



**T.C. İSTANBUL T CARET
 NİVERSİTESİ**

FEN B LİMLERİ ENSTİT S 

**MAKİNE   RENME İYLE SA LIK S GORTALARINDA K MELEME
VE SINIFLANDIRMA**

 mer G L

**Dan şman
Do . Dr. Mustafa Cem KASAPBA I**

**Y KSEK LİSANS TEZİ
BİLGİSAYAR M HENDİSLİ İ ANABİLİM DALI
İSTANBUL - 2023**

KABUL VE ONAY SAYFASI

Ömer GÜL tarafından hazırlanan **Makine Öğrenmesiyle Sağlık Sigortalarında Kümeleme ve Sınıflandırma** adlı tez çalışması 12/06/2023 tarihinde aşağıdaki jüri üyeleri önünde başarı ile savunularak, İstanbul Ticaret Üniversitesi Fen Bilimleri Enstitüsü **Bilgisayar Mühendisliği Anabilim Dalı'nda YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Danışman **Doç. Dr. Mustafa Cem KASAPBAŞI**
İstanbul Ticaret Üniversitesi

Jüri Üyesi **Doç. Dr. Buket DOĞAN**
Marmara Üniversitesi

Jüri Üyesi **Dr. Öğr. Üyesi Ayşe TURAN ŞERBETÇİ**
İstanbul Ticaret Üniversitesi

Onay Tarihi : 16.06.2023

İstanbul Ticaret Üniversitesi, Fen Bilimleri Enstitüsünün 16.06.2023 tarih ve 2023/389 numaralı Yönetim Kurulu Kararının 2. maddesi gereğince, ders yüklerini ve tez yükümlülüğünü yerine getirdiği belirlenen “Ömer GÜL” adlı öğrencinin mezun olmasına oy birliği ile karar verilmiştir.

Prof. Dr. Doğan KAYA
Enstitü Müdürü

AKADEMİK VE ETİK KURALLARA UYGUNLUK BEYANI

İstanbul Ticaret Üniversitesi, Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada,

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

16.06.2023

Ömer GÜL

İÇİNDEKİLER

	Sayfa
İÇİNDEKİLER.....	i
ÖZET	ii
ABSTRACT	iii
TEŞEKKÜR.....	iv
ŞEKİLLER	v
ÇİZELGELER.....	vi
SİMGELER VE KISALTMALAR.....	vii
1. GİRİŞ	1
2. LİTERATÜR ÖZETİ.....	7
3. MATERYAL VE METHOD.....	14
3.1 Veri Setinin Hazırlanması	14
3.2 Aykırı Değerlerin Bulunması	23
3.3 Dirsek Yöntemi (Elbow Method)	24
3.4 K-Means Algoritması	25
3.5 Rastgele Orman (Random Forest) Algoritması	26
3.6 Bilgi Kazancı (Gain Information)	27
3.7 Bilgi Kazancı Oranı (Gain Information Ratio – GIR)	28
3.8 Gini İndeksi (Gini Index)	29
3.9 Gradyan Yükseltme (Grandient Boosting) Algoritması.....	29
3.10 Aşırı Gradyan Yükseltme (Extrem Grandient Boosting) Algoritması	30
3.11 Çok Katmanlı Algılayıcı (Multilayer Perceptron – MLP)	31
3.12 Knime	31
3.13 Çalışma Kapsamı ve Akışı	32
4. ARAŞTIRMA BULGULARI VE TARTIŞMA.....	34
5. SONUÇ VE ÖNERİLER.....	50
KAYNAKLAR	51
ÖZGEÇMİŞ.....	55

ÖZET

Yüksek Lisans Tezi

MAKİNE ÖĞRENMESİYLE SAĞLIK SİGORTALARINDA KÜMELEME VE SINIFLANDIRMA

Ömer GÜL

İstanbul Ticaret Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Doç. Dr. Mustafa Cem KASAPBAŞI

2023, 55 sayfa

Birçok sektörde olduğu gibi, sigortada da en büyük hedef şirketin karlılığını maksimize etmektir. Bu hedefe ulaşmanın temel unsurlarından biri, karlı olabilecek yeni müşteriler çekmektir. Sigorta sektöründe sağlık branşı, bir sigorta şirketine hem kar hem de zarar getirebilen bir branştır. Bu nedenle, hedeflenen müşteriler için doğru fiyatları sunarak şirketin taleplerini minimize etmek gereklidir. Sağlık branşında, müşteri bir poliçe satın alabilir ve o poliçe kapsamında birden fazla kişi sigortalı olabilir. Bu alandaki önceki akademik çalışmalar, sigorta şirketlerinin genellikle karlılıklarını poliçe düzeyinde hesapladıklarını göstermektedir. Bu nedenle, bir müşterinin sigortalı grubunda ne kadar karlı olduğu belirsiz kalmaktadır. Ayrıca, faydalı bir müşteriye doğru fiyatın verilmediği tespit edilmiştir ve bu da müşterinin kaybedilmesine neden olmaktadır. Fiyatlandırma ve satış süreçlerindeki kilit değer doğru fiyatın yani sigorta şirketi için toplamda kar edebileceği satış tutarının belirlenmesidir. Ancak bu nokta çok kolay hesaplanamamakla birlikte gelir gider dengesini kurgulamak da çok meşakatlidir. Bu durumun zorluğundan kaynaklı olarak, düşük fiyattan yapılan bir poliçeden zarar edilebileceği gibi yüksek kar elde edilebilecek bir müşteriye de yüksek fiyat verilerek müşteri kaybedilebilir. Bu çalışma kapsamında, karlılık hesabı sigortalı düzeyine indirgenmiştir. Sigorta şirketinin portföyündeki müşterilerin karlılığı, K-ortalama algoritması ile kümelenmelere ayrılmıştır. Sigortalı müşterinin demografik bilgilerini, doktorlar tarafından girilen teşhis bilgilerini, ödenen primleri, talep edilen tutarı ve karlılık miktarlarını içeren yaklaşık 480.000 poliçe sahibinin verileri kümelenmelerde kullanılmıştır. Yeni gelen müşteriler, makine öğrenimi algoritmasıyla işlem görür ve benzer özelliklere sahip müşterilerin kümesine düşer. Bu şekilde, müşterinin şirket için karlı olup olmayacağı öngörülür. Müşterinin karlılığına göre, şirketin kaybını minimize etmek için fiyatlandırmada indirim veya fazlalık tanımlanabilir. Yapılan araştırmalara neticesinde ulusal yada uluslar arası sigortacılık alanında bu çalışmaya benzer bir çalışmaya rastlanmamıştır. Dolayısıyla yapılan çalışma sektör düzeyinde önem kazanmaktadır.

Anahtar Kelimeler: Fiyatlandırma, karlılık, makine öğrenimi, sigorta.

ABSTRACT

M.Sc. Thesis

SINIFING AND CLASSIFICATION IN HEALTH INSURANCE WITH MACHINE LEARNING

Ömer GÜL

**İstanbul Commerce University
Graduate School of Applied and Natural Sciences
Department of Computer Engineering**

Supervisor: Assoc. Prof. Dr. Mustafa Cem KASAPBAŞI

2023, 55 pages

As in many sectors, the biggest target in insurance is to maximize the company's profitability. One of the essential elements of achieving this goal is to attract new customers that can be profitable. In the insurance sector, the health branch is a branch that can bring profit as well as loss to an insurance company. For this reason, minimizing the company's claim is necessary by offering the right prices for the targeted customers. In the health branch, the customer can purchase a policy and have more than one person covered by that policy. Past academic studies in this field show that insurance companies usually calculate their profitability at the policy level. Therefore, it stays unclear how profitable a customer is at the insured group. Also, it has been determined that the correct price is not given to a beneficial customer, resulting in the loss of the customer. The key value in pricing and sales processes is determining the correct price, which is the total sales amount that the insurance company can make a profit from. However, this point is not easy to calculate, and building a balance between income and expenses is also very challenging. Due to the difficulty of this situation, it is possible to incur losses from a policy sold at a low price, and a customer who can provide high profits can also be lost by offering a high price. Within the scope of this study, the profitability account is reduced to the insured level. The customers' profitability in the insurance company's portfolio is divided into Sınıflar with the K-means algorithm. The data of approximately 480.000 policyholders, which include the features demographic information of the insured customer, the diagnosis information entered by the doctors, the premiums paid, the amount of claim and the profitability amounts, have been used in Sınıflar. New incoming customers pass through the machine learning algorithm and fall into the Sınıf of customers with similar characteristics. In this way, it is predicted whether the customer will be profitable for the company. According to the customer's profitability, a discount or surplus can be defined for the pricing to minimize the company's loss.

Keywords: Insurance, machine learning, pricing, profitability.

TEŞEKKÜR

Bu araştırma için beni yönlendiren, karşılaştığım zorlukları bilgi ve tecrübesi ile aşmamda, literatür araştırmalarımnda yardımcı olan değerli değerli Danışman Hocam Doç. Dr. Mustafa Cem KASAPBAŞI'na teşekkürlerimi sunarım.

Araştırmanın yürütülmesinde maddi ve manevi yardımlarını gördüğüm Vedat GÜNEŞ ve Sedat CEBESÖY'a olmak üzere tüm Anadolu Sigorta personeline teşekkür ederim.

Tezimi AR-GE Projesi olarak hayata geçirmemi destekleyen Anadolu Sigorta Veri Yönetimi ve Analitik Çözümler Müdürlüğüne teşekkür ederim.

Tezimin her aşamasında beni yalnız bırakmayan aileme sonsuz sevgi ve saygılarımı sunarım.

Ömer GÜL
İSTANBUL, 2023

ŞEKİLLER

	Sayfa
Şekil 3.1 Şehir ve ilçe bilgisi boş olan sigortalı sayısı	21
Şekil 3.2 Aykırı değer hesaplama	23
Şekil 3.3 Küme sayısı	25
Şekil 3.4 K Değerine göre kümeleme örneği	26
Şekil 3.5 Ağaç yapısı	27
Şekil 3.6 Veri setinde entropi.....	28
Şekil 3.7 Çok katmanlı ağ yapısı	31
Şekil 3.8 Çalışma kapsamı akışı.....	33
Şekil 4.1 KNIME üzerinde tasarlanan akış	34
Şekil 4.2 Elbow testi sonuç grafiği	35
Şekil 4.3 TSS kümeleme sonucu.....	35
Şekil 4.4 ÖSS kümeleme sonucu.....	36

ÇİZELGELER

	Sayfa
Çizelge 1.1 Sigortalı Bazında karlılık örneği	6
Çizelge 3.1 Tüfe oranları	18
Çizelge 3.2 Örnek KPI tutarları.....	18
Çizelge 3.3 Demografik bilgileri boş olan sigortalılar.....	21
Çizelge 3.4 Hastalık grupları bazında provizyon sayıları.....	22
Çizelge 4.1 TSS sınıf açıklaması.....	36
Çizelge 4.2 ÖSS sınıf açıklaması	37
Çizelge 4.3 TSS RF – Gini index karışıklık matrisi.....	37
Çizelge 4.4 TSS RF– Gini index genel istatistikler	38
Çizelge 4.5 TSS RF – Gain ratio karışıklık matrisi	38
Çizelge 4.6 TSS RF – Gain ratio genel istatistikler.....	39
Çizelge 4.7 TSS RF – Gain karışıklık matrisi	39
Çizelge 4.8 TSS RF – Gain genel istatistikler.....	39
Çizelge 4.9 TSS– GBoosted Trees karışıklık matrisi	40
Çizelge 4.10 TSS – GBoosted trees genel istatistikler	40
Çizelge 4.11 TSS – XGBoost karışıklık matrisi	41
Çizelge 4.12 TSS – XGBoost genel istatistikler	41
Çizelge 4.13 TSS – Neural Netwok - MLP karışıklık matrisi.....	42
Çizelge 4.14 TSS – Neural Netwok - MLP genel istatistikler	42
Çizelge 4.15 ÖSS RF – Gini index karışıklık matrisi.....	43
Çizelge 4.16 ÖSS RF – Gini index genel istatistikler	43
Çizelge 4.17 ÖSS RF – Gain ratio karışıklık matrisi	44
Çizelge 4.18 ÖSS RF – Gain ratio genel istatistikler.....	44
Çizelge 4.19 ÖSS RF – Gain karışıklık matrisi	45
Çizelge 4.20 ÖSS RF – Gain genel istatistikler.....	45
Çizelge 4.21 ÖSS – GBoosted Trees karışıklık matrisi	46
Çizelge 4.22 ÖSS – GBoosted Trees genel istatistikler	46
Çizelge 4.23 ÖSS – XGBoost karışıklık matrisi.....	47
Çizelge 4.24 ÖSS – XGBoost genel istatistikler	47
Çizelge 4.25 ÖSS – Neural Network - MLP karışıklık matrisi.....	48
Çizelge 4.26 ÖSS – Neural Network - MLP genel istatistikler	48
Çizelge 4.27 Algortimaların sonuç karşılaştırması.....	48

SİMGELER VE KISALTMALAR

ICD	International Classification of Disease
IQR	Inter Quantile Range / Çeyrekler Arası Açıklık
KNIME	Konstanz Information Miner
KPI	Key Performance Indicator / Temel Performans Göstergesi
MLP	Multilayer Perceptron / Çok Katmanlı Algılayıcı
ÖSS	Özel Sağlık Sigortası
PVOP	Pozitif Vergi Öncesi Prim
RF	Random Forest / Rastgele Orman
SOM	Self Organizin Map
SVM	Support Vector Machine / Destek Vektör Makineleri
TKZ	Teknik Kar Zarar
TSS	Tamalayıcı Sağlık Sigortası
VOP	Vergi Öncesi Prim
WCSS	Within Sınıfs Sum of Qquare / Kümeler İçi Kareler Toplamı
XGBoost	Extrem Grandient Boosting / Aşırı Gradyan Yükseltme

1. GİRİŞ

Sigorta, insanların hayatı boyunca kendinde ya da eşyalarında oluşabilecek hasarlara karşı finansal olarak korunabilmesi için sigorta şirketleri tarafından belirli bir prim karşılığında yapılan sözleşmedir. İnsanlar, günlük yaşamlarından birçok riskli olayla karşılaşmaktadırlar. Kişiler kendilerini güvende hissetmek ve yaşam koşullarını yükseltmek için bu risklerde korunmak isterler. (Yayla, 2019). Sigorta, gelecekte oluşabilecek risklere karşı kişileri ya da varlıkları güvence altına almak için yapılan mali bir araçtır. Günümüzde birçok insan kendilerini ya da varlıklarını sigortalayarak güvence altına almaktadır. Sigorta firmaları da Sigorta satışlarıyla finansal karlarını yükseltmektedir.

Kar kavramı, işletme faaliyetlerinin sürekliliğini sağlamak için temel bir unsurdur. Şirketlerin, varlıklarını sürdürmek ve sahiplerine gelir sağlamak amacıyla her dönem sonunda kar elde etme çabası doğaldır. Bunun için şirketler, maliyetleri minimize etmeyi, satış fiyatlarını maksimize etmeyi ve yatırım fırsatlarını ve daha uygun üretim alanlarını keşfetmeyi amaçlayan stratejiler ararlar (Sabuncuoğlu ve Tokol, 2001).

Diğer sektörlerde de olduğu gibi sigorta firmalarının en büyük hedefi maksimum karlılığı sağlayabilmektir. Finansal karlılık bir şirketin gelişimini gösterdiği gibi gelecekte ne kadar büyük yatırımlar yapabileceğinin bir göstergesidir. Karlılığı yüksek tutmak için maksimum gelir elde etmek, giderleri de minimuma indirmek gerekir.

Sigorta sektöründe, bireysel ve kurumsal müşterilere birçok branşta poliçeler satılmaktadır. Trafik sigortası, kasko, yangın, dask, sağlık poliçeleri bir sigorta şirketinde en çok satılan ürünler arasında yer almaktadır. Bu ürünlerden trafik sigortası ve dask fiyatları Hazine ve Maliye Bakanlığı tarafından belirlenmektedir. Aracı ya da evi olan herkes bu sigortayı yaptırmak zorundadır. Bu ürünlerin fiyatı Hazine ve Maliye Bakanlığı tarafından belirlendiği için sigorta şirketleri bu ürünlerden genelde kar elde edemez. Tam aksine bu ürünlerde tavan fiyat uygulaması olduğu için hasar oluşturma oranı bakımından sigorta şirketleri için

zararlı ürünlerdir. Diğer branşlarda yer alan ürünlerin fiyatları sigorta şirketleri tarafından belirlenmekte ve müşterilerin isteği doğrultusunda satılmaktadır. Bu fiyatların belirlenmesinde müşteriye ve sigortalıya (sigortalı bir kişi ya da bir eşya olabilir) göre birçok farklı yol izlenmektedir.

Fiyatlandırma ve satış süreçlerindeki kilit değer doğru fiyatın yani sigorta şirketi için toplamda kar edebileceği satış tutarının belirlenmesidir. Ancak bu nokta çok kolay hesaplanamamakla birlikte gelir gider dengesini kurgulamak da çok meşakatlidir. Bu durumun zorluğundan kaynaklı olarak, düşük fiyattan yapılan bir poliçeden zarar edilebileceği gibi yüksek kar elde edilebilecek bir müşteriye de yüksek fiyat verilerek müşteri kaybedilebilir. Sigortacılıkta karlılık verisini etkileyen önemli parametreler vardır. Bunlardan bazıları aşağıdaki gibidir;

Hasar Oranları: Sigorta şirketleri, poliçe sahiplerine hasar ödemeleri yapmak zorunda kalabilir. Hasar oranları, sigorta şirketinin ödenen hasarların prim gelirine oranını gösterir. Yüksek hasar oranları, sigorta şirketleri için karlılığı olumsuz etkileyebilir (Koçkayalar vd., 2016).

Prim Gelirleri: Sigorta primleri, sigorta şirketinin satın aldığı risk karşılığında ödenen bedellerdir ve aynı zamanda genel giderler, sigorta şirketinin karşılamak zorunda olduğu maliyetler ve komisyonlar da primlere dahil edilir. Sigorta hizmetiyle sağlanan kapsamın tüm koşulları karşılansa bile, primlerin belirlenen süreçlere göre ödenmemesi durumunda sigorta sözleşmesi askıya alınır ve sigorta koruma hizmetinden faydalanma imkanınız olumsuz etkilenir. (Aksigorta, 2023).

Sigorta şirketleri, prim gelirlerini yatırımlara yönlendirerek gelir elde edebilir. Bu yatırımların getirisi, şirketin karlılığını önemli ölçüde etkileyebilir. Yatırım getirileri düşükse, sigorta şirketlerinin karlılığı da etkilenebilir (Asunakutlu, 1998).

Reasürans: Sigorta şirketleri, hayat ve hayat dışı alanlarda karşılayabilecekleri risklerin tamamını üstlenme yeteneğine sahip değillerdir. Bu nedenle, prim

karşılığında üstlendikleri risklerin belirli bir bölümünü başka bir sigorta şirketine devretme yoluna giderler. Bu sürece 'reasürans işlemi' denir ve reasürans işlemini üstlenen şirkete 'reasürans şirketi', devreden şirkete ise 'sedan şirket' denir. Reasürans işlemleri, sigorta şirketleri tarafından veya reasürans işlemlerinde aracılık yapan uzman reasürans aracıları tarafından gerçekleştirilir. Reasürans aracıları, riskle ilgili bilgiler sağlayarak hizmetlerinin karşılığında komisyon alır ve konularında uzmandırlar (Gücenme ve İşverenoğlu, 2005). Sigorta şirketleri, risklerini yaymak için reasürans şirketlerinden sigorta satın alabilir. Reasürans, sigorta şirketinin riskini paylaşarak karlılığını korumasına yardımcı olabilir.

Rekabet: Sigorta sektörü oldukça rekabetçidir. Rekabet, prim fiyatlarının düşük olmasına ve kar marjlarının azalmasına neden olabilir. Yoğun rekabet, sigorta şirketlerinin karlılığını etkileyebilir. Türkiye'de, rekabetin etkili olduğu bir sosyal ve ekonomik yapı bulunmaktadır. Örneğin, sigorta bilincinin zayıf olduğu ve gelir düzeyinin düşük olduğu az gelişmiş bölgelerde, aracılar arasında birleşmeler, ürün çeşitliliğini ve prim üretimini artırmak için mümkün olabilir. Öte yandan, yoğun nüfusa sahip büyük şehirlerde, sigorta piyasasının daha derin ve geniş olmasıyla birlikte artan rekabet, yoğunlaşmanın temel itici gücü olacaktır (Tunay, 2014).

Tüm bu faktörler göz önüne alındığında, sigorta sektöründe karlılık genel olarak dalgalanabilir. Bazı yıllar sigorta şirketleri için oldukça karlı geçebilirken, diğer yıllarda hasar oranları veya yatırım getirileri gibi etkenler nedeniyle karlılık azalabilir. Her sigorta şirketinin karlılık durumu, iş modeli, pazar payı ve risk yönetimi stratejilerine bağlı olarak farklılık gösterecektir.

Sağlık sigortalarında "sigortalı" terimi, bir sigorta poliçesinin sahibi veya kapsamında sağlık sigortası olan kişiyi ifade eder. Sigortalı, sigorta şirketi ile yapılan sözleşme uyarınca belirli bir prim ödemesi karşılığında sağlık sigortası kapsamında belirlenen hizmetlerden yararlanma hakkına sahip olan kişidir. Sigortalı, sağlık sigortası poliçesinde belirtilen şartlara ve kapsama dahil olan hizmetlere tabi olur. Sigorta şirketi, sigortalıya sağlık hizmetlerinin maliyetini

karşlamak amacıyla belirli bir prim talep eder. Sigortalı, poliçe kapsamında belirli bir süre boyunca sigorta şirketine prim öder ve belirli bir süre içinde belirli koşullar altında sağlık hizmetlerinden faydalanabilir. Sigortalı, sigorta poliçesi ile belirlenen hastalık veya kaza durumunda tıbbi tedavi, ilaç masrafları, hastaneye yatış, ameliyatlar, laboratuvar testleri, muayene ücretleri gibi sağlık hizmetlerinden yararlanma hakkına sahiptir. Sigortalı, sigorta şirketi ile yaptığı sözleşme kapsamında sağlık sigortası teminatlarından faydalanabilir ve tedavi masraflarının bir kısmını veya tamamını sigorta şirketinin ödemesini talep edebilir (Varoğlu, 2013).

Sağlık branşında bir poliçe birden fazla sigortalıyı kapsayabilmektedir. Karlılık modelleri poliçe seviyesinde karlılık (kar zarar durumu) hesaplaması yapmaktadır. Bu çalışma kapsamında sigortalı seviyesinde karlılığı takip edebilmek ve bu verileri kullanarak yeni gelecek olan müşterilerin karlı bir müşteri olup olmayacağını tahminlenmesi amaçlanmaktadır.

Sigortalı seviyesinde bir müşterinin ne kadar karlı olup olmadığı net olarak anlaşılmamaktadır. Bundan dolayı karlı bir müşteri için doğru fiyat verilmeyerek müşterinin kaybedilmesine sebep olunduğu tespit edilmiştir. Yeni gelen müşteri için kar zarar durumunu gösterecek referans bir veri bulunmamaktadır. Bu sebeple zarar potansiyeli yüksek müşterilere düşük fiyattan teklifler verilebilmektedir ve karlılığın düşmesine sebep olunabilmektedir. Bu çalışma sonucunda müşterinin kar zarar durumuna göre doğru fiyatlandırma yapılması beklenmektedir. Sigortacılık sektöründe sağlık branşında kar marjının daha yukarıya taşınabilmesi için bu çalışma önem arz etmektedir.

Çalışma kapsamında poliçe teknik karlılık hesaplamasında kullanılan kpi'ların poliçenin alt kırılımı olan sigortalı bazında detaya indirilmesi ve hesaplamalarının yapılması ve iş konusu bazında belirlenecek yeniden hesaplama dönemlerine (Günlük/Aylık) göre güncellenmesidir.

Sağlık sigortalarında, mevcut sigortalıların poliçelerinin süresinin dolmasında veya dolmasına yakın poliçesinin yenilenmesi için müşteriyle irtibata geçilir.

Müşterinin poliçeyi yenilemeyi istemesi durumunda sigorta şirketi tarafından eski poliçesi üzerinden yeni bir fiyatlandırma ve poliçe kapsamında değişiklikler yapılabilir. Yenileme dönemlerinde en önemli faktörlerden bir tanesi de ömür boyu yenileme garantisi (ÖBYG)'dir. ÖBYG'nin amacı, sigortalının poliçesini belirli bir süre boyunca düzenli yenileme hakkını sağlamaktır. ÇBYG, sigortalının belirli bir yaşa geldiğinde veya sağlık durumu kötüleştiğinde dahi sigortasını yenileme hakkını korur (Süzel, 2018). Bu garantiyle sigortalının yaşlanması veya ciddi bir sağlık sorunuyla karşılaşması durumunda prim artışı veya poliçe iptali gibi durumlarla karşılaşma riskini azaltır.

Sağlık sigortalarında yeni müşterilerin poliçe yaptırma süreçleri sigorta şirketlerine göre farklılık gösterebilir. Genellikle müşteri hangi sağlık poliçesi tipini yaptırmak istediğini seçtikten sonra başvuru formu doldurulması istenir. Bu formda müşterinin bilgileri, sağlık geçmişi, sigorta kapsamına alınmak istenen diğer şahısların bilgileri istenir. Daha sonrasında sigortalıların sağlık durumlarının değerlendirilmesi yapılır. Değerlendirme sürecinde, sigortalıların detaylı geçmiş sağlık bilgisi, mevcut durumu hakkında bilgi istenebilir. Riskleri değerlendirmek için bir anket, sağlık muayenesi veya ameliyat sonuç raporu, doktor muayene raporu gibi tıbbi raporlar talep edilebilir. Bu değerlendirme sonucunda sigorta şirketi bazı sağlık teminatlarını kapsam dışı bırakabilir veya sigorta yapılması uygun bulmayabilir. Teminatların kapsam dışı bırakılması veya sigortalının ödeyeceği primi doğrudan etkiler. Tüm bu süreçlerin yapılması sigorta şirketin ileride doğabilecek hasar ödemelerinden kaynaklı zarar etmesinin önüne geçilmek istenmesidir.

Mevutta çalışan karlılık verileri poliçe üzerinden olduğu için fiyatlandırma çalışmalarında poliçenin kar zarar durumuna göre yeni fiyat teklifi hesaplanmaktadır. Yeni yapılan çalışmayla sigortalı seviyesinde kar zarar durumu görülebildiği için daha isabetli bir fiyat çalışması yapılabilecektir. Bir poliçenin altında birden fazla sigortalı yer alabilmektedir. Yapılan çalışmayla her bir sigortalı için ödenen prim, sigortalı için ödenen hasar tutarı ve elde edilen karlılık tutarı sigortalı bazında gözlemlenebilmektedir. Çizelge 1.1 de görüldüğü üzere poliçeden -5.895,57 Türk Lirası zarar edilmiş. Poliçe sigortalı bazında

incelendiğinde, 2 tane sigortalıdan kar edildiği ama diğer sigortalıdan kaynaklı yüksek miktarda ödenen hasar olduğu için zarar ettiğini görebiliyoruz. Bu poliçenin yenileme dönemi geldiği zaman 88888 sigortalı numaralı sigortalının ödeneceği prim buradaki karlılık tutarı göz önünde bulundurularak revize edilecektir.

Çizelge 1.1 Sigortalı bazında karlılık örneği

Poliçe No	Sigortalı No	Prim	Ödenen Hasar	Karlılık Tutarı
1111111	99999	4.881,57	310,09	3.659,61
1111111	88888	9.897,57	18.262,93	-10.214,22
1111111	77777	5.699,79	3.976,04	659,04
Toplam		20.478,93	22.549,06	-5.895,57

Müşteri ilk defa sağlık poliçesi yaptırılacaksa; kar zarar durumu bilinmediği için yüksek bir fiyatlandırma yapılabilir, bu durumda müşteri kaybına sebebiyet verebilir. Sigortalı karlılığı hesaplamaları sonucunda oluşacak olan karlılık verileri yardımıyla kümeleme algoritması kullanılarak analitik bir model oluşturuldu. Bu model sayesinde yeni müşterilerin hangi kümede yer alabileceği görülmüş olacak. Yeni müşterinin yer alabileceği kümeye göre müşteriye ekstra indirim uygulanabilir ya da daha yüksek fiyattan bir teklif verilebilecektir.

Yapılan araştırmalara neticesinde ulusal ya da uluslararası sigortacılık alanında bu çalışmaya benzer bir çalışmaya rastlanmamıştır. Dolayısıyla yapılan çalışma sektör düzeyinde önem kazanmaktadır. Sigortalı seviyesinde karlılık takibinin yapılabilmesi ve yeni müşteriler için kar zarar durumunu referans gösterecek bir veri oluşturulması hedeflenmektedir. Hesaplanan karlılık verisi müşterilerimize teklif hazırlanma sürecinde, poliçe yapma adımlarında ve poliçe yenileme kademelerinde girdi olarak kullanılacaktır. Bu sayede şirketimizin müşteri seviyesinde karlılığının arttırılması hedeflenmektedir.

2. LİTERATÜR ÖZETİ

Yapılan araştırmalar sonucunda bu çalışma kapsamını doğrudan konu alan ulusal ve uluslararası çalışma tespit edilememiştir. Çalışma içerisinde kullanılan bazı yöntemlerin kullanıldığı çalışmalar incelenmiştir.

S. Balaji ve S. K. Srivatsa tarafından gerçekleştirilen bir çalışmada, müşterilerin hayat sigortası poliçelerinde tercih ettikleri poliçe türleri ve bu poliçelerin getirileri incelenmiştir. Çalışma, müşterilerin demografik bilgileri ile poliçe tercihleri arasındaki ilişkiyi anlamak ve müşteri segmentasyonu yapmak amacıyla gerçekleştirilmiştir. Çalışma için müşterilerin yaş, cinsiyet, medeni durum, çocuk sayısı gibi demografik bilgileri toplanmıştır. Ayrıca, müşterilerin hangi poliçe türlerini tercih ettikleri (Birikim, Koruma, Emeklilik, Çocuk) de kaydedilmiştir. Bu veriler analiz edilerek, müşteriler A, B, C, D gibi sınıflara atanmıştır. Daha sonra, etiketlenmiş veriler kullanılarak Naïve Bayes sınıflandırma algoritması uygulanmıştır. Bu algoritma, müşterilerin demografik özelliklerini ve poliçe tercihlerini temel alarak, bir müşterinin hangi sınıfa ait olduğunu tahmin etmek için kullanılmıştır. Bu tahminle birlikte, her müşterinin hangi sınıfa ait olduğu belirlenmiştir. Elde edilen sonuçlar, müşterilerin hangi sınıfa ait olduğunu göstermiştir. Bu bilgi, sigorta şirketlerine pazarlama stratejileri geliştirmede ve pazarlama aksiyonları almakta yardımcı olmuştur. Müşterilerin bulundukları sınıfların getirilerine göre, sigorta şirketleri hedeflenen pazarlama aksiyonlarını belirleyebilmiştir. Örneğin, bir sınıfta bulunan müşterilere özellikle birikim veya emeklilik poliçeleriyle ilgili teklifler sunulabilirken, başka bir sınıfta bulunan müşterilere ise koruma poliçeleriyle ilgili teklifler sunulabilir. Bu çalışma, sigorta şirketlerinin müşteri segmentasyonunu ve pazarlama stratejilerini iyileştirme konusunda önemli bir adım olmuştur. Demografik bilgiler ve poliçe tercihleri gibi verilerin analizi, sigorta şirketlerine müşteri davranışlarını daha iyi anlama ve hedeflenmiş pazarlama stratejileri oluşturma imkanı sağlamaktadır. Böylece, şirketler müşterilere daha uygun ve kişiselleştirilmiş teklifler sunarak, müşteri memnuniyetini artırabilir ve satış performansını geliştirebilir (Balaji ve Srivatsa, 2012).

Sigortaya olan talebin artması ve sigorta şirketlerinin işlerini akıllı çözümler kullanarak başarılı sonuçlar elde edebilmesi için müşteri verileri ve analitik çalışmalar önemli bir rol oynamaktadır. Bu bağlamda, Noorhannah Boodhun tarafından yapılan bir çalışma, müşteri analitiği, risk tahmini ve dolandırıcılık üzerine analitik çalışmaların faydalarını ve kullanılabilecek algoritmaları incelemiştir. Müşteri analitiği, sigorta şirketlerine müşterilerini daha iyi anlama ve onlara daha iyi hizmet sunma imkanı sağlar. Bu analitik çalışmalar, müşteri verilerinin incelenmesi ve işlenmesi yoluyla gerçekleştirilir. Müşterilerin demografik özellikleri, geçmiş sigorta talepleri, tercihleri ve davranışları gibi bilgiler kullanılarak müşteri segmentasyonu yapılabilir. Bu sayede, şirketler müşterilere daha uygun ürünleri sunabilir, müşteri memnuniyetini artırabilir ve sadakati sağlayabilir. Risk tahmini, sigorta şirketlerinin potansiyel riskleri belirleyerek önlemler almasına yardımcı olur. Müşteri analitiği ve risk tahmini birlikte kullanıldığında, şirketler müşterilerin gelecekteki sigorta taleplerini ve risklerini tahmin edebilir. Karar ağaçları ve sinir ağları gibi sınıflandırma modelleri, müşteri kaybı gibi risklerin tahmin edilmesinde sıklıkla kullanılan yöntemlerdir. Bu modeller, müşterileri farklı kümelerde sınıflandırarak, potansiyel riskleri belirlemeye ve buna göre stratejiler oluşturmaya yardımcı olur. Dolandırıcılık tespiti de sigorta şirketleri için önemli bir konudur. Dolandırıcılık olaylarının tespiti, geleneksel analitik yöntemlerle gerçekleştirilebilir. Kümeleme, sınıflandırma ve lineer regresyon gibi yöntemler, dolandırıcılık şüphelerini belirlemek ve incelemek için kullanılabilir. Bu yöntemler, algoritmalara dayalı analizlerle şüpheli faaliyetleri tespit etmeye ve önlem almaya yardımcı olabilir. Sonuç olarak, müşteri verilerini kullanarak yapılan analitik çalışmalar, sigorta şirketlerine müşterilerini daha iyi anlama, riskleri tahmin etme ve dolandırıcılık gibi sorunları tespit etme imkanı sağlar. Müşteri analitiği, risk tahmini ve dolandırıcılık tespiti gibi alanlarda kullanılan algoritmalar ve analitik modeller, sigorta şirketlerinin iş süreçlerini optimize etmelerine ve başarılı sonuçlar elde etmelerine yardımcı olur. Bu sayede, sigorta şirketleri müşteri memnuniyetini artırabilir, riskleri yönetebilir ve iş süreçlerini verimli bir şekilde yürütebilir (Boodhun, 2017).

Kuangnan Fang, Yefei Jiang ve Malin Song tarafından gerçekleştirilen çalışmada, sigorta sektöründe büyük veri analitiği yöntemleri kullanarak müşteri karlılığının hesaplanmasının önemi vurgulanmıştır. Büyük veri analitiği, sigorta şirketlerine müşterileri daha iyi anlama ve pazarlama stratejilerini optimize etme imkanı sağlayan güçlü bir araçtır. Çalışma, büyük veri analitiği için Rastgele Orman (Random Forest) modelinin diğer yaygın kullanılan modellerden (Lineer Regresyon, Karar Ağacı, Destek Vektör Makinası gibi) daha iyi performans sergilediğini göstermiştir. Rastgele Orman modeli, birçok ağaçtan oluşan bir orman yapısıyla çalışarak, karmaşık ilişkileri ve etkileşimleri tespit etme yeteneğiyle bilinir. Bu model, sigorta şirketlerine müşteri karlılığını hesaplama ve tahmin etme konusunda daha doğru ve güvenilir sonuçlar sunmaktadır. Çalışma, pazarlama alanında müşteri karlılığının önemine de vurgu yapmıştır. Sigorta şirketleri için değerli müşterileri belirlemek ve onlara öncelik vermek büyük önem taşır. Değerli müşteriler, şirket için daha yüksek prim geliri ve daha düşük talep oranı gibi avantajlar sunabilir. Büyük veri analitiği, müşterileri segmentlere ayırma, davranışlarını analiz etme ve karlılık potansiyellerini tahmin etme konusunda yardımcı olur. Bu sayede, sigorta şirketleri değerli müşterilere özel teklifler sunabilir, müşteri sadakatini artırabilir ve müşteri yaşam döngüsünü yönetebilir. Ayrıca, çalışmada değersiz müşterilerin belirlenmesinde de büyük veri analitiğinin önemi vurgulanmıştır. Değersiz müşteriler, şirket için yüksek maliyetlere ve düşük gelire neden olabilir. Büyük veri analitiği, bu müşterileri belirleyerek, pazarlama kaynaklarının daha etkin bir şekilde kullanılmasını sağlar. Şirketler, değersiz müşterilere yönelik pazarlama çabalarını azaltabilir veya onları daha karlı hale getirmek için stratejiler geliştirebilir. Sonuç olarak, sigorta sektöründe büyük veri analitiği, müşteri karlılığının hesaplanması ve değerli/değersiz müşterilerin belirlenmesi konularında önemli bir rol oynamaktadır. Rastgele Orman gibi güçlü analitik modeller, sigorta şirketlerine daha doğru kararlar almak ve müşterileriyle daha etkili bir şekilde etkileşimde bulunmak için değerli bir araç sunmaktadır. Bu sayede, sigorta şirketleri müşteri odaklı stratejilerini geliştirerek rekabet avantajı elde edebilir ve sürdürülebilir büyüme sağlayabilir (Fang vd., 2016).

Mebeki Haile Kabeta'nın yüksek lisans tez çalışması, tıbbi sigorta müşterilerini yüksek talep oranına göre sınıflandırarak kar sağlayan müşterilerin korunması üzerine odaklanmıştır. Tıbbi sigortadan kaynaklanan yüksek maliyetler, sigorta şirketleri için önemli bir sorun oluşturur. Bu nedenle, sigorta şirketleri maliyetlerini düşürmek ve karlılığı artırmak için kar sağlayan müşterileri belirlemek ve zarar ettiren müşterileri ayırtmak amacıyla bir model geliştirmeyi hedeflemiştir. Çalışma kapsamında, Destek Vektör Makinası (Support Vector Machine - SVM), Naïve Bayes ve Lojistik Regresyon gibi tahmin modelleri kullanılmıştır. Bu modeller, tıbbi sigorta müşterilerini farklı sınıflara ayırmak ve müşterilerin karlılık potansiyellerini tahmin etmek için kullanılan algoritmalar olarak bilinir. Bu modeller, müşterilerin geçmiş talepleri, maliyetleri, tıbbi durumları ve diğer demografik bilgileri gibi verileri kullanarak kar sağlayan ve zarar ettiren müşterileri belirlemeye yardımcı olur. Çalışmanın sonuçlarına göre, tıbbi sigorta müşterilerinin sınıflandırılmasında en iyi sonuçları Destek Vektör Makinası (SVM) modeli vermiştir. SVM modeli, müşterileri karlılık potansiyellerine göre sınıflandırmada yüksek doğruluk oranları sağlamıştır. Bu sonuçlar, sigorta şirketlerine, müşterilerini daha iyi anlama ve karlılık potansiyellerine göre pazarlama stratejilerini belirleme konusunda önemli bir rehberlik sağlamaktadır. Bu çalışma, tıbbi sigorta şirketlerinin maliyetleri yönetme ve karlılıklarını artırma konusundaki çabalarına ışık tutmaktadır. Modelleme teknikleri ve algoritmaları kullanarak, sigorta şirketleri, kaynaklarını doğru şekilde tahsis edebilir ve kar sağlayan müşterilere odaklanarak finansal performanslarını iyileştirebilir. Ayrıca, müşterilerin korunması ve memnuniyetlerinin sağlanması da önemli bir hedef olmaktadır (Haile, 2021).

Nilüfer Dalkılıç ve Merve Esra Gülcemal tarafından hayat dışı sigorta şirketlerinin performanslarının değerlendirilmesi üzerine bir çalışma yapılmıştır. Çalışma kapsamında 38 farklı hayat dışında sigorta hizmeti veren şirketlerin 2019 yılına ait finansal değerleri üzerinden kümeleme çalışması yapılmıştır. Çalışmada kullanılan en önemli parametreler prim, net kar ve zararı, varlıkların toplam değeri, özsermayedir. Şirketlerin mali değerlerine bakılarak kümeleme çalışması yapılmıştır. Çalışma SPSS uygulaması üzerinde yapılmıştır. Çalışmada

mesafe matrisini bulabilmek için Öklidyen mesafesi kullanılmıştır. Kümeleme yöntemi olarak K-Means algoritması kullanılmıştır (Dalkılıç ve Gülcemal, 2022).

Buket Doğan, Ali Buldu, Önder Demir ve Bahar Erol Ceren tarafından Türkiye’de bulunan özel bir sigortacılık şirketi verileri kullanarak sigortacılık sektöründe müşterilerin kümelenmesi üzerine bir çalışma yapılmıştır. Müşterilere ait veriler bir araya getirilmiş, müşterilerin karakteristik özelliklerini tanımlayabilmek için Veri Madenciliği teknikleri kullanılarak analiz çalışmaları yapılmıştır. Analiz çalışmaları tamamlandıktan sonra K-Means kümeleme algoritması yöntemiyle müşteriler kümelerine ayrılmıştır. Elde edilen bu çıktılar yardımıyla benzer müşterilerin tespiti kolaylaşmıştır. Müşterileri belirli kümelerine ayırdıktan sonra bu kümelerine göre hizmet vermek, taleplerini karşılamak müşteri memnuniyetinin artması için son derece önemlidir. Firmalara müşterilerini öğrenip, karlı ve karsız müşterileri ayırtılabildikten sonra onlara bekledikleri yönde pazarlama hizmeti verebilmektedir. Bu çalışma kapsamında da her bir kümedeki müşterilere uygun olarak yeni pazarlama yöntemlerinin geliştirilmesi hedeflenmiştir (Doğan vd., 2017).

Wafa Qadadeh ve Sherief Abdallah tarafından bir sigorta şirketinin verileri kullanarak müşteri segmentasyonu üzerine çalışma yapılmıştır. Çalışmanın amacı karavan sigortası yaptırabilecek müşterileri tahminlemektir. Çalışmada ana hedef CRM verisi üzerinde veri madenciliği tekniği kullanılarak müşteri verilerini analiz etmek ve anlamlı kümeler oluşturmaktır. İlk yöntem olarak K-Means algoritması kullanılmıştır. Demografik ve davranışsal veriler kullanarak yapılan bu çalışmada müşteriler anlamak ve çapraz satış kampanyaları düzenlemek için faydalı olmuştur. İkinci yöntem olarak Self Organizing Map (SOM) algoritması uygulanmıştır. SOM algoritması K-Means algoritmasına göre çok daha iyi sonuç verdiği gözlemlenmiştir. Elde edilen bu bilgiler doğrultusunda müşterilere yeni pazarlama stratejileri uygulanması hedeflenmiştir (Qadadeh ve Abdallah, 2018).

Saeedodin Ghoreyshi ve Javad Hosseinkhani tarafından bir sigorta şirketinde yapılan çalışma, sağlık poliçeleri üzerinde veri madenciliği ve kümeleme analizi

üzerine odaklanmıştır. Bu çalışma, sigorta şirketlerinin farklı müşteri gruplarının özelliklerini kullanarak müşterilere özel hizmetler sunabilmesi için veri kümelerinin analizini yapmayı amaçlamaktadır. Bu çalışmada, sağlık sigortasına sahip müşteriler üzerinde K-Means kümeleme algoritması kullanılarak müşteriler kümelerine ayrılmıştır. Bu algoritma, müşterilerin benzer özelliklere sahip oldukları kümelerle atanmalarını sağlamaktadır. Elde edilen sonuçlara göre, dört farklı küme oluşmuştur. Her bir küme, sigorta şirketi için farklı değerler sunmaktadır. Bu kümeler, müşterilerin sağlık sigortası ile ilgili farklı ihtiyaçlarını ve davranışlarını yansıtmaktadır. Bu çalışma sayesinde, sigorta şirketi müşterilerinin farklı gruplarını daha iyi anlayarak, müşterilere daha uygun ve kişiselleştirilmiş hizmetler sunma potansiyeline sahip olmaktadır (Ghoreyshi ve Hosseinkhani, 2015).

Irum Matloob, Shoab Ahmad Khan, Farhan Hussain, Wasi Haider Butt ve arkadaşları tarafından sağlık sigortaları paketlerinin makine öğrenmesi algoritmasıyla optimize edilmesi üzerine bir çalışma yapılmıştır. Optimizasyon, müşterilerin sağlık verilerine göre sağlık sigortası paketlerinin tasarlanması üzerinedir. Geleneksel metodlarla sigorta paketlerinin ve ücretlerinin hesaplanması zaman almakta ve maliyeti yüksek olmaktadır. Sigorta paketlerinin içeriği müşterilerin ihtiyaçlarına göre bir paketten başka pakete geçiş yapmasıyla değişiklik gösterebilir. Önerilen yöntemle, tanımlanan parametrelere dayalı olarak optimum sigorta paketleri tanımlanabilmektedir. Çalışma kapsamında ilk olarak tıbbi kayıtlar üzerinden K-Means algoritması kullanılarak kümeleme çalışması yapılmıştır. Daha sonra, bu paketlerden tıbbi fayda optimizasyonu, beş yıllık sigorta kayıtlarına olasılık dağılımı modeli uygulanarak elde edilmiştir. Bu yöntem, gerçek bir şirketin verilerine uygulandığında %25 oranında bir iyileşme olduğunu gözlemlenmiştir (Matloob vd., 2021) .

Shuzlina Abdul-Rahman, Nurin Faiqah Kamal Arifin, Mastura Hanafiah ve Sofianita Mutalib tarafından sigorta sektöründe segmentasyon ve sınıflandırma çalışması yapılmıştır. Müşteri segmentasyonu, müşterileri daha iyi anlama, kaliteli hizmet verebilmek ve müşteri ilişkileri yöntemlerini geliştirebilmek adına bir çok firma için önemlidir. Çalışma kapsamında K-Means kümeleme algoritması

ve Karar Ağacı sınıflandırma algoritması kullanılmıştır. Kümeleme çalışmasında, müşteriler potansiyellerine göre ilgisiz, düşük değerli ve yüksek değerli müşteriler olarak 3 kümeye ayrıldı. Sınıflandırma çalışmasında Karar Ağacı algortimasında Gini modeli uygulandı ve çapraz doğrulama sonucu %81,30 ile en uygun model olduğu gözlemlendi. Çalışmadan elde edilen veriler, pazarlama ekiplerinin pazarlama planlarının belirlenmesinde önemli bir rol oynayacaktır (Rahman vd, 2021).



3. MATERYAL VE METHOD

3.1 Veri Setinin Hazırlanması

Çalışma kapsamında özel bir sigortacılık şirketinin verilerinden yararlanılmıştır. Veriler 2013 - 2023 yılları arasında sağlık branşında yer alan Tamamlayıcı Sağlık Sigortası (TSS) ve Özel Sağlık Sigortası (ÖSS) poliçeleri kullanılarak hazırlanmıştır. Tek kişi için yapılan poliçeler için veriler sigortalı seviyesinde, diğerleri poliçe seviyesinde alınabilmektedir.

Bazı veriler sigortalı seviyesinde, bazıları police seviyesinde alınabilmektedir. Sigortalı seviyesinde alınabilen KPI (Key Performance Indicator / Temel Performans göstergesi) lar;

- Vergi Öncesi Prim (VOP)
- CRK : Kazanılmamış prim karşılığı ifade eder.
- Ertelenen Komisyon: Acentelere ödenen komisyonun ertelenen tutarlarını ifade eder.
- Muallak Hasar: Ödenmemiş hasar tutarını ifade eder.
- Ödenen Hasar: Ödenmiş olan hasar tutarını ifade eder.
- Brüt Hasar: Vergiler dahil ödenen hasarı ifade eder.
- Brüt Hasar Yaslı: Geçmişte ödenen vergiler dahil hasar tutarının tüfe oranınca günümüzde değerini ifade eder.
- Kazanılmış Prim : Vergi Öncesi Prim – CRK
- Pozitif VOP (PVOP) : Pozitif olan vergi öncesi primler
- Acente Komisyonu: Acente kanalından yapılan üretim için ödenen tutardır.
- Prim Yaslı: Geçmişte ödenen primlerin tüfe oranınca günümüzde değerini ifade eder.

Aşağıdaki KPI lar poliçe seviyesinde alınmaktadır. Bu KPI lar bazı formül ile sigortalı seviyesine kırılmaktadır;

Formül 1;

Poliçenin kazanılmış prim rakamı sıfıra eşit veya küçükse sigortalının poliçede ki PVOP oranı alınır değilse sigortalının poliçede ki kazanılmış prim oranı alınır. Bulunan oran KPI tutarıyla çarpılarak bir sigortalının ilgili KPI için tutarı bulunur. KPI lar;

- Arydovk: Arydovk bir uygulamanın ismidir. Bu uygulama üzerinden acentelere ödenen ekstra komisyonlar kesilmektedir. Buradaki tutarlar Arydovk olarak ifade edilmektedir.
- Aspara: Aspara bir uygulamanın adıdır. Bu uygulama üzerinden kampanyalar düzenlenmekte ve kampanya sonucunda acenteler yaptıkları poliçeler üzerinden puan kazanmaktadır. Kazandıkları her 2 puan 1 TL ye denk gelmektedir. Buradaki tutar Aspara yı ifade etmektedir.
- Ertelenen MSU Teşvik Prim: Maksimum sigorta uzmanlarına ödenmesi gereken olan ödül tutarlarının ertelenen kısmını ifade etmektedir.
- İşbank Mensup Ödül: İş bankası tarafından hazırlanıp, üretimi iş bankası personeli aracılığı ile yapılan poliçelere ait ödenen ödül tutarını ifade etmektedir.
- Kredi Kartı Diğer Teknik: Kredi kartı ile gerçekleştirilmiş ödemeler için, ilgili bankaya ödenen komisyonu ifade etmektedir.
- Kredi Kartı Bonus Gideri: Banka ile yapılan anlaşma gereği kredi kartı tahsilatına ilişkin banka tarafından müşteriye verilen bonus tutarıdır.
- Medline Gideri: Medline firmasına ödenen sağlık gereksinimine istinaden ödenen asistans hizmeti bedellerini temsil etmektedir.
- MSU Teşvik Primi: Müşteri satış uzmanlarına teşvik için verilen tutarı ifade eder.
- FGODULMD: Ödül programı ile ödenen ek komisyonu ifade eder.
- UC XX: Ayrıca ödenen bir ek komisyon tipidir.
- Komisyon Son: Acenteye ödenen ek komisyonu ifade eder.
- MSU Payı: Maksimum Sigorta Uzmanları üzerinden gerçekleştirilen üretim için ödenen ödül tutarlarıdır.
- ICM Komisyonu: Acentelere belirli dönemlerle yapılan kampanyalar için ödenen ödül komisyonlarını ifade eder.

- Oper Faaliyet Gideri: Poliçelendirme sürecinde yapılan operasyonel maaliyet tutarını ifade eder.
- Bölüşmeli Prim Repayı: Reasürans anlaşması gereği primin reasürör firmaya devredilen kısmıdır.
- Rekomisyon: Reasürans anlaşması gereği reasürör firmadan alınan komisyonudur.

Formül 2;

Sigortalının brüt hasar tutarının poliçenin brüt hasar tutarına oranıyla ilgili KPI ların değeri çarpılarak sigortalının değeri bulunur. KPI lar;

- Hasarın Repayı: Ödenen hasarın reasürans payını ifade eder.
- Hasarın Repayı Yaşlı: Geçmişte ödenen hasarın reasürans pay tutarının tüfe oranınca çarpılarak günümüzdeki tutarını gösterir.

Formül 3;

Sigortalının muallak hasar tutarının poliçenin muallak hasar tutarına oranıyla ilgili Muallak Repayı (Rizikonun gerçekleşmesi sonrası henüz fiili olarak ödenmemiş hasar tutarının reasüransa devredilecek olan karşılığıdır) ın değeri çarpılarak sigortalının değeri bulunur.

Formül 4;

Aşağıdaki KPI lar sigortalının IBNR Oranıyla çarpılarak sigortalının ilgili KPI değeri bulunur. KPI lar;

- IBNR Değeri: Branşlar itibariyle meydana gelmiş, ancak bu hasarların varlığı konusunda sigorta şirketinin henüz bilgisinin bulunmadığı hasarların tutarını ifade eder.
- IBNR Repay Değeri: Branşlar itibariyle meydana gelmiş, ancak bu hasarların varlığı konusunda sigorta şirketinin henüz bilgisinin bulunmadığı hasarların, reasüransa devredilecek olan kısmıdır.

Formül 5;

Aşağıdaki KPI lar sigortalının CRK/VOP Oranıyla çarpılarak sigortalının ilgili KPI değeri bulunur. KPI lar;

- CRK Repayı: Reasüransa devredilen primin kazanılmamış kısmını ifade eder.
- Ertelenen ICM Komisyonu: Acentelere ödenen ve poliçe bazında üretim ağırlığına dağıtımı gerçekleştirilen ICM Ödül Komisyon tutarlarının esasınca ertelenen kısımlarını ifade eder.
- Ertelenen Rekomisyon: Reasürans komisyonunun kazanılmamış kısmını ifade eder.

Formül 6;

Reprim = VOP – Bölüşmeli Prim Repayı – CRK + CRK Repayı

Yukarıda belirtilen KPI lar kullanılarak aşağıdaki KPI ları hesaplanmaktadır ;

- TKZ: Teknik kar zarar tutarını ifade eder.
- TKZ Yaşlı: Geçmiş dönemlere ait rakamların tüfe oranınca günümüzdeki değerinin bulunduktan sonra teknik kar zarar tutarını ifade eder.
- TKZ FA: Reasürans etkisi olmaksızın hesaplanan faaliyet gideri etkili teknik karlılık bilgisidir.
- TKZ FA Yaşlı: Geçmiş TKZ FA değerinin tüfe oranınca çarpılarak günümüzdeki rakamını gösterir.

Verilerin hazırlanması sürecinde IBM DataStage ürünü kullanılmıştır. Verilerin temizlenmesi, poliçe seviyesinde olan verilerin sigortalı seviyesine kırılması, nihai karlılık rakamının hesaplanması gibi tüm işlemler bu ürün üzerinde modellenmiştir. Bir sigortalının geçmiş yıldaki tutarları tüfe oranlarına yaşlandırılarak günümüzdeki tutarsal değerleri bulunmuştur. Bir poliçenin başlangıç yılı ve ayındaki tüfe değeri bulunur. Bulunan değer poliçelinin ilgili kpi değeri ile çarpılarak günümüzdeki değeri bulunur. Veri setimizde yaşlı ifadesini içeren KPI lar bu mantıkta hesaplanmıştır. Aşağıdaki Çizelge 3.1 de 2022 yılına ait tüfe oranları gösterilmiştir.

Çizelge 3.1 Tüfe oranları

Yıl	Ay	Tüfe Oranı
2022	1	1,214842863
2022	2	1,198200171
2022	3	1,183270335
2022	4	1,168102256
2022	5	1,152002066
2022	6	1,139425481
2022	7	1,125168099
2022	8	1,109711129
2022	9	1,094315998
2022	10	1,079886973
2022	11	1,067377751
2022	12	1,054823246

Uygulanan tüm formüller sonucunda bir sigortalının ne kadar kar getirdiğinin tutarı hesaplanmıştır. Çizelge 3.2 bir sigortalıya ait KPI değerleri paylaşılmıştır.

Çizelge 3.2 Örnek KPI tutarları

KPI	Tutar
BRUT_HASAR	25337,45
ODENEN_HASAR	25337,45
MUALLAK_HASAR	0
UC_XX	0
ARYDOVK	0
KOMISYON_SON	0
MSU_PAYI	0
MSU_TESVIK_PRIMI	0
ISBANK_MENSUP_ODUL	0
OPER_FAALİYET_GIDERİ	716,23
MEDLINE_GIDERİ	3,03
KREDİ_KARTI_DİĞER_TEKNİK	0
KREDİ_KARTI_BONUS_GIDERİ	0
ASPARA	78,10
BOLUSMELİ_PRİM_REPAYI	0
CRK_REPAYI	0
REKOMİSYON	0
ERTELENEN_REKOMİSYON	0
HASARIN_REPAYI	0
MUALLAK_REPAYI	0
CRK	0
ERTELENEN_KOMİSYON	0
ERTELENEN_MSU_TESVIK_PRİM	0
BRUT_HASAR_YASLI	37760,42

ICM_KOMISYONU	0
ERTELENEN_ICM_KOMISYONU	0
ACENTE_KOMISYONU	3473,42
IBNR_REPAYI	0
IBNR_TUTARI	6,62
POZITIF_VOP	23098,10
REPRIM	23098,10
TKZ	-6569,33
TKZ_FA	-6569,33
TKZ_FA_YASLI	-8312,72
TKZ_YASLI	-8312,72
VOP	23098,10
PRIM_YASLI	36236,92
PMRODULMD	52,55
HASARIN_REPAYI_YASLI	0

Analitik modelde kullanılmak üzere sigortalının karlılık verileri üzerinden bir veri seti hazırlanmıştır. Veri setine poliçe başlangıç yılı 2019 ve sonrası olan poliçeler dahil edilmiştir. Bunun sebebi, poliçeler üzerinde yer alan demografik bilgilerin doğruluğu, provizyon bilgilerinde yer alan ICD kodlarının 2019 ve sonrası yıllarda daha güvenilir olmasından kaynaklanmaktadır. Bu konularla ilgili ilerleyen adımlarda tekrar bilgi verilecektir. Bir sigortalının bir veya daha fazla poliçesi olabilir. Sigortalının 2 yıl veya daha fazla süredir poliçesi olabilir. Bu poliçelerin toplam rakamlarını alarak veri setine eklemek 1 yıllık poliçesi olan sigortalının rakamlarıyla arasında çok fark çıkartacaktır. Bu durumdan kaynaklı analitik model sonuçlarımız doğru sonuçlar üretmeyecektir. Bu durumu aşabilmek adına sigortalının poliçeleri üzerinden bir sigortalının ortalama primi, ortalama ödenen hasar tutarı, ortalama muallak hasar tutarı, ortalama tkz (teknik kar zarar) bulunmuştur. Ortalama rakamlar hesaplanırken prim yaşlı, ödenen hasar yaşlı, muallak hasar ve tkz KPI ları üzerinden hesaplanmıştır. Bu sayede 1 yıllık bir sigortalı ile 2 yıl ve daha fazladır aktif olan sigortalının ortalama rakamları yakınlık gösterebilmektedir.

Veri setimiz TSS ve ÖSS kırılımı eklenmiştir. Bunun sebebi, TSS ve ÖSS poliçelerinin fiyatlandırmalarının farklılık göstermesidir. TSS genellikle bir ülkenin kamusal sağlık sigortaları kapsamında yer alan boşlukları ele almaktadır. Bu sigorta türü, kamusal sağlık sisteminin cevap veremediği ya da sistemin finansal desteklere ihtiyaç duyduğu alanları göz önüne alarak sınırlı bir kapsam

sunmaktadır (Orhaner ve Yavaş, 2023). TSS poliçeleri, sigortalarının SGK (Sosyal Güvenlik Kurumu) anlaşmalı özel hastanelerde hiçbir ek ücret ödemeden tedavi hizmetini, muayene masraflarını ve tahlil giderlerini karşılayan sağlık sigortası türüdür. Bu poliçeler SGK ya ek olarak yapıldığı için daha uygun fiyatlara yaptırılmaktadır.

ÖSS, sigorta süresi boyunca poliçe şartlarına göre belirlenen durumlar sonucunda meydana gelen kazalar ve hastalıklarla ilgili tedavi masraflarını kapsayan bir sigorta türüdür. Sigorta şirketi, poliçede belirtilen limit dahilinde sigortalının katılım payı oranında ödeme yapar ve sigortalının muafiyet tutarı dışında kalan kısmı karşılar (Ateş, 2011). ÖSS, beklenmedik durumlarda ortaya çıkan hastalıkların özel hastanelerde uygulanan tedavi sürecinden doğan ücretlerin karşılanması için yaptırılmaktadır. Poliçeler özel hastanelerde geçerli olduğu için daha yüksek fiyatlardan yaptırılmaktadır. Bu sebeplerden dolayı TSS ve ÖSS poliçeleri arasında çok fazla fiyat farkı çıkmakta, sigorta şirketleri tarafından ödenen hasar tutarları farklılık gösterebilmektedir. Analitik modelin daha verimli olabilmesi adına bu adımda TSS ve ÖSS verileri ayrı ayrı kullanılmaktadır.

Bir sigortalının en güncel poliçesi bulunarak demografik bilgileri bu police üzerinden alınmıştır. Çalışma kapsamında kullanılacak olan demografik bilgiler aşağıdaki gibidir;

- Şehir : Sigortalının ikametgah ettiği ili gösterir.
- Yaş
- Cinsiyet
- Doğum Muafiyeti
- Medeni Durum

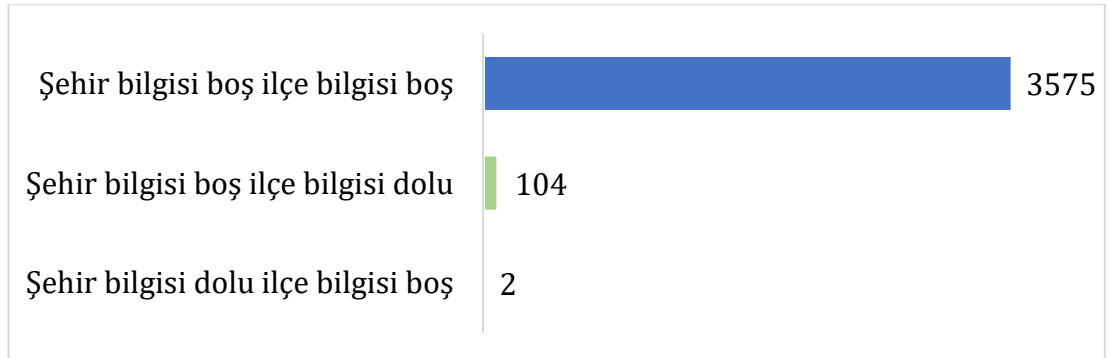
Poliçe üzerinde bulunan VIP, boy, kilo, vücut kitle indeksi, ilçe gibi demografik bilgiler veri kaletisinin kötü olmasından dolayı kullanılmamıştır. Veri setimizde yer alan sigortalıların boy, kilo, vücut kitle indeksi bilgilerinin büyük bir kısmı boş olduğu tespit edilmiştir ve veri setinden bu bilgiler tamamen çıkartılmıştır. Yukarıda belirtilen demografik bilgileri boş olan ve poliçesi aktif olan sigortalılar

tespit edilmiştir. Bilgileri boş olan sigortalılar aktif ve pasif olarak Çizelge 3.2 de gösterilmiştir.

Çizelge 3.3 Demografik bilgileri boş olan sigortalılar

BILGI	AKTİF	HAYIR	TOPLAM
BOY	%88,610	%87,369	%82,969
DOGUM_YERI	%0,276	%0,885	%0,579
ILCE	%0,033	%0,029	%0,029
KILO	%0,086	%0,143	%0,111
MEDENI_DURUM	%9,827	%10,579	%9,666
SEHIR	%1,167	%0,987	%1,007
VIP_MI	%0,000	%0,007	%0,004
VUCUT_KITLE_INDEX	%0,000	%0,002	%0,001

Şehir ve ilçe bilgileri incelendiğinde bazı veri problemleri tespit edilmiştir. Tespit edilen bilgilerin sigortalı adetleri Şekil 3.1 de gösterilmiştir.



Şekil 3.1 Şehir ve ilçe bilgisi boş olan sigortalı sayısı

İlçe bilgisinde çok detay olması sınıflandırma modelini olumsuz etkileyeceği için veri setine dahil edilmedi. Şehir bilgi boş olan kayıtlar veri setinden çıkartıldı.

Medeni durum bilgisinde boş ve bilinmeyen veriler tespit edilmiş ve veri setinden çıkartılmıştır. Medeni durum alanında evli ve bekar bilgisi dışında başka bilgiler

olduğu tespit edilmiştir. Evli ve bekar dışındaki bilgiler bekar olarak kabul edilmiştir. Tespit edilen bu bilgilerin sigortaları adetleri Çizelge 3.3 de gösterilmiştir.

Demografik bilgilerde tespit edilen veri kalitesi problemleri düzeltilmesi için iş birimleriyle paylaşıldı.

Çalışma kapsamında kullanılmak üzere sigortalıların geçmiş provizyonları incelenmiştir. Provizyonlar üzerinde muane oldukları doktorlar tarafından hastalık tanıları girilmektedir. Bu tanıların uluslar arası sağlık sınıflandırılması bulunmaktadır. Bu tanıların kod karşılığına ICD-10 [8] kodu (International Statistical Classification of Diseases and Related Health Problems) denmektedir. ICD kodları , 22 ana sağlık sorunu başlığı altında toplanmaktadır. Sigortalıların geçmiş provizyon adetleri 22 başlık altında işaretlendi. Çizelge 3.4 te 22 başlıktaki toplam provizyon adetleri gösterilmiştir.

Çizelge 3.4 Hastalık grupları bazında provizyon sayıları

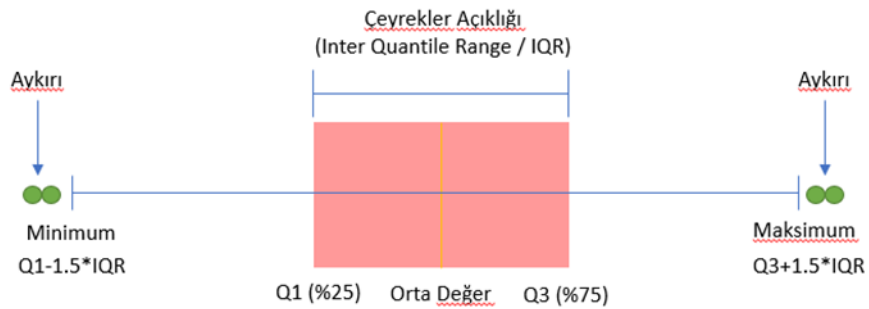
Hastalık Grup Adı	Provizyon Dağılımı
Solunum sistemi hastalıkları	%19,11
Semptomlar, belirtiler ve anormal klinik ve laboratuvar bulguları, başka yerde sınıflandırılmamış	%16,61
Endokrin, beslenme ve metabolik hastalıklar	%11,92
Genitoüriner sistem hastalıkları	%9,55
Kas-iskelet sistemi ve bağ dokusu hastalıkları	%8,30
Sindirim sistemi hastalıkları	%8,04
Kan ve kan oluşturan organların hastalıkları ve bağışıklık mekanizmasını ilgilendiren bazı bozukluklar	%5,80
Göz ve adneks hastalıkları	%3,95
Deri ve deri altı doku hastalıkları	%3,25
Bazı bulaşıcı ve paraziter hastalıklar	%2,97
Dolaşım sistemi hastalıkları	%2,42
Yaralanma, zehirlenme ve dış nedenlerin diğer bazı sonuçları	%2,27

Kulak hastalıkları ve mastoid süreç	%1,45
Hamilelik, doğum ve lohusalık	%1,32
Sinir sistemi hastalıkları	%1,31
neoplazmalar	%1,12
Morbidite ve mortalitenin dış nedenleri	%0,24
Zihinsel ve davranışsal bozukluklar	%0,20
Konjenital malformasyonlar, deformasyonlar ve kromozomal anormallikler	%0,16
Perinatal dönemden kaynaklanan bazı durumlar	%0,02

Sigortalıların karlılık verileri, demografik bilgileri ve hastalık bazında provizyon sayıları birleştirilerek analitik modelde kullanılacak nihai dataset oluşturuldu.

3.2 Aykırı Değerlerin Bulunması

Aykırı değerler, veri kümelerinin her iki ucunda bulunan çok düşük yada çok yüksek değerlerdir. Aykırı değerlerin veri setine dahil edilmesi, analitik modelinizin sonucunu olumsuz etkileyebilir veya size yanıltabilir. Bu sebepten dolayı aykırı değerlerin bulunması ve veri setinden çıkartılması gerekmektedir. Aykırı değerlerin bulunması için Çeyrekler Açıklığı (Inter Quantile Range / IQR) yöntemi kullanılmaktadır (Singhg, 2020). Çeyrekler açıklığı sıralanmış bir veri dizisinin orta yarısını (%50sini) kapsayan ve üçüncü dörttebirlik ve birini dörttebirlik aralığı veya farkını (yani $Q3-Q1$) gösteren bir istatistiksel yayılma ölçüsüdür(Wikipedia, 2023).



Şekil 3.2 Aykırı değer hesaplama (Singhg, 2020)

İlk olarak veri setinin orta değeri bulunur. Veri seti orta değer altında ve üstünde olarak ikiye ayrılır. Orta değer altında kalan veri setinin kendi içerisindeki orta değeri bulunur değere Q1 adı verilir. Aynı şekilde orta değer üzerindeki veri setinin de kendi içerisinde orta değeri bulunur ve bu değere Q3 denir. Şekil 3.2 de yer alan değer aralıklarının nasıl bulunduğunun formülü aşağıda belirtilmiştir (Thakur);

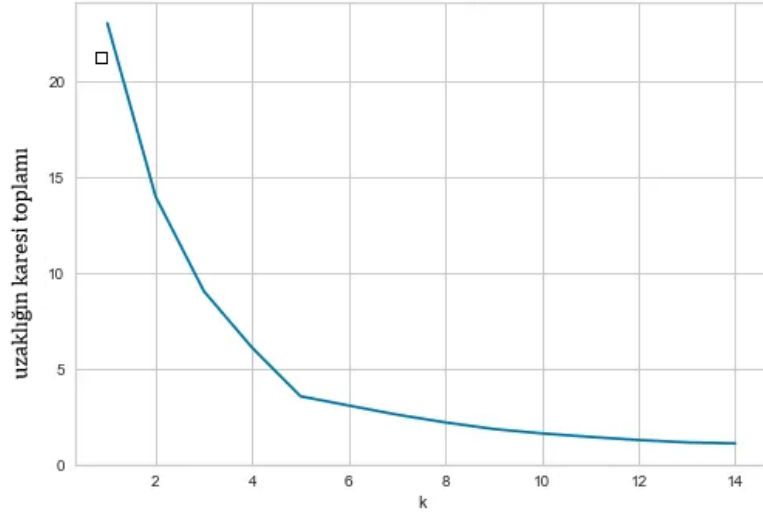
$$\text{Çeyrek Aralığı (IQR)} = Q3 - Q1$$

$$\text{Çok Düşük Aykırı Değer} = Q1 - (1.5 * IQR)$$

$$\text{Çok Yüksek Aykırı Değer} = Q3 - (1.5 * IQR)$$

3.3 Dirsek Yöntemi (Elbow Method)

Kümeleme algoritmalarının temel mantığı birbirine benzeyen verilerin aynı kümede olsun, birbirine benzemeyen verilerin birbirinden uzak kümelerde olmasıdır. Bu mantık doğrultusunda verilerin kaç kümeye ayrılması gerektiğini belirlemek için Dirsek Yöntemi (Elbow Method) uygulanır. Dirsek Yöntemi sonucunda bir metrik değeri oluşur. Bu metrik değerine kareler toplamı (Within Clusters Sum of Square / WCSS) denir. Çalışma mantığı; veri setinin bölüneceği küme sayısı belirlenir. Veri setindeki her bir nokta bir kümeye düşer. Herbir noktanın düşmüş olduğu kümenin merkezine olan uzaklığını karelerinin toplamı bulunur ve toplanır. Tüm kümelerin toplamı ne kadar düşükse kümeleme o kadar iyi çalışmış demektir (Bholowalia ve Kumar, 2014). Örnek bir WCCS grafiği Şekil 3.3 te gösterilmiştir;

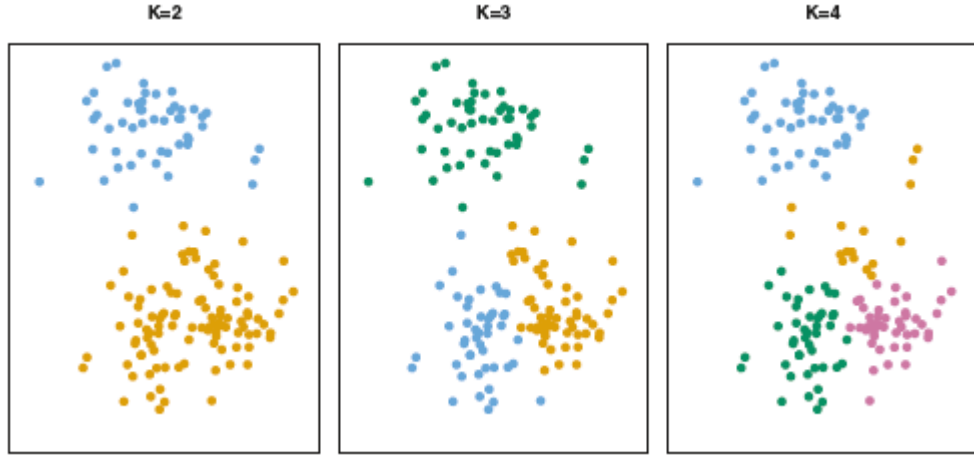


Şekil 3.3 Küme sayısı (Tusyakdiah, 2020)

3.4 K-Means Algoritması

K-means algoritması, n adet veri nesnesinden oluşan bir veri setini, k adet giriş parametresi sayısına göre kümeler halinde ayırmak için kullanılır. Bu algoritmanın temel amacı, bölümlleme işlemi sonucunda elde edilen kümelerin içindeki benzerliklerin maksimum düzeyde olmasını ve kümeler arasındaki benzerliklerin minimum düzeyde olmasını sağlamaktır (Erdoğan ve Durdağ, 2016. Genowesse, 2023).

K-Means algoritmasındaki en önemli noktalardan birtanesi k değerinin belirlenmesi yani model çıktısında verimizin kaç adet kümeye bölüneceğinin belirlenmesidir. Şekil 3.4 de k değerinin değişimine göre kümelerin nasıl değiştiğine bir örnek verilmiştir.

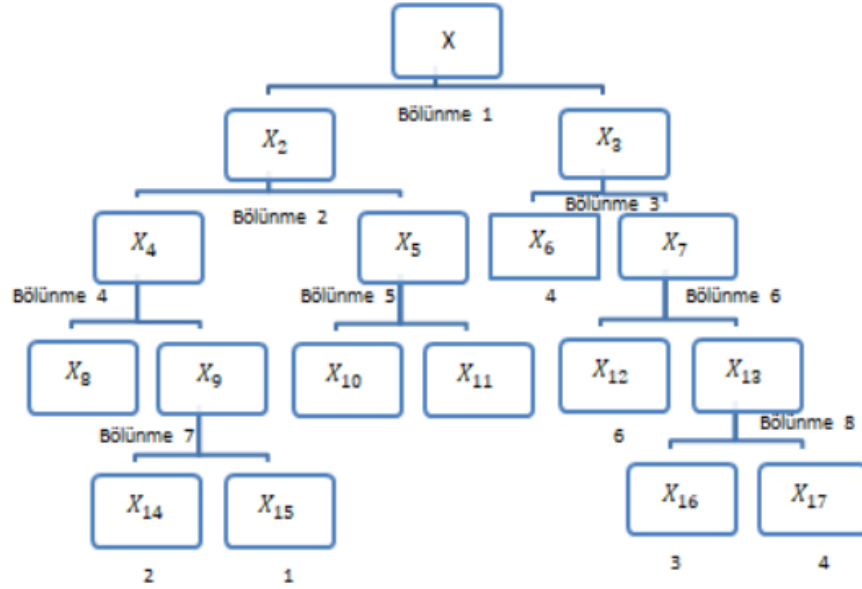


Şekil 3.4 K Değerine göre kümeleme örneği (Mesci, 2020)

3.5 Rastgele Orman (Random Forest) Algoritması

Rastgele Orman algoritması, hem sınıflandırma hem de regresyon işlemleri için çok kullanılan bir algoritmadır. Algoritma, her bir eğitim setinin bir alt kümesini ve özelliklerin bir alt kümesi üzerinde eğitilen bir karar ağacı ormanı olarak çalışır. Eğitim sürecinde, ormandaki her bir ağaç, budama olmaksızın maksimum derinliğe kadar büyütülür. Son tahmin ormandaki tüm ağaçların çoğunluk oyu alınarak yapılır. Rastgele Orman algoritması hem küçük hem de büyük veri setlerinde iyi bir doğruluk verir. Algoritma, karar ağaçlarının eğitim kümesine aşırı uyum sağlama eğilimini de düzeltir (Wessel, 2023).

İkili ağaç tabanlı sınıflandırıcılar veya daha doğru bir ifadeyle karar ağaçları, bir XX kümesinin alt kümelere bölünerek tekrar tekrar inşa edildiği bir yapıya sahiptir (Loh, 2011). Bu süreç altı sınıflı bir ağaç için Şekil 1’de gösterilmiştir.

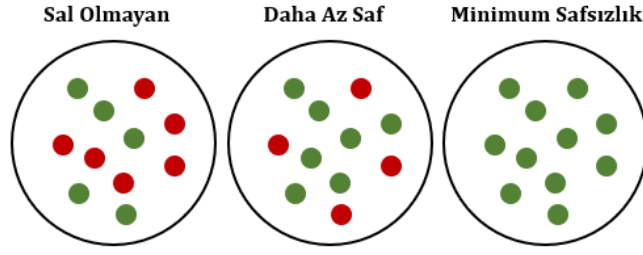


Şekil 3.5 Ağaç yapısı (Kormaz vd., 2018)

3.6 Bilgi Kazancı (Gain Information)

Bilgi Kazancı, bir hedef değişkenini öngörmeye bir veri kümesinde belirli bir özellik tarafından sağlanan bilgi miktarını ölçmek için makine öğrenimi ve veri madenciliği alanında kullanılan bir ölçüdür. Bilgi Kazancı, hedef değişken açısından mümkün olan en yüksek homojenliğe sahip alt kümeler oluşturmak için veri kümesini bölmeyi amaçlayan karar ağacı algoritmalarında sıkça kullanılır. Algoritma, durma kriteri karşılanana kadar, en yüksek Bilgi Kazancı'na sahip özelliği veri kümesini bölmek için yinelenen şekilde seçer (Wikipedia, 2022).

Karar ağaçlarını oluşturulurken entropi ve bilgi kazanımından faydalanılır. Entropi, bir dizi örnekteki safsızlık, düzensizlik veya belirsizliğin ölçüsüdür. Entropi, bir Karar Ağacının verileri bölmeye nasıl karar verdiğini kontrol eder. Şekil 3.6 da her bir setin safsızlık seviyesi gösterilmektedir (Gopalan, 2020).



Şekil 3.6 Veri setinde entropi (Gopalan, 2020)

Entropi formülü (Amelia, 2019) aşağıda belirtilmiştir.

$$\text{Entropy} = - \sum_{i=1}^n \log_2 (P_i)$$

n = özellik sayısı

i = özellik

P = i olasılığı

Bilgi kazancı hesaplamasının detaylı öğrenimi için Azika Amelia tarafından yapılan örneği inceleyebilirsiniz (Amelia, 2019).

3.7 Bilgi Kazancı Oranı (Gain Information Ratio - GIR)

Bilgi Kazanı Oranı, özellik seçiminde kullanılan bir karar ağacı algoritması ölçüsü olan Bilgi Kazancı ölçüsünün bir varyantıdır. Bilgi Kazancı Oranı ölçüsü, özellikteki kategorilerin sayısını ve her kategorideki örneklerin sayısını dikkate alarak Bilgi Kazancı'nu ayarlar. Bu, eğitim verilerinde iyi performans gösteren ancak yeni verilerde kötü performans gösteren aşırı uyum (overfitting) sorununu önlemeye yardımcı olur. Bilgi Kazancı Oranı formülü ;

$$\text{Bilgi Kazancı Oranı} = (\text{Bilgi Kazancı}) / (\text{Bölme Bilgisi})$$

Burada Bölme Bilgisi, bir özellikteki örneklerin dağılımını temsil etmek için gereken bilgi miktarının bir ölçüsüdür. Bilgi Kazancı Oranı değerleri 1'e yakın olanlar sınıflandırma için daha faydalı bir özellik gösterir (Wikipedia, 2022).

3.8 Gini İndeksi (Gini Index)

Karar ağaçlarında, Gini indeksi sonuçtaki alt kümelerin saf olmayanlığını ölçerek bir bölmenin kalitesini değerlendirmek için kullanılır. Gini indeksi 0 ile 1 arasında değişir, 0 tamamen homojen bir dağılımı gösterirken, 1 tamamen heterojen bir dağılımı gösterir. Hedef, mümkün olan en homojen alt kümeleri sağlayan bölümleri seçerek Gini indeksini minimize etmektir.

Gini indeksi, hedef değişkeni tahmin etmek için en bilgilendirici özellikleri belirlemek için kullanılabilen yararlı bir araçtır. Ayrıca, karar ağaçlarının birleştirilerek tahmin doğruluğunu artırdığı topluluk yöntemleri gibi rastgele ormanlarda da yaygın olarak kullanılır (Tyagi, 2020). Şekil 3.9 da Gini İndeks formülü (Patel, 2021) gösterilmiştir.

$$GI = 1 - \sum_{i=1}^n (P)^2$$

3.9 Gradyan Yükseltme (Gradient Boosting) Algoritması

Algoritma, eğitim verileri üzerinde karar ağaçları gibi bir dizi zayıf modeli eğiterek, ardından bunları birleştirerek tek bir güçlü model oluşturarak çalışır. Dizideki her model, önceki modelin artıkları üzerinde eğitilir, bu da önceki modellerin kötü performans gösterdiği alanlara odaklanmasına olanak tanır. Esnektir ve geniş bir sorun yelpazesinde kullanılabilir, eksik verileri ve aykırı değerleri iyi işleyebilir ve genellikle hem eğitim hem de test verilerinde yüksek doğruluğa sahiptir. Ancak, eğitimi yavaş olabilir ve büyük miktarda bellek gerektirebilir (Wikipedia, 2023). Gradyan Yükseltme algoritmasının çalışma mantığı aşağıda gibidir;

3.10 Aşırı Gradyan Yükseltme (Extrem Gradient Boosting) Algoritması

Extrem Gradient Boosting (XGBoost), geleneksel gradient boosting algoritmasına ek olarak daha hızlı, daha doğru ve daha ölçeklenebilir hale getiren birkaç yeni özellik ekleyerek inşa edilmiştir. XGBoost'un bazı temel özellikleri şunları içerir:

- Aşırı öğrenmeyi önlemek için kullanılan karar ağaçlarını budamak için bir yöntem içerir.
- Hiperparametreleri optimize etmeye ve aşırı öğrenmeyi önlemeye yardımcı olmak için dahili çapraz doğrulama desteği sağlar.
- Makinedeki tüm çekirdekleri kullanarak çoklu işlem yapabilir ve bu, eğitim sürecini büyük ölçüde hızlandırır.

GBoost hesaplama açısından maliyetli olabilir ve büyük miktarda bellek gerektirebilir (Wikipedia, 2023).

XGBoost, Gradyan Yükseltme algoritmasının aksine benzerlik puanı ve kazancının en iyi düğüm bölünmelerinin belirlendiği kendi ağaç oluşturma yöntemini kullanır. XGBoost algoritmasının formülü gösterilmiştir (Dobilas,2020).

$$\text{Benzerlik Skoru} = \frac{(\sum_{i=1}^n \text{kalan}_i)^2}{\sum_{i=1}^n [\text{önceki olasılık}_i * (1 - \text{önceki olasılık}_i)] + \lambda}$$

Kalan : Gerçek Değer – Tahminlenen Değer

Önceki Olasılık : Bir önceki adımda hesaplanan bir olayın olasılığıdır. İlk ağacı oluşturmak için kullanılan her gözlem için başlangıç olasılığının 0,5 olduğu varsayılır. Sonraki ağaçlar için önceki olasılık, süreç haritasında gösterildiği gibi ilk tahmine ve önceki tüm ağaçlardan gelen tahminlere dayalı olarak yeniden hesaplanır.

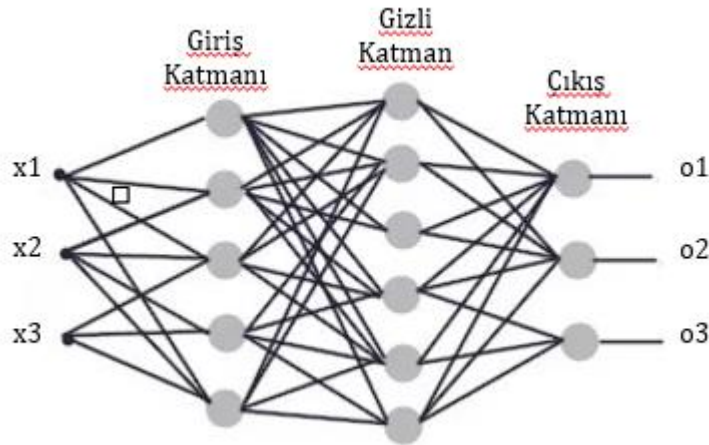
λ : Bir düzenleme parametresidir. Artan lambda orantısız bir şekilde küçük yaprakların (birkaç gözleme sahip olanlar) etkisini azaltırken, daha büyük

yapraklar (birçok gözleme sahip olanlar) üzerinde yalnızca küçük bir etkiye sahip olur.

3.11 Çok Katmanlı Algılayıcı (Multilayer Perceptron – MLP)

MLP, girdi katmanı, bir veya daha fazla gizli katman ve bir çıkış katmanına organize edilmiş, birbirine bağlı düğümlerden veya nöronlardan oluşan yapay sinir ağlarından bir türdür. Ağdaki her nöron, önceki katmandaki nöronlardan girdi alır, bir aktivasyon fonksiyonunu kullanarak girdiyi işler ve çıktıyı bir sonraki katmandaki nöronlara aktarır. MLP'lerin öğrenme süreci, geri yayılım gibi bir eğitim algoritması kullanarak nöronların ağırlıklarını ve sapmalarını ayarlayarak tahmin edilen çıktı ile istenen çıktı arasındaki farkı en aza indirmeye işlemidir (Niceguyede, 2023).

MLP özellikle sınıflandırma ve genelleme yapma durumlarında etkin çalışır. Çok Katmanlı Ağ'ların yapısı Şekil 3.12 de gösterilmiştir.



Şekil 3.7 Çok katmanlı ağ yapısı (Işıkhani, 2020)

3.12 Knime

Knime (Konstanz Information Miner), verilerin uçta uca işlenebildiği, görsel olarak kolay bir şekilde birleştirebildiği, açık kod kaynaklı bir platformdur (Berthold vd., 2009). KNIME üzerinde makine öğrenimi ve veri madenciliği için

kullanıma hazır araçlar bulunur. Verilerin görselleştirilmesi, veri modeli oluşturulması, verilerin analiz edilmesi gibi temel veri önışleme işlemlerinde kullanılmaktadır (Wikipedia, 2022).

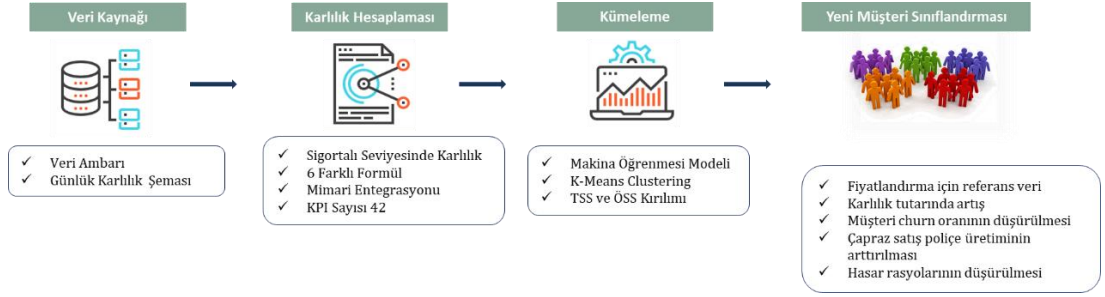
Günümüzde veri biliminin yaygınlaşmasıyla birlikte KNIME kullanımı artmıştır. KNIME üzerinde bulunan hazır algoritmalar sayesinde hızlı bir şekilde analitik çözümler sunulabilmektedir. Bunun yanında veri bilimi sürecini otomatikleştirmesi, birçok işletim sistemi ile entegre olması ve raporlama fırsatı sunması da KNIME'in özelliklerindendir. KNIME da sürükle bırak yöntemi ile çalıştığı için son kullanıcılar için çok kolaylık sağlamaktadır (Alabas, 2020).

3.13 Çalışma Kapsamı ve Akışı

Bu çalışmada özel bir sigorta şirketinin verileri kullanılmıştır. Çalışma kapsamında kullanılan veriler sigorta şirketinin veri ambarından alınmıştır. Veri ambarında yer alan poliçeye ve sigortalılara ait bazı bilgiler ve karlılık verileri alınmıştır. Karlılık hesaplaması adımıında, sigortalı seviyesinde alınamayıp poliçe seviyesinde alınabilen KPI (Key Performance Indicator / Temel Performans göstergesi) lar 6 farklı formül ile sigortalı seviyesine indirgenmiştir. Sigortalı seviyesinde oluşan karlılık rakamları iş birimlerinin analiz edebilmesi için raporlama ortamının mimari tasarımına entegre edilmiştir. Kümeleme adımıında, sigortalıların teknik kar zarar tutarına göre K-Means algoritması kullanılarak kümeleme çalışması yapılmıştır. Kümeleme çalışması özel sağlık sigortası ve tamamlayıcı sağlık sigortaları için ayrı ayrı uygulanmıştır. Yeni gelen müşteriler için referans bir veri oluşturmak için kümeleme çalışmasının sonucu üzerinden sınıflandırma çalışması yapılmıştır. Sınıflandırma çalışmasında 6 farklı model denenmiştir. Denenen modeller aşağıdaki gibidir;

- Rastgele Orman algoritmasının bilgi kazancı özelliği
- Rastgele Orman algoritmasının bilgi kazancı oranı özelliği
- Rastgele Orman algoritmasının indeks özelliği
- Gradyan Yükseltme algoritması
- Aşırı Gradyan Yükseltme algoritması
- Çok Katmanlı Algılayıcı algoritması

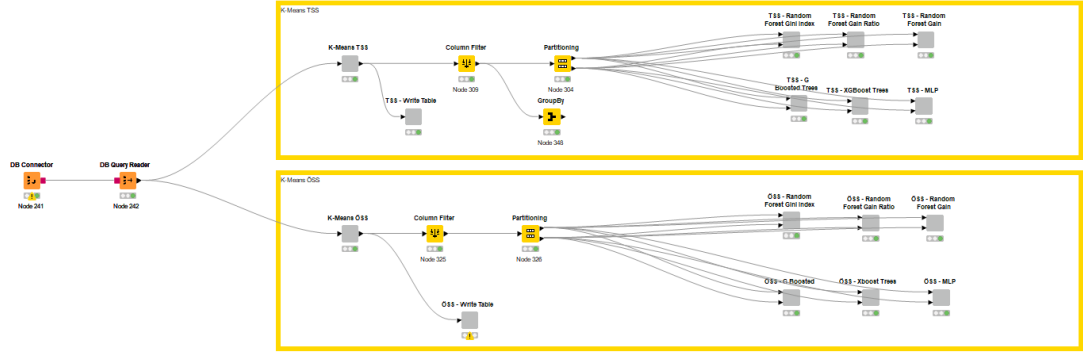
Çalışmada uygulanan adımların akışı Şekil 3.8 de gösterilmiştir.



Şekil 3.8 Çalışma kapsamı akışı

4. ARAŞTIRMA BULGULARI VE TARTIŞMA

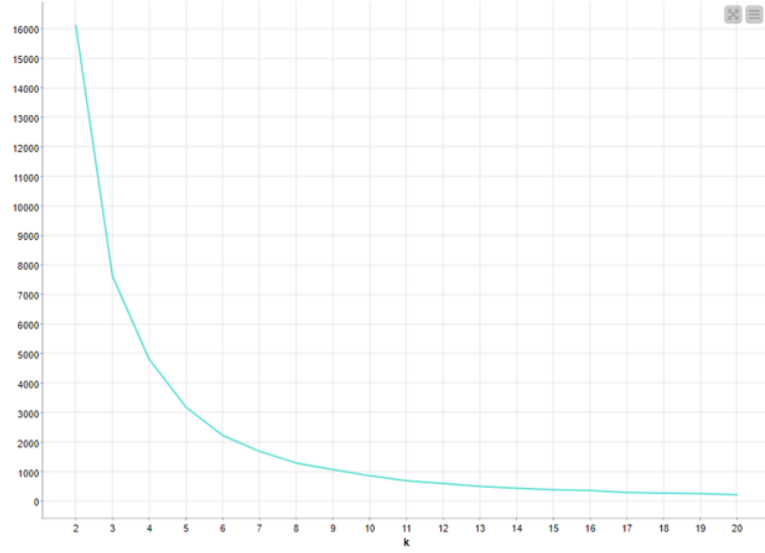
Çalışma kapsamında ulaşmak istediğimiz ana hedefimiz yeni gelen sigortalıların karlı bir sigortalı olup olmayacağını tahminleyebilmektir. Bu sayede sigortalı için indirim uygulanabilir. Zarar edilebilecek bir sigortalıysa fiyat yükselterek zararı minimize edebiliriz. Bu tahminleme işlemini yapabilmek için doğru kurgulanmış bir makina öğrenmesi modeline ihtiyacımız bulunmaktadır. Makine öğrenmesi algoritmaları, bir problemi ilgili veri setine göre modelleyen algoritmaların genel ismidir (Atalay ve Çelik, 2017). Veri seti hazırlandıktan sonra aykırı değerlerin bulunması ve analitik modelin kurgulanması süreci KNIME ürünü üzerinde yapılmıştır. Tasarlanan akış Şekil 4.1 de gösterilmiştir.



Şekil 4.1 KNIME üzerinde tasarlanan akış

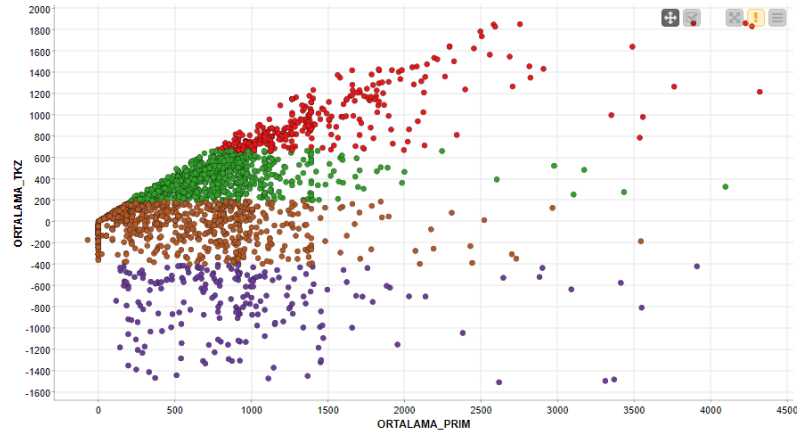
Birinci adım olarak veri setimizde yer alan aykırı değerleri temizleyebilmek için IQR Yönetimi uygulandı. IQR yöntemi bir sigortalının 1 yıldaki ortalama teknik kar zarar rakamına göre uygulanmıştır. Formülde yer alan 1.5 olan k değeri kullanılan veri setinin geniş tutabilmek adına 3 olarak belirlenmiştir. Bu yöntem sonucunda veri setimizin %5 lik kısmının outlier olduğu tespit edilerek veri setinden çıkartıldı.

Kümeleme çalışması yapmadan önce verilerin kaç kümeye ayrılacağına bulabilmek için Elbow testi yapıldı. Elbow testi sonucunda oluşan grafik Şekil 4.2 de gösterilmiştir.

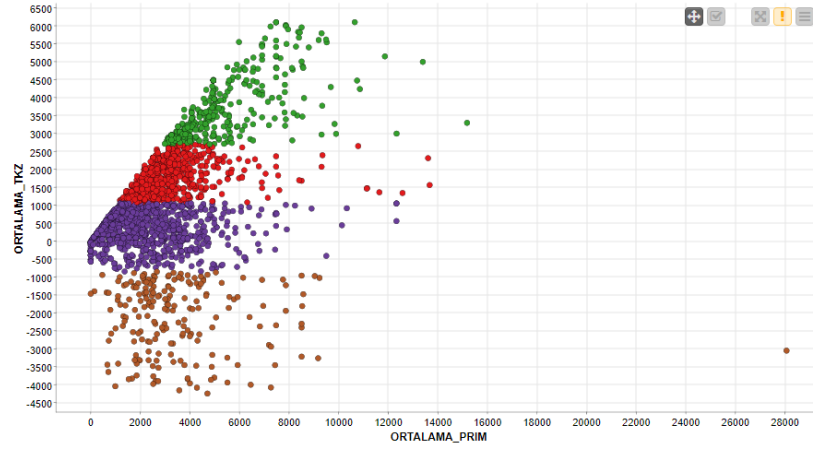


Şekil 4.2 Elbow testi sonuç grafiği

Elbow testi sonucunda en büyük değişimin k değerinin 4 ten 5 e geçti kırılım olduğu gözlemlendiğinden k=4 olarak belirlendi. K değeri belirlendikten sonra TSS ve ÖSS sigortalıları için ayrı ayrı K-Means algoritması kullanılarak kümeleme çalışması yapıldı. Kümeleme sonuçları Şekil 4.3 de ve Şekil 4.4 te gösterilmiştir.



Şekil 4.3 TSS kümeleme sonucu



Şekil 4.4 ÖSS kümeleme sonucu

Şekil 4.3 ve Şekil 4.4 te yer alan grafilerimizin en üst grubunda yer alan sigortalılar şirketin en çok kar ettiği sigortalı grubudur. Bu grupta yer alan sigortalıları kaybetmemek adına özel kampanyalar düzenlenmesi ve fiyat indirimleri sağlanması planlanmaktadır. Grafiğin en alt grubunda yer alan sigortalılar şirketin en çok zarar eden sigortalı grubudur. Bu grupta yer alan sigortalıların yenileme dönemlerinde yüksek fiyat verilmesi, poliçe kapsamlarının değiştirilmesi yada daha geniş kapsamlı poliçe paketleri teklif edilmesi planlanmaktadır. Grafiğin 2. ve 3. Sırasında yer alan sigortalılar şirkete minimum kar getiren yada minimum zarar ettiren sigortalı grubudur. Bu grupta yer alan sigortalılar en çok kar getiren gruba çıkartabilmek için kampanyalar, paket değişiklikleri gibi öneriler sunulması planlanmaktadır.

TSS için kümeleme sonucunda oluşan sınıfların açıklamaları Çizelge 4.1 de belirtilmiştir.

Çizelge 4.1 TSS sınıf açıklaması

Sınıf Adı	Açıklama
Sınıf 0	Düşük Karlı Sigortalılar
Sınıf 1	Karlı Sigortalılar
Sınıf 2	Düşük Zararlı Sigortalılar
Sınıf 3	Zararlı Sigortalılar

ÖSS için kümeleme sonucunda oluşan sınıfların açıklamaları Çizelge 4.2 de belirtilmiştir.

Çizelge 4.2 ÖSS sınıf açıklaması

Sınıf Adı	Açıklama
Sınıf 0	Karlı Sigortalılar
Sınıf 1	Düşük Karlı Sigortalılar
Sınıf 2	Zararlı Sigortalılar
Sınıf 3	Düşük Zararlı Sigortalılar

Bu çalışmanın en önemli adımlarından birisi yeni müşterilerin oluşturulan 4 kümeden hangisine düşeceğinin tahmin edilmesidir. Kümeleme sonucunda oluşan veri setinin %80 'i öğrenmeye, %20' si modelin doğruluğunda kullanılmıştır. Tahminleme çalışmasında 6 farklı algoritma denenmiştir. Denenmiş olan algortimaların sonuçları aşağıdaki çizelgelerde gösterilmiştir.

Çizelge 4.2 de TSS sigortalıları için Random Forest algoritmasının Gini Index özelliği kullanılarak elde edilen sonuçların sınıflar bazında detayları paylaşılmıştır. Çizelde 4.2 incelendiğinde doğruluk oranı en yüksek sınıfın, Sınıf 2 olarak tahminlenen sigortalıların %93,01 oranla gerçekte de Sınıf 2 olduğu gözlemlenmiştir. Gerçekte Sınıf 1 olan sigortalıların doğru tahminlenmesinin diğer sınıflara göre daha düşük olduğu gözlemlenmiştir. Sınıf 2 ve Sınıf 3 arasında geçişmeler olduğu gözlemlenmiştir.

Çizelge 4.3 TSS RF – Gini Index karışıklık matrisi

	Sınıf_0 (Tahmin)	Sınıf_1 (Tahmin)	Sınıf_2 (Tahmin)	Sınıf_3 (Tahmin)	Gerçek Değerin Doğru Tahminlenme Olasılığı
Sınıf_0 (Gerçek)	10962	1	1114	5	%90,73
Sınıf_1 (Gerçek)	25	13749	21	2075	%86,64
Sınıf_2 (Gerçek)	790	5	47785	2466	%93,61
Sınıf_3 (Gerçek)	52	1619	2458	47329	%91,98
Tahminin Gerçek Olma Olasılığı	%92,67	%89,43	%93,01	%91,24	

Çizelge 4.4 de TSS sigortalıları için Random Forest algoritmasının Gini Index özelliğinin genel başarısının (F1 Skoru) %91,85 olduğu paylaşılmıştır.

Çizelge 4.4 TSS RF– Gini Index genel istatistikler

Doğruluk	F1 Skoru	Hassasiyet	Kesinlik
%95,93	%91,85	%91,85	%91,85

Çizelge 4.5 de TSS sigortalıları için Random Forest algoritmasının Gain Ratio özelliği kullanılarak elde edilen sonuçların sınıflar bazında detayları paylaşılmıştır. Çizelge 4.5 incelendiğinde doğruluk oranı en yüksek sınıfın, Sınıf 2 olarak tahminlenen sigortalıların %94,10 oranla gerçekte de Sınıf 2 olduğu gözlemlenmiştir. Gerçekte Sınıf 1 olan sigortalıların doğru tahminlenmesinin diğer sınıflara göre daha düşük olduğu gözlemlenmiştir. Sınıf 2 ve Sınıf 3 arasında geçişmeler olduğu gözlemlenmiştir.

Çizelge 4.5 TSS RF – Gain Ratio karışıklık matrisi

	Sınıf_0 (Tahmin)	Sınıf_1 (Tahmin)	Sınıf_2 (Tahmin)	Sınıf_3 (Tahmin)	Gerçek Değerin Doğru Tahminlenme Olasılığı
Sınıf_0 (Gerçek)	11232	1	844	5	%92,96
Sınıf_1 (Gerçek)	6	14087	7	1770	%88,76
Sınıf_2 (Gerçek)	721	8	47864	2453	%93,77
Sınıf_3 (Gerçek)	19	1657	2151	47631	%92,56
Tahminin Gerçek Olma Olasılığı	%93,77	%89,42	%94,10	%91,85	

Çizelge 4.6 de TSS sigortalıları için Random Forest algoritmasının Gain Ratio özelliğinin genel başarısının (F1 Skoru) %92,61 olduğu paylaşılmıştır.

Çizelge 4.6 TSS RF – Gain Ratio genel istatistikler

Doğruluk	F1 Skoru	Hassasiyet	Kesinlik
%96,30	%92,61	%92,61	%92,61

Çizelge 4.7 de TSS sigortalıları için Random Forest algoritmasının Gain özelliği kullanılarak elde edilen sonuçların sınıflar bazında detayları paylaşılmıştır. Çizelge 4.7 incelendiğinde doğruluk oranı en yüksek sınıfın, Sınıf 2 olarak tahminlenen sigortalıların %93,27 oranla gerçekte de Sınıf 2 olduğu gözlemlenmiştir. Gerçekte Sınıf 1 olan sigortalıların doğru tahminlenmesinin diğer sınıflara göre daha düşük olduğu gözlemlenmiştir. Sınıf 2 ve Sınıf 3 arasında geçişmeler olduğu gözlemlenmiştir.

Çizelge 4.7 TSS RF – Gain karışıklık matrisi

	Sınıf_0 (Tahmin)	Sınıf_1 (Tahmin)	Sınıf_2 (Tahmin)	Sınıf_3 (Tahmin)	Gerçek Değerin Doğru Tahminlenme Olasılığı
Sınıf_0 (Gerçek)	11084	1	992	5	%91,74
Sınıf_1 (Gerçek)	20	14008	25	1817	%88,27
Sınıf_2 (Gerçek)	828	6	47613	2599	%93,27
Sınıf_3 (Gerçek)	42	1736	2327	47353	%92,02
Tahminin Gerçek Olma Olasılığı	%92,57	%88,93	%93,44	%91,46	

Çizelge 4.8 de TSS sigortalıları için Random Forest algoritmasının Gain özelliğinin genel başarısının (F1 Skoru) %92,03 olduğu paylaşılmıştır.

Çizelge 4.8 TSS RF – Gain genel istatistikler

Doğruluk	F1 Skoru	Hassasiyet	Kesinlik
%96,01	%92,03	%92,03	%92,03

Çizelge 4.9 de TSS sigortalıları için GBoosted Trees algoritması kullanılarak elde edilen sonuçların sınıflar bazında detayları paylaşılmıştır. Çizelge 4.9 incelendiğinde doğruluk oranı en yüksek sınıfın, Sınıf 2 olarak tahminlenen sigortalıların %94,36 oranla gerçekte de Sınıf 2 olduğu gözlemlenmiştir. Gerçekte Sınıf 1 olan sigortalıların doğru tahminlenmesinin diğer sınıflara göre daha düşük olduğu gözlemlenmiştir. Sınıf 2 ve Sınıf 3 arasında geçişmeler olduğu gözlemlenmiştir.

Çizelge 4.9 TSS- GBoosted Trees karışıklık matrisi

	Sınıf_0 (Tahmin)	Sınıf_1 (Tahmin)	Sınıf_2 (Tahmin)	Sınıf_3 (Tahmin)	Gerçek Değerin Doğru Tahminlenme Olasılığı
Sınıf_0 (Gerçek)	11295	0	780	7	%93,49
Sınıf_1 (Gerçek)	0	14226	3	1641	%89,64
Sınıf_2 (Gerçek)	623	8	48166	2249	%94,36
Sınıf_3 (Gerçek)	3	1526	2121	47808	%92,91
Tahminin Gerçek Olma Olasılığı	%94,75	%90,27	%94,31	%92,46	

Çizelge 4.10 de TSS sigortalıları için GBoosted Trees algoritmasının genel başarısının (F1 Skoru) %93,13 olduğu paylaşılmıştır.

Çizelge 4.10 TSS – GBoosted Trees genel istatistikler

Doğruluk	F1 Skoru	Hassasiyet	Kesinlik
%96,57	%93,13	%93,13	%93,13

Çizelge 4.11 de TSS sigortalıları için XGBoost algoritması kullanılarak elde edilen sonuçların sınıflar bazında detayları paylaşılmıştır. Çizelge 4.11 incelendiğinde doğruluk oranı en yüksek sınıfın, Sınıf 2 olarak tahminlenen sigortalıların %94,89 oranla gerçekte de Sınıf 2 olduğu gözlemlenmiştir. Gerçekte Sınıf 1 olan

sigortalıların doğru tahminlenmesinin diğer sınıflara göre daha düşük olduğu gözlemlenmiştir. Sınıf 2 ve Sınıf 3 arasında geçişmeler olduğu gözlemlenmiştir.

Çizelge 4.11 TSS – XGBoost karışıklık matrisi

	Sınıf_0 (Tahmin)	Sınıf_1 (Tahmin)	Sınıf_2 (Tahmin)	Sınıf_3 (Tahmin)	Gerçek Değerin Doğru Tahminlenme Olasılığı
Sınıf_0 (Gerçek)	11393	0	682	7	%94,30
Sınıf_1 (Gerçek)	0	14571	1	1298	%91,81
Sınıf_2 (Gerçek)	574	8	48436	2028	%94,89
Sınıf_3 (Gerçek)	1	1256	1918	48283	%93,83
Tahminin Gerçek Olma Olasılığı	%95,20	%92,02	%94,90	%93,54	

Çizelge 4.11 da TSS sigortalıları için XGBoost algoritmasının genel başarısının (F1 Skoru) %94,04 olduğu paylaşılmıştır.

Çizelge 4.12 TSS – XGBoost genel istatistikler

Doğruluk	F1 Skoru	Hassasiyet	Kesinlik
%97,02	%94,04	%94,04	%94,04

Çizelge 4.13 de TSS sigortalıları için Neural Netwok - MLP algoritması kullanılarak elde edilen sonuçların sınıflar bazında detayları paylaşılmıştır. Çizelge 4.13 incelendiğinde doğruluk oranı en yüksek sınıfın, Sınıf 2 olarak tahminlenen sigortalıların %94,04 oranla gerçekte de Sınıf 2 olduğu gözlemlenmiştir. Gerçekte Sınıf 1 olan sigortalıların doğru tahminlenmesinin diğer sınıflara göre daha düşük olduğu gözlemlenmiştir. Sınıf 2 ve Sınıf 3 arasında geçişmeler olduğu gözlemlenmiştir.

Çizelge 4.13 TSS – Neural Netwok - MLP karışıklık matrisi

	Sınıf_0 (Tahmin)	Sınıf_1 (Tahmin)	Sınıf_2 (Tahmin)	Sınıf_3 (Tahmin)	Gerçek Değerin Doğru Tahminlenme Olasılığı
Sınıf_0 (Gerçek)	11439	0	635	8	%94,68
Sınıf_1 (Gerçek)	0	13898	0	1972	%87,57
Sınıf_2 (Gerçek)	969	6	47440	2631	%92,94
Sınıf_3 (Gerçek)	3	2453	2988	46014	%89,42
Tahminin Gerçek Olma Olasılığı	%92,17	%84,97	%92,90	%90,89	

Çizelge 4.14 de TSS sigortalıları için Neural Netwok - MLP algoritmasının genel başarısının (F1 Skoru) %91,06 olduğu paylaşılmıştır.

Çizelge 4.14 TSS – Neural Netwok - MLP genel istatistikler

Doğruluk	F1 Skoru	Hassasiyet	Kesinlik
%95,53	%91,06	%91,06	%91,06

Çizelge 4.15 de ÖSS sigortalıları için Random Forest algoritmasının Gini Index özelliği kullanılarak elde edilen sonuçların sınıflar bazında detayları paylaşılmıştır. Çizelge 4.15 incelendiğinde doğruluk oranı en yüksek sınıfın, Sınıf 2 olarak tahminlenen sigortalıların %93,65 oranla gerçekte de Sınıf 2 olduğu gözlemlenmiştir. Gerçekte Sınıf 3 olan sigortalıların doğru tahminlenmesinin diğer sınıflara göre daha düşük olduğu gözlemlenmiştir. Sınıf 0 ile Sınıf 2 arasında geçişmeler olduğu gözlemlenmiştir.

Çizelge 4.15 ÖSS RF – Gini Index karışıklık matrisi

	Sınıf_0 (Tahmin)	Sınıf_1 (Tahmin)	Sınıf_2 (Tahmin)	Sınıf_3 (Tahmin)	Gerçek Değerin Doğru Tahminlenme Olasılığı
Sınıf_0 (Gerçek)	27126	62	2173	1051	%89,20
Sınıf_1 (Gerçek)	45	7216	811	2	%89,37
Sınıf_2 (Gerçek)	2893	944	44696	607	%90,96
Sınıf_3 (Gerçek)	1191	19	49	9317	%88,10
Tahminin Gerçek Olma Olasılığı	%86,79	%87,56	%93,65	%84,88	

Çizelge 4.16 de ÖSS sigortalıları için Random Forest algoritmasının Gini Index özelliğinin genel başarısının (F1 Skoru) %89,97 olduğu paylaşılmıştır.

Çizelge 4.16 ÖSS RF – Gini Index genel istatistikler

Doğruluk	F1 Skoru	Hassasiyet	Kesinlik
%94,99	%89,97	%89,97	%89,97

Çizelge 4.17 de ÖSS sigortalıları için Random Forest algoritmasının Gain Ratio özelliği kullanılarak elde edilen sonuçların sınıflar bazında detayları paylaşılmıştır. Çizelge 4.17 incelendiğinde doğruluk oranı en yüksek sınıfın, Sınıf 2 olarak tahminlenen sigortalıların %94,99 oranla gerçekte de Sınıf 2 olduğu gözlemlenmiştir. Gerçekte Sınıf 0 olan sigortalıların doğru tahminlenmesinin diğer sınıflara göre daha düşük olduğu gözlemlenmiştir. Sınıf 0 ile Sınıf 2 arasında geçişmeler olduğu gözlemlenmiştir.

Çizelge 4.17 ÖSS RF – Gain Ratio karışıklık matrisi

	Sınıf_0 (Tahmin)	Sınıf_1 (Tahmin)	Sınıf_2 (Tahmin)	Sınıf_3 (Tahmin)	Gerçek Değerin Doğru Tahminlenme Olasılığı
Sınıf_0 (Gerçek)	27481	40	1780	1111	%90,36
Sınıf_1 (Gerçek)	40	7474	557	3	%92,57
Sınıf_2 (Gerçek)	2964	912	44635	629	%90,83
Sınıf_3 (Gerçek)	985	11	15	9565	%90,44
Tahminin Gerçek Olma Olasılığı	%87,32	%88,59	%94,99	%84,59	

Çizelge 4.18 da ÖSS sigortalıları için Random Forest algoritmasının Gain Ratio özelliğinin genel başarısının (F1 Skoru) %90,79 olduğu paylaşılmıştır.

Çizelge 4.18 ÖSS RF – Gain Ratio genel istatistikler

Doğruluk	F1 Skoru	Hassasiyet	Kesinlik
%95,39	%90,79	%90,79	%90,79

Çizelge 4.19 de ÖSS sigortalıları için Random Forest algoritmasının Gain özelliği kullanılarak elde edilen sonuçların sınıflar bazında detayları paylaşılmıştır. Çizelge 4.19 incelendiğinde doğruluk oranı en yüksek sınıfın, Sınıf 2 olarak tahminlenen sigortalıların %94,15 oranla gerçekte de Sınıf 2 olduğu gözlemlenmiştir. Gerçekte Sınıf 0 olan sigortalıların doğru tahminlenmesinin diğer sınıflara göre daha düşük olduğu gözlemlenmiştir. Sınıf 0 ile Sınıf 2 arasında geçişmeler olduğu gözlemlenmiştir.

Çizelge 4.19 ÖSS RF – Gain karışıklık matrisi

	Sınıf_0 (Tahmin)	Sınıf_1 (Tahmin)	Sınıf_2 (Tahmin)	Sınıf_3 (Tahmin)	Gerçek Değerin Doğru Tahminlenme Olasılığı
Sınıf_0 (Gerçek)	27182	65	2015	1150	%89,38
Sınıf_1 (Gerçek)	41	7310	719	4	%90,54
Sınıf_2 (Gerçek)	2938	999	44571	632	%90,70
Sınıf_3 (Gerçek)	1045	18	35	9478	%89,62
Tahminin Gerçek Olma Olasılığı	%87,11	%87,11	%94,15	%84,14	

Çizelge 4.19 de ÖSS sigortalıları için Random Forest algoritmasının Gain özelliğinin genel başarısının (F1 Skoru) %90,16 olduğu paylaşılmıştır.

Çizelge 4.20 ÖSS RF – Gain genel istatistikler

Doğruluk	F1 Skoru	Hassasiyet	Kesinlik
%95,08	%90,16	%90,16	%90,16

Çizelge 4.21 da ÖSS sigortalıları için GBoosted Trees algoritması kullanılarak elde edilen sonuçların sınıflar bazında detayları paylaşılmıştır. Çizelge 4.21 incelendiğinde doğruluk oranı en yüksek sınıfın, Sınıf 2 olarak tahminlenen sigortalıların %94,10 oranla gerçekte de Sınıf 2 olduğu gözlemlenmiştir. Gerçekte Sınıf 0 olan sigortalıların doğru tahminlenmesinin diğer sınıflara göre daha düşük olduğu gözlemlenmiştir. Sınıf 0 ile Sınıf 2 arasında geçişmeler olduğu gözlemlenmiştir.

Çizelge 4.21 ÖSS – GBoosted Trees karışıklık matrisi

	Sınıf_0 (Tahmin)	Sınıf_1 (Tahmin)	Sınıf_2 (Tahmin)	Sınıf_3 (Tahmin)	Gerçek Değerin Doğru Tahminlenme Olasılığı
Sınıf_0 (Gerçek)	27175	37	2193	1007	%89,36
Sınıf_1 (Gerçek)	29	7412	628	5	%91,80
Sınıf_2 (Gerçek)	2684	816	45054	586	%91,68
Sınıf_3 (Gerçek)	1045	6	5	9520	%90,02
Tahminin Gerçek Olma Olasılığı	%87,85	%89,61	%94,10	%85,63	

Çizelge 4.22 de ÖSS sigortalıları için GBoosted Trees algoritmasının genel başarısının (F1 Skoru) %90,79 olduğu paylaşılmıştır.

Çizelge 4.22 ÖSS – GBoosted Trees genel istatistikler

Doğruluk	F1 Skoru	Hassasiyet	Kesinlik
%95,40	%90,79	%90,79	%90,79

Çizelge 4.23 de ÖSS sigortalıları için XGBoost algoritması kullanılarak elde edilen sonuçların sınıflar bazında detayları paylaşılmıştır. Çizelde 4.23 incelendiğinde doğruluk oranı en yüksek sınıfın, Sınıf 2 olarak tahminlenen sigortalıların %95,46 oranla gerçekte de Sınıf 2 olduğu gözlemlenmiştir. Gerçekte Sınıf 0 olan sigortalıların doğru tahminlenmesinin diğer sınıf lara göre daha düşük olduğu gözlemlenmiştir. Sınıf 0 ile Sınıf 2 arasında geçişmeler olduğu gözlemlenmiştir.

Çizelge 4.23 ÖSS – XGBoost karışıklık matrisi

	Sınıf_0 (Tahmin)	Sınıf_1 (Tahmin)	Sınıf_2 (Tahmin)	Sınıf_3 (Tahmin)	Gerçek Değerin Doğru Tahminlenme Olasılığı
Sınıf_0 (Gerçek)	27644	34	1700	1034	%90,90
Sınıf_1 (Gerçek)	27	7602	441	4	%94,15
Sınıf_2 (Gerçek)	2631	845	45163	501	%91,91
Sınıf_3 (Gerçek)	817	6	8	9745	%92,14
Tahminin Gerçek Olma Olasılığı	%88,83	%89,57	%95,46	%86,36	

Çizelge 4.24 de ÖSS sigortalıları için XGBoost algoritmasının genel başarısının (F1 Skoru) %91,80 olduğu paylaşılmıştır.

Çizelge 4.24 ÖSS – XGBoost genel istatistikler

Doğruluk	F1 Skoru	Hassasiyet	Kesinlik
%95,90	%91,80	%91,80	%91,80

Çizelge 4.25 de ÖSS sigortalıları için Neural Network - MLP algoritması kullanılarak elde edilen sonuçların sınıflar bazında detayları paylaşılmıştır. Çizelge 4.25 incelendiğinde doğruluk oranı en yüksek sınıfın, Sınıf 2 olarak tahminlenen sigortalıların %92,78 oranla gerçekte de Sınıf 2 olduğu gözlemlenmiştir. Gerçekte Sınıf 0 olan sigortalıların doğru tahminlenmesinin diğer sınıflara göre daha düşük olduğu gözlemlenmiştir. Sınıf 0 ile Sınıf 2 arasında geçişmeler olduğu gözlemlenmiştir.

Çizelge 4.25 ÖSS – Neural Network - MLP karışıklık matrisi

	Sınıf_0 (Tahmin)	Sınıf_1 (Tahmin)	Sınıf_2 (Tahmin)	Sınıf_3 (Tahmin)	Gerçek Değerin Doğru Tahminlenme Olasılığı
Sınıf_0 (Gerçek)	27042	80	2479	811	%88,92
Sınıf_1 (Gerçek)	15	7138	921	0	%88,41
Sınıf_2 (Gerçek)	3757	1212	43796	375	%89,12
Sınıf_3 (Gerçek)	4703	10	10	5853	%55,34
Tahminin Gerçek Olma Olasılığı	%76,14	%84,57	%92,78	%83,15	

Çizelge 4.26 de ÖSS sigortalıları için Neural Network - MLP algoritmasının genel başarısının (F1 Skoru) %95,36 olduğu paylaşılmıştır.

Çizelge 4.26 ÖSS – Neural Network - MLP genel istatistikler

Doğruluk	F1 Skoru	Hassasiyet	Kesinlik
%92,68	%85,36	%85,36	%85,36

6 farklı algoritmanın F1 skor sonuçları Çizelge 4.27 de gösterilmiştir.

Çizelge 4.27 Algoritmaların sonuç karşılaştırması

Algoritma	TSS (F1 Skoru)	ÖSS (F1 Skoru)
RF – Gini Index	%91,85	%89,97
RF – Gain Ratio	%92,63	%90,79
RF – Gain	%92,03	%90,16
GB	%93,13	%90,79
XGBoost	%94,04	%91,80
MLP	%91,06	%85,36

Çalışmada kullanılan veri seti özel bir sigorta şirketinin gerçek verilerinden oluşmaktadır. Veri seti orijinal bir veri seti olduğu için ulusal ya da uluslararası karşılaştırılabilecek bir çalışma bulunmamaktadır. Sigorta sektöründe, ulusal ve

uluslararası alıřmalar incelendiĐinde yapmıř olduĐumuz alıřmaya benzer bir alıřmaya rastgelinmemiřtir. Müşteri verileriyle kümeleme alıřmaları yapıldıĐı gözlemlenmiřtir. İncelenen alıřmalar literatür taramasında belirtilmiřtir.



5. SONUÇ VE ÖNERİLER

Yapılan tüm çalışmalar sayesinde yenilemesi gelen poliçelerde sigortalıların karlılık durumuna bakılarak bir fiyatlandırma çalışması yapılabilir duruma gelindi. Yeni müşteriler için referans bir veri oluşturuldu. Yeni gelen müşterilerin hangi kümeye dahil olacağını tahminlemek için kullanılan 6 farklı algoritmanın F1 skor sonuçları karşılaştırıldığında hem TSS hem de ÖSS için en başarılı algortimanın XGBoost olduğu gözlemlenmiştir. Teklif ve fiyatlandırma biriminde çalışan arkadaşlar, sigortalı karlılığı takip edilebileceği için daha isabetli fiyatlandırma çalışmaları yapılabilir. Müşteriye daha düşük ya da daha yüksek fiyat verebilir. Bu şekilde şirketin bu poliçeden zarar etme ihtimali minimize edilebilir. Müşterilerin diğer sigorta şirketlerine yüksek fiyattan dolayı kaçırılmasının önüne geçilmesi planlanmaktadır.

Çalışmanın kapsamını genişletmek ve başarı oranını artırmak için bazı ek çalışmalar yapılabilir. Bunlara örnek vermek gerekirse;

- Yeni gelen müşterilerin geçmiş sağlık bilgileri sigortalının beyanına göre yer almamaktadır. Müşterilerin sağlık geçmişini gösteren veriler merkezi bir sistemden alınabilirse çalışmanın detayı arttırılabilir.
- Sigortalının provizyonları üzerinde yer alan tanı bilgilerinde eksiklik ya da olmayan tanı kodları olduğu gözlemlenmiştir. Tanı kodlarının doğruluğu modelin başarısını arttırabilmektedir.
- Bu çalışma kapsamında sigortalıların eczane bilgileri yer almamaktadır. Eczaneden almış oldukları ilaç bilgileri alınabilirse kategorize edilerek sisteme dahil edilebilir.

KAYNAKLAR

- Aksigorta, 2023. Prim. Erişim Tarihi: 17 Mayıs 2023.
<https://www.aksigorta.com.tr/yarim-merkezi/sigorta-sozlugu/prim>
- Alabas, M., 2020. KNIME İle Veri Analizi. Erişim Tarihi: 26 Mayıs 2023
<https://medium.com/datarunner/knime-i%CC%87le-veri%CC%87-anali%CC%87zi%CC%87-51add5411831>
- Amelia, A., 2019. Decision tree: Part 2/2. Erişim Tarihi: 24 Mayıs 2023
<https://towardsdatascience.com/decision-tree-part-2-34b31b1dc328>
- Asunakutlu, T., 2016. Türkiye’ de Sigorta İşletmelerinin Yatırımları. Dergipark, 13, 2, 113 – 126
- Atalay, M., Çelik, E., 2017. Büyük Veri Analizinde Yapay Zeka ve Makine Öğrenmesi Uygulamaları - Artificial Intelligence and Machine Learning Applications in Big Data Analysis. Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 9(22) , 155-172. DOI: 10.20875/makusobed.309727
- Ateş, M., 2011. Sağlık Sistemleri. Beta Yayıncılık, İstanbul
- Balaji S., Srivatsa S. K., 2012. Naïve Bayes Classification Approach for Mining Life Insurance Databases for Effective Prediction of Customer Preferences over Life Insurance Products. International Journal of Computer Applications (0975 – 8887) 51
- Berthold, M., Cebron, N., Dill, F., Gagliel, T.R. vd. 2009. KNIME - the Konstanz information miner: version 2.0 and beyond. ACM (Association for Computing Machinery). <https://doi.org/10.1145/1656274.1656280>
- Bholowalia, P., Kumar, A., 2014. EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN. International Journal of Computer Applications (0975 – 8887). 105
- Boodhun, N., 2017. A Review of Data Analytical Approaches in the Insurance Industry. Journal of Applied Technology and Innovation. 1, 1, 58-73
- Dalkılıç, N., Gülcemal, M. E., 2022. Hayat Dışı Sigorta Şirketlerinin Finansal Performanslarının Sınıflandırılması. The Journal of Social Science. DOI: 10.30520/tjsosci.1055973
- Dobilas, S., 2021. XGBoost: Extreme Gradient Boosting — How to Improve on Regular Gradient Boosting?. Erişim Tarihi : 25 Mayıs 2023.
<https://towardsdatascience.com/xgboost-extreme-gradient-boosting-how-to-improve-on-regular-gradient-boosting-5c6acf66c70a#:~:text=Extreme%20Gradient%20Boosting%20is%20a,the%20algorithm's%20use%20for%20classification>

- Doğan, B., Buldu, A., Demir, Ö., Ceren, B. E., 2017. Sigortacılık Sektöründe Müşteri İlişki Yönetimi İçin Kümeleme Analizi. Karaelmas Fen ve Mühendislik Dergisi, DOI: 10.7212%2Fzkufbd.v7i2.664
- Erdoğmuş, P., Çolak, B., Durdağ, Z., 2016. K-Means Algoritması İle Otomatik Kümeleme . El-Cezeri . 3 (2) , 0-0 . DOI: 10.31202/ecjse.264195
- Fang, K., Jiang, Y., Song, Malin., 2016. Customer profitability forecasting using Big Data analytics: A case study of the insurance industry. Computers & Industrial Engineering. 101. 554-564
- Genowesse, V., 2023. K-Means Kümeleme. Erişim Tarihi: 9 Nisan 2023 Wikipedia. https://tr.wikipedia.org/wiki/K-means_k%C3%BCmeleme
- Ghoreyshi, S., Hosseinkhani, J., 2015. Developing a Clustering Model based on K-Means Algorithm in order to Creating Different Policies for Policyholders in Insurance Industry. International Journal of Advanced Computer Science and Information Technology (IJACSIT). 4. 2. 2015.46-53, ISSN: 2296-1739
- Gopalan, B., 2020. What is Entropy and Information Gain? How are they used to construct decision trees?. Erişim Tarihi: 24 Mayıs 2023. <https://www.numpyninja.com/post/what-is-entropy-and-information-gain-how-are-they-used-to-construct-decision-trees>
- Gücenme, Ü., İşverenoğlu, G., 2005. Sigorta Şirketlerinde Finansal Başarının Ölçülmesi ve Bir Uygulama. Uludağ Üniversitesi
- Haile, M., 2021. Application of Data Mining to Classify Medical Insurance Customers Based on Claim Experience: The Case of Awash Insurance Company S.C. ST. Mary's University.
- Işıksan, E., 2020. Multi Layer Perceptron (MLP) Nedir?. Erişim Tarihi: 25 Mayıs 2023. <https://isikhanelif.medium.com/multi-layer-perceptron-mlp-nedir-4758285a7f15>
- Koçkaya, G., Atikeler, K., Yenilmez, F., 2016. Türkiye Özel ve Kamu Sağlık Sigortacılığı Prim/Hasarsağlık Harcaması Değerlendirmesi. Dergipark , 0 (9) , 82-101
- Kormaz, D., Çelik, H. E., Kapar, M., 2018. Sınıflandırma ve Regresyon Ağaçları ile Rastgele Orman Algoritması Kullanarak Botnet Tespiti: Van Yüzüncü Yıl Üniversitesi Örneği, Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi, (3): 297-307
- Loh, W. Y., 2011. Classification and regression trees. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1 (1): 14-23.

- Matloob, I., Khan, S. A., Hussain, F., Butt, W. H., vd., Need-Based and Optimized Health Insurance Package Using Clustering Algorithm, Applied Sciens. <https://doi.org/10.3390/app11188478>
- Mesci, Y., 2020. Kümeleme Teknikleri ve Python'da K-Means. Erişim Tarihi: 23 Mayıs 2023. <https://medium.com/deep-learning-turkiye/k%C3%BCmeleme-teknikleri-ve-pythonda-k-means-62d93c624e5c>
- Niceguyede, 2023. Multilayer Perceptron. Erişim Tarihi: 9 Nisan 2023. Wikipedia, https://en.wikipedia.org/wiki/Multilayer_perceptron
- Orhaner, E., Yavaş, E. (2023). Tamamlayıcı Sağlık Sigorta Kullanıcılarının Bilgi ve Memnuniyet Düzeylerinin Değerlendirilmesi: Ankara İlinde Bir Uygulama . Uluslararası Sağlık Yönetimi ve Stratejileri Araştırma Dergisi , 9 (1) , 12-24
- Patel, P., 2021. Decision Tree with Gini index. Erişim Tarihi: 24 Mayıs 2023. <https://purav-patel.medium.com/decision-tree-with-gini-index-4268f277817a>
- Qadadeh, W., Abdallah, S., 2018. C ustomers Segmentation in the Insurance Company (TIC) Dataset. ScienceDirect. Procedia Computer Science 144 (2018) 277–290
- Rahman, S. A., Arifin, N. F. K., Hanafiah, M., Mutalib, S., 2021. Customer Segmentation and Profiling for Life Insurance using K-Modes Clustering and Decision Tree Classifier. International Journal of Advanced Computer Science and Applications, 12, 9
- Sabuncuoğlu, Z., Tokol, T., 2001. İşletme. Ezgi Yayınları
- Singhg A., 2020. Detecting and Removing Outliers. Erişim Tarihi: 16 Mayıs 2023. <https://medium.com/analytics-vidhya/detecting-and-removing-outliers-7b408b279c9>
- Süzel, C., 2018. Özel Sağlık Sigortası Sözleşmesinin Niteliği. Tarafları ve Ömür Boyu Yenileme Garantisi . Marmara Üniversitesi Hukuk Fakültesi Hukuk Araştırmaları Dergisi. 24 (2) . 931-965 . DOI: 10.33433
- Thakur, M., 2023. Outlier Formula. Erişim Tarihi: 9 Nisan 2023 <https://www.educba.com/outliers-formula/>
- Tunay, N., 2014. Sigorta Sektöründe Aracılar Arası Rekabet ve Bileşenleri: Yeni Yönelimler Işığında Türkiye İçin Çıkarsamalar. Dergipark. 3 (6) . 65-70

- Tusyakdiah, H., 2020. Introduction to K-Means Clustering Algorithm in Python. Erişim Tarihi: 18 Mayıs 2023. <https://halimatusyak.medium.com/introduction-to-k-means-clustering-algorithm-in-python-ee6ddf6ef2c0>
- Tyagi, N., 2020. Understanding the Gini Index and Information Gain in Decision Trees. Erişim Tarihi: 9 Nisan 2023. Medium. <https://medium.com/analytics-steps/understanding-the-gini-index-and-information-gain-in-decision-trees-ab4720518ba8>
- Umargono, E., Suseno, J. E., Gunawan S.K, V., 2019. K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula. Proceedings of the 2nd International Seminar on Science and Technology (ISSTEC 2019). 474
- Varoğlu, F., 2013. Tamamlayıcı Sağlık Sigortasının Genel Görünümü ve Türkiye’deki Gelişim Süreci . Sosyal Güvence , 0 (3) , 20-40 . DOI: 10.21441
- Wessel, N., 2023. Random Forest. Erişim Tarihi: 9 Nisan 2023. Wikipedia. https://en.wikipedia.org/wiki/Random_forest
- Wikipedia, 2022. Information Gain (Decision tree). Erişim Tarihi: 9 Nisan 2023. [https://en.wikipedia.org/wiki/Information_gain_\(decision_tree\)](https://en.wikipedia.org/wiki/Information_gain_(decision_tree))
- Wikipedia, 2022. Information Gan Ratio. Erişim Tarihi: 9 Nisan 2023. https://en.wikipedia.org/wiki/Information_gain_ratio
- Wikipedia, 2023. Gradient Boosting. Erişim Tarihi: 9 Nisan 2023. https://en.wikipedia.org/wiki/Gradient_boosting
- Wikipedia, 2023. XGBoost. Erişim Tarihi: 9 Nisan 2023. <https://en.wikipedia.org/wiki/XGBoost>
- Wikipedia, 2023. Çeyrekler Açıklığı. Erişim Tarihi: 9 Nisan 2023. https://tr.wikipedia.org/wiki/%C3%87eyrekler_a%C3%A7%C4%B1kl%C4%B1%C4%9F%C4%B1
- Wikipedia, 2023. Elbow Method (Clustering). Erişim Tarihi: 9 Nisan 2023. [https://en.wikipedia.org/wiki/Elbow_method_\(clustering\)](https://en.wikipedia.org/wiki/Elbow_method_(clustering))
- Wikipedia, 2022. KNIME, Erişim Tarihi: 26 Mayıs 2023. <https://tr.wikipedia.org/wiki/KNIME>,
- Yayla Ş. O. , 2019. Sigortacılık ve Türkiye’de Sigorta Sektörünün Durumu. Liberal Düşünce Dergisi. 24(94),.107-205

ÖZGEÇMİŞ

Adı Soyadı : Ömer GÜL

Eğitim Durumu

Lise : Aydıncık Çok Programlı Lisesi, 2009

Lisans : İstanbul Ticaret Üniversitesi, Mühendislik ve Tasarım
Fakültesi, Bilgisayar Mühendisliği Bölümü

Yayınlar

Gül, Ö., Kasapbaşı, M.C., 2023. Makine Öğrenmesiyle Sağlık Sigortalarında
Kümeleme ve Sınıflandırma. 8. Asia Congress Proceedings Book