıp değerlerinin azalma eğilimi, modelin öğrenme yeteneğini ve veriye genelleştirme kapasitesini yansıtmaktadır.

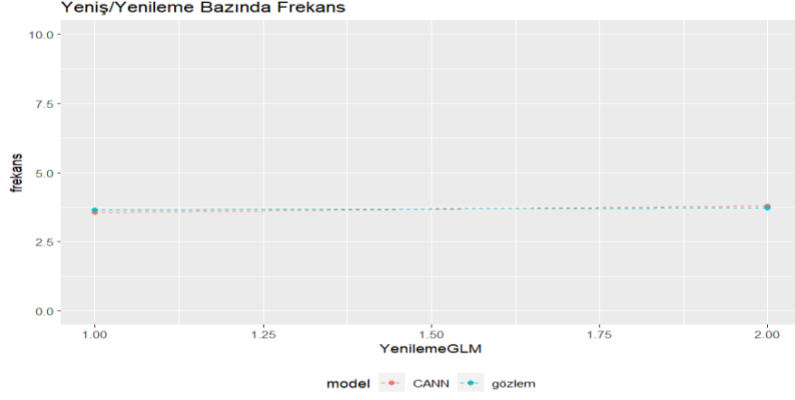
Eğitim ve doğrulama setlerindeki kayıp eğrileri, modelin epoch ilerledikçe hem hızlı hem de tutarlı bir şekilde geliştiğini ve belli bir düzeyden sonra istikrar kazandığını göstermektedir. Bu, modelin veriyi anlamada ve olası aşırı uyum (overfitting) sorunlarını önlemede başarılı olduğunu işaret etmektedir.

Kayıp eğrilerinin düşük ve sabit seyri, CANN modelinin TSS verilerindeki karmaşık yapıları ve ilişkileri kavrayabildiğini ve güvenilir tahminler üretebildiğini gösterirken, doğrulama setindeki kayıp değerlerinin eğitim setine yakın seyretmesi, modelin görülmemiş veriler üzerinde de benzer performans sergileyebileceğini düşündürmektedir.

Sigorta sektöründe iflas riskini azaltmayı amaçlayan bu çalışmada, CANN modelinin gelişmiş öğrenme algoritmaları kullanarak, GLM'ye kıyasla daha doğru prim tahminleri yapabileceği ve böylece şirketlerin risk yönetim stratejilerini daha etkili bir şekilde şekillendirebileceği sonucuna varılmıştır.

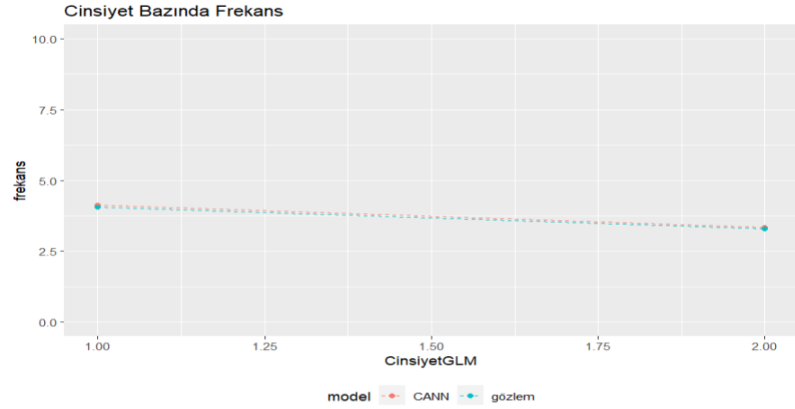
5.3.7. CANN Modelinin Çeşitli Demografik Özelliklere Göre Frekans Tahmin Performansı

Şekil 5.39’da CANN modelinin yen iş/yenileme bazında frekans tahmin performansını göstermektedir.



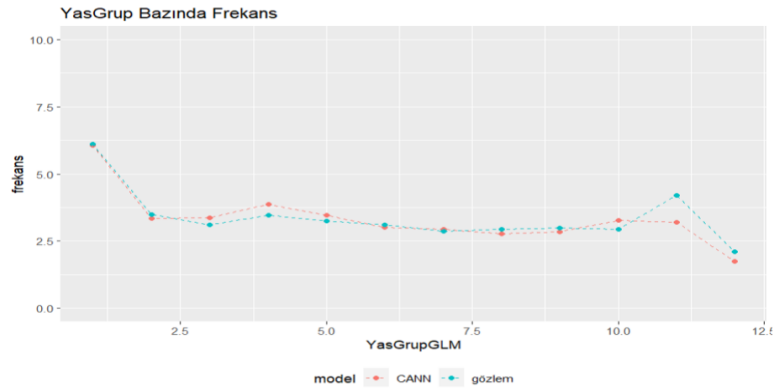
Şekil 5.39. CANN modelinin yen iş/yenileme bazında frekans tahmin performansı

Şekil 5.40. CANN modelinin cinsiyet bazında frekans tahmin performansını göstermektedir.



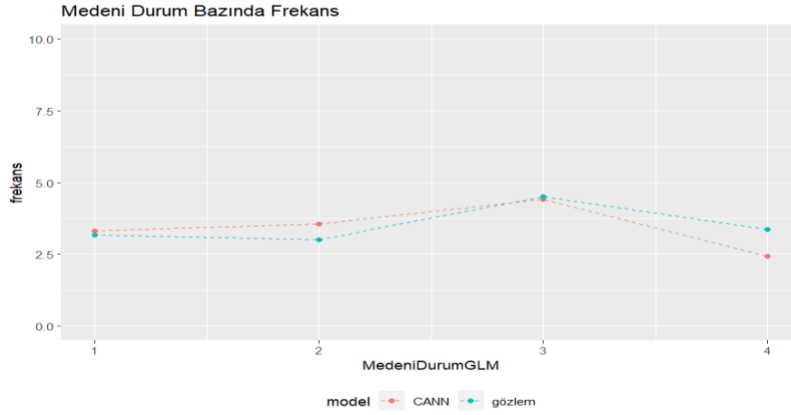
Şekil 5.40. CANN modelinin cinsiyet bazında frekans tahmin performansı

Şekil 5.41. CANN modelinin yaşgrup bazında frekans tahmin performansını gösterilmektedir.



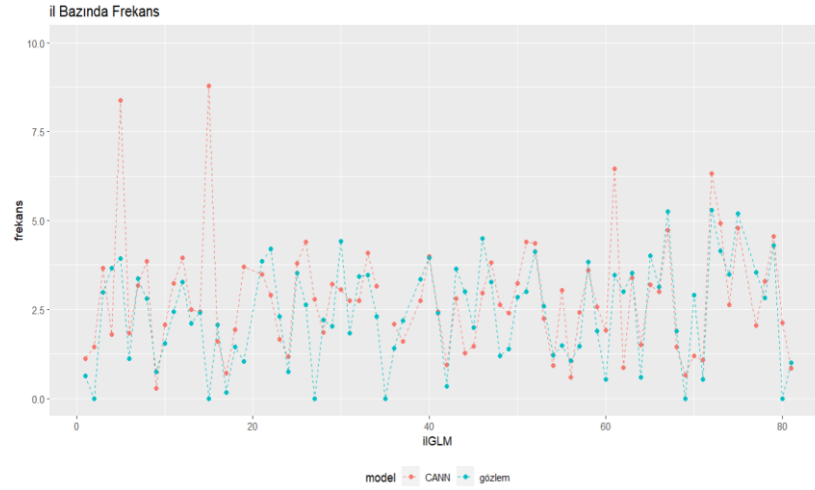
Şekil 5.41. CANN modelinin yaşgrup bazında frekans tahmin performansı

Şekil 5.42’de CANN modelinin medeni durum bazında frekans tahmin performansı gösterilmektedir.



Şekil 5.42. CANN modelinin medeni durum bazında frekans tahmin performansı

Şekil 5.43’te CANN modelinin çeşitli demografik özelliklere göre frekans tahmin performansı gösterilmektedir.



Şekil 5.43. CANN modelinin çeşitli demografik özelliklere göre frekans tahmin performansı

Şekil 5.43’teki CANN modelinin çeşitli demografik özelliklere göre frekans tahmin performansı kapsamlı bir şekilde incelenmiştir. Görsel analizler, Şekil 5.39’da Yenileme/Yeni İş, Şekil 5.40’ta Cinsiyet; Şekil 5.41’de Yaş ve Grubu; Şekil 5.42’de Medeni Durum gibi değişkenlerin model tarafından nasıl işlendiğini ve tahminlerin bu demografik özelliklere göre nasıl değiştiğini gözler önüne sermiştir. Yenileme/Yeni İş ve Cinsiyet bazında, CANN modelinin tahminleri neredeyse sabit bir seviyede kalmış ve bu değişkenlerin model tahminleri üzerinde önemli bir etkisi olmadığını göstermiştir. Medeni Durum ve Yaş Grubu gibi değişkenler için ise, CANN modeli ile elde edilen tahminlerin,

geleneksel GLM modeline kıyasla daha fazla varyansa sahip olduğu ve bu değişkenlerin tahmin performansı üzerinde daha belirgin bir etkisi olduğu belirlenmiştir. Şekil 5.44'te CANN şiddet modelinin akış diyagramı gösterilmiştir.



Şekil 5.44. CANN şiddet modelinin akış diyagramı

Model oluřturma sreci, gerekli paketlerin yklenmesiyle bařlar ve bu paketlerin yklenmesi bařarıyla tamamlandıktan sonra, model iin global parametreler tanımlanır. Veri n iřleme ařamasında, verilerin sayısallařtırılması ve veri matrisinin oluřturulması gerekmektedir. Ayrıca, eksik veya kategorik veriler iin dummy deęiřkenler oluřturulur ve bu sre bařarıyla tamamlandıęında, modelin girdileri tamamen sayısal hale getirilmiř olur. Srecin devamında, Gamma Deviance kayıp fonksiyonu ve 'adam' optimizasyon algoritması kullanılarak modelin kayıp fonksiyonu tanımlanır. Model, ReLU aktivasyon fonksiyonu kullanılarak oluřturulur ve eęitilir. Model eęitimi tamamlandıktan sonra, modelin tahmin performansı deęerlendirilir ve sonular Excel'e yazdırılarak sre tamamlanır.

5.3.7. CANN Hasar řiddetinin Modellenmesi

Veri setinin %80'lik bir kısmını ieren bir eęitim seti ve %20'sini kapsayan bir test seti kullanılarak geliřtirilmiřtir. Ayrıca modelimiz drt gizli katmandan oluřmaktadır. Bu katmanlarda sırasıyla 40, 35, 30 ve 25 nron kullanılmıřtır.

Modelimizde kullanılan lekleme fonksiyonu, srekli deęiřkenleri belirli bir aralıęa leklemek iin kullanılmaktadır. Bu iřlem, modelin veri setindeki farklı leklere sahip deęiřkenleri daha etkin bir řekilde iřlemesini saęlar. zellikle, modelin eęitim srecini hızlandırır ve farklı zellikler arasında daha iyi bir uyum saęlamıřtır.

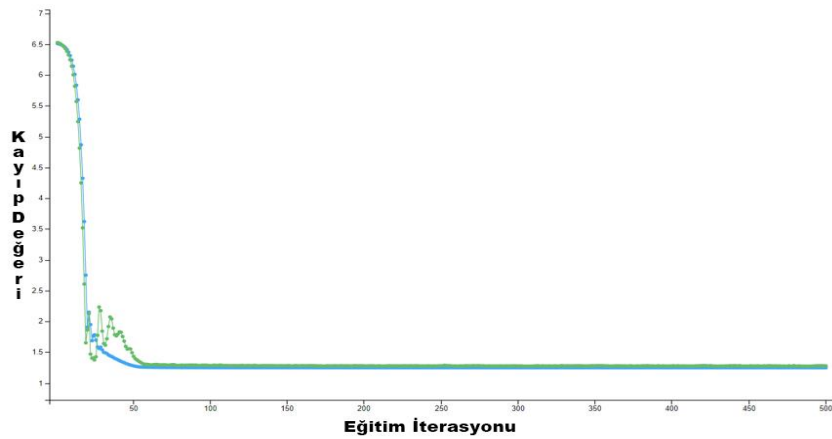
Modelde ayrıca, kategorik deęiřkenler iin dummy (gsterge) stunlar oluřturulmuřtur. Bu stunlar, kategorik deęiřkenlerin sayısal temsillerini saęlar ve modelin bu tr deęiřkenleri iřleyebilmesine olanak tanır. rneęin, bir kategorik deęiřkenin farklı kategorileri, ayrı stunlarda 0 ve 1 deęerleri ile ifade edilir. Bu yaklařım, modelin kategorik deęiřkenlerdeki desenleri ve iliřkileri ęrenmesini kolaylařtırır ve modelin genel doęruluęunu artırmaktadır. Kategorik deęiřkenlerin bu řekilde iřlenmesi, modelin daha karmařık ve eřitli veri yapılarını anlamasını saęlar. Bu, zellikle saęlık sigortası verileri gibi eřitli ve kategorik zelliklere sahip veri setlerinde nemlidir. Dummy stunlar, modelin her bir kategoriye zg desenleri tanımasına ve bu bilgileri tahminlerde kullanmasına imkn tanıtılmıřtır. Modelimizde, doęrusal olmayan iliřkileri modellemek iin aktivasyon fonksiyonu tercih edilmiřtir. TSS verilerinde bu fonksiyonların kullanımı, veri setindeki karmařık ve non-lineer iliřkileri daha iyi yakalamak iin nemli olmuřtur.

Model, verilerin zerindeki karmařıklıęı ve desenleri yeterince ęrenebilmesi iin 500 iterasyon (epoch) boyunca eęitilmiřtir. Her bir iterasyonda, veri setinden rastgele seilen

10,000 örneklilik bir grup (batch) üzerinde çalışılmıştır. Bu yaklaşım, modelin veri setinin genel yapısını daha iyi öğrenmesine olanak tanıtılmıştır. Eğitim sürecinde, modelin genelleme yeteneğini değerlendirmek ve aşırı uyuma (overfitting) karşı korumak için %20 oranında bir doğrulama seti ayrılmıştır. Modelde, hasar şiddet tahminleri için Gamma dağılımı temelli bir GLM kullanılmıştır. Bu modelin seçilmesinin sebebi, Gamma dağılımının sürekli ve pozitif verilerle uyumlu olması ve sağlık sigortası verilerindeki hasar şiddetinin doğasını iyi bir şekilde temsil etmektedir. Model derleme aşamasında optimizasyon algoritması ve gama kayıp fonksiyonu kullanılmıştır. Modelde kullanılan gamma kayıp fonksiyonu ve logaritmik kayıp fonksiyonunu içerir ve modelin tahmin hatalarını değerlendirmede kullanılır. Bu fonksiyon, modelin hedef değişken üzerindeki hatalarını azaltmayı amaçlar ve özellikle sürekli ve pozitif değerler için etkili bir optimizasyon yöntemi sunar. Gamma kayıp fonksiyonunun kullanımı, modelin hasar şiddeti gibi sürekli ve değişken verileri daha doğru tahmin etmesine yardımcı olmaktadır. 'adam', veri setindeki çeşitli özelliklerin hızlı ve etkili bir şekilde öğrenilmesini sağlarken, Gamma kayıp fonksiyonu, hasar şiddeti tahminlerinin doğruluğunu artırır. Modelin performansını değerlendirmek için gamma sapması metriği kullanılmıştır. Modelin performansını değerlendirmek için kullanılan gamma metriği, modelin tahminlerinin gerçek değerlere ne kadar yakın olduğunu ölçer. Bu metrik, özellikle sürekli ve pozitif veriler için uygun olan Gamma dağılımına dayanır ve modelin verileri ne kadar iyi temsil ettiğini sayısal olarak ifade eder. Bu metrik, modelin gerçek verilere uygunluğunu ve tahminlerinin doğruluğunu etkili bir şekilde değerlendirilmektedir.

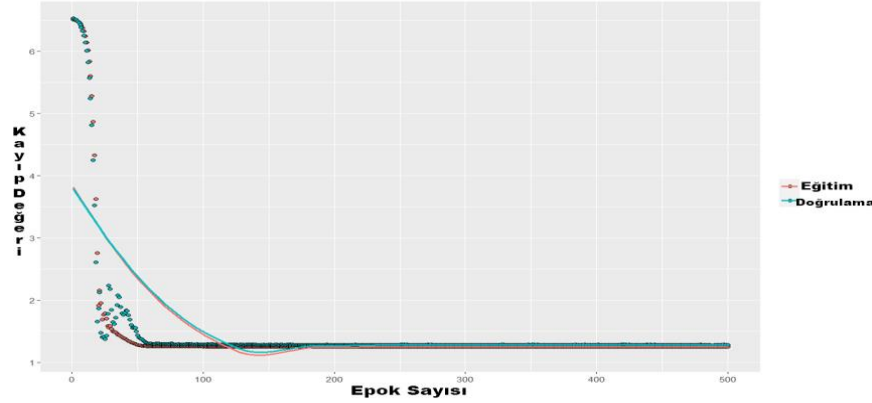
5.3.8. CANN Eğitim Sürecinde Hasar Şiddetinin Değerlerinin Değişimi

Şekil 5.45. CANN eğitim sürecinde hasar şiddet değerlerinin değişim analizini göstermektedir.



Şekil 5.45. CANN eğitim sürecinde hasar şiddet değerlerinin değişim analizi-1

Şekil 5.46. CANN eğitim sürecinde hasar şiddet değerlerinin değişim-2 analizini göstermektedir.

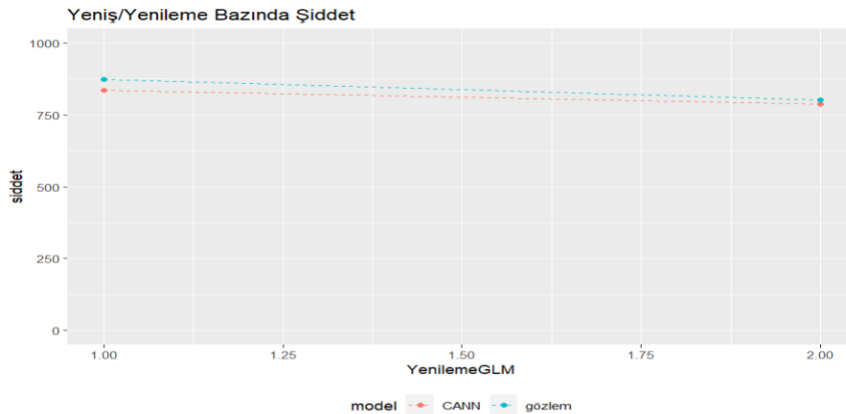


Şekil 5.46. CANN eğitim sürecinde hasar şiddet değerlerinin değişim analizi-2

Şekil 5.45. ve Şekil 5.46'daki CANN modelinin eğitim süreci, hasar şiddeti tahminlerinin doğruluğunu ve modelin genelleştirme kapasitesini anlamamız açısından kritik öneme sahiptir. Eğitim sürecinde elde edilen kayıp değerlerinin incelenmesi, modelin hem eğitim hem de doğrulama veri setleri üzerindeki öğrenme eğilimini ve kararlılığını gözler önüne serer. Görsel analizler, modelin eğitim sürecinde kayıp değerlerinin hızla azaldığını ve belirli bir epoch sayısından sonra düşük ve istikrarlı bir seviyede sabitlendiğini ortaya koymaktadır. Bu, modelin veri üzerindeki öğrenme kabiliyetini ve aşırı uydurmaya (overfitting) karşı dirençli olduğunu gösterir. Doğrulama kaybının da benzer bir eğilim göstermesi, modelin görülmemiş verilere genelleme yapabilme potansiyelini işaret etmektedir.

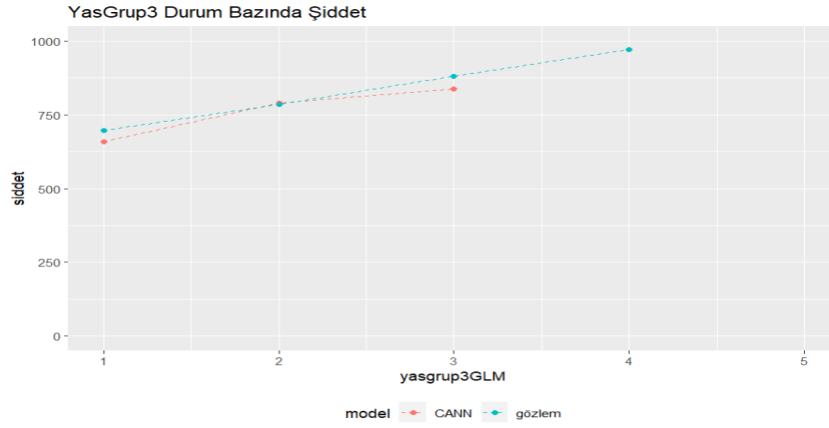
5.3.9. CANN Eğitim Sürecinde Hasar Değerlerinin Performansı

Şekil 5.47'de CANN eğitim sürecinde yen iş/yenileme bazında şiddet değerlerinin değişiminin analizi göstermektedir.



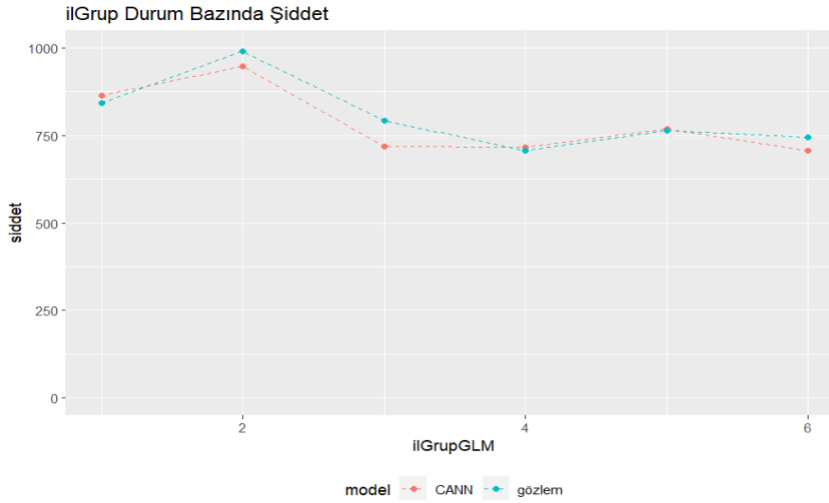
Şekil 5.47. CANN eğitim sürecinde yen iş/yenileme bazında şiddet değerlerinin değişimi analizi

Şekil 5.48. CANN eğitim sürecinde yasgrup bazında şiddet değerlerinin değişim analizini göstermektedir.



Şekil 5.48. CANN eğitim sürecinde yasgrup bazında şiddet değerlerinin değişim analizi

Şekil 5.49. CANN eğitim sürecinde ilGrup bazında şiddet değerlerinin değişim analizini göstermektedir.

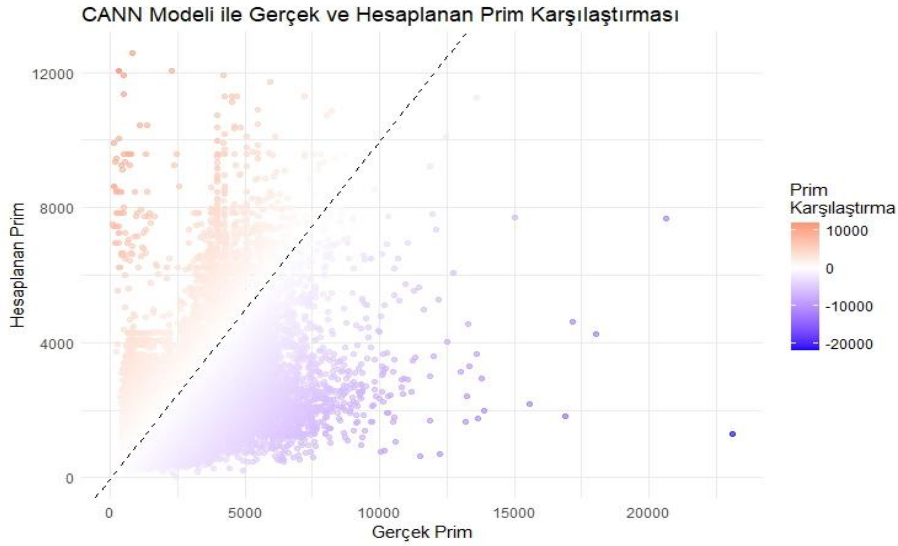


Şekil 5.49. CANN eğitim sürecinde ilGrup bazında şiddet değerlerinin değişim analizi

Şekil 5.47'deki Yenileme/Yeni İş; Şekil 5.48'deki Yaş Grubu ve Şekil 5.49'daki İl Grubu gibi değişkenlerin bazında, hasar şiddeti tahminlerini karşılaştırmak amacıyla CANN ve geleneksel GLM modellerinin performansları detaylı bir şekilde incelenmiştir. Analiz sonuçları, CANN modelinin, farklı gruplarda tutarlı ve dengeli tahminler üretme eğiliminde olduğunu ortaya koymaktadır. Özellikle, Yenileme/Yeni İş durumuna ve farklı yaş gruplarına dayalı tahminlerde CANN modeli, GLM modeline kıyasla daha sabit bir tahmin çizgisi sergileyerek, modelin bu kategorilerde daha az varyansla daha istikrarlı sonuçlar verdiğini göstermiştir. İl Grubu bazında ise, her iki modelin tahminleri arasında daha dar bir uyum olduğu gözlemlenmiştir. Bu bulgular, CANN modelinin karmaşık

aktüeryal veri setlerinde daha güvenilir ve dengeli tahminler sağlayarak sigorta şirketlerinin risk değerlendirme süreçlerinde etkin bir araç olabileceğini göstermektedir.

Şekil 5.50’de CANN ile sigorta primi tahmin modeline ilişkin prim karşılaştırmasını göstermektedir.



Şekil 5.50. CANN ile sigorta primi tahmin modeline ilişkin prim karşılaştırması

Şekilde 5.50’deki, CANN modeli, genel olarak gerçek prim değerlerine oldukça yakın tahminler üretmektedir. Noktaların büyük bir kısmının ideal çizgiye (yani hesaplanan primin gerçek prim ile eşit olduğu çizgiye) yakın bir dağılım göstermesi, modelin tutarlı ve doğru tahminler yapma eğiliminde olduğunu göstermektedir. Renk gradyanındaki dağılım, modelin bazı durumlarda gerçek primden sapma gösterse de, bu sapmaların geniş bir aralıkta olmadığını ve çoğu tahminin kabul edilebilir bir hata marjı içinde olduğunu ortaya koymaktadır. Bu sonuçlar, CANN modelinin prim tahminleri konusunda güvenilir olduğunu ve gerçek prim değerlerini iyi bir hassasiyetle yakaladığını gösterir.

5.4. CANN ve GLM Modellerinin Karşılaştırılması

Bu bölümde CANN ve GLM Modelleri sırasıyla performans ölçütleri ve iflas durumuna ilişkin etkileri açısından karşılaştırılmıştır.

5.4.1. CANN ve GLM Performans Ölçütlerinin Karşılaştırılması

Bu bölümde CANN ve GLM Modellerinin karşılaştırılmasında MAE (Ortalama Mutlak Hata), MAPE (Ortalama Yüzdesele Mutlak Hata), RMSE (Kök Ortalama Kare Hata) ve MSE (Ortalama Kare Hata) performans ölçütlerinden yararlanılmıştır. CANN ve GLM modelleri sırasıyla hem frekans hem şiddet hem de frekans ve şiddetin birlikte alındığı prim

tahmin deęerleri yönünden karşılaştırılmıştır. Her bir performans ölçütünde minimum deęeri sağlayan modelin daha tercih edilebilir olduęu varsayılmıştır. Çizelge 5.11’de CANN ve GLM performans metrikleri karşılaştırması gösterilmiştir.

Çizelge 5.11. CANN ve GLM performans metrikleri

Model	MAE	MAPE	RMSE	MSE
CANN - Prim	1261.61	85.11306	1691.483	2861114
GLM - Prim	2282.616	191.4105	2571.831	6614315
CANN-Frekans	2.855919	86.29869	3.679427	13.53818
CANN - Şiddet	468.6141	82.09753	1017.664	1035641
GLM - Frekans	2.084245	83.9086	2.544044	6.472159
GLM - Şiddet	468.166	82.73709	1024.754	1050121

CANN modeli MAE, MAPE, RMSE ve MSE performans ölçütlerine göre GLM modeline göre daha iyi sonuçlar vermiştir. Daha düşük hata metrikleri, CANN prim modeli tahmin performansının GLM prim modeline kıyasla genel olarak daha iyi olduğunu göstermektedir.

Özellikle, MAE ve RMSE deęerleri, modelin ortalama hata miktarını ve hata dağılımının tutarlılığını göstermesi açısından önemlidir; bu durumda CANN modeli daha az hata ile tahmin yapmıştır.

Genel olarak, CANN modelinin performans metrikleri, bu modelin tahminlerinin GLM modeline göre daha tutarlı ve doğru olduğunu göstermektedir.

Bu, özellikle kompleks veri yapıları ve ilişkiler mevcut olduğunda, derin öğrenme modellerinin geleneksel istatistiksel modellere kıyasla avantaj sağlayabileceğini düşündürmektedir.

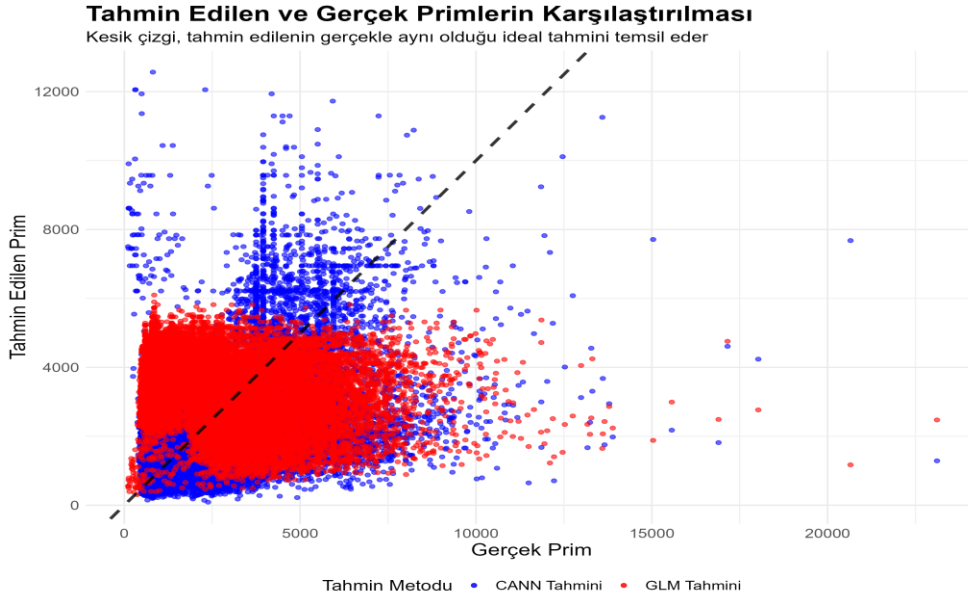
Ayrıca, MAPE metriğine göre, CANN modelinin tahminlerinin gözlemlenen verilere oranla yüzdeler bir sapma açısından GLM modeline kıyasla daha tutarlı olduğu anlaşılmaktadır.

Bu bulgular, CANN modelinin hasar tahmininde GLM modeline göre daha iyi performans gösterdiğini ve bu nedenle bu tür veri setlerinde tercih edilebileceğini öne sürmektedir.

5.4.2. İflas Durumunun (Hasar-Prim Yeterliliği) Değerlendirilmesi

Bu bölümde CANN ile GLM Modellerinden elde edilen tahmin primleri ile gerçek primlerin hasarları karşılamaadaki yeterlilikleri incelenmiştir.

Şekil 5.51'deki grafik gerçek prim değerlerinin (x ekseninde) hem GLM (Kırmızı Noktalar) hem de CANN (Mavi Noktalar) tahmin modelleri ile elde edilen prim değerlerinin (y ekseninde) karşılaştırmasını göstermektedir.



Şekil 5.51. CANN ve GLM karşılaştırma grafiği

Şekil 5.51'deki, her iki modelin tahminleri, gerçek prim değerlerine olan yakınlığına göre değerlendirilmektedir. Tahmin edilen prim değerleri gerçek prim değerleriyle tam olarak örtüşecek ve tüm noktalar grafikteki kesik çizgi ($y=x$) üzerinde yer almaktadır. Kesik çizgi, modelin mükemmel tahminlerini temsil etmektedir. GLM ve CANN modellerinin tahminleri, bu ideal çizgiye ne kadar yakın veya uzak oldukları ile değerlendirilmektedir. Gözlemlenen büyük kırmızı alan, GLM modelinin genel olarak daha düşük prim tahminleri yaptığını ve bu tahminlerin çoğunun gerçek değerlerden anlamlı derecede düşük olduğunu gösterir. Mavi noktaların dağılımı, CANN modelinin tahminlerinin gerçek prim değerlerine daha yakın olduğunu, ancak yüksek gerçek prim değerlerinde bazı sapmaların gözlemlendiğini göstermektedir. Kesikli çizgi, modellerin mükemmel tahminler yapabildiği ideal senaryoyu temsil ederken, bu çizgiden sapmalar, potansiyel olarak alt veya üst tahmin edilmiş riskleri ve dolayısıyla prim yetersizliğini göstermektedir. GLM modelinin (Kırmızı Noktalar) genel olarak daha düşük prim tahminleri yapması, özellikle düşük gerçek prim değerlerinde, sigorta şirketlerinin yeterli rezerv ayırmamasına ve dolayısıyla iflas riskinin

artmasına yol açabilecek bir eğilimi işaret etmektedir. CANN modeli (Mavi Noktalar) daha dengeli bir tahmin yelpazesi sunmakta ve özellikle yüksek primli poliçelerde GLM'ye kıyasla daha gerçekçi bir risk değerlendirmesi sunmaktadır. Ayrıca analizde, CANN ve GLM modellerine göre tahmin edilen prim tutarlarının gerçekleşen hasar tutarlarını karşılamadaki yeterlilikleri sayısal olarak incelenmiştir. Prim ve hasar tutarlarına ilişkin karşılama yeterliliğinin belirlenmesinde iflas olasılığın temelini oluşturan klasik iflas olasılığının bakış açısı kullanılmıştır. Başlangıç sermayesinin bulunmadığı durum için, her iki modele göre, her bir poliçe başına tahmin edilen prim tutarları ile hasar ödemeleri arasındaki farkın negatif olduğu durumlara ilişkin oranlar toplam portföy üzerinden belirlenmiştir. Bu oranlar yeterlilik karşılamama oranı olarak adlandırılmıştır. Tahmin edilen prim tutarlarındaki değişkenliğin gösterildiği Şekil 5.51 ile benzer olacak şekilde, CANN priminin GLM prim tutarından yüksek tahmin edildiği durumlarda, poliçelerin prim ve hasar dengesinin negatif olduğu oranın; CANN tahmin primleri kullanıldığında %13 iken; GLM tahmin primleri kullandığında %32 olduğu gözlenmiştir. GLM priminin CANN priminden daha yüksek hesaplandığı durumlarda ise; poliçelerin prim ve hasar dengesinin negatif olduğu oran GLM tahmin primleri kullanıldığında %18; CANN tahmin primleri kullanıldığında ise %42 olarak belirlenmiştir. CANN ve GLM prim tahmin değerlerinin birbirine yakın olarak tahmin edildiği durumlarda ise (en fazla 500 TL fark varsayılmıştır) poliçelerin prim ve hasar dengesinin negatif olduğu oranların her iki model için yakın çıktığı gözlenmiştir (GLM %28, CANN %30).

Analizde ayrıca, veriye ilişkin şirket tarafından hali hazırda kullanılmakta olan prim tutarlarına bağlı olarak karşılaştırma da yapılmıştır. CANN modeline ilişkin tahmin edilen prim tutarları ile şirketin gerçek prim tutarlarına ilişkin poliçelerin prim ve hasar dengesinin negatif olduğu oranlar da belirlenmiştir. CANN priminin, şirketin kullanmakta olduğu gerçek prim tutarından yüksek tahmin edildiği durumlarda (poliçelerde) şirkete ilişkin gerçek prim tutarlarının kullanılması durumunda prim ve hasar dengesinin negatif olduğu oranın %55 olarak hesaplandığı ancak CANN tahmin primlerinin kullanılması durumunda bu oranın %30'a düştüğü gözlenmiştir.

6. SONUÇ VE ÖNERİLER

Aktüerya bilimleri alanında, özellikle de sağlık sigortası alanında, sigorta ürünlerinin finansal istikrarının ve sürdürülebilirliğinin sağlanmasında fiyatlandırma büyük önem taşımaktadır. Genelleştirilmiş Doğrusal Modellerin (GLM) ve makine öğrenimi tekniklerinin uygulanması, fiyatlandırma metodolojilerinde yeni bir yaklaşım sağlayarak aktüerlerin sigorta primlerini doğru bir şekilde değerlendirmek ve tahmin etmek için geniş veri kümelerinden ve karmaşık risk faktörlerinden yararlanmasına olanak sağlamıştır. GLM çeşitli risk faktörleri ile hasar frekansları veya şiddetleri arasındaki karmaşık ilişkilerin modellenmesini sağlayarak fiyatlandırma kararları için sağlam bir çerçeve sunar. Dahası, makine öğrenimi algoritmaları, verilerdeki ince kalıpları ortaya çıkararak fiyatlandırma hassasiyetini artırır, böylece risk değerlendirmelerini iyileştirir ve sigortacıların primleri bireysel poliçe sahiplerine daha etkili bir şekilde uyarlamasını sağlar. Sonuç olarak, GLM'lerin ve makine öğreniminin aktüeryal fiyatlandırmada benimsenmesi yalnızca sigortacılar için kârlılığı artırmakla kalmaz, aynı zamanda poliçe sahipleri için adil fiyatlandırma sağlar ve nihayetinde daha esnek ve eşitlikçi bir sağlık sigortası ortamını teşvik eder.

Birleştirilmiş Aktüeryal Yapay Sinir Ağları (CANN) modelinin ortaya çıkışı, geleneksel GLM'lere göre belirgin avantajlar sunarak aktüerya biliminde önemli bir sıçramayı temsil etmektedir. Doğrusal varsayımları ve verilerdeki karmaşık doğrusal olmayan ilişkileri yakalayamamaları nedeniyle sınırlı olan GLM'lerin aksine CANN, karmaşık kalıpları ve bağımlılıkları verimli bir şekilde modellemek için sinir ağlarının gücünden yararlanmaktadır. CANN, aktüeryal ilkeleri gelişmiş sinir ağı mimarileriyle entegre ederek, sigorta fiyatlandırması ve risk değerlendirmesinde yaygın olarak karşılaşılan çeşitli ve yüksek boyutlu veri kümelerini etkili bir şekilde ele alabilir. Ayrıca CANN, ham verilerden ilgili özellikleri otomatik olarak çıkararak özellik öğrenmede üstünlük sağlar, böylece manuel özellik mühendisliği ihtiyacını azaltır ve tahmin doğruluğunu artırır. Bu uyarlanabilirlik, değişkenler arasındaki etkileşimleri yakalama yeteneği ile birleştiğinde CANN'ı aktüeryal modelleme için üstün bir seçenek olarak konumlandırır ve sigortacıların daha bilinçli kararlar almasını ve sürekli gelişen sigorta ortamında riski daha iyi yönetmesini sağlayacaktır.

Bu tezde, tamamlayıcı sağlık sigortası verileri kullanılarak GLM ve CANN modelleri arasında bir karşılaştırma sunulmuştur. Optimal prim hesaplamaları, gelecekteki risk

tahminlerinde ve şirketlerin mevcut iflas olasılığının değerlendirilmesinde kritik bir rol oynar. CANN modeli, GLM'in sunduğu geleneksel yaklaşımlara alternatif olarak karmaşık veri yapılarını ve ilişkileri işleyebilme kapasitesiyle dikkat çekmektedir. Frekans ve şiddet modelleri, sırasıyla bir poliçenin belirli bir sürede karşılaşılabileceği hasar sayısı ve bu hasarların maliyetini tahmin etmeye odaklanmıştır. Analizde, frekans modellemesinde Poisson dağılımı kullanılırken, şiddet modellemesinde Gamma dağılımı tercih edilmiştir. Her iki modelde de kayıp fonksiyonları ve en uygun şekilde sokma algoritmaları (PoissonDeviance ve GammaDeviance) kullanılarak modellerin performansı iyileştirilmiştir.

Elde edilen sonuçlar, CANN modelinin prim tahminlerinde GLM'e göre daha dengeli ve tutarlı sonuçlar verdiğini göstermektedir. Özellikle, prim farklarının dağılımı ve gerçek ile hesaplanan primlerin karşılaştırmasında, CANN modelinin, mevcut primlere göre potansiyel iflas riskini azalttığını ortaya koymaktadır. Gerçek primlere kıyasla hesaplanan primlerin daha düşük çıkması durumunda sigorta şirketleri, yetersiz prim toplama riski ile karşı karşıya kalabilirken, aşırı yüksek tahminler müşteri kaybına yol açabilir. Analiz sonuçlarına göre, GLM modelinde genel olarak daha düşük prim tahminleri yapması, özellikle düşük gerçek prim değerlerinde, sigorta şirketlerinin yeterli rezerv ayırmamasına ve dolayısıyla iflas riskinin artmasına yol açabilecek bir eğilimi işaret etmektedir. CANN modeli daha dengeli bir tahmin yelpazesi sunmakta ve özellikle yüksek primli poliçelerde GLM'ye kıyasla daha gerçekçi bir risk değerlendirmesi sunmaktadır. Her iki modelin de yüksek gerçek prim değerleri karşısında gösterdiği performans, iflas riskinin yönetilmesindeki hassasiyetin önemini vurgulamaktadır.

CANN modeli, sigorta şirketlerine yönelik risk değerlendirme ve prim politikası belirleme süreçlerinde önemli katkılar sağlamaktadır. Modelin gelişmiş öğrenme algoritmaları sayesinde, GLM'e kıyasla daha doğru prim tahminlerinin elde edilmesi ve sigorta şirketlerinin risk yönetim stratejilerinin daha etkin biçimde şekillendirilmesi ve yeterli prim ve hasar dengesinin kurulması mümkün olabilmektedir.

Bu çalışma, CANN ve GLM modelleri için hem hasar frekansı hem de hasar şiddeti açısından farklı dağılım varsayımlarına bağlı olarak incelenebileceği gibi, bu tez çalışmasında uygulaması yapılan tamamlayıcı sağlık sigortası branşı yerine farklı sigorta branşlarında denenerak de sonuçların değerlendirilmesi mümkündür.

KAYNAKÇA

- [1] Türkiye Sigorta Birliđi, TSB sektör Raporu, https://www.tsb.org.tr/content/Broadcasts/TSB_SEKTOR_raporu_final.pdf, erişim tarihi: 01.01.2023, **2022**.
- [2] J.A. Nelder, R.W.M. Wedderburn, Generalized Linear Interactive Models, Journal of the Royal Statistical Society. Series A (General), 135(3), 370-384, **1972**.
- [3] S. Haberman, A.E. Renshaw, Generalized Linear Models and Actuarial Science, The Statistician, 45(4), 407-436, **1996**.
- [4] P. De Jong, G.Z. Heller, Generalized Linear Models for Insurance Data, Cambridge University Press, London, 196, **2008**.
- [5] E.W. Frees, Regression Modeling with Actuarial and Financial Applications, Cambridge University Press, New York, 565, **2010**.
- [6] E. Ohlsson, B. Johansson, Non-Life Insurance Pricing with Generalized Linear Models. Springer, Berlin, 174, **2010**.
- [7] M. Denuit, D. Hainaut, J. Trufin, Effective Statistical Learning Methods For Actuaries I, GLMs and Extensions, Springer Publishing, Switzerland, 441, **2019**.
- [8] A. Gabrielli, R. Richman, M. V. Wüthrich., Neural Network Embedding of the Over-Dispersed Poisson Reserving Model, Scandinavian Actuarial Journal, 1-29, **2020**.
- [9] C. Dugas, Y. Bengio, N. Chapados, P. Vincent, G. Denoncourt, and C. Fournier, Statistical Learning Algorithms Applied to Automobile Insurance Ratemaking, Casualty Actuar Soc Forum 1(1), 179–214, **2003**.
- [10] C. Mano, E. Rasa. A Discussion of Modeling Techniques for Personal Lines Pricing, Trans 27th ICA, **2002**.
- [11] C. Blier-Wong, H. Cossette, L. Lamontagne, E. Marceau. Machine Learning in P&C Insurance: A Review for Pricing and Reserving, Risks, 9(1), 4, **2021**.
- [12] M.V. Wüthrich, M. Merz, Editorial: Yes, We Cann!, ASTIN Bulletin. 49(1), 1-3, **2019**.
- [13] J.Schelldorfer, M.V. Wüthrich, Nesting Classical Actuarial Models into Neural Networks, Available at SSRN 3320525, **2019**.

- [14] A. G. Laporta, A Neural Network Approach to Measure Health Insurance Risk, Ph.D. thesis. Sapienza – University of Rome, **2022**.
- [15] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, D. Pedreschi, A survey of methods for explaining black box models. ACM computing surveys (CSUR), 22;51(5):1-42, **2018**.
- [16] C. Molnar. Interpretable Machine Learning A Guide for Making Black Box Models Explainable, Creative Commons License, Available online: <https://christophm.github.io/interpretable-ml-book>, **2020**.
- [17] E. W. Frees, R. A. Derrig, G. Meyers, Predictive Modeling Applications in Actuarial Science. volume 1. Predictive modeling , Cambridge University Press, **2014**.
- [18] M. V. Wüthrich, M. Michael, Statistical Foundations of Actuarial Learning and its Applications, Springer Nature, 605, **2023**.
- [19] R. Richman, Ai in Actuarial Science–A Review Of Recent Advances–Part 1, Annals of Actuarial Science, 15(2), pp.207-229, **2021**.
- [20] Kaas, R., Goovaerts, M., Dhaene, J., Denuit, M., Modern Actuarial Risk Theory: Using R, Springer Science & Business Media, **2008**.
- [21] İstatistiksel Tablolar, Tüketici Fiyat endeksleri, Harcama gruplarına göre endeks sonuçları, <https://data.tuik.gov.tr/Bulten/Index?p=T%C3%BCketici-Fiyat-Endeksi-Mart-2023-49652&dil=1>, erişim tarihi: 15.10.2023, **2023**.
- [22] R Core Team R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>,(2022).

EKLER

Ek 1 - Genelleştirilmiş Lineer Modellerle Hasar Sıklığı Tablosu

		Tahmin	Std. Hata	Z Degeri	Pr(> z)
GLM1- Hepsi	Sabit Katsayı	1.909669	0.015054	126.854	< 2e-16
	YasGrupGLM07-20 Arası	-0.33624	0.020567	-16.349	< 2e-16
	YasGrupGLM21-25 Arası	-0.40657	0.02374	-17.126	< 2e-16
	YasGrupGLM26-30 Arası	-0.30787	0.021744	-14.159	< 2e-16
	YasGrupGLM31-35 Arası	-0.39789	0.021818	-18.236	< 2e-16
	YasGrupGLM36-40 Arası	-0.4815	0.022355	-21.539	< 2e-16
	YasGrupGLM41-45 Arası	-0.52739	0.023065	-22.866	< 2e-16
	YasGrupGLM46-50 Arası	-0.54108	0.02457	-22.022	< 2e-16
	YasGrupGLM51-55 Arası	-0.52993	0.026691	-19.854	< 2e-16
	YasGrupGLM56-60 Arası	-0.48605	0.030456	-15.959	< 2e-16
	YasGrupGLM61-65 Arası	-0.43887	0.07935	-5.531	3.19E-08
	YasGrupGLM65+	-0.93182	0.213917	-4.356	1.32E-05
	CinsiyetGLM Erkek	-0.20649	0.00633	-32.62	< 2e-16
	ilGLMADANA	-1.4857	0.082962	-17.908	< 2e-16
	ilGLMADIYAMAN	-1.16714	0.250076	-4.667	3.05E-06
	ilGLMAFYONKARAHİSAR	-0.68858	0.140211	-4.911	9.06E-07
	ilGLMAĞRI	-0.49843	0.13262	-3.758	0.000171
	ilGLMAKSARAY	-0.1156	0.082461	-1.402	0.160972
	ilGLMAMASYA	-0.71498	0.060032	-11.91	< 2e-16
	ilGLMANKARA	-0.19342	0.010323	-18.736	< 2e-16
	ilGLMANTALYA	-0.32029	0.023501	-13.629	< 2e-16
	ilGLMARDAHAN	-0.80525	0.174218	-4.622	3.80E-06
	ilGLMARTVİN	-0.87677	0.108022	-8.117	4.80E-16
	ilGLMAYDIN	-0.47649	0.060808	-7.836	4.65E-15
	ilGLMBALIKESİR	-0.31725	0.046057	-6.888	5.65E-12
	ilGLMBARTIN	-0.67217	0.162352	-4.14	3.47E-05
	ilGLMBATMAN	-0.20827	0.058418	-3.565	0.000364
	ilGLMBAYBURT	-0.21087	0.204248	-1.032	0.30187
	ilGLMBİLECİK	-0.72642	0.196217	-3.702	0.000214
	ilGLMBİNGÖL	-1.78821	0.242622	-7.37	1.70E-13
	ilGLMBİTLİS	-1.21494	0.169153	-7.182	6.85E-13
	ilGLMBOLU	-1.09057	0.164559	-6.627	3.42E-11
	ilGLMBURDUR	-1.67811	0.500093	-3.356	0.000792
	ilGLMBURSA	-0.16363	0.012145	-13.473	< 2e-16
	ilGLMÇANAKKALE	-0.12591	0.046466	-2.71	0.006734
	ilGLMÇANKIRI	-0.64423	0.213342	-3.02	0.00253
	ilGLMÇORUM	-1.33335	0.218339	-6.107	1.02E-09
	ilGLMDENİZLİ	-0.10102	0.025659	-3.937	8.26E-05
	ilGLMDİYARBAKIR	-0.324	0.036179	-8.955	< 2e-16
	ilGLMDÜZCE	-0.74541	0.185824	-4.011	6.04E-05
	ilGLMEDİRNE	-0.59361	0.129262	-4.592	4.38E-06

ilGLMELAZIĞ	-0.61835	0.074762	-8.271	< 2e-16
ilGLMERZİNCAN	-0.0629	0.080817	-0.778	0.436419
ilGLMERZURUM	-0.40636	0.056634	-7.175	7.22E-13
ilGLMESKİŞEHİR	-0.05989	0.043441	-1.379	0.168014
ilGLMGAZİANTEP	0.050571	0.016844	3.002	0.00268
ilGLMGİRESUN	-0.33663	0.036379	-9.253	< 2e-16
ilGLMGÜMÜŞHANE	-1.09549	0.316324	-3.463	0.000534
ilGLMHAKKARİ	0.049832	0.169183	0.295	0.768341
ilGLMHATAY	-0.95631	0.062034	-15.416	< 2e-16
ilGLMISPARTA	-0.28282	0.108	-2.619	0.008826
ilGLMİZMİR	-0.38593	0.023118	-16.694	< 2e-16
ilGLMKAHRAMANMARAŞ	-1.20556	0.134986	-8.931	< 2e-16
ilGLMKARABÜK	-0.13554	0.066944	-2.025	0.042908
ilGLMKARAMAN	-0.42833	0.140282	-3.053	0.002263
ilGLMKARS	-0.29012	0.218559	-1.327	0.184376
ilGLMKASTAMONU	-0.06453	0.058651	-1.1	0.27125
ilGLMKAYSERİ	-0.0347	0.022174	-1.565	0.117572
ilGLMKİLİS	-0.31987	0.140243	-2.281	0.02256
ilGLMKIRIKKALE	-0.98707	0.185863	-5.311	1.09E-07
ilGLMKIRKLARELİ	-0.41996	0.084126	-4.992	5.97E-07
ilGLMKİRŞEHİR	0.019742	0.156347	0.126	0.899516
ilGLMKOCAELİ	0.134935	0.01062	12.706	< 2e-16
ilGLMKONYA	-0.41331	0.027123	-15.238	< 2e-16
ilGLMKÜTAHYA	-1.08175	0.156318	-6.92	4.51E-12
ilGLMMALATYA	-0.54858	0.072045	-7.614	2.65E-14
ilGLMMANİSA	-1.07174	0.090373	-11.859	< 2e-16
ilGLMMARDİN	-0.58658	0.062074	-9.45	< 2e-16
ilGLMMERSİN	-0.23595	0.024275	-9.72	< 2e-16
ilGLMMUĞLA	-1.02767	0.106192	-9.677	< 2e-16
ilGLMMUŞ	-0.67267	0.316345	-2.126	0.033471
ilGLMNEVŞEHİR	-0.17794	0.098287	-1.81	0.070234
ilGLMNİĞDE	-1.68392	0.378051	-4.454	8.42E-06
ilGLMORDU	-0.20321	0.028543	-7.119	1.08E-12
ilGLMOSMANİYE	-0.71385	0.160249	-4.455	8.40E-06
ilGLMRİZE	-0.12194	0.05739	-2.125	0.033609
ilGLMSAKARYA	-0.21671	0.027531	-7.871	3.51E-15
ilGLMSAMSUN	0.250631	0.016469	15.218	< 2e-16
ilGLMŞANLIURFA	-1.0067	0.101732	-9.896	< 2e-16
ilGLMSİİRT	-1.91661	0.577388	-3.319	0.000902
ilGLMSİNOP	-0.41403	0.131551	-3.147	0.001648
ilGLMŞIRNAK	-0.9168	0.154457	-5.936	2.93E-09
ilGLMSİVAS	0.133044	0.04363	3.049	0.002293
ilGLMTEKİRDAĞ	0.08017	0.019797	4.05	5.13E-05
ilGLMTOKAT	-0.05532	0.047549	-1.163	0.244663
ilGLMTRABZON	0.288531	0.016345	17.652	< 2e-16
ilGLMTUNCELİ	-0.62985	0.301653	-2.088	0.036799

	ilGLMUŞAK	-0.45172	0.085969	-5.255	1.48E-07
	ilGLMVAN	-0.2128	0.053774	-3.957	7.58E-05
	ilGLMYALOVA	-0.10281	0.050893	-2.02	0.043364
	ilGLMYOZGAT	-0.63674	0.23581	-2.7	0.00693
	ilGLMZONGULDAK	-0.70236	0.162376	-4.325	1.52E-05
	YenilemeGLMYenileme	-0.12831	0.008267	-15.521	< 2e-16
	MedeniDurumGLMBekar	-0.1421	0.012919	-10.999	< 2e-16
	MedeniDurumGLMBoşanmış	0.050962	0.01836	2.776	0.005509
	MedeniDurumGLMDul	0.018796	0.040907	0.459	0.645883
	MeslekGLMBeyaz Yaka	-0.48421	0.01927	-25.127	< 2e-16
	MeslekGLMÇalışmayan	-0.19947	0.030627	-6.513	7.37E-11
	MeslekGLMDiğer	-0.06762	0.018101	-3.736	0.000187
	MeslekGLMEMekli	-0.04176	0.050043	-0.834	0.404059
	MeslekGLMMavi Yaka	-0.26295	0.017221	-15.269	< 2e-16
	MeslekGLMMemur	-0.10793	0.019502	-5.534	3.12E-08
	MeslekGLMÖğrenci	-0.23519	0.022955	-10.246	< 2e-16
	MeslekGLMÖğretmen	NA	NA	NA	NA
	VKIGrup2GLMBebek	NA	NA	NA	NA
	VKIGrup2GLMKilolu	-0.00639	0.01976	-0.323	0.746437
	VKIGrup2GLMObez	0.143013	0.057755	2.476	0.013279
	VKIGrup2GLMRiskli	0.10286	0.01538	6.688	2.26E-11
		Tahmin	Std. Hata	Z Degeri	Pr(> z)
	Sabit Katsayı	1.878839	0.015	125.259	< 2e-16
	YasGrupGLM07-20 Arası	-0.54575	0.011459	-47.626	< 2e-16
	YasGrupGLM21-25 Arası	-0.64382	0.017043	-37.777	< 2e-16
	YasGrupGLM26-30 Arası	-0.53798	0.014642	-36.742	< 2e-16
	YasGrupGLM31-35 Arası	-0.61526	0.015313	-40.178	< 2e-16
	YasGrupGLM36-40 Arası	-0.70183	0.015951	-44	< 2e-16
	YasGrupGLM41-45 Arası	-0.74926	0.016905	-44.321	< 2e-16
	YasGrupGLM46-50 Arası	-0.77177	0.018752	-41.156	< 2e-16
	YasGrupGLM51-55 Arası	-0.76176	0.021237	-35.87	< 2e-16
	YasGrupGLM56-60 Arası	-0.70744	0.025786	-27.434	< 2e-16
	YasGrupGLM61-65 Arası	-0.58381	0.078052	-7.48	7.45E-14
	YasGrupGLM65+	-1.03531	0.213896	-4.84	1.30E-06
	CinsiyetGLMERkek	-0.21501	0.006309	-34.079	< 2e-16
	ilGLMADANA	-1.49771	0.082954	-18.055	< 2e-16
	ilGLMADIYAMAN	-1.26136	0.250053	-5.044	4.55E-07
	ilGLMAFYONKARAHİSAR	-0.75059	0.140188	-5.354	8.60E-08
	ilGLMAĞRI	-0.49283	0.132602	-3.717	0.000202
	ilGLMAKSARAY	-0.07401	0.08244	-0.898	0.369345
	ilGLMAMASYA	-0.67426	0.060017	-11.235	< 2e-16
	ilGLMANKARA	-0.12029	0.009438	-12.746	< 2e-16
	ilGLMANTALYA	-0.29026	0.023465	-12.37	< 2e-16
	ilGLMARDAHAN	-0.71707	0.174196	-4.116	3.85E-05
	ilGLMARTVİN	-0.83683	0.107994	-7.749	9.27E-15
	ilGLMAYDIN	-0.45716	0.060768	-7.523	5.35E-14

GLM2-
Meslek
yok

ilGLMBALIKESİR	-0.28823	0.046014	-6.264	3.75E-10
ilGLMBARTIN	-0.66353	0.16234	-4.087	4.37E-05
ilGLMBATMAN	-0.18188	0.058404	-3.114	0.001844
ilGLMBAYBURT	-0.18214	0.204248	-0.892	0.372511
ilGLMBİLECİK	-0.76613	0.196208	-3.905	9.43E-05
ilGLMBİNGÖL	-1.76699	0.242616	-7.283	3.26E-13
ilGLMBİTLİS	-1.2973	0.169133	-7.67	1.72E-14
ilGLMBOLU	-1.14827	0.164537	-6.979	2.98E-12
ilGLMBURDUR	-1.81368	0.500068	-3.627	0.000287
ilGLMBURSA	-0.14	0.012061	-11.607	< 2e-16
ilGLMÇANAKKALE	-0.06005	0.046374	-1.295	0.195331
ilGLMÇANKIRI	-0.57924	0.21332	-2.715	0.00662
ilGLMÇORUM	-1.42561	0.218319	-6.53	6.58E-11
ilGLMDENİZLİ	-0.08648	0.025603	-3.378	0.000731
ilGLMDİYARBAKIR	-0.31262	0.036159	-8.645	< 2e-16
ilGLMDÜZCE	-0.75774	0.185816	-4.078	4.54E-05
ilGLMEDİRNE	-0.52974	0.129234	-4.099	4.15E-05
ilGLMELAZIĞ	-0.63997	0.074752	-8.561	< 2e-16
ilGLMERZİNCAN	-0.02459	0.0808	-0.304	0.760924
ilGLMERZURUM	-0.40978	0.056622	-7.237	4.58E-13
ilGLMESKİŞEHİR	-0.0178	0.043425	-0.41	0.681973
ilGLMGAZİANTEP	0.100896	0.016734	6.029	1.65E-09
ilGLMGİRESUN	-0.31329	0.036343	-8.62	< 2e-16
ilGLMGÜMÜŞHANE	-1.03538	0.316297	-3.273	0.001062
ilGLMHAKKARİ	0.059701	0.16917	0.353	0.724159
ilGLMHATAY	-0.95305	0.062028	-15.365	< 2e-16
ilGLMISPARTA	-0.26411	0.107993	-2.446	0.014461
ilGLMİZMİR	-0.38263	0.023111	-16.556	< 2e-16
ilGLMKAHRAMANMARAŞ	-1.26182	0.134971	-9.349	< 2e-16
ilGLMKARABÜK	-0.12655	0.066903	-1.892	0.058554
ilGLMKARAMAN	-0.43589	0.140238	-3.108	0.001882
ilGLMKARS	-0.1865	0.218403	-0.854	0.393159
ilGLMKASTAMONU	-0.01438	0.058596	-0.245	0.806169
ilGLMKAYSERİ	0.004518	0.022138	0.204	0.838281
ilGLMKİLİS	-0.25789	0.140201	-1.839	0.065849
ilGLMKIRIKKALE	-0.94138	0.185833	-5.066	4.07E-07
ilGLMKIRKLARELİ	-0.39346	0.084111	-4.678	2.90E-06
ilGLMKIRŞEHİR	0.139288	0.156308	0.891	0.372872
ilGLMKOCAELİ	0.187059	0.010496	17.821	< 2e-16
ilGLMKONYA	-0.41532	0.027086	-15.333	< 2e-16
ilGLMKÜTAHYA	-1.16816	0.15629	-7.474	7.76E-14
ilGLMMALATYA	-0.54145	0.072034	-7.517	5.62E-14
ilGLMMANİSA	-1.11971	0.090362	-12.391	< 2e-16
ilGLMMARDİN	-0.53905	0.062042	-8.688	< 2e-16
ilGLMMERSİN	-0.2143	0.024238	-8.842	< 2e-16
ilGLMMUĞLA	-1.01061	0.106172	-9.519	< 2e-16

İLGLMMUŞ	-0.67735	0.316328	-2.141	0.03225
İLGLMNEVŞEHİR	-0.08573	0.098237	-0.873	0.38286
İLGLMNIĞDE	-1.66543	0.378022	-4.406	1.05E-05
İLGLMORDU	-0.17436	0.02851	-6.116	9.62E-10
İLGLMOSMANİYE	-0.70295	0.160244	-4.387	1.15E-05
İLGLMRİZE	-0.09833	0.057368	-1.714	0.08654
İLGLMSAKARYA	-0.19622	0.027515	-7.132	9.92E-13
İLGLMSAMSUN	0.295735	0.016413	18.018	< 2e-16
İLGLMŞANLIURFA	-1.05865	0.101718	-10.408	< 2e-16
İLGLMSİİRT	-1.94176	0.577382	-3.363	0.000771
İLGLMSİNOP	-0.42552	0.131557	-3.235	0.001218
İLGLMŞIRNAK	-0.91502	0.154446	-5.925	3.13E-09
İLGLMSİVAS	0.175783	0.043598	4.032	5.53E-05
İLGLMTEKİRDAĞ	0.113088	0.01975	5.726	1.03E-08
İLGLMTOKAT	-0.01687	0.047527	-0.355	0.722648
İLGLMTRABZON	0.321314	0.016292	19.723	< 2e-16
İLGLMTUNCELİ	-0.67128	0.301647	-2.225	0.026055
İLGLMUŞAK	-0.456	0.085935	-5.306	1.12E-07
İLGLMVAN	-0.20132	0.053755	-3.745	0.00018
İLGLMYALOVA	-0.07569	0.050875	-1.488	0.136813
İLGLMYOZGAT	-0.7097	0.235789	-3.01	0.002613
İLGLMZONGULDAK	-0.68254	0.162367	-4.204	2.63E-05
YenilemeGLMYenileme	-0.11109	0.008221	-13.514	< 2e-16
MedeniDurumGLMBekar	-0.13975	0.012869	-10.86	< 2e-16
MedeniDurumGLMBoşanmış	0.052683	0.01835	2.871	0.004092
MedeniDurumGLMDul	0.020356	0.040888	0.498	0.618598
VKIGrup2GLMBebek	NA	NA	NA	NA
VKIGrup2GLMKilolu	0.033824	0.0197	1.717	0.085996
VKIGrup2GLMObez	0.175573	0.057737	3.041	0.002359
VKIGrup2GLMRiskli	0.114935	0.015375	7.475	7.69E-14
	Tahmin	Std. Hata	Z Degeri	Pr(> z)
Sabit Katsayı	1.87803	0.014995	125.246	< 2e-16
YasGrupGLM07-20 Arası	-0.50654	0.010172	-49.798	< 2e-16
YasGrupGLM21-25 Arası	-0.63651	0.017012	-37.416	< 2e-16
YasGrupGLM26-30 Arası	-0.53453	0.01463	-36.537	< 2e-16
YasGrupGLM31-35 Arası	-0.61148	0.015298	-39.972	< 2e-16
YasGrupGLM36-40 Arası	-0.69834	0.015937	-43.818	< 2e-16
YasGrupGLM41-45 Arası	-0.74531	0.016885	-44.141	< 2e-16
YasGrupGLM46-50 Arası	-0.76808	0.018729	-41.011	< 2e-16
YasGrupGLM51-55 Arası	-0.75902	0.021217	-35.775	< 2e-16
YasGrupGLM56-60 Arası	-0.70416	0.02577	-27.325	< 2e-16
YasGrupGLM61-65 Arası	-0.58129	0.078044	-7.448	9.46E-14
YasGrupGLM65+	-1.02139	0.213892	-4.775	1.79E-06
CinsiyetGLMERkek	-0.21542	0.006293	-34.233	< 2e-16
İLGLMADANA	-1.49693	0.082953	-18.045	< 2e-16
İLGLMADIYAMAN	-1.25405	0.25003	-5.016	5.29E-07

GLM3-
Meslek,
VKI yok

İGLMAFYONKARAHİSAR	-0.74464	0.140175	-5.312	1.08E-07
İGLMAĞRI	-0.50054	0.132597	-3.775	0.00016
İGLMAKSARAY	-0.06701	0.08241	-0.813	0.41616
İGLMAMASYA	-0.67398	0.060017	-11.23	< 2e-16
İGLMANKARA	-0.12005	0.009437	-12.721	< 2e-16
İGLMANTALYA	-0.29095	0.023465	-12.399	< 2e-16
İGLMARDAHAN	-0.72665	0.174185	-4.172	3.02E-05
İGLMARTVİN	-0.83741	0.107989	-7.755	8.86E-15
İGLMAYDIN	-0.45523	0.060766	-7.491	6.81E-14
İGLMBALIKESİR	-0.2871	0.04601	-6.24	4.38E-10
İGLMBARTIN	-0.65845	0.162338	-4.056	4.99E-05
İGLMBATMAN	-0.18207	0.058403	-3.117	0.001824
İGLMBAYBURT	-0.18396	0.204248	-0.901	0.367765
İGLMBİLECİK	-0.76259	0.196207	-3.887	0.000102
İGLMBİNGÖL	-1.76772	0.242615	-7.286	3.19E-13
İGLMBİTLİS	-1.30177	0.169132	-7.697	1.40E-14
İGLMBOLU	-1.14024	0.164506	-6.931	4.17E-12
İGLMBURDUR	-1.82553	0.500065	-3.651	0.000262
İGLMBURSA	-0.13946	0.012059	-11.565	< 2e-16
İGLMÇANAKKALE	-0.05224	0.046331	-1.128	0.259474
İGLMÇANKIRI	-0.57912	0.213319	-2.715	0.006631
İGLMÇORUM	-1.42388	0.218311	-6.522	6.92E-11
İGLMDENİZLİ	-0.08089	0.025589	-3.161	0.001571
İGLMDİYARBAKIR	-0.3124	0.036159	-8.639	< 2e-16
İGLMDÜZCE	-0.75811	0.185815	-4.08	4.51E-05
İGLMEDİRNE	-0.53322	0.129224	-4.126	3.69E-05
İGLMELAZIĞ	-0.63464	0.074746	-8.491	< 2e-16
İGLMERZİNCAN	-0.02436	0.0808	-0.302	0.763013
İGLMERZURUM	-0.40856	0.056621	-7.216	5.36E-13
İGLMESKİŞEHİR	-0.02148	0.043423	-0.495	0.620799
İGLMGAZİANTEP	0.102397	0.016732	6.12	9.38E-10
İGLMGİRESUN	-0.31257	0.036342	-8.601	< 2e-16
İGLMGÜMÜŞHANE	-1.03746	0.316296	-3.28	0.001038
İGLMHAKKARİ	0.060116	0.169165	0.355	0.722313
İGLMHATAY	-0.95379	0.062028	-15.377	< 2e-16
İGLMİSPARTA	-0.26259	0.107992	-2.432	0.015036
İGLMİZMİR	-0.38218	0.02311	-16.537	< 2e-16
İGLMKAHRAMANMARAŞ	-1.26355	0.13497	-9.362	< 2e-16
İGLMKARABÜK	-0.12682	0.066903	-1.896	0.058025
İGLMKARAMAN	-0.44144	0.140235	-3.148	0.001645
İGLMKARS	-0.19045	0.218403	-0.872	0.3832
İGLMKASTAMONU	-0.01006	0.058593	-0.172	0.863684
İGLMKAYSERİ	0.005596	0.022137	0.253	0.800448
İGLMKİLİS	-0.25319	0.1402	-1.806	0.070935
İGLMKIRIKKALE	-0.94298	0.185833	-5.074	3.89E-07
İGLMKIRKLARELİ	-0.39302	0.084109	-4.673	2.97E-06

İLGLMKİRŞEHİR	0.146322	0.156305	0.936	0.349206
İLGLMKOCAELİ	0.187147	0.010493	17.836	< 2e-16
İLGLMKONYA	-0.41359	0.027081	-15.272	< 2e-16
İLGLMKÜTAHYA	-1.17202	0.156288	-7.499	6.43E-14
İLGLMMALATYA	-0.53948	0.072017	-7.491	6.84E-14
İLGLMMANİSA	-1.11864	0.090361	-12.38	< 2e-16
İLGLMMARDİN	-0.54259	0.06204	-8.746	< 2e-16
İLGLMMERSİN	-0.21327	0.024237	-8.799	< 2e-16
İLGLMMUĞLA	-1.01232	0.106172	-9.535	< 2e-16
İLGLMMUŞ	-0.64963	0.316297	-2.054	0.03999
İLGLMNEVŞEHİR	-0.08422	0.098232	-0.857	0.391225
İLGLMNİĞDE	-1.67086	0.378021	-4.42	9.87E-06
İLGLMORDU	-0.17427	0.028504	-6.114	9.72E-10
İLGLMOSMANİYE	-0.70769	0.160242	-4.416	1.00E-05
İLGLMRİZE	-0.09106	0.057336	-1.588	0.112231
İLGLMSAKARYA	-0.19328	0.027511	-7.026	2.13E-12
İLGLMSAMSUN	0.296776	0.016402	18.094	< 2e-16
İLGLMŞANLIURFA	-1.0601	0.101718	-10.422	< 2e-16
İLGLMSİİRT	-1.949	0.577381	-3.376	0.000737
İLGLMSİNOP	-0.42442	0.131553	-3.226	0.001254
İLGLMŞIRNAK	-0.91373	0.154439	-5.916	3.29E-09
İLGLMSİVAS	0.176433	0.043598	4.047	5.19E-05
İLGLMTEKİRDAĞ	0.114067	0.01975	5.776	7.66E-09
İLGLMTOKAT	-0.01373	0.047507	-0.289	0.772606
İLGLMTRABZON	0.321345	0.016289	19.728	< 2e-16
İLGLMTUNCELİ	-0.67251	0.301647	-2.229	0.025784
İLGLMUŞAK	-0.45607	0.085935	-5.307	1.11E-07
İLGLMVAN	-0.20077	0.053755	-3.735	0.000188
İLGLMYALOVA	-0.07103	0.050871	-1.396	0.162641
İLGLMYOZGAT	-0.70657	0.235782	-2.997	0.002729
İLGLMZONGULDAK	-0.68738	0.162365	-4.234	2.30E-05
YenilemeGLMYenileme	-0.10979	0.008218	-13.36	< 2e-16
MedeniDurumGLMBekar	-0.13946	0.012858	-10.847	< 2e-16
MedeniDurumGLMBoşanmış	0.051174	0.018347	2.789	0.005283
MedeniDurumGLMDul	0.020883	0.040887	0.511	0.609523

Ek 2 - Genelleştirilmiş Linear Modellerle Hasar Güncel Tutar Sıklığı Tablosu

		Tahmin	Std. Hata	Z Degeri	Pr(> z)
GLM1 - Hepsi - Şiddet	Sabit Katsayı	6.524915	0.029233	223.206	< 2e-16
	yasgrup3GLM17 - 31	0.186433	0.029686	6.28	3.43E-10
	yasgrup3GLM32 - 40	0.29127	0.033554	8.681	< 2e-16
	yasgrup3GLM41 - 49	0.424705	0.035484	11.969	< 2e-16
	yasgrup3GLM50 - 75	0.503406	0.039916	12.612	< 2e-16
	ilGrupGLM2	0.095466	0.021642	4.411	1.03E-05
	ilGrupGLM3	-0.06468	0.020724	-3.121	0.001804
	ilGrupGLM4	-0.14609	0.02254	-6.481	9.26E-11
	ilGrupGLM5	-0.07933	0.025706	-3.086	0.002031
	ilGrupGLM6	-0.16395	0.018525	-8.851	< 2e-16
	YenilemeGLMYenileme	0.119014	0.016513	7.207	5.87E-13
	CinsiyetGLMERkek	0.006798	0.013279	0.512	0.608696
	MedeniDurumGLMBekar	0.032618	0.023221	1.405	0.160129
	MedeniDurumGLMBoşan mış	0.016445	0.037431	0.439	0.660411
	MedeniDurumGLMDul	0.001162	0.085396	0.014	0.989141
	MeslekGLMBeyaz Yaka	-0.04832	0.035146	-1.375	0.169189
	MeslekGLMÇalışmayan	0.029895	0.054607	0.547	0.584073
	MeslekGLMDiğer	-0.11112	0.03088	-3.598	0.000321
	MeslekGLMEmekli	-0.03272	0.096704	-0.338	0.735136
	MeslekGLMMavi Yaka	0.006938	0.029348	0.236	0.813118
	MeslekGLMMemur	-0.06909	0.036391	-1.899	0.057633
	MeslekGLMÖğrenci	-0.07537	0.033206	-2.27	0.023222
	MeslekGLMÖğretmen	-0.17218	0.043481	-3.96	7.52E-05
	VKIGrup2GLMBebek	NA	NA	NA	NA
VKIGrup2GLMKilolu	0.05562	0.039311	1.415	0.157121	
VKIGrup2GLMObez	-0.04375	0.122532	-0.357	0.721089	
VKIGrup2GLMRiskli	0.029497	0.031278	0.943	0.345653	
GLM2-VKI yok - Şiddet		Tahmin	Std. Hata	Z Degeri	Pr(> z)
	Sabit Katsayı	6.524849	0.029231	223.217	< 2e-16
	yasgrup3GLM17 - 31	0.177253	0.027696	6.4	1.58E-10
	yasgrup3GLM32 - 40	0.281651	0.031706	8.883	< 2e-16
	yasgrup3GLM41 - 49	0.415471	0.033712	12.324	< 2e-16
	yasgrup3GLM50 - 75	0.495208	0.03826	12.943	< 2e-16
	ilGrupGLM2	0.095385	0.021641	4.408	1.05E-05
	ilGrupGLM3	-0.06445	0.020708	-3.112	0.001858
	ilGrupGLM4	-0.14586	0.022536	-6.472	9.83E-11
	ilGrupGLM5	-0.07985	0.025706	-3.106	0.001897
	ilGrupGLM6	-0.16346	0.018513	-8.829	< 2e-16
	YenilemeGLMYenileme	0.119162	0.016513	7.216	5.50E-13
	CinsiyetGLMERkek	0.007379	0.013237	0.557	0.577232
	MedeniDurumGLMBekar	0.032231	0.023193	1.39	0.164627
	MedeniDurumGLMBoşan mış	0.014739	0.037423	0.394	0.693698

	MedeniDurumGLMDul	-0.00094	0.085385	-0.011	0.991193
	MeslekGLMBeyaz Yaka	-0.03772	0.033114	-1.139	0.254613
	MeslekGLMÇalışmayan	0.045225	0.052785	0.857	0.391576
	MeslekGLMDiğer	-0.09875	0.028361	-3.482	0.000499
	MeslekGLMEmekli	-0.02168	0.095941	-0.226	0.821256
	MeslekGLMMavi Yaka	0.019065	0.026763	0.712	0.47624
	MeslekGLMMemur	-0.05868	0.03448	-1.702	0.088786
	MeslekGLMÖğrenci	-0.06324	0.0308	-2.053	0.040068
	MeslekGLMÖğretmen	-0.16007	0.041755	-3.834	0.000127
GLM3- Cinsiyet, Medeni Durum, VKI yok - Şiddet		Tahmin	Std. Hata	Z Degeri	Pr(> z)
	Sabit Katsayı	6.56099	0.01724	380.483	< 2e-16
	yasgrup3GLM17 - 31	0.15874	0.02473	6.419	1.40E-10
	yasgrup3GLM32 - 40	0.25382	0.02501	10.15	< 2e-16
	yasgrup3GLM41 - 49	0.38753	0.02711	14.297	< 2e-16
	yasgrup3GLM50 - 75	0.4666	0.03198	14.591	< 2e-16
	ilGrupGLM2	0.0954	0.02166	4.405	1.06E-05
	ilGrupGLM3	-0.0649	0.02072	-3.132	0.001737
	ilGrupGLM4	-0.14613	0.02255	-6.479	9.39E-11
	ilGrupGLM5	-0.07961	0.02572	-3.095	0.001971
	ilGrupGLM6	-0.16378	0.01852	-8.842	< 2e-16
	YenilemeGLMYenileme	0.11948	0.01652	7.233	4.87E-13
	MeslekGLMBeyaz Yaka	-0.03932	0.03309	-1.188	0.234764
	MeslekGLMÇalışmayan	0.04583	0.05283	0.867	0.385701
	MeslekGLMDiğer	-0.10083	0.02834	-3.558	0.000375
	MeslekGLMEmekli	-0.02263	0.09597	-0.236	0.813593
	MeslekGLMMavi Yaka	0.01574	0.0267	0.59	0.555509
	MeslekGLMMemur	-0.06025	0.03449	-1.747	0.080664
	MeslekGLMÖğrenci	-0.05926	0.03067	-1.932	0.053386
MeslekGLMÖğretmen	-0.16291	0.04175	-3.902	9.58E-05	
GLM4- Meslek, Cinsiyet, Medeni Durum, VKI yok - Şiddet		Tahmin	Std. Hata	Z Degeri	Pr(> z)
	Sabit Katsayı	6.54899	0.015	436.472	< 2e-16
	yasgrup3GLM17-31	0.14257	0.01786	7.983	1.49E-15
	yasgrup3GLM32-40	0.23078	0.01687	13.678	< 2e-16
	yasgrup3GLM41-49	0.36867	0.01988	18.541	< 2e-16
	yasgrup3GLM50-75	0.44579	0.02534	17.595	< 2e-16
	ilGrupGLM2	0.08379	0.0193	4.341	1.42E-05
	ilGrupGLM3	-0.0719	0.02043	-3.519	0.000433
	ilGrupGLM4	-0.13623	0.0223	-6.108	1.02E-09
	ilGrupGLM5	-0.075	0.02558	-2.932	0.00337
	ilGrupGLM6	-0.16124	0.01827	-8.825	< 2e-16
YenilemeGLMYenileme	0.11277	0.01642	6.869	6.64E-12	