

KARAYOLU TAŞIMACILIĞI SEKTÖRÜNDE MÜŞTERİ ANALİTİĞİ: BİR VAKA ÇALIŞMASI



Makale Gönderim Tarihi: 17.10.2019 Yayına Kabul Tarihi: 10.05.2020

Kafkas Üniversitesi
İktisadi ve İdari Bilimler
Fakültesi
KAÜİBFD
Cilt, 11, Sayı 21, 2020
ISSN: 1309 – 4289
E – ISSN: 2149-9136

Ayşenur BUDAK
Dr. Öğr. Üyesi
Gebze Teknik Üniversitesi
Mühendislik Fakültesi
Türkiye
abudak@gtu.edu.tr
ORCID ID: 0000-0002-6961-
2414

ÖZ | Müşteri analitiği, hızlı değişen pazarlarda ve kar marjlarının küçüldüğü alanlarda istatistiksel analiz yöntemleri ile kârlı müşterilerin elde tutulması için firmaya daha hızlı, dinamik ve isabetli kararlar almalarına yardımcı olan bir araştırma alanıdır. Karayolu taşımacılığı sektörü büyüyen, gelişen ve önemini arttıran bir alan konumundadır. Bu yüzden lojistik firmaları anlık ve hızlı bir şekilde veriyi kullanarak doğru kararları almak istemektedirler. Lojistik firması müşterilerine, taşıyıcıdan elde ettiği fiyatın üstüne bir kar marjı koyarak sunmaktadır. Bu kapsamda müşteri bazlı bir kar marjı öneri ve tahminleme sistemi geliştirilmiştir. Bu çalışmada amaç; karayolu taşımacılığında, hangi müşteriye hangi kar marjı uygulanacağını tahmin ve analiz edilmesini kapsamaktadır. Bu kapsamda müşteri analitiği çalışması gerçekleştirilmiştir. Uygulama Türkiye’de karayolu taşımacılığı yapan bir firma üzerinden gerçekleştirilmiştir. Tasarlanan bu metodoloji kestirimsel analitik (predictive analytics) yöntemlerini içermektedir. Öncelikle ilk aşamada müşteri segmentasyonu gerçekleştirilmiştir. İkinci aşamada kar marjı tahminlemede, CHAID, CART ve doğrusal regresyon yöntemleri kullanılmıştır. Elde edilen en iyi küme senaryosu, değişken olarak bu modellere eklenip senaryolar oluşturulmuştur. Bu sayede tahmin performansları karşılaştırılmıştır. CHAID yönteminin kar marjı tahminlemede en iyi sonucu verdiği belirlenmiştir.

Anahtar Kelimeler: Müşteri Analitiği, Karayolu Taşımacılığı, Veri Analitiği

JEL Kodu: C8, D40, L91
Alanı: İşletme, Lojistik, Veri bilimi
Türü: Araştırma

DOI: 10.36543/kauibfd.2020.004

Atıfta bulunmak için: Budak, A. (2020). Karayolu taşımacılığı sektöründe müşteri analitiği: bir vaka çalışması. *KAÜİBFD*, 11(21), 85-117.

CUSTOMER ANALYTICS IN TRUCKLOAD TRANSPORTATION SECTOR: A CASE STUDY



Kafkas University
Economics and Administrative
Sciences Faculty
KAUJEASF
Vol. 11, Issue 21, 2020
ISSN: 1309 – 4289
E – ISSN: 2149-9136

Article Submission Date: 17.10.2019 Accepted Date: 10.05.2020

Ayşenur BUDAK
Assistant Professor
Gebze Technical University
Faculty of Engineering
Turkey
abudak@gtu.edu.tr
ORCID ID: 0000-0002-
6961-2414

ABSTRACT | Customer analytics is a field of research that helps the firm to make faster, dynamic and accurate decisions to obtain profitable customers through statistical analysis methods in rapidly changing markets and areas. Truckload transportation sector is an area that is growing, developing and increasing. Therefore, logistics companies want to make the right decisions by using the data instantly and quickly. In this context, a customer-based profit margin suggestion and prediction system is developed. The purpose in this developed system is; to determine which profit margin should provide to its customers who wish to receive transportation services. Implementation is carried out through a company engaged in truckload transportation in Turkey. This methodology is designed to include predictive analytical methods. Customer segmentation is carried out in the first stage. In the second stage, to predict profit margin, classification algorithms such as CHAID, CART and linear regression methods are used. It is determined that CHAID method gives the best results in profit margin prediction.

Keywords: *Customer Analytics, Truckload Transportation, Data Analytics*

Jel codes: *C8, D40, L91*

Scope: *Business Administration, Logistics, Data Science*

Type: *Research*

Cite this Paper: Budak, A. (2020). Customer analytics in truckload transportation sector: a case study. *KAUJEASF*, 11(21), 85-117.

1. GİRİŞ

Son yıllarda, iş dünyasında değişim hızı önemli ölçüde artmıştır. Müşteri ilişkileri yönetimi (MİY) doğru bir şekilde kullanılmasıyla birlikte şirketler, müşterilerinin davranışlarını analiz etmeye odaklandılar, böylece müşterilerinin özellikleri ve istekleriyle ilgili akıllı pazarlama hizmetleri sağlamaya başlamışlardır. Akıllı pazarlama faaliyetleri hem müşteriler hem de şirketler için yüksek değer yaratma potansiyeline sahiptir. Analitik MİY ile şirketin stratejilerine yardımcı olmak için müşterilerin davranışlarını analiz edilmesi sağlanır. Örneğin, analitik MİY, kaynakları daha etkin bir şekilde kullanmak ve en kazanç sağlanan müşteri profilini daha iyi ayırt etmek amacıyla kullanılabilir (Ngai, Xiu & Chau,2009).Bu nedenle, MİY analizleri, şirketlere yardımcı olmaktadır. Müşterilerin ihtiyaçlarını doğru bir şekilde belirlemek ve doğru müşteri profilleri belirlemek için şirketler analitik uygulamalara odaklanır (Chung & Chen,2016).

Firmalar stratejik karar verme sürecinde, doğru ve değer yaratan stratejiler geliştirmeyi ve müşteri karlılığını arttırmayı amaçlamaktadır. Müşteri profillerinin doğru belirlenmesi hem müşterilerin değişken ihtiyaçlarının doğru değerlendirmesi hem de uzun vadede firmaların sürdürülebilirliği açısından önemlidir (Khajvand, Zolfaghar, Ashoori & Alizadeh ,2011). Bu açıdan bakıldığında büyük veri ambarlarının analiz edilmesi son yıllarda önemli bir araştırma alanı haline gelmiştir.

Birçok farklı sektörde işletmelere ait veri ambarlarında tutulan müşteriler ile ilgili pek çok veri üzerinde farklı yöntemler uygulanarak müşteri davranışları ile ilgili modeller kurulmaya ve çıkarımlar yapılmaya çalışılmaktadır. Bu veri ambarı içinden, müşteri değeri ya da hangi müşterinin gelecekte kaybedilme durumunun oluşabileceğinin tahminlenmesi gibi farklı bilgiler de elde edilebilmektedir (Hadden, Tiwari, Roy, & Ruta, 2007). Kestirimsel analitik (predictive analytics), geçmişte ne olduğunu bilmenin ötesinde, bir sonraki adımda olacakları tahmin etmelerine yardımcı olan bir veri analitiği aşamasıdır. Bunun için veri madenciliği (VM), makine öğrenme algoritmaları, yapay zeka yöntemleri, istatistiksel teknikler kullanılmaktadır (Xu & Chu,2015).

MİY kapsamında, günümüzde lojistik sektöründe, mevcut koşulları anlamak için verilerin nasıl kullanılacağı önemli bir araştırma konusu olarak görülmektedir. Lojistik sektörünün performansını artırmak için veri analitiği yöntemlerinin kullanımı literatürde yeni yeni gündeme alınmaktadır. Aslında, veri analitiği neredeyse tüm veri odaklı karar vermenin ayrılmaz bir parçasıdır. Veri analitiği, firmalarının daha akıllı kararlar ve daha iyi iş sonuçları ile sonuçlanan işlem

yapılabilir bilgiler edinmesine yardımcı olmaktadır. Bu sebeple, veri analitiği, lojistik firmaları için oldukça cazip bir konu haline gelmektedir (Jain, 2017).

Bu çalışmada amaç; karayolu taşımacılığında, hangi müşteriye hangi kar marjı uygulanmasının tahmin ve analiz edilmesini kapsamaktadır. Bu kapsamda müşteri analitiği çalışması gerçekleştirilmiştir. Türkiye’de karayolu taşımacılığında faaliyet gösteren lojistik firması taşıyıcı ve müşterilerle işbirliği içerisindedir. Müşteri hizmet almak istediğinde firma taşıyıcıdan fiyat alır ve belli bir oranda kar marjı üstüne ekleyerek müşteriye teklif eder. Buradaki problem hangi müşteriye hangi kar marjının verilmesinin belirlenmesidir. Verilecek kar marjı pek çok değişkene göre değişebilmektedir. Müşteri tipi, ürün, araç ve yük yükleme tipine, rota özelliklerine göre kar marjı değişkenlik gösterebilmektedir. Bu açıdan bakıldığında müşteri profilinin analiz edilmesi gereklidir. Bu çalışma, karayolu taşımacılığı sektöründe müşteri bazlı bir kar marjı tahminleme sistemi geliştirmektedir.

Bu kapsamda öncelikle, kümeleme analizi yöntemi ele alınmıştır. Kümeleme analizinin çıktısı tahminleme algoritmalarında girdi olarak kullanılmıştır. Şimdiye kadar yapılan çalışmalara bakıldığında, RFM müşteri segmentasyonuna dayalı analizler, çoğunlukla yalnızca RFM endekslerine dayanmaktadır. Başka bir deyişle, RFM kullanılarak kapsamlı bir müşteri kümeleme modeli oluşturmak için ağırlıklandırma faktörleri ve diğer müşteri değişkenleri gibi diğer özellikler göz ardı edilmemelidir. Bu çalışma için, kümeleme analizi kapsamında müşterilerin ağırlıklandırılmış RFM (WRFM) analiz çıktıları ve diğer müşteriye ait veri değişkenlerini kullanarak farklı senaryolar oluşturulmuştur. Bu sayede en iyi ve doğru kümeler kullanılarak müşteri segmentasyonu yapılabilmektedir. RFM analizini kullanarak kapsamlı bir müşteri kümeleme yaklaşımını formüle etmek için ağırlıklandırma faktörleri ve diğer müşteri değişkenleri dikkate alınmıştır. WRFM analizi sayesinde, müşteriye verilecek olan kar marjını en iyi tahmin edebilecek güçlü kümenin bulunması amaçlanmıştır.

Kar marjının tahmini için kestirimsel analitik yöntemlerinden CHAID (Ki-kare Otomatik Etkileşim Dedektörü), CART (Sınıflandırma ve Regresyon Ağaçları), ve doğrusal regresyon kullanılmıştır. İki farklı yaklaşımla hata oranları karşılaştırılmıştır. Kullanılan yöntemlere en iyi kar marjını ifade eden müşteri küme değişkeni eklenerek modellerin performansındaki değişimler analiz edilmiştir. Çalışma Türkiye’de gerçek bir veri üzerinde gerçekleştirilmiştir. Yöntemlerin performansları MAPE (Ortalama mutlak yüzde sapması) ve MSE

(Ortalama Kare Hata) hata oranları ile kıyaslanmıştır. Bölüm 2’de veri analitiği ile alakalı tekniklerin kullanımı ile ilgili literatür özeti verilmiştir. Bölüm 3’te kullanılan yöntemler hakkında bilgi verilmiştir. Bölüm 4’te tasarlanan metodoloji açıklanmıştır. Bölüm 5 ve 6’da uygulama gerçekleştirilmiştir. Bölüm 7 makalenin sonuç ve önerilerini kapsamaktadır.

2. LİTERATÜR

Literatürde veri analitiği ile müşteri pazarlama stratejileri üzerine çalışmalar mevcuttur. Veri analitiğinin bir parçası olan kestirimsel analitiğinin (predictive analytics) temel amacı bir sonraki aşamanın tahmin edilmesi üzerinedir. Müşteri analitiği çalışmalarında tek algoritma yerine pek çok algoritmanın bir arada kullanıldığı hibrit modeller oluşturulmaktadır. Bu modeller nitelik seçiminden, tahminlemeye kadar gerçekleşen süreçte pek çok algoritma barındırabilir. Yöntemler kümeleme analizleri, veri madenciliği yöntemleri ile sınıflandırma, makine öğrenme algoritmaları olabilir.

Aşağıda bazı RFM ve veri madenciliği yöntemleri, istatistiksel ve yapay zeka analizleri ile ilgili çalışmalara müşteri ilişkileri yönetimi kapsamında örnek verilmiştir.

Paker ve Vural (2016), marina müşterilerinin segmentasyonu için faktör analizini uygulamışlardır. Liu ve Ong (2008), k-ortalamalar analizi ile müşterileri kümelemiştir. Kullanılacak değişkenleri belirlemek amacıyla genetik algoritmayı kullanan bir model önermişlerdir. Coussement *vd.*, (2014) veri analizinde sınıflandırma doğruluğun performansını ölçmek için karar ağaçları, RFM analizini ve lojistik regresyon olmak üzere üç farklı teknik ile araştırmışlardır. Değerlendirme sonucunda, müşteri segmentasyonu için karar ağaçlarının üstün performans verdiğini belirtmişlerdir. McCarty ve M. Hastak (2007) , RFM analizi, CHAID ve lojistik regresyonu, bir pazarlama sektöründe müşteri segmentasyonu için iki farklı veri seti kullanarak analitik yöntemleri araştırmışlardır. Bu çalışma, RFM'in CHAID ve lojistik regresyona benzer sonuçlar verebileceğini göstermiştir. RFM' nin segmentasyonda kabul edilebilir bir seviyede performans gösterebileceği sonucuna varmıştır. Chan (2008), otomobil satış pazarını segmentlere ayırmak için RFM analizi ve genetik algoritma ile bir model geliştirmiştir. Sunulan metodoloji sayesinde en karlı müşteriler için özel teklifler belirlenmiştir. Hwang ,Jung ve Suh (2004), telekomünikasyon alanında lojistik regresyon kullanarak müşteri sadakatini belirlenmişler ve üç tür müşteri değerine ilişkin müşterilere özel kampanyalar

geliştirmişlerdir. Hong ve Kim (2012), bir şirketin çevrimiçi mağazalarının müşterilerini segmentlere ayırmak için öz düzenleyici haritalar (SOM) yöntemini ve k-ortalamlar yöntemini kullanarak gerçekleştirmişlerdir. Wei, Zhang ,Shi, Xia ,Pan ,Wu ,Han ve Zhao, (2018),yeni gelen müşterileri kümelemek amacıyla, iki aşamalı yöntemden oluşur. Birincisi, k-ortalamlar yöntemi ile müşteriler RFM analizi ile farklı gruplara ayrılır. İkincisi, müşteri demografik verileri kullanarak, her küme tekrar yeni kümeler bölünmüştür. Müşterileri segmentlere ayırmak ve buna bağlı olarak pazarlama stratejileri geliştirmek için Tayvan'daki bir güzellik salonuna modeli uygulayarak müşteri demografik özelliklerinin araştırılmasının farklı pazarlama sonuçlarına yol açabileceğini öne sürmüşlerdir.

Singh ve Tucker (2017) müşteri yorumlarını sınıflandırarak yorumların ürünlerin satışlarına etkisini inceleyen bir metodoloji geliştirmişlerdir. Uygulama, Amazon.com sitesinden, 3 farklı model cep telefonu için yapılmış yorumlardan elde edilmiş olup, sınıflandırma işi Naif Bayes, Destekçi Vektör Makineleri ve Karar Ağaçları yöntemleri ile yapılmıştır. Azeem, Usman, ve Fong (2017), kayıp müşteri sınıflandırma analizi için yapay sinir ağları, doğrusal regresyon, C4.5 ve destek vektör makinaları tekniklerini kullanıp ,performanslarını karşılaştırmıştır. Tanaka, Hamaguchi, Saigo ve Tsuda (2017), RFM analizi ile lojistik regresyondan oluşan bir hibrid metodoloji kullanarak, Japonya'daki bir süpermarket zinciri müşterilerini, önem derecelerine göre sınıflandırılmıştır.

Nakano ve Kondo (2018) , yaklaşık 2500 Japon müşteriye ait mağaza içi alışveriş, internet üzerinden alışveriş, internet tıklama verileri, sosyal medya verileri ve demografik verileri kullanarak müşterileri Gizli Sınıf kümeleme analizi ile segmente etmiştir. Vincent, Makinde, Salako ve Oluwafemi (2018), 1000 müşteriye ait verileri k-ortalamlar tekniği kullanılarak sınıflandırmışlardır. Ortaya çıkan sonuçlar müşterilerin yeni siparişlerinde ne tür bir tasarım istedikleri hususunda firmalara yardımcı olacaktır. Liang-Lin ve ve Lung-Fan (2019), çalışmalarında CHAID, CART ve QUEST yöntemini inşaat sektöründe oluşabilecek hata risklerini tahmin etmek için kullanmışlardır. En iyi tahmin yönteminin CHAID olduğu belirlenmiştir. Nyitrai ve Virag (2019), iflas tahminlerini CHAID metoduyla tahmin etmişlerdir.Wei, Lin,Yang ve Wu (2019), veterinerlik hizmeti almak isteyen müşterileri segment etmek için iki aşamalı RFM modeli kullanarak sınıflandırma yapmışlardır. Boix ve Moreno (2018), mobil telefon şebekelerinin için pazarlama tekniklerinde kullanılacak veri analitiği yöntemleirni kullanmışlardır.Böylelikle, müşteri kullanım detaylarına ilişkin önemli değişkenlerin seçilmesi ve müşteri kayıp analizinin gerçekleştirilmesi amaçlanmıştır.

Müşterilere atanan kar marjı tahmini ile alakalı çalışmalara bakıldığında, kar tahmininde, çoğu çalışma finansal hizmetler enerji ve elektrik sektörüyle ilgili (Chen & Hao,2017). Bazı çalışmalar aşağıda detaylı olarak verilmektedir. Martínez, Schmuck, Pereverzyev , Pirker ve Haltmeier (2018), finans sektöründe önümüzdeki ay için müşteri karlılık tahminini ele almışlardır. Gradyan artırma (gradient boosting), makine öğrenme algoritmaları teknikleri arasında daha iyi performans göstermiştir. Larivière ve Van den Poel, (2005) finansal hizmetlerde müşteriden elde edecek karlılığını tahmin etmek için rastgele orman (random forest) tekniklerini kullanmıştır. Chang, Hung, Pai ve Lin (2013), müşterilere verilecek olan projelerin karını tahmin etmek için iki aşamalı hibrit modeli kullanıldı. Öncelikle, bulanık yaklaşımlarla veri kümelendi, ardından en küçük kareler destek vektör regresyonunu (LSSVR) yöntemi kullanılarak kar marjları tahmin edilmiştir.

Yapılan çalışmalara bakıldığında, enerji, kimya, telekom, finans alanında tahminleme ve müşteri analitiği yöntemleri için sunulan farklı alanlarda birçok çalışma olduğu görülmektedir. Bu çalışma aşağıda sıralanan maddelere göre orijinallik içermekte ve literatüre yenilik getirmektedir.

1. Lojistik sektörü karayolu taşımacılığında müşteri bazlı kar tahmini çalışmasına literatürde rastlanmamaktadır. Bu nedenle ağırlıklandırılmış RFM analizi yöntemi ve kestirimsel analitik yöntemleri bir arada kullanılarak yeni bir hibrit model yaklaşımı geliştirilmiştir. Makale bu açıdan literatürdeki boşluğu doldurmaktadır.
2. WRFM çalışmasına literatürde çok az rastlanmaktadır. Aynı şekilde RFM analizin tek başına değil müşteriye ait diğer farklı veri gruplarıyla birlikte analizi ile lojistik sektörde farklı kümelerin performansının kar marjı tahmin odaklı değerlendirmesiyle ilgili çalışmaya rastlanmamıştır.
3. Tahminleme algoritmalarına eklenen küme değişkeninin tahmin performansını ne derece etkilediği karşılaştırılmıştır. Tahminlemede önemli ve anlamlı olan değişkenler ağırlıklarıyla belirlenmiştir.
4. Türkiye'deki karayolu taşımacılığı yapan bir lojistik firmasının veri tabanındaki verileri işleyerek, müşterilere uygulanacak kar marjı araştırılmıştır. Bu çalışma, lojistik sektörde geliri arttırarak doğru müşteriye doğru kar marjını sunan bir öneri sistemi olarak görülebilir. Bu amaçla, mevcut müşterilerinden veya yeni gelen bir müşterinin hangisinin, ne gibi koşullarda hangi kar marjının atanması araştırılmıştır.

Karayolu taşımacılığı sektöründe geliştirilen bu öneri sistemi literatüre ayrı bir yenilik katmaktadır.

3. YÖNTEMLER

Veri analitiği, bir veri hacminden anlamlı çıkarımlar ve tahminler elde etmek için en gelişmiş hesaplama veri madenciliği ve istatistiksel teknikleri birleştiren bir alan olarak son yıllarda günümüzde büyük önem kazanan araştırma alanlarından biri olarak kabul edilmektedir. Bu analitik yöntemin avantajı, yüksek hacimli verileri daha iyi, doğru, anlık ve hızlı kararlar vermek için kullanılacak faydalı bilgilere dönüştürecek analitik yöntemleri içermesidir. Günümüz iş dünyasında şirketler, daha iyi hizmetler için üretkenlik, karlılık ve sürdürülebilir iş süreçlerinde rekabet avantajı elde etmek için veri analitiğine giderek daha fazla önem vermektedir (Raut, Mangla, Narwane, Gardas, Priyadarshinee & Narkhede, ,2019). Son yıllarda ,veri tabanı pazarlama teknikleri, RFM analizinden (yenilik, sıklık ve parasallık analiz) karar ağaçları (CHAID,CART) ve lojistik regresyon gibi istatistiksel tekniklere dönüşmüştür. Özellikle son zamanlarda, yapay sinir ağları (ANN) modelleri, birçok sektörde, müşteri bilgilerini analiz ederek veri analitiği kapsamında kullanılmaktadır (Cheng & Chen ,2009).

Çalışma kapsamında kullanılan analiz yöntemlerinin açıklamaları aşağıda belirtilmiştir. Sınıflandırma algoritmaları ve kümeleme, iki ana makine öğrenme aracı ve veri madenciliğinin en önemli alanlarından ikisidir (Xu & Chu, 2015). Uygulama alanları sektöre göre farklılık gösterebilmektedir. Bu çalışmada lojistik sektörü, karayolu taşımacılığında müşteriye verilecek olan kar marjını tahmin edilmesinde kullanılmıştır. Genellikle, MİY için hiçbir veri analitiği aracı mükemmel değildir, her yöntem kendi içinde avantaj ve dezavantajlara sahiptir (Cheng & Chen ,2009). Bu çalışma kapsamında RFM analizi K-ortalamalar kümeleme analizi ile gerçekleştirilip, CART, CHAID, doğrusal regresyon yöntemlerinde de kullanılmıştır. Bu sayede veri madenciliği yöntemlerinin performanslarının değerlendirilmesine yönelik bir hibrit yaklaşım geliştirilmiştir.

3.1 Kümeleme Analizi

Kümeleme yöntemleri, verilerin anlamlı bir şekilde sınıflandırılmak veya gruplandırılmak için kullanılmaktadır. Verileri anlamlı gruplara kümelemek hem yapay zekânın hem de veri madenciliğinin önemli bir görevidir. Kümeleme yöntemleri, denetlenmeyen yapıdaki verileri sınıflandırmak için kullanılmaktadır. Sonuçların performansı genellikle kullanılan algoritmaya bağlıdır. Kümeleme analizinde, aynı kümedeki elemanlar birbirleriyle benzerlik

gösterirken başka kümelerin elemanlarından farklıdır (Bock, 2008).

3.1.1 K- ortalamalar analizi

K-ortalamalar analizi genellikle küme sayısı hakkında bir ön bilginin olması veya küme sayısının önceden belirli olduğu durumlarda kullanılır. Hiyerarşik olmayan kümeleme yönteminde en çok tercih edilen yöntemlerin başında gelmektedir (MacQueen 1967).

Burada, k adet özellik vektörüne ve d adet değişkene sahip bir x veri seti düşünelim.

Bu veri setinde k . özellik vektörü Eş. 1 'deki gibi tanımlanabilir.

$$x_k = [x_{k1} x_{k2} \dots x_{kd}], x_k \in R^d \quad (1)$$

K-Ortalamlar yönteminde, veri seti alt kümelere ayrılırken Eş. 2'deki amaç fonksiyonunun en küçüklenmesi amaçlanmaktadır. Öklit mesafe ölçütü Eş. 3'de gösterilmektedir.

$$J(S: X) = \sum_{i=1}^C \sum_{k=1}^N d_{ik}^2(x_k, s_i) \quad (2)$$

$$d_{ik}^2 = \left\| x_k^{(i)} - s_i \right\|^2 \quad (3)$$

$$s_i = \frac{1}{N} \sum_{k=1}^N x_k^{(i)} \quad (4)$$

Bölme yöntemlerinde, d veri tabanındaki nesne sayısı ve k oluşturulacak küme sayısı olarak kabul edilir. Bölme algoritması d adet nesneyi, k adet kümeye böler. Burada asıl amaç, aynı kümedeki nesnelere birbirlerine benzerken, farklı kümedeki nesnelere farklılık gösterecek şekilde kümelerin oluşturulmasıdır. Birinci aşama, kümelere ait merkezlerin belirlenmesidir ,ikinci olarak her bir değişkenin benzerlik ölçütü yardımıyla en yakın küme merkezine atanır. Değişkenlerin kümelere atanmasından sonra, her bir küme için küme merkezleri yeniden hesaplanmaktadır. Bu süreçler küme merkez değerlerinde herhangi bir değişim olmayana kadar tekrarlanır (MacQueen (1967)).

3.1.2 RFM analizi

RFM, müşteri analizinin detaylandırılması için kullanılan bir pazarlama tekniğidir. Bir müşterinin en son ne zaman satın aldığı (recency), ne sıklıkta satın aldığı (frequency) ve müşterinin yaptığı harcama tutarı (monetary) gibi müşteri davranışlarını dikkate alır. Gelecekte müşterileri çeşitli gruplara ayırarak ve yapılacak olan kampanya ve promosyonlara cevap verme eğiliminde olan müşterileri belirleyerek müşteri segmentasyonunu iyileştirmek için yararlı bir

yöntemdir. RFM analizi, gelecekte müşterilerin satın alma eğilimlerinden etkilenen satın alımla ilgili üç önemli değişken olan güncellik (R), frekans (F) ve parasal (M) ölçümlerine bağlıdır (Sellers, J., & Hughes, A. (1994)).

RFM analizi, bir müşterinin en son ne zaman satın aldığı (güncellik), ne kadar sıklıkta satın aldığı (frekans) ve müşterinin ne kadar harcadığı (parasal) gibi verileri kullanarak müşteri davranışını analiz etmek için tasarlanan bir pazarlama tekniğidir. MİY’de müşteri segmentasyonunu iyileştirmek için oldukça önemli bir yöntemdir. RFM analizi, müşterinin satın alımıyla ilgili üç önemli değişken olan güncellik (R), frekans (F) ve parasal (M) ölçümlerine bağlıdır. İlgili hesaplamalar aşağıdaki Eş. 5,6 ve 7 ‘de gösterilmiştir (Cheng & Chen, 2009):

$$R(C_i) = \frac{R_i - R_{min}}{R_{max} - R_{min}} \quad (5)$$

$$F(C_i) = \frac{F_i - F_{min}}{F_{max} - F_{min}} \quad (6)$$

$$M(C_i) = \frac{M_i - M_{min}}{M_{max} - M_{min}} \quad (7)$$

Burada $R_{(C_i)}$, i müşterisinin güncellik değeri; R_i , müşterinin son satın aldığı tarihten bu yana geçen süre; R_{Min} , müşterinin son satın alınmasından bu yana geçen en az süre; R_{Max} , bir müşterinin son satın alınmasından bu yana geçen en uzun süre; $F_{(C_i)}$, müşterinin frekans değeri; F_i , yani müşteri satın alımlarının sıklığı; F_{Min} , müşterinin minimum satın alma sıklığı; F_{Max} , müşterinin maksimum satın alma sıklığı; $M_{(C_i)}$, müşterinin parasal değeri; M_i , müşterinin sisteme bıraktığı miktar; M_{Min} , müşterinin parasal değerinin minimum değeri; ve M_{Max} , müşterinin parasal değerinin maksimum değeridir.

3.2 Kestirimsel Analitik Yöntemleri

Kestirimsel analitik , analitik endüstrisindeki en popüler tahmin modelleme teknikleri arasındadır. CHAID, CART, destek vektör makinaları, Naive Bayes, Doğrusal Regresyon, K en yakın komşuluk algoritması en bilinen algoritmalarıdır (Abbott, 2014). Bu çalışma kapsamında CHAID, CART ve doğrusal regresyon kullanılmıştır. Bu yöntemlerin küme değişkeni modele eklenerek performanslarının kıyaslanması amaçlanmıştır. Tahmine dayalı olan kestirimsel analitik yaklaşımı, hedef (bağımlı) değişkeni açıklayıcı (bağımsız) değişkenlerin bir işlevi olarak ifade eden bir modeli oluşturmaktadır. Öngörülen ve gerçek değerler arasındaki farklı minimize etmek , kestirimsel analitik

yöntemlerinin amacı olarak kabul edilir. Bu yöntemlerin seçilmesinin nedeni, bu metotlar hem kategorik hem sürekli değişkenlerin bir arada kullanılmasına olanak sağlar. Aynı zamanda kolayca analiz edilmesini ve yorumlanmasını olanak sunar. Çalışmada gerçek bir firmadan ele alınan veri farklı değişken tiplerini bir arada içerdiğinden dolayı bu üç avantajlı yöntem seçilmiştir. CHAID ile verilerin dağılımında normallik ve homojenlik sağlanabilmektedir. Ağaç diyagramı diğer karar ağaçlarına göre daha görsel anlamda kolay yorumlanabilmektedir. Çıktıların anlaşılabilirliği kolaydır. CART, önemli olan bağımsız değişkenleri otomatik olarak ağaç yapısına dahil eder ve büyük veri gruplarını basit kurallar tanımlayarak hızlı bir şekilde sınıflandırmaktadır. Doğrusal regresyon yöntemi ise kestirimsel analitik kapsamında uygulama kolaylığı, çabuk sonuç verme, sürekli değişkenlerin kullanılabilmesi nedeni ile tercih edilmiştir. Bu yöntemler sadece bağımlı değişken ile bağımsız değişken arasındaki ilişkinin yapısını araştırmakla kalmayıp, aynı zamanda bağımsız değişkenlerin önem derecelerini de sıralamaya olanak verir (Abbott, 2014). Bu üç yöntemin kendi içinde avantajları kullanılarak en iyi tahmin performansı veren yöntem ve yaklaşım belirlenecektir.

3.2.1 CHAID algoritması

CHAID, verilerin bölünmesi için ki-kare istatistiğini kullanmaktadır. CHAID analizi, değişkenlerin verilen bağımlı değişkenin sonucunu açıklamak için en iyi nasıl birleştirileceğini belirlemeye yardımcı olmak için tahmini bir model veya ağaç oluşturur. Bu analizde, yaklaşık eşit sayıda gözlemlerle kategorilere ayrıldığı nominal, sıralı ve sürekli veriler kullanılabilir. CHAID, en iyi sonuca ulaşılan ve daha fazla bölünme gerçekleştirilinceye kadar her bir kategorik öngörücü için olası tüm çapraz tabloları oluşturur. CHAID tekniğinde, bölünmüş değişkenler ile ağaçtaki ilişkili faktör arasındaki ilişkileri görsel olarak görebiliriz. Kararın veya sınıflandırma ağacının geliştirilmesi, hedef değişkeni veya bağımlı değişkeni tanımlamakla başlar; kök olarak kabul edilir. CHAID analizi, hedefi başlangıç veya üst düğüm olarak adlandırılan iki veya daha fazla kategoriye ayırır ve daha sonra düğümler istatistiksel algoritmalar kullanılarak alt düğümlere bölünür. Regresyon analizinden farklı olarak, CHAID tekniği verilerin normal olarak dağıtılmasını gerektirmez (Abbott,2014).

3.2.2 CART algoritması

Hem sayısal hem de nominal veri türlerini, girdi ve kestirimsel değişken olarak kabul edebilen CART algoritması, sınıflandırma ve regresyon problemlerinde bir

çözüm olarak kullanılabilir. CART karar ağacı, ikili olarak özyinelemeli biçimde bölünen bir yapıya sahiptir. Dallanma kriteri olarak Gini indeksinden yararlanmaktadır. Başlangıç olarak herhangi bir durma kuralı olmaksızın sürekli olarak bölünerek büyümektedir. Bölünme durduğu zamanda, uçtan köke doğru budama işlemi başlatılır. Olası en başarılı karar ağacı, her budama işlemi sonrası elde edilen bağımsız seçilen bir test verisi ile değerlendirme yapılarak tespit edilmeye çalışılır (Chen & Hao ,2017).

Analiz yapılmak istenen bağımlı değişken kategorik ise yöntem sınıflama ağaçları (Classification Tree), sürekli ise regresyon ağaçları (Regression Tree) olarak adlandırılmaktadır (Deconinck, Hancock, Coomans, Massart & Vander Heyden, 2005).

3.2.3 Doğrusal regresyon

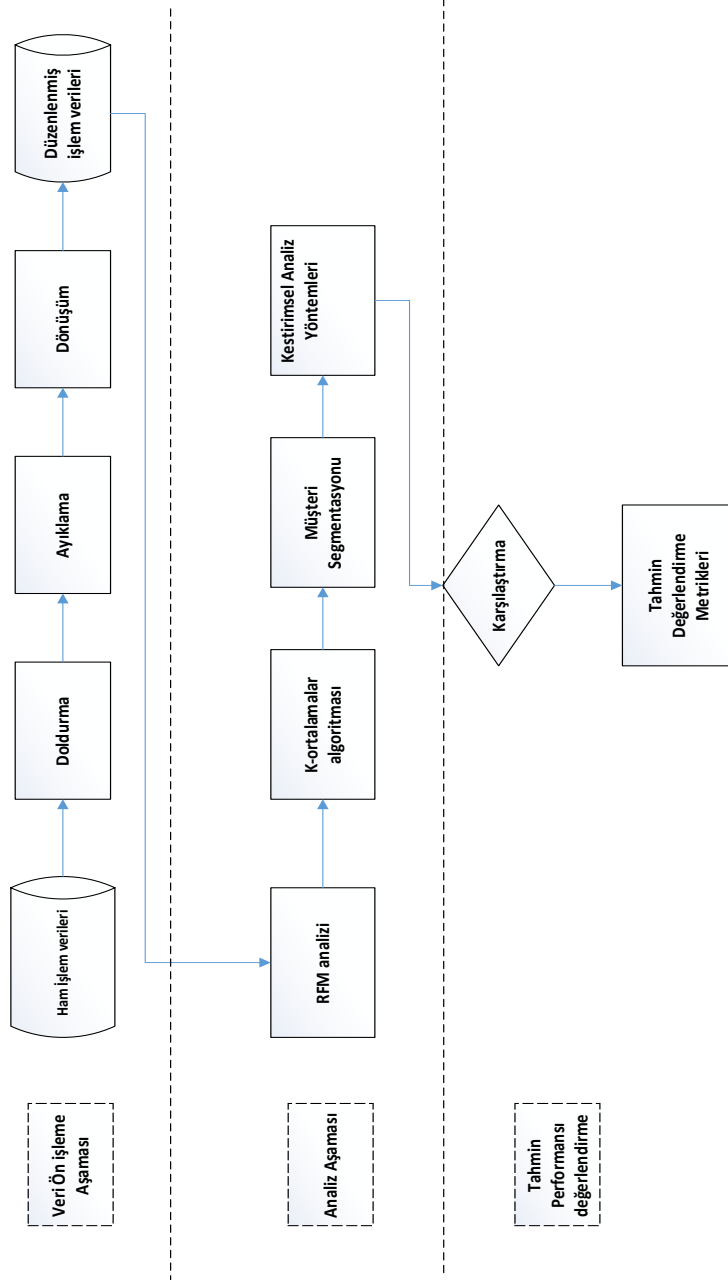
y bir bağımlı tesadüfi değişken ve x_1, x_2, \dots, x_p açıklayıcı değişkenler olarak belirlensin. Bir regresyon işlemi, x_1, x_2, \dots, x_p değerleri verildiğinde y 'nin beklenen değerini tahmin etmede kullanılan bir istatistik yöntem olarak ifade edilebilir.

Doğrusal Regresyon, çıktının girdiyle ilgili eğimini bulmayı amaçlayan en bilinen regresyon algoritmasıdır. Bu fikir ayrıca, değişkenlerden hangisinin hedef değışkene ne kadar katkıda bulunduğunu da vurgulamaktadır. Her bir özelliğe w_i (ağırlık) faktörü belirlenir (Abbott, 2014). Eş. 8 ile doğrusal regresyon ifade edilebilir.

$$y_i = w_0 + w_1x_1 + w_2x_2 + \dots \dots \dots \dots w_nx_n \quad (8)$$

4. TASARLANAN METODOLOJİ

Kurulan model 15.842 veri ve 630 müşteri değışkenini içermektedir. Veri 2016-2017 arasında yapılan işlemlerden oluşmaktadır. Ele alınan veriler verilen teklifi kabul eden müşteriler baz alınarak oluşturulmuştur. Gerçek veriden ele alınan teklifler göz önüne alınarak müşteri bazlı kar marjı tahmininde kullanılmıştır. İşlem verisinde 13 tane bağımsız değışken bulunmaktadır. Kar marjı tahmininde kullanılan değışkenler aşağıdaki gibidir. Parantez içinde düzenlenen değışkenlerin tipleri tanımlanmıştır. Farklı algoritmalar ile modelin tahmin güçleri belirlenmiştir. Kullanılan algoritmalar ve bu algoritmaların doğrulukları karşılaştırılmıştır. Çalışma kapsamında SPSS Clementine programı kullanılmıştır. Şekil 1 lojistik sektörde müşteriye uygulanacak olan kar marjı tahmini için tasarlanan metodolojiyi göstermektedir.



Şekil 1: Karayolu taşımacılığında müşteri analitiği için tasarlanan metodoloji.

4.1 Veri ön İşleme Aşaması

Veri ön işleme aşaması, ham verilerin kullanılabilir veriye dönüştürülmeden önce hazırlanması için gereken ilk aşamadır. Bu aşamada ham veriler üzerinde doldurma, ayıklama ve dönüştürme işlemleri gerçekleştirilir. Bu dönüşüm aşaması, değerlerin düzenlenmesi, aykırı değer tespiti, normalizasyon gibi yöntemlerin kullanılmasını içerir. Bu aşama ile tahmin doğruluğu artar ve modellerin uygulanma sürelerinde düşüş gözlemlenir.

4.2 Analiz Aşaması

Veri temizleme ve dönüşümünden sonraki adım modelleme adıdır. Bu adımda farklı modeller veri kümesi üzerinde denenerek doğruluğu en yüksek olan model seçilir. Analiz aşamasında öncelikle RFM analizi gerçekleştirilmiştir. RFM skor değerleri oluşturularak ağırlıkları değiştirilerek farklı senaryolar belirlenmiştir.

K- ortalamalar analizi ile kümeleme gerçekleştirilmiştir. Daha sonra tahminleme için kestirimsel analitik yöntemleri olarak CHAID , CART ve doğrusal regresyon kullanılmıştır.

4.3 Tahmin Performans Değerlendirme Aşaması

Analiz aşamasının sonunda, kümeleme analizinde farklı senaryolardan farklı kurallar elde edilmiştir. Kar marjı tahmininde etki seviyesi en yüksek küme belirlenmiştir. Sınır ağları ile kar marjı hedefli en etki seviyesi yüksek küme seçilmiştir. Bu küme değişkenini kullanarak, kestirimsel analitik tekniklerinin performansları değerlendirilmiştir. Bu aşamada kar marjını tahmin eden en iyi yöntem ve senaryo MAPE ve MSE ile karşılaştırılmıştır.

4.4 Kullanılan Değişkenler

1. Rota kodu: Taşımanın yapılacağı iller arasındaki kodu belirtir. Çıkış şehir kodu ve varış şehir kodu olmak üzere ikiye ayrılır. (Nominal)
2. Plan Km: Sefere ait toplam km.(Sürekli)
3. Araç Tipi: kamyon- kırkayak- tır şeklinde 3 farklı araç cinsi bulunmaktadır. (Nominal)
4. Tonaj düzeyi: Taşınan yük miktarına göre ordinal veri olarak dönüşümünü içeren değişkendir. Yük seviyeleri üçe ayrılmıştır.(Ordinal)
5. Ürün tipi: Yükün cinsi. Saç, tarım ürünü, boya, gıda ve cam olmak üzere nominal olarak belirlenmiştir.(Nominal)

6. Yükü yükleme tipi: Yük palet mi yoksa dökme şeklinde mi araca yüklenecek olarak nominal değişkene dönüştürülmüştür. (Nominal)
7. Müşteri ölçek seviyesi: Müşterileri büyük, orta küçük ölçekli olarak ayıran ordinal değişkeni tanımlanmıştır. (Ordinal)
8. Müşteri parasallık değeri: Müşterinin sistemde ne kadarlık para harcadığını ölçen bir sürekli değişken olarak tanımlanmıştır. Sistemde bıraktığı toplam miktarı ifade eder. Müşteri bağlılığını ölçen bir değişken olarak tanımlanabilir.(sürekli)
9. Müşteri sıklığı: Müşterinin firmayla yaptığı sipariş sayısını ifade etmektedir.(sürekli)
10. Müşteri güncelliği: Müşterinin firmayla yakın zamanda yaptığı siparişleri göstermektedir. En son ne zaman sipariş verdiği bilgisini içerir.(Sürekli)
11. RFM skoru: RFM analizine göre elde edilen puan değişkenidir. Her bir değer için 1-5 arasında bir skor oluşturulmuştur. 5 en yüksek seviyeyi 1 en düşük seviye olarak görülmektedir.
12. Müşteri segmentasyonu: Kümeleme analizi sonucu elde edilen kar marjını en iyi tahmin eden kümeyi ifade eden değişkendir.(Nominal)
13. Taşıyıcı Popülasyonu: Teklif geldiğinde o rota için mevcut durumda olan taşıyıcı sayısını ifade eder.(Sürekli)

Tablo 1’de bir müşteri verisinin örneği verilmiştir.

Tablo 1: Bir müşteri verisinin örneği.

Değişken	Değeri
Çıkış şehir kodu	34100
Varış Şehir kodu	54200
Verilen Kar marjı	0,095%
Ürün tipi	Gıda
Müşteri Sıklık Değeri	235
Müşteri Güncelliği	61
Müşteri Parasallık Değeri	1,418,155 (TL)
RFM skoru	550
Yük yükleme tipi	Palet
Müşteri segmentasyonu	2
Planlanan km	145 (km)
Tonaj Düzeyi	2
Araç tipi	Kamyon
Taşıyıcı Popülasyonu	27
Müşteri Ölçek Seviyesi	3

5. RFM VE KÜMELEME ANALİZİ UYGULAMASI

Uygulama Türkiye’de karayolu taşımacılığı yapan bir firmada gerçekleştirilmiştir. Firma taşıyıcılarla iş birliği içerisinde. Müşteri taşıma ile ilgili hizmet almak istediğinde bu firmaya başvurmaktadır. Firma taşıyıcıdan taşıma için fiyat alır buna bir kar marjı ekleyerek müşteriye sunar. Karayolu taşımacılığı ülke içinde taşımacılık alanında büyük bir yüzdeye sahiptir. Bu anlamda bakıldığında öneri sisteminin tasarlanması stratejik bir planın yapılması olası zararlardan korunma açısından çok önemlidir. Şekil 2 karayolu taşımacılığındaki tüm paydaşları özetlemektedir. Uygulama kümeleme ve RFM analizi ile başlamaktadır, kestirimsel analitik yöntemleri ile devam etmektedir.



Şekil 2. Karayolu taşımacılığındaki tüm paydaşlar

5.1 Kümeleme Analizi

Kümeleme algoritmaları ile küme içindeki benzerliğin maksimize, kümeler arası benzerliğin minimize edilmesi temeline dayanır. Bu açıdan bakıldığında geniş verileri için tanımlayıcı veriler belirlenir ve yığındaki benzer özellikli kümelerin elde edilmesi amaçlanır. RFM analizi ve müşteri değişkenleri dikkate alarak değişik senaryolar oluşturulmuştur, hangi senaryonun yani hangi kümenin kar marjı tahmininde üstün olduğu tahmin edilmeye çalışılmıştır. Çalışma kapsamında K-ortalama yöntemini kullanılmıştır. Dunn indeksi kullanılarak küme sayısı belirlenmiştir. Dunn indeksi (DI), sıklık teşhisini ve ayrık küme kullanımını önerir. Buna göre en optimum küme sayısı 4 olarak belirlenmiştir. Küme sayısı hesaplanırken yoğun ve iyi dağılmış olması beklenmektedir. Dunn indeksi ile hesaplanan D katsayısı büyüdükçe, küme kalitesi artmaktadır ve D değeri yükseldikçe en uygun küme sayısına yaklaşılmaktadır (Pakhira, Bandyopadhyay&Maulik,2004). Tablo 2’de Dunn indeks değerleri gösterilmektedir.

Tablo 2. K-Ortalama algoritması için küme sayısının Dunn İndeks ile belirlenmesi

Küme sayısı	Dunn
K=2	0,35
K=3	0,45
K=4	0,72
K=5	0,32

5.2 RFM Analizi

Çalışmada segmentasyon için RFM (Recency, Frequency and Monetary) analizinden faydalanılmıştır. Bu sayede müşterilerin hangi kar marjı seviyelerine eğilimli olduğu gözlemlenebilir. RFM analizi müşterilerin geçmişte yapmış oldukları alışveriş tarihinin yakınlığına, yapmış oldukları alışveriş miktarına ve harcama tutarına göre müşterilerin olası karı kabul etme eğilimlerini belirlenmiştir. Bu üç etkenin kullanılması ile bir RFM değeri hesaplanır. Daha sonra bu değerler sınıflandırarak belirli skorlara dönüştürülür. RFM skorları 1 ile 5 arasında değişmektedir. RFM skoru için güncellik, frekans ve parasallık skoru oluşturulur. Her bir değişken ağırlıklandırılmıştır. Ağırlıklar hem skor değerlerine hem de RFM değerine verilmektedir. Güncellik, Frekans ve Parasal skor ve değerlerinin ağırlıkları değiştirilerek farklı 52 küme senaryoları belirlenmiştir. RFM skor ve değer ağırlıkları, 100,10 ve 1 olarak verilmiş, değişkenlerin farklı ağırlıklarına göre senaryolar oluşturulmuştur.

5.3 Senaryoların Oluşturulması

Bu çalışma kapsamında, müşteri özelliklerini ele alan kümeleme analizi yapılmıştır. Daha sonrasında yapay sinir ağlarında kümenin kar marjını tahmin etme gücü test edilerek, kar marjını tahmininde en güçlü küme elde edilmeye çalışılmıştır. Daha sonraki tahminleme aşamasında segmentasyon analizi bir değişken olarak ele alınarak kestirimsel analitik metotlarında kullanılmıştır. Tablo 3 'de kümeleme analizinde kullanılan senaryolar ve küme etki seviye değerleri belirtilmiştir.

RFM değeri ve skoruna ek olarak müşteri ölçek seviyesi değişkeni modele eklenerek senaryolar oluşturulmuştur. Buna göre en iyi senaryonun 37 numaralı senaryo olduğu gözlemlenmektedir.

5.4 Kümeleme Analizi Performansının Değerlendirilmesi

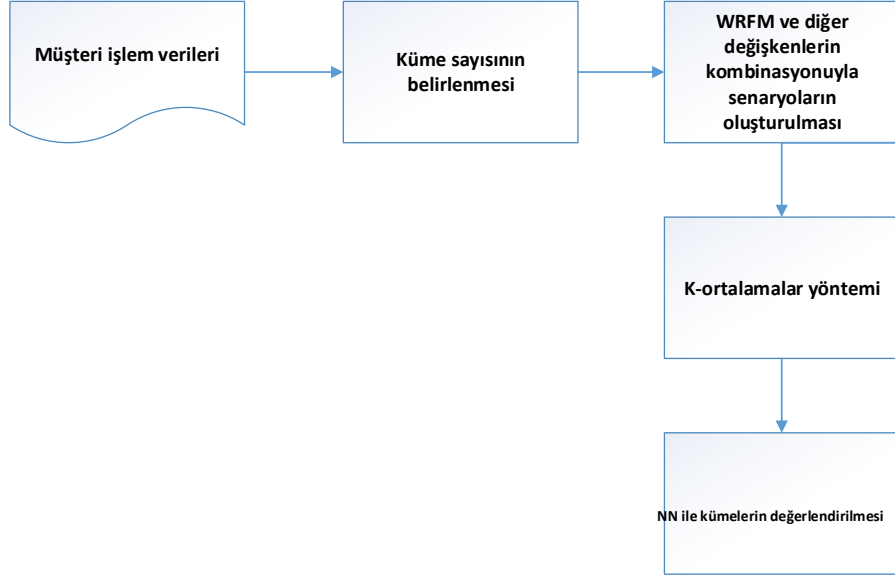
Bu aşamada kümeleme analizinin senaryoların performansını değerlendirilmesi amaçlanmıştır. Böylelikle oluşturulan birçok kümeleme analizi senaryolarının hangisinin müşteriye verilecek olan kar marjını tahmin etmede üstünlük sağladığı hesaplanabilir. Bu seçilen küme tahminleme yöntemlerinde değişken olarak kullanılarak tahmin performansı artırılabilir. Sinir Ağları (NN) mantığına dayanan bu aşama sayesinde kümeleme performansı sonuçlarımızı karşılaştırmamızı sağlar.

NN'ler sinir sisteminin temel birimleri, tipik olarak katmanlar halinde

düzenlenmiş nöronlardır. Bazen çok katmanlı bir algılayıcı olarak adlandırılan bir NN, temel olarak çok sayıda birbirine bağlı nöronlardan oluşan birimlerini simüle ederek çalışır. İşleme birimleri katman halinde düzenlenmiştir. Bir NN'de tipik olarak üç bölüm vardır: giriş katmanı; bir veya daha fazla gizli katman; ve son olarak çıkış katmanından oluşur. NN analizinin rolü, farklı faktörlere göre belirlenen kümelerin belirlenen hedef doğrultusunda anlamlı etkilere sahip olup olmadığının belirlenmesidir.

Başka bir deyişle, NN, belirtilen bir veri setinin müşterisine verilen kar marjlarını tahmin etmede en önemli faktörleri bulma avantajlarına sahiptir. Çalışmanın müşteri kümeleme analizi kısmının amacı, müşteriye verilecek kar marjını etkileyen en önemli özellik ve nitelikleri bulmaktır. Bu arada, farklı özelliklerden faydalanan farklı kümelerin etkilerini kar marjına olan etkilerini ölçebiliriz. Etki seviyesi olarak NN'dan çıkan sonuçlar ile en güçlü performans gösteren kümeler elde edilebilmektedir (Sarvari, Ustundag & Takci 2016).

Bu yaklaşımı kullanmak, farklı senaryolardan oluşturulmuş farklı kümelerin etkilerini ölçmeye olanağı sağlar. Küme etki derecelerinin yüksekse, kümelerimizin yeterince doğru değişken kombinasyonlarına sahip olduğuna emin olabiliriz; Aksi takdirde, küme gücünün tahminlemede düşük olduğunu doğru değişken kombinasyonlarına sahip olmadığını söyleyebiliriz. Şekil (3) kümeleme analizi aşamalarını göstermektedir.



Şekil 3. Kümeleme analizi aşamaları.

Tablo 3. WRFM analizi, oluşturulan senaryolar ve etki analizi sonuçları

	RFM ağırlıkları	RFM skor	RFM değeri	Müşteri Ölçek seviyesi Değişkeni	Etki Seviyeleri
S1	$W_R > W_F > W_M$	√			0,0212
S2	$W_F > W_R > W_M$	√			0,0253
S3	$W_M > W_F > W_R$	√			0,0156
S4	$W_R > W_M > W_F$	√			0,0197
S5	$W_M > W_R > W_F$	√			0,0098
S6	$W_F > W_M > W_R$	√			0,0343
S7	$W_R = W_F = W_M$	√			0,0165
S8	$W_R = W_F > W_M$	√			0,0188
S9	$W_F = W_M > W_R$	√			0,0454
S10	$W_R = W_M > W_F$	√			0,0344
S11	$W_R > W_F = W_M$	√			0,0576
S12	$W_M > W_F = W_R$	√			0,0465

S13	$W_F > W_R = W_M$	√			0,0322
S14	$W_R > W_F > W_M$		√		0,0567
S15	$W_F > W_R > W_M$		√		0,0354
S16	$W_M > W_F > W_R$		√		0,0634
S17	$W_R > W_M > W_F$		√		0,0776
S18	$W_M > W_R > W_F$		√		0,0222
S19	$W_F > W_M > W_R$		√		0,0633
S20	$W_R = W_F = W_M$		√		0,0465
S21	$W_R = W_F > W_M$		√		0,0155
S22	$W_F = W_M > W_R$		√		0,0468
S23	$W_R = W_M > W_F$		√		0,0688
S24	$W_R > W_F = W_M$		√		0,0058
S25	$W_M > W_F = W_R$		√		0,0561
S26	$W_F > W_R = W_M$		√		0,0248
S27	$W_R > W_F > W_M$	√		√	0,0681
S28	$W_F > W_R > W_M$	√		√	0,0576
S29	$W_M > W_F > W_R$	√		√	0,0766
S30	$W_R > W_M > W_F$	√		√	0,0297
S31	$W_M > W_R > W_F$	√		√	0,0335
S32	$W_F > W_M > W_R$	√		√	0,0765
S33	$W_R = W_F = W_M$	√		√	0,0534
S34	$W_R = W_F > W_M$	√		√	0,0118
S35	$W_F = W_M > W_R$	√		√	0,0667
S36	$W_R = W_M > W_F$	√		√	0,0326
S37	$W_R > W_F = W_M$	√		√	0,0983
S38	$W_M > W_F = W_R$	√		√	0,0944
S39	$W_F > W_R = W_M$	√		√	0,0324
S40	$W_R > W_F > W_M$		√	√	0,0543
S41	$W_F > W_R > W_M$		√	√	0,0898
S42	$W_M > W_F > W_R$		√	√	0,0433
S43	$W_R > W_M > W_F$		√	√	0,0098
S44	$W_M > W_R > W_F$		√	√	0,0132
S45	$W_F > W_M > W_R$		√	√	0,0787

S46	$W_R=W_F=W_M$		√	√	0,0034
S47	$W_R=W_F>W_M$		√	√	0,0054
S48	$W_F=W_M>W_R$		√	√	0,0065
S49	$W_R=W_M>W_F$		√	√	0,0031
S50	$W_R>W_F=W_M$		√	√	0,0910
S51	$W_M>W_F=W_R$		√	√	0,0906
S52	$W_F>W_R=W_M$		√	√	0,0970

Buna göre kar marjları tahmininde kullanılan en güçlü küme senaryo 37 olarak belirlenmiştir. Bu küme, müşteri segmentasyonu olarak ayrı bir değişken olarak ele alınıp, diğer tahminleme yöntemlerinde kullanılmıştır.

5.5 En iyi Müşteri Küme Senaryosunun Analizi

Müşteri kümeleme analizi sonucuna göre 4 küme oluşturulmuştur. Tablo 4’de müşteri segmentlerinin kar marjı ortalamaları verilmiştir. Buna göre en yüksek kar marjına eğilimli küme 4 numaralı olarak belirlenmiştir. Müşteri bilgilerinin ortalama değerleri Tablo 4’de belirtilmiştir. Buna göre en yüksek kar marjı uygulanan 4 numaralı küme hem sistemde en aktif, en fazla parasal değerde firma ile işbirliği yapan, en yakın zamanda sipariş vermiş müşteri grubu olduğunu söylemek mümkündür. Bu müşteri grubuna yüksek kar marjı atanmaktadır.

Tablo 4. Etki seviyesi en yüksek müşteri kümesi ve özellikleri.

Küme	Ortalama kar marjı değeri (%)	Yüzdesi %	Yenilik Skor Değeri	Sıklık Skor Değeri	Parasallık Skor Değeri
1	0,072	26	3,3	2,8	4,2
2	0,051	28,7	2,7	2,4	2,8
3	0,035	29,8	1	1	2,2
4	0,105	15,5	5	5	5

6. KESTİRİMSSEL ANALİTİK YÖNTEMLERİ

Kestirimsel analitik (Predictive Analytics) modelleri uygulamasında, segmentasyon aşamasında elde edilen müşteriye özgü küme değişkeni modele eklenerek modellerin performanslarının karşılaştırılması yapılmıştır. Böylelikle küme değişkeninin model performanslarına etkisi var mı analiz edilmiştir.

Yöntem olarak CART ve CHAID, ve doğrusal regresyon kullanılmıştır. Müşterilere uygulanması beklenen kar marjı tahminleri için modelde mevcut verinin %70'si eğitim %30'si test verisi olarak kullanılmış ve modellerin geçerliliği test edilmiştir. Tahmin modellerinin oluşturulmasında SPSS Clementine yazılımında kestirimsel analiz algoritmalarından, CHAID karar ağacı, CART karar ağacı ve doğrusal regresyon kullanılmıştır. Kullanılan veri 630 müşteriye ait 15.742 veriden oluşmaktadır. Veri pek çok değişken içermektedir. Kar marjını tahmini için en önemli verilerin seçimi SPSS Clementine programındaki özellik seçimi modülü ile gerçekleştirilmiştir. Özellik seçimi (diğer adıyla nitelik seçimi veya değişken seçimi), kullanılan algoritmaya göre özellikleri değerlendirerek veri setindeki n adet özellik arasından en iyi k adet özelliği seçme işlemidir.

Şekil 4 'te özellik seçimi modülü kullanılarak kar marjına etki eden önemli değişkenler seçilmiştir. Bu analiz sonunda uğrak noktası değişkeni önemli bulunmamıştır. Bu yüzden modellemede kullanılmamıştır.

Rank	Field	Type	Importance	Value
1	CustomerScaleCode	Ordered Set	Important	1,0
2	DepartureCityCode	Set	Important	1,0
3	FreightTypeCode	Set	Important	1,0
4	ArrivalCityCode	Set	Important	1,0
5	CustomerFrequency	Range	Important	1,0
6	TruckersPopulation	Range	Important	1,0
7	CustomerRevenue	Range	Important	1,0
8	EstimatedDistance	Range	Important	1,0
9	TonnageLevel	Ordered Set	Important	1,0
10	VehicleTypeCode	Set	Important	1,0
11	FreightLoadingTypeCode	Set	Important	0,994
12	CustomerRecency	Range	Important	0,98
13	WayStop	Ordered Set	Unimportant	0,2

Selected fields:12 Total fields available:16

> 0,95 <= 0,95 < 0,9

3 Screened Fields

Şekil 4. Özellik seçimi ile belirlenen önemli değişkenler.

Modelin çıktısı müşteriye verilen kar marjıdır. Tüm seçilen değişkenler girdi olarak kullanılmıştır. Buna ek olarak müşteri kümeleme analizinden elde edilen küme değişkeni de modele eklenmiştir. Küme değişkeni eklenmiş ve eklenmemiş

senaryoların farklı modellerdeki tahmin performansı karşılaştırılmıştır. Her bir model için değişken önem dereceleri hesaplanmış, o model için anlamlı olan ve kullanılan değişkenler grafiksel olarak gösterilmiştir. Modeller iki farklı yaklaşımda, iki farklı hata oranı ile karşılaştırılmıştır.

6.1 Modellerin Karşılaştırılması

Modeller iki farklı şekilde karşılaştırılmıştır. Bunlar MSE ve MAPE'dir.

6.1.1 Ortalama kare hata

Ortalama kare hata (MSE) aşağıdaki Eş. 9 ile hesaplanır (Hyndman & Koehler, 2006).

$$MSE = \frac{1}{n} \sum_{j=1}^n e_j^2 \quad (9)$$

Ortalama kare hata ile makine öğrenme modelinin tahminleyicinin performansını ölçmede oldukça etkilidir. Bu hata oran değerlerinin sıfıra yakın olması tahminleyicinin daha iyi performansa sahip olduğunu gösterir.

6.1.2 Ortalama mutlak yüzde hata

Ortalama mutlak yüzde hata (MAPE) aşağıdaki Eş. 10 ile hesaplanır.

$$MAPE = \frac{100}{N} \sum_{i=1}^N \left| \frac{x_i - \hat{x}_i}{x_i} \right| \quad (10)$$

- $\{x_i\}$: Gerçek gözlem verisi
- $\{\hat{x}_i\}$: Tahmin edilen veya öngörülen veri
- N : Veri sayısı

Modellerin performansları karşılaştırılırken test verisi üzerindeki hata oranları dikkate alınmıştır. Tablo 5 ve 6 'da kullanılan tahminleme algoritmalarının farklı senaryolara göre karşılaştırmaları gösterilmektedir.

Tablo 5. Küme değişkeni dahil edilen tahminleme algoritmalarının karşılaştırılması.

	MAPE (%)	MSE (%)
Doğrusal Regresyon	14,17	14,42
CHAID	6,59	7,51
CART	8,29	9,87

Tablo 6. Küme değişkeni dahil edilmeyen tahminleme algoritmalarının karşılaştırılması.

	MAPE (%)	MSE (%)
Doğrusal Regresyon	16,21	17,45
CHAID	8,02	9,02
CART	9,34	10,81

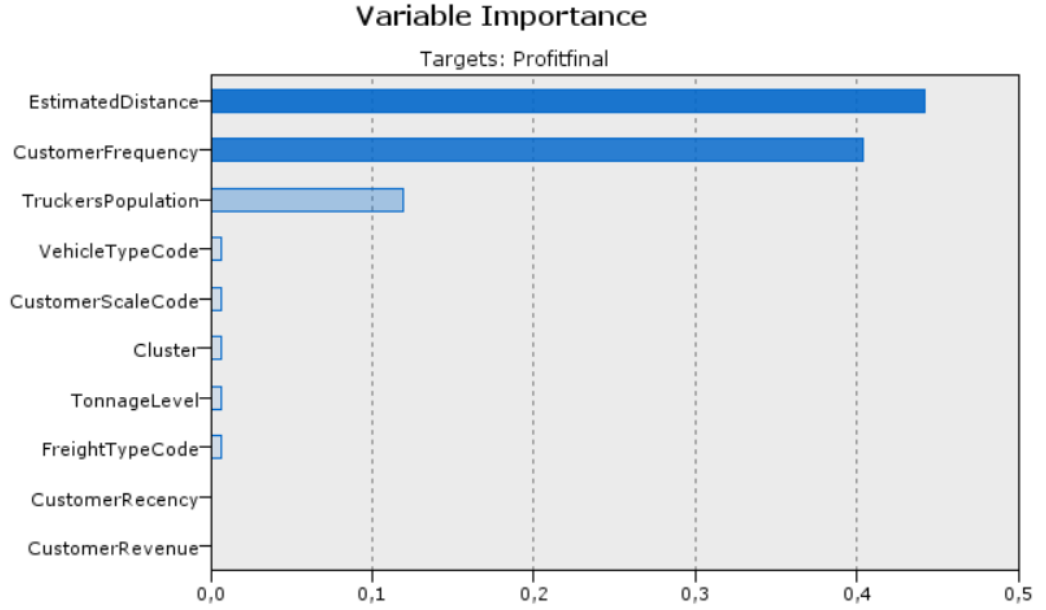
Tüm modeller için MAPE değerleri MSE değerlerinden düşük çıkmıştır. İki farklı yaklaşımdaki tüm modellere baktığımızda tahminlemedeki en iyi yaklaşımın kümeleme değişkeninin de kullanıldığı CHAID ile olduğu belirlenmiştir.

Küme değişkeninin dahil edilmediği durumda en yüksek tahmin gücüne sahip algoritma CHAID algoritmasıdır. Küme değişkenini dahil ettiğimiz durumda en iyi algoritmanın yine CHAID olduğunu görmekteyiz. Küme değişkeninin modele eklenmesi tüm algoritmalarda tahmin gücü performansını arttırmıştır. Bu yüzden küme değişkeni eklenmiş olan yaklaşım detaylı olarak bölüm 6.2’de çıktılarıyla analiz edilmiştir. Müşterilere belli özelliklerle kar marjını atamada hedef odaklı en iyi kümenin seçilip tahminlemede kullanılması modelin açıklayıcılığını yani tahmin doğruluğunu arttırmaktadır. Yine tahminleme algoritmalarında kar marjı tahminini dikkate alarak değişkenlerin seçilimi ile modelin hata oranını azaltmaktadır. Bu durum kümeleme analizinin doğru bir kombinasyonla yapılmasının önemini vurgulamaktadır. Kümelemede doğru değişken kombinasyonlarını belirleyerek, tahmin edilecek modele dahil etmek tahminlemede iyileştirilmiş sonuçlar verebilmektedir. Bu yöntemler için kar marjı tahmininde etkili ve anlamlı olan değişkenler ve onların ağırlıkları da

belirlenmiştir.

6.2 Doğrusal Regresyon

Doğrusal Regresyon algoritması ile kar marjı tahmininde Şekil 5'teki değişkenler kullanılmıştır ve bu değişkenler kar marjı tahmininde farklı önem derecesine sahiptir. Bunlar planlanan mesafe, müşteri sıklığı ,taşıyıcı popülasyonu ,araç tipi, müşteri ölçek seviyesi, küme değişkeni, tonaj seviyesi, ürün tipidir. Değişkenlerin önem derece katsayıları Tablo 7'da verilmiştir.



Şekil 5. Doğrusal regresyon yöntemi değişken önem dereceleri.

Doğrusal regresyon denklemi ve katsayıları aşağıda belirtilmiştir.

Küme değişkeni * 0,001778 + Araç tipi * -0,002228 + Ürün tipi * 0,000001676 + Taşıyıcı Popülasyonu * 0,00001239 + Planlanan km * -0,00002303 + Tonaj seviyesi * -0,003041 + Müşteri ölçek seviyesi* 0,002154 + Müşteri geliri * 0,00000001692 + Müşteri Güncelliği* - 0,00000686 + Müşteri Sıklığı * 0,00003434 + 0,08314

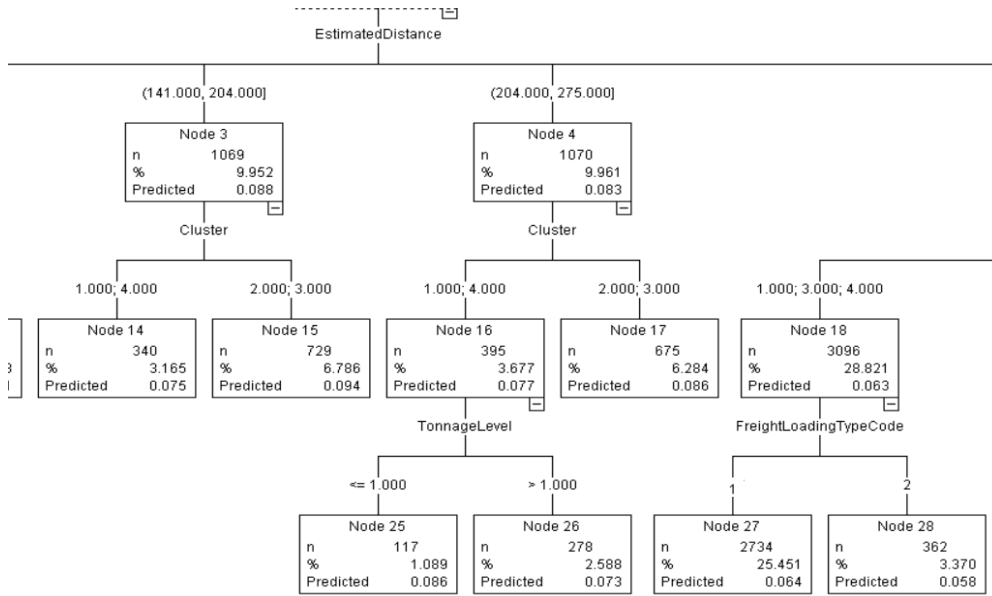
Tablo 7. Doğrusal regresyon yöntemine göre değişkenlerin önem derece

katsayıları.

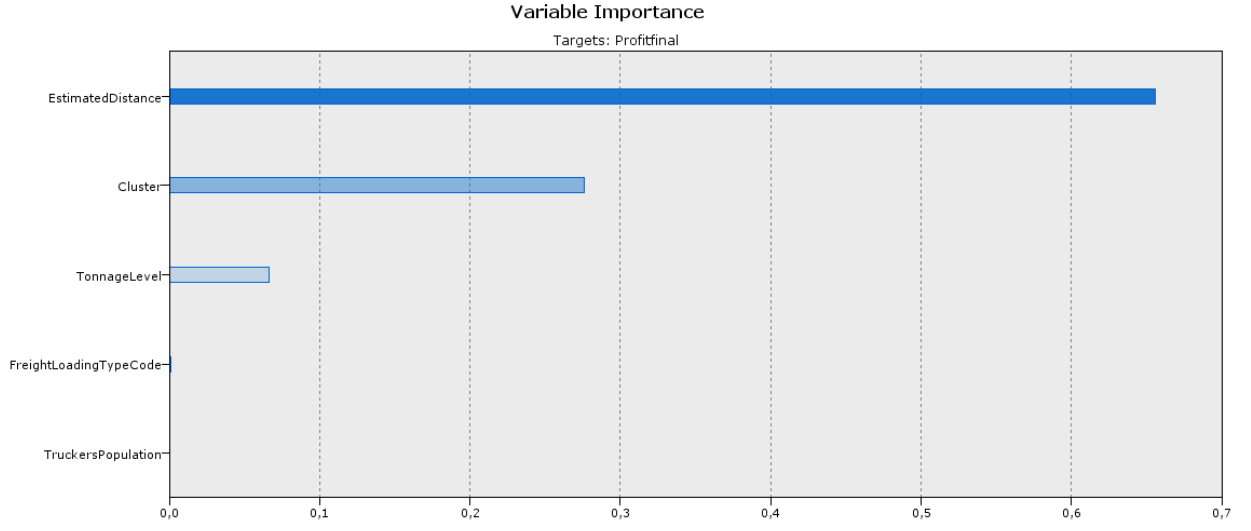
Değişken	Önem Derecesi
Yük tipi	0,007
Tonaj Seviyesi	0,007
Küme Değişkeni	0,007
Müşteri Ölçek Seviyesi	0,007
Araç Tipi	0,007
Taşıyıcı Popülasyonu	0,1193
Müşteri Sıklığı	0,404
Planlanan km	0,4418

6.3. CHAID

CHAID karar ağacı algoritmasında kar marjı tahmininde Şekil 7’teki değişkenler önem derecesine sahiptir. Küme değişkeni burada önemli bir ağırlığa sahiptir. Şekil 6 ‘da karar ağacının bir kısmı gösterilmektedir.



Şekil 6. Kar marjı tahmini için CHAID karar ağacının bir kısmı.



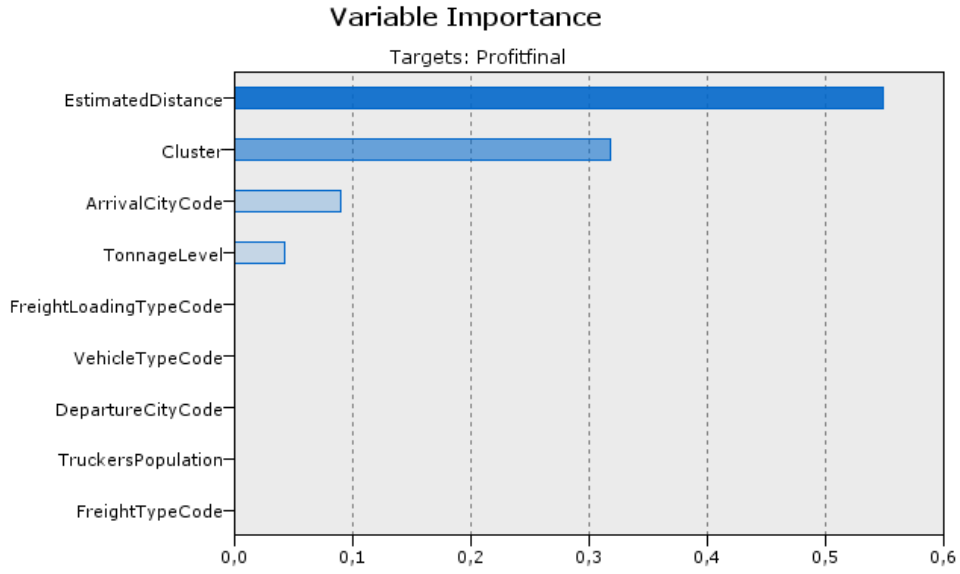
Şekil 7. CHAID yöntemi değişken önem dereceleri.

Tablo 8. CHAID yöntemine göre değişkenlerin önem derece katsayıları.

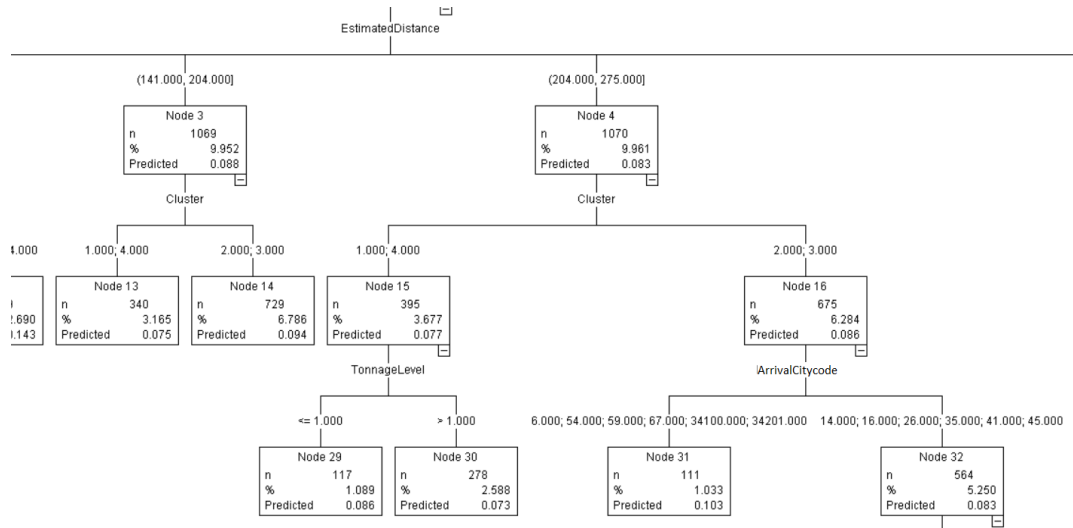
Değişkenler	Önem Derecesi
Yük yükleme tipi	0,0012
Tonaj Seviyesi	0,0665
Küme Değişkeni	0,2763
Planlanan km	0,656

6.3 CART

CART karar ağacı algoritmasında, kar marjı tahmininde önemli olan ve modelde kullanılan değişkenler şekil 8 'deki gibidir. Bunlar planlanan km, küme değişkeni, varış şehri ve tonaj seviyesidir. Karar ağacının bir kısmı şekil 9 'da gösterilmektedir.



Şekil 8. CART yöntemi değişken önem dereceleri.



Şekil 9 : Kar marjı tahmini için CART karar ağacının bir kısmı.

Tablo 9. CART yöntemine göre değişkenlerin önem derece katsayıları.

Değişkenler	Önem Derecesi
Tonaj Seviyesi	0,0424
Variş şehri	0,0898
Küme Değişkeni	0,3189
Planlanan km	0,5489

7. TARTIŞMA VE SONUÇ

Lojistik sektörü hızlı büyüyen bir konumdur. Karayolu taşımacılığı ise lojistik sektöründe ciddi bir paya sahiptir. Firmaların müşterilerini analiz etmesi, gerekli analitik yöntemlerinin kullanılması karlılığı arttırma ve uzun vadede strateji geliştirme için çok önemlidir. Bu çalışmada müşteriye uygulanacak olan kar marjı tahminleme problemi dikkate alınmıştır.

Bu çalışmada lojistik sektörü, karayolu taşımacılığında müşteri analitiği çalışması gerçekleştirilmiştir. Kümeleme ve veri madenciliğinin yöntemleri kullanılmıştır. Müşteri profilini ve müşteri türlerini sistematik olarak incelemeye imkanı sağlayan bir metodoloji geliştirilmiştir. Böylece her bir müşteri türü için pazarlama stratejileri geliştirmek kolaylaşmaktadır. Bu sayede değerli müşterileri etkin bir şekilde hedefleme ve farklı pazarlama stratejileri uygulama konusunda fayda sağlar.

Geliştirilen hibrit metodolojide öncelikle müşteri segmentasyonu aşaması ilk basamak olarak tanımlanmıştır. Segmentasyon performansını zenginleştiren RFM analizi ve müşteri değişkenleri kullanılmıştır. K-ortalamlar yöntemi ile kümeleme analizi gerçekleştirilmiştir. Farklı senaryolar tasarlanmış, uygulanmış ve küme etki sonuçları sinir ağları kullanılarak değerlendirilmiştir. Müşteri bazlı kar marjı tahminleme için kestirimsel analiz yöntemleri kullanılmıştır. CHAID, CART ve doğrusal regresyon kullanılmıştır. Yöntemler iki farklı yaklaşımla uygulanmıştır. Bunlardan birincisi küme değişkeni kullanılmadan, ikincisi ise küme değişkeninin modele dahil edilmesiyle kar marjını tahmin performansları karşılaştırılmıştır.

Bu çalışma, WRFM analizinin ile müşteri segmentasyonunun, tahminleme performansını olumlu yönde etkilediğini göstermektedir. Ayrıca, daha doğru ve etkili müşteri segmentleri elde etmek için, WRFM ve diğer müşteri değişkenlerinin kombinasyonlarının gerçekleştirilmesinin önemi ortaya çıkartılmıştır. Buna göre, önerilen hibrit metodolojide en iyi tahmin

performansını küme değişkeni eklenen CHAID modelinde olduğu belirlenmiştir. Ayrıca doğru WRFM ve müşteri verilerinin kombinasyonu kullanılarak oluşturulan kümenin önemi kanıtlanmıştır.

Yöneticiler bu hibrit metodolojiyi temel alarak müşteri odaklı, toplam kar marjını arttıran etkili kampanyalar hazırlayabilir ve daha doğru müşteri profilleri belirleyebilir. Karayolu taşımacılığı sektöründe müşteri analitiği uygulaması sayesinde müşteri kaybı minimize edilebilecek, buna bağlı olarak uygulanacak olan kar marjı belirlenmesi sistematik kurallara bağlanacaktır. Geliştirilen model sayesinde, işletmeler müşterileri bireysel olarak tanıyıp onlara yönelik kar marjını teklifini geliştirme fırsatı sunacaklardır. Yöneticiler, geliştirilen model ile lojistik sektöründeki kar eğilimlerini gözlemleyip bu bilgi doğrultusunda pazarlama yapabileceklerdir. Müşteri analitiği sayesinde müşteri istek ve ihtiyaçları hakkındaki bilgilerin detaylı analizi ile müşterilerin ihtiyaç ve isteklerine özgü kar marjlarının tahmini mümkün olmaktadır. Müşteri hakkında doğru bilgilerin elde edilmesi, müşteri kaybını en azlayacaktır. Doğru müşteri profillerinin belirlenmesi ile firmalar uzun vadede karlılık politikasını analiz edebilecektir. Türkiye’de karayolu taşımacılığının önemi günden güne artmaktadır. Bu önerilen hibrit sistem karayolu taşımacılığında kar marjının belirlenmesi için bir öneri sistemi olarak uygulanabilir.

İleriki çalışmalarda, tahminlemede rastgele orman (random forest) , XGboost, destek vektör makinaları (SVM), ANN gibi makine öğrenme algoritmaları ile farklı kümeleme teknikleri (iki aşamalı, Kohonen- SOM) kullanılabilir.

TEŞEKKÜR

Bu çalışmanın bir kısmı TÜBİTAK tarafından 1059B141600618 kodu ile desteklenmiştir.

8. KAYNAKÇA

- Abbott, D. (2014). *Applied predictive analytics: Principles and techniques for the professional data analyst*. John Wiley & Sons, 213-223.
- Ambler, T., Bhattacharya, C. B., Edell, J., Keller, K. L., Lemon, K. N., & Mittal, V. (2002). Relating brand and customer perspectives on marketing management. *Journal of Service Research*, 5(1), 13-25.
- Azeem, M., Usman, M., & Fong, A.C.M. (2017). A churn prediction model for prepaid customers in telecom using fuzzy classifiers. *Telecommunication Systems*, 66 (4), 603-614.

- Boix, M.P., Moreno, J.L.M. (2018). A social model based on customers' profiles for analyzing the churning process in the mobile market of data plans. *Physica A: Statistical Mechanics and its Applications*, 496, 571-592.
- Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D. T., ... & Ma, J. (2017). A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena*, 151, 147-160.
- Cheng, C. H., & Chen, Y. S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. *Expert systems with applications*, 36(3), 4176-4184.
- Chan, C. C. H. (2008). Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer. *Expert systems with applications*, 34(4), 2754-2762.
- Chang, P. T., Hung, L. T., Pai, P. F., & Lin, K. P. (2013). Improving project-profit prediction using a two-stage forecasting system. *Computers & Industrial Engineering*, 66(4), 800-807.
- Chen, Y., & Hao, Y. (2017). A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction. *Expert Systems with Applications*, 80, 340-355.
- Chung, Y. C., & Chen, S. J. (2016). Study on customer relationship management activities in Taiwan tourism factories. *Total Quality Management & Business Excellence*, 27(5-6), 581-594.
- Coussement, K., Van den Bossche, F. A., & De Bock, K. W. (2014). Data accuracy's impact on segmentation performance: Benchmarking RFM analysis, logistic regression, and decision trees. *Journal of Business Research*, 67(1), 2751-2758.
- Deconinck, E., Hancock, T., Coomans, D., Massart, D. L., & Vander Heyden, Y. (2005). Classification of drugs in absorption classes using the classification and regression trees (CART) methodology. *Journal of Pharmaceutical and Biomedical Analysis*, 39(1-2), 91-103.
- Hadden, J., Tiwari, A., Roy, R., & Ruta, D. (2007). Computer assisted customer churn management: State-of-the-art and future trends. *Computers & Operations Research*, 34(10), 2902-2917.
- Hans-Hermann, B. O. C. K. (2008). Origins and extensions of the k-means algorithm in cluster analysis. *Journal Electronique d'Histoire des Probabilités et de la Statistique Electronic Journal for History of Probability and Statistics*, 4(2).
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4), 679-688.

- Hong, T., & Kim, E. (2012). Segmenting customers in online stores based on factors that affect the customer's intention to purchase. *Expert Systems with Applications*, 39(2), 2127-2131.
- Jain, V. K. (2017). *Big Data and Hadoop*. Khanna Publishing.
- Hwang, H., Jung, T., & Suh, E. (2004). An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. *Expert systems with applications*, 26(2), 181-188.
- Khajvand, M., Zolfaghar, K., Ashoori, S., & Alizadeh, S. (2011). Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study. *Procedia Computer Science*, 3, 57-63.
- Larivière, B. , & Van den Poel, D. (2005). Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert Systems with Applications*, 29 (2), 472-484.
- Liu, H. H., & Ong, C. S. (2008). Variable selection in clustering for marketing segmentation using genetic algorithms. *Expert Systems with Applications*, 34(1), 502-510.
- Lin, C. L., & Fan, C. L. (2019). Evaluation of CART, CHAID, and QUEST algorithms: a case study of construction defects in Taiwan. *Journal of Asian Architecture and Building Engineering*, 18(6), 539-553.
- Martínez, A., Schmuck, C., Pereverzyev Jr, S., Pirker, C., & Haltmeier, M. (2018). A machine learning framework for customer purchase prediction in the non-contractual setting. *European Journal of Operational Research*.
- McCarty, J. A., & Hastak, M. (2007). Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression. *Journal of business research*, 60(6), 656-662.
- MacQueen. (1967) Some methods for classification and analysis of multivariate observations. *Proc. 5th Berkeley Symp. Math. Stat. Prob.* 281.
- Ngai, E. W., Xiu, L., & Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications*, 36(2), 2592-2602.
- Nakano, S., & Kondo, F. N. (2018). Customer segmentation with purchase channels and media touchpoints using single source panel data. *Journal of Retailing and Consumer Services*, 41, 142-152.
- Nyitrai, T., & Virág, M. (2019). The effects of handling outliers on the performance of bankruptcy prediction models. *Socio-Economic Planning Sciences*, 67, 34-42.
- Oğuzlar, A. (2010). Cart analizi ile hanehalki işgücü anketi sonuçlarının

- özetlenmesi. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 18(3-4).
- Paker, N., & Vural, C. A. (2016). Customer segmentation for marinas: Evaluating marinas as destinations. *Tourism Management*, 56, 156-171.
- Pakhira, M.K., Bandyopadhyay, S. and Maulik, U. (2004), "Validity index for crisp and fuzzy clusters", *Pattern Recognit*, Vol. 37, pp. 487-501.
- Raut, R. D., Mangla, S. K., Narwane, V. S., Gardas, B. B., Priyadarshinee, P., & Narkhede, B. E. (2019). Linking big data analytics and operational sustainability practices for sustainable business management. *Journal of Cleaner Production*, 224, 10-24.
- Sarvari, P. A., Ustundag, A., & Takci, H. (2016). Performance evaluation of different customer segmentation approaches based on RFM and demographics analysis. *Kybernetes*, 45(7), 1129-1157.
- Singh, A., & Tucker, C. S. (2017). A machine learning approach to product review disambiguation based on function, form and behavior classification. *Decision Support Systems*, 97, 81-91.
- Tanaka, T., Hamaguchi, T., Saigo, T., & Tsuda, K. (2017). Classifying and Understanding Prospective Customers via Heterogeneity of Supermarket Stores. *Procedia Computer Science*, 112, 956-964.
- Vincent, O. R., Makinde, A. S., Salako, O. S., & Oluwafemi, O. D. (2018). A self-adaptive k-means classifier for business incentive in a fashion design environment. *Applied Computing and Informatics*, 14(1), 88-97.
- Wei, Y., Zhang, X., Shi, Y., Xia, L., Pan, S., Wu, J., & Zhao, X. (2018). A review of data-driven approaches for prediction and classification of building energy consumption. *Renewable and Sustainable Energy Reviews*, 82, 1027-1047.
- Wei, J. T., Lin, S. Y., Yang, Y. Z., & Wu, H. H. (2019). The application of data mining and RFM model in market segmentation of a veterinary hospital. *Journal of Statistics and Management Systems*, 22(6), 1049-1065.
- Xu, L., & Chu, H. C. (2016). Big Data Analytics Toward Intelligent Mobile Service Provisions of Customer Relationship Management in ECommerce. *Journal of Computers*, 26(4).