# Multinominal Lojistik Regresyon Analizi ile

# Temerrüt Olasılığı Tahminlerine

# Yapılandırmanın Dâhil Edilmesi

**Seda Bağdatlı Kalkan[[1]](#footnote-1) & Celil Taşkın[[2]](#footnote-2)**

### Özet

Bu çalışmada özel bir bankanın verileri kullanılarak bir borçlunun sonraki bir yıl içinde temerrüt mü edeceği, yapılandırılmaya konu mu olacağı veya ödemelerini düzenli mi yapacağının tahminine yönelik istatistiksel bir model geliştirilmiştir. İstatistikî testler, modelin iyi bir tahmin gücünün olduğunu göstermiştir. Araştırma sonuçları göstermiştir ki …. Bu model, veri sağlayan bankanın yapılandırma hususunda etkin bir risk yönetimi faaliyeti icra etmesine yardımcı olacaktır. Yapılandırılan krediler iyi yönetilebilirse sorunlu kredi portföyünün büyümesinin önüne geçilebilir.

Anahtar Kelimeler: Multinominal Lojistik Regresyon, yapılandırma, temerrüt olasılığı

###

### Abstract

Keywords: Multinominal Logistic Regression, re-structuring, probability of default

# 1. GİRİŞ

Bir kredi müşterisinin kredi sözleşmesi şartları doğrultusunda mevcut kaynaklarıyla borcunu ödemeyeceğine dair uyarı sinyalleri gelmeye başladığında bankaların önünde iki ana alternatif bulunmaktadır. Birincisi, yapılan analiz ve incelemeler ve görüşmelerden sonra, kredinin temerrüt edeceğine dair kanaat baskın gelirse yasal takip başlatılacaktır. İkincisi, sözleşme şartlarında değişiklik yapılıp müşteriye farklı bir geri ödeme planı sunulduğu takdirde müşterinin borcunu ödeyebileceğine dair görüş ağırlık kazanırsa kredi yeniden yapılandırılacaktır. Yapılandırma seçeneği çokça tercih edilmektedir. Bankaların internet sitelerinde yayınladıkları denetim raporları incelendiğinde ‘Sözleşme Koşullarında Değişiklik Yapılan’ kredi tutarlarının ‘Takipteki Krediler’ tutarlarından genellikle yüksek olduğu görülür. Yapılandırmaların sıkça yapılıyor olması, risk yönetimi açısından bazı zorluklara yol açabilmektedir. ‘Sözleşme Koşullarında Değişiklik Yapılan’ portföyde doğası gereği Standart Nitelikli kredilere kıyasla daha yüksek bir temerrüt oranı gerçekleşecektir. Bu yüzden bu portföye dair modelleme çalışmaları önem arz etmektedir.

Müşterinin sonraki bir yılda yüzde kaç ihtimalle temerrüt edeceğini tahmin etmek yeterli değildir. Yapılandırmanın temerrütle birlikte ele alınması gerekmektedir. Çünkü nispeten benzer özelliklere sahip ve dışsal bir müdahale olmadığı durumda temerrüt edecek krediler arasından bazıları yapılandırılmakta ve dolayısıyla modelleme çalışmalarına temerrüt vakası olarak girmesi gereken bazı vakalar canlı kredi vakası olarak tahmin modelinde yer alacaktır. Modellerin ayrıştırma gücünün bu durumdan etkileme ihtimali yüksektir. Tahmin modellerine yapılandırmanın dâhil edilmesiyle yapılandırma kararlarının temerrüt tahminlerini olumsuz etkilemesinin önüne geçilebilir.

Yapılandırmanın tahmin modellerine dâhil edilmesinin yollarından birisi, temerrüt tanımının içine yapılandırmanın dâhil edilmesidir. Bir kredi yapılandırıldığında temerrüt etmiş kabul edilebilir. Lakin Türkiye’deki mevzuat, yapılandırmayı zorunlu olarak temerrüt nedenleri arasında saymamaktadır. Bu yüzden bu seçenek Türkiye’de uygulanabilir değildir. İkincisi, bağımlı değişkenin ikiden fazla (canlı, yapılandırma ve temerrüt) kategoriye sahip olduğu bir tahmin modelinin geliştirilmesidir. Bu çalışmada gerçek bir banka verisi ile böyle bir model kurulacaktır.

# 2. YÖNTEM

Lojistik Regresyon denildiğinde ikili lojistik regresyon kast edilir ve temerrüt olasılığı modellemesinde sıkça kullanılır. Türkçe literatürde veri tabanlarında yapılan taramalarda multinominal lojistik regresyon [bundan sonra MLR olarak kısaltılacaktır] kullanılarak yapılandırma tahmini üzerine yapılan bir çalışmaya rastlayamadık. MLR’nin anlaşılabilmesi için öncelikle lojistik regresyon hakkında bilgi verilecek ve ardından MLR analizinin uygulandığı çeşitli makalelere değinilerek MLR üzerinde detaylıca durulacaktır.

## 2.1 Lojistik Regresyon

Lojistik regresyon analizinde bağımlı (yordanan) ve bağımsız (yordayıcı) değişkenler arasındaki ilişki modellenir ve kategorik bağımlı değişkenin aldığı değerlere ilişkin bir üyelik tahmini yapılır (Lojistik regresyonun detaylı anlatımı için bkz., Harrell, 2001; Kleinbaum and Klein, 2002). Lojistik Regresyon, kendi içinde bağımlı değişkenin sahip olduğu kategori sayısı ve kategoriler arasındaki ilişkiye göre 3’e ayrılır: a) İkili (binary) b) Sıralı (ordinal) ve c) Multinominal (Çok Terimli). İlk türde, bağımlı değişken 0 – 1 gibi (yani temerrüt etmiş – etmemiş gibi) iki değer alır. İkinci türde, bağımlı değişken en az 3 kategoriden oluşur ve kategoriler arasında bir hiyerarşi vardır. Örneğin, beğeniyorum, az beğeniyorum, beğenmiyorum, hiç beğenmiyorum gibi opsiyonlar vardır. Üçüncü türde ise bağımsız değişken gene en az 3 kategoriden oluşur. Fakat kategoriler arasında üstünlük-altlık yani bir hiyerarşi yoktur (Çokluk vd, 2014: 50, 60).

Lojistik regresyonun temel kavramlarını ve mantığını şöyle özetleyebiliriz:

*Olasılık*, p sembolü ile gösterilir. Bir olayın olma olasılığı *p* ise, olmama olasılığı de *1-p*’dir. Örneğin hastalığa yakalanma ihtimali % 42 ise, yakalanmama ihtimali 1 – 0.42 = 0.58 dir, yani % 58’dir. Olasılık, 0 ile 1 arasında bir değer alabilir. *Odds,* ‘p’ ile ifade edilen olma, başarı ya da görülme olasılığının, ‘1-p’ ile ifade edilen olmama, başarısızlık ya da görülmeme olasılığına “1-­p” oranıdır. Matematiksel olarak gösterimi şöyledir: odds = p / (1-p). Odds ları olasılığa dönüştürmek ise şöyle olur: P = odds / (1+odds). Odds, 0 ile pozitif sonsuz arasında bir değer olabilir. Simetrik bir yapıya sahip değildir. *Odds Ratio* [OR, eB veya Exp(B) olarak da kısaltılır], Üstünlük Oranı yani iki odds’un birbirine oranıdır. Diğer bir ifadeyle, birşeyin odds unun başka birşeyin odds una bölünmesidir. Odds ratio, 0 ile pozitif sonsuz arasında bir değer olabilir. *Logit,* odds’un doğal logaritmasıdır. Odds, asimetriktir. Doğal logaritması alınarak simetrik hale dönüştürülür. Odds un doğal logaritması alındığında (taban e, Euler değeri: 2.718), yani ln odds, - ∞ ile + ∞ arasında bir değer alabilir. İhtimaller 0 ile 1 aralığında sınırlı olduğu için doğrusal değildirler ama ln odds doğrusal olabilir. Bağımlı değişkenin logit dönüştürmeye tabi tutulmuş hali ile bağımsız değişkenler arasında doğrusal ilişki olduğu varsayılır.

Lojistik regresyondaki logit dönüştürmeyi biraz daha açmak gerekirse, olasılık odds’a dönüştürüldüğünde monotonik bir dönüştürme yapılmış olur. Yani, olasılık arttıkça odds da artar, yada olasılık azaldıkça odds da azalır. İkinci aşamada odds’lar log odds’a dönüştürülebilir. Bu dönüştürme de monotoniktir. Odds’lar arttıkça log odds’larda artar, yada odds azaldıkça log odds da azalır. Bu dönüştürme işlemine logit dönüştürme denir. Bu dönüştürmelerin esas nedeni şöyledir: Sınırlı bir aralıkta değer alan bir değişkeni modellemek çok zordur. Olasılık, 0 ile 1 arasında sınırlı bir aralıkta değer alır. İlk dönüştürme ile sınırlı aralık genişletilmiş olur. Odds, 0 ile pozitif sonsuz arasında bir değer alır. Ama elde edilen sonuç, simetrik bir yapıda değildir. Ardından odds un doğal logaritmasını alınarak eksi sonsuz ile artı sonsuz arasında değerler elde edilir. Böylece bağımlı ve bağımsız değişkenler arasında doğrusal bir yapı/ilişki tesis edilebileceği için modelleme yapmak kolaylaşır.

## 2.2 Multinominal Lojistik Regresyon

MLR, ikili lojistik regresyonun bağımlı değişkenin ikiden fazla kategorisi olduğu durumlar için geliştirilmiş halidir (Hosmer ve Lemeshow, 2000: 260). Bağımlı değişkenin iki değil de üç ve daha fazla kategoriden oluştuğu ve kategoriler arasında bir hiyerarşi olmadığı durumlarda MLR uygulanır. MLR’nin en önemli özelliği bağımlı değişkenin bu çok kategorili olma durumudur. Bu olgu İngilizce literatürde farklı kavramlarla ifade edilmiştir. Örneğin, polytomous logistic regression, multiclass logistic regression, multinomial logit, multicategory logistic model gibi. Benzer şekilde bu olguyu ifade etmek için Türkçe literatürde de birçok farklı kavram kullanılmıştır. Örneğin, çok kategorili/düzeyli isimsel lojistik regresyon, multinominal lojistik regresyon, multinominal (çok terimli) lojistik regresyon, multinomial logit model, multinomiyal logistik regresyon gibi.

Sağlıklı bir bilimsel tartışma için aynı dili konuşuyor olmak gerekir. Kullanılan kavramların içeriğinin net olması, aynı dili konuşmaya katkı sağlar. Bilimsel çalışmaların bir handikabı farklı bilim insanlarının aynı kavrama farklı anlamlar yüklemesidir. Bu yüzden, kavramların içeriği netleştirilmeden bir konu anlatıldığında anlatılmak istenenle anlaşılan arasında bir uçurum oluşabilir. Çünkü karşı taraf sizin kullandığınız kavrama farklı bir anlam yüklüyor olabilir. Bu yüzden makalede kullanılan kavramların net tanımlarına yer verilmiştir. Sıkça karşılaşılan diğer bir sorun aynı olgunun farklı kavramlarla ifade edilebiliyor olmasıdır. MLR de bu durum söz konusudur. Araştırma yapacaklara kolaylık sağlaması için literatürdeki farklı kullanımlara özellikle vurgu yapılacaktır.

MLR’de bağımsız değişkenin bir kategorisi referans kategori olarak seçilir ve diğer kategoriler bu referans kategoriye göre yorumlanıp analiz edilir (MLR nin nasıl yapıldığı hususunda bkz., Petrucci, 2009; Kleinbaum and Klein, 2002). Bu çalışmada referans kategori, ‘canlı’ yani müşterinin kredilerini düzenli şekilde ödemeye devam etmesidir. Bağımlı değişkenin sahip olduğu kategori sayısının bir eksiği kadar model kurulmaktadır (Bayram, 2011: 67; Hosmer ve Lemeshow, 2000: 261). Bu çalışmada bağımlı değişkenin 3 tane kategorisi olduğu için 2 tane model kurulacaktır. Birinde yapılandırma modellenecektir. Diğerinde temerrüt modellenecektir. Kurulacak olan lojistik fonksiyonlar şöyledir (Hosmer ve Lemeshow, 2000: 261):



Bağımlı değişkenin her bir kategorisinin koşullu olasılıkları ise şöyle elde edilir (Hosmer ve Lemeshow, 2000: 261-2):



Her istatistikî model bazı varsayımlara dayanmaktadır. Model kurmaya başlamadan önce bu varsayımların karşılanıp karşılanmadığı sınanmalıdır. Böylece kullanılmak istenen regresyon analizinin doğru analiz türü olup olmadığına kesin şekilde karar verilebilir. Varsayımlar karşılanmadığı durumlarda ya mevcut veri setine uygun başka bir analiz yöntemi seçilmeli ya da veri setinde bazı dönüşümler yapılması gerekmektedir.

Varsayımlar karşılanmadığı durumlarda elde edilen sonuçlar, geçerli sonuçlar olmayacaktır. Veri seti, ilgili varsayımları karşılamadığı halde karşıladığı varsayılıp analizler yapıldığında ulaşılan sonuçların pratik değeri de olmayacaktır. Örneğin bu çalışmada, varsayımlar karşılanmadan bir model kurulsa, pratikte hiçbir banka karar alırken bu modeli kullanmak istemeyecektir.

MLR varsayımlar açısından ikili lojistik regresyon ile aynı varsayımlara dayanmaktadır. Bu varsayımları şöyle özetleyebiliriz:

Birincisi, MLR’de parametre tahmininde en çok olabilirlik yöntemi kullanıldığı ve bu yöntem büyük örneklemler gerektirdiği için mümkün oldukça büyük örneklem ile çalışılması önerilmektedir. Örneklem büyüklüğü hususunda Çokluk (2010: 1361)’un önerisi şöyledir: Bağımlı değişken ikiden fazla kategoriden oluşuyorsa her bir bağımsız değişkende en az 50 kişilik bir grup büyüklüğüne gerek vardır. Bu çalışmada kullanılan örneklem oldukça büyük olduğu için bu varsayımınn gereği karşılanmıştır. İkincisi, çoklu bağlantı (multicollinearity) olmaması gerekir. Bağımsız değişkenler arasındaki yüksek korelasyona oldukça duyarlıdır. Aralarında çoklu bağıntı olan değişkenlerin olması durumunda değişkenlerin azaltılması yani sorunu çözmeye yetecek kadar değişkenin çıkarılması gerekir. Bu varsayım uygulama bölümünde ayrıntılı şekilde ele alınacaktır. Üçüncüsü, lojistik model, uç değerlere son derece duyarlıdır. Bağımsız değişkenlerdeki uç değerlin iyi analiz edilmesi gerekir. Çoğu istatistiksel işlem, ortalamadan olan sapmaların karesine dayanır. Eğer bir gözlem, dağılımın geri kalanından çok uzakta ise yani ortalamadan uzakta ise sapma değeri büyüyecektir. Bu varsayıma veri seti anlatılırken değinilecektir. Dördüncüsü, uygun tüm bağımsız değişkenler modele dâhil edilmelidir. Açıklayıcılık gücü olan değişkenlerin modele dâhil edilmemesi modelin yetersizliğine neden olabilir. Bu varsayıma veri seti anlatılırken değinilecektir. Beşincisi, veri analize başlanmadan önce kayıp değerler açısından kontrol edilmeli ve kayıp değerlere nasıl bir uygulamanın yapılacağı belirlenmelidir. Bu varsayıma da veri seti anlatılırken değinilecektir. Altıncısı, gözlemler birbirlerinden bağımsız olmalıdır. Olayların meydana gelmeleri arasında hiçbir ilişki bulunmamalıdır. Bir gözlem, başka bir gözlem hususunda bir bilgi sunuyor olmamalıdır. Yani bir gözlemin gerçekleşmesi başka bir gözlemin gerçekleşme ihtimalini azaltmamalı veya arttırmamalıdır. Bazı gözlemler arasında bağımlılık söz konusu ise rassallık ilkesi ihlal edilmiş olmaktadır. Bir gözlemin gerçekleşme ihtimali başka bir gözlemin gerçekleşmesine bağlı ise tüm gözlemlerin aynı gerçekleşme ihtimaline sahip olduğu söylenemez. Yedincisi, sürekli bağımsız değişkenler ile bağımlı değişkenin logit dönüşümü arasında doğrusal bir ilişki olmalıdır. İkili logistik regresyonda

# 3. VERİ SETİ

Başvuru modelleri, ilk defa bankaya başvuran ve bankanın tam anlamıyla tanımadığı müşterilere yöneliktir. Bu yüzden müşterinin başvurduğu bankadaki ödeme davranışı ile ilgili iç kaynaklara dayanan bir ödeme verisine yer verilememektedir. Ancak KKB, Risk Merkezi (eski memzuç) gibi dış kaynaklı ödeme davranış bilgilerini de dâhil edebilir. Davranışsal modeller ise, başvuru modelinde yer alan verilere ilaveten müşterinin çalıştığı bankadaki ödeme davranışına ait verileri de içermektedir. Davranışsal modeller, iç kaynaklara dayanan ödeme verisini/performansını içerdiği için bankalar için daha kullanışlıdır. Bu çalışmada bankanın tanımadığı ve ilk defa limit tahsisi yapacağı müşteriler hedef kitleyi oluşturmamaktadır. Modellemeye konu müşteri kitlesi, banka ile kredili çalışıyor olan müşterilerdir.

Bu çalışmada bir bankanın gerçek verileri kullanılarak bir yıllık süre zarfında kredilerin temerrüt etme, yapılandırılma veya canlı kalma ihtimallerini hesaplamaya yönelik bir regresyon modeli kurulacaktır. Bağımlı değişken 3 kategoriden oluşacaktır: a) Sonraki bir yıl içinde temerrüt etmiş, b) sonraki bir yıl içinde yapılandırma yapılmış (diğer bir ifadeyle müşteri borcunu ödeyemediği için sözleşme koşullarında değişiklik yapılmış) ve c) sonraki bir yılda sorunsuzca ödemeler devam etmiş.

31 Aralık 2009 tarihinden başlayıp 31 Aralık 2015 tarihine kadar süren 6 aylık periyotlar ile 13 kesit tarihi için veri çıkılmıştır. Kesit tarihleri sırasıyla şöyledir: 31/12/2009, 30/06/2010, 31/12/2010, 30/06/2011, 31/12/2011, 30/06/2012, 31/12/2012, 30/06/2013, 31/12/2013, 30/06/2014, 31/12/2014, 30/06/2015 ve 31/12/2015. 6,5 yıllık bir veri seti ile çalışılacağı için kurulan modelin banka müşterilerinin genel yapısını yansıtma imkanı daha fazladır.

## 3.1 Tanımların Netleştirilmesi

Temerrüt tanımı tüm araştırmayı etkileyeceği için etraflıca tanımlanmalıdır. Müşterinin kesit tarihinde riski devam eden projelerinde veya performans dönemi içinde yeni kullandığı projelerinden herhangi bir tanesinde aşağıdaki durumlardan birisi gerçekleşirse müşteri temerrüt etmiş olarak kabul edilir: (a) ödemesini 90 günden fazla geciktirirse veya (b) geç gün sayısı 90 gün olmadan donuk alacak olarak sınıflandırılıp takip hesaplarına virmanlanırsa. Temerrüt tarihi olarak hangisi önce gerçekleşmişse o kabul edilecektir. Yapılandırma, finansal güçlük içinde olan bir borçluya sözleşme koşullarında değişiklik yapılarak bir imtiyaz tanınmasıdır. Temerrüt ve yapılandırma tanımları kapsamına girmeyen tüm krediler geri ödemesinde sorun olmayan canlı krediler olarak kategorize edilecektir.

## 3.2 Modellemeye Konu Müşteriler

Kesit tarihi itibariyle kredi riski olan, gecikmesi olmayan ve 6 aydan fazla süredir banka ile çalışıyor olan (yani eski müşteri) işletme segmenti müşterileri için modelleme yapılmıştır. İşletme segmenti, yıllık geliri 3 ila 30 milyon TL arasında olan veya bankanın 1 ila 10 milyon TL limit açtığı müşterilerden oluşur. Gecikmesi olmayan ibaresini biraz daha detaylandırmak modellemeye konu kitlenin anlaşılmasını daha kolaylaştıracaktır. Kesit tarihi itibariyle 1 gün bile olsa gecikmesi olan müşteriler modelleme kapsamına alınmamıştır. Bu hususta kesit tarihinden önceki 6 veya 12 ayda belli bir geç günden fazla –örneğin 60 gün- gecikme yaşanmışsa modellemeden çıkarılması gibi farklı yaklaşımlar da söz konusudur. Bu hususta daha muhafazakâr yaklaşımlar da vardır. Örneğin Sıddıqi (2006: 3) son 24 ayda 60 gün üzeri gecikmenin ikiden fazla yaşandığı müşterilerin modelleme kapsamından çıkarılabileceğinden bahseder. Gecikmesi olan müşterilerin temerrüt veya yapılandırmaya konu olma ihtimali gecikmesi olmayanlara göre daha yüksektir. Gecikmesi olan ve olmayan müşterileri aynı anda modellemeye çalıştığımızda homojen bir yapı ile çalışmıyoruz demektir. Ayrıca, bir müşteri kredi ödemesini zamanında yapmayarak ciddi bir erken uyarı sinyali vermiş olmaktadır. Bankalar zaten gecikme yaşandığında çeşitli aksiyonlar almaktadırlar. Geliştireceğimiz model ile gecikme yaşanmadan daha erken bir dönemde müşteriler hakkında tahmin modeli kurulması hedeflenmektedir.

Geliştirilecek olan model davranışsal bir model olacağı için müşterilerin bankada belli bir geçmişinin olması gerekmektedir. Piyasa uygulamaları doğrultusunda, banka ile 6 aydır çalışan müşteriler eski müşteri kabul edildi. Çalışmada kullanılan eski müşteri tanımı şöyledir: Kesit tarihinde açık olan (riski devam eden) nakdi veya gayri nakdi projelerinden herhangi birisinin kullandırım tarihi 6 aydan eski olan müşteri eski müşteri olarak kabul edilir. İlaveten, kesit tarihi itibariyle riski olan projelerin hepsi yeni -yani son 6 ay içinde yapılmış- olsa bile müşterinin son 12 ayda kapanan herhangi bir nakdi projesi varsa da eski müşteri olarak kabul edilmiştir. Sürekli 3-4 aylık kısa vadeli kredi kullanan müşterilerin 6 ay kriterinden dolayı kapsam dışında kalmasının önüne geçilebilmesi için bu ilave kriter eklenmiştir.

Veri setinin ilk halinin dağılımı şöyledir:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Segment Adı** | **Gözlem Sayısı** | **Canlı Müşteri Sayısı** | **Temerrüt Eden Müşteri Sayısı** | **Yapılandırma Yapılan Müşteri Sayısı** |
| **İşletme** | 63,068 | **60,664** | **1,195** | **1,209** |

## 3.3 Modellemeye Konu Örneklemin Oluşturulması

Veri seti, kesit bazlı oluşturulduğu için modellemeye başlanmadan önce çoklama sorununun giderilmesi maksadıyla bazı elemelerin yapılması gerekmektedir. Bir kesit tarihi baz alınıp sonraki 12 ayda temerrüt veya yapılandırmanın gerçekleşip gerçekleşmediğine bakıldığı için aynı olay birden fazla kesit tarihine kaydedilebilir. Örneğin müşteri Kasım 2014’te temerrüt etmiş olsun. 31 Aralık 2013 kesit tarihinde sonraki 12 ayda temerrüt veya yapılandırma yapılmış mı diye baktığımızda temerrüt etmiş olarak kaydedeceğiz. 30/06/2014 kesit tarihinde de aynı olay nedeniyle müşteriyi temerrüt olarak kaydetmiş olacağız. Bu örnekte görüldüğü üzere aynı olayı iki defa dikkate alındığı için bir çoklama sorunu mevcuttur.

Öncelikle her bir kesit tarihi için performans döneminde (yani sonraki 12 ay) hangi olay gerçekleşmişse onu gösteren bir etiket tanımlandı. Bir müşteri birden fazla yapılandırılabilir veya temerrüt edebilir. Çoklu yapılandırma veya temerrüt durumlarında modellemeye sadece ilk yapılandırma veya temerrüt alındı. Bunun için en küçük temerrüt veya yapılandırma tarihi müşterinin temerrüt veya yapılandırma tarihi olarak seçildi. Ardından bu belirlenen olayın farklı dönemlerde kaydedilmesinden kaynaklı çoklama sorununun çözümüne geçildi.

Çoklamaya konu Temerrütlerin Elenmesi: Müşterinin temerrüt etiketi ile işaretlendiği kesit tarihleri belirlendi. Aynı temerrüt olayından dolayı birden fazla kesit tarihinde işaretleme yapıldığı durumlarda ilgili en küçük kesit tarihi alındı ve diğerleri silindi.

Çoklamaya konu Yapılandırmaların Elenmesi: Müşterinin yapılandırma etiketi ile işaretlendiği kesit tarihleri belirlendi. Aynı yapılandırma olayından dolayı birden fazla kesit tarihinde işaretleme yapıldığı durumlarda ilgili en küçük kesit tarihi alındı ve diğerleri silindi.

Çoklamaya konu Canlı Kredilerin Elenmesi: Müşterinin canlı olarak etiketlendiği kesit tarihleri belirlendi. Müşteriler kesit tarihlerine göre sıralandılar. Yeni bir bilgi içermeyen kesit tarihleri silindi. Bir örnekle anlatmak gerekirse, bir müşteri tüm kesit tarihlerinde canlı olarak etiketlenmiş olsun. 31/12/2009 kesit tarihinde sonraki 12 aya baktığımızda 31/12/2010’a kadar bakmış oluruz. 31/12/2010’da da 31/12/2011’e kadar olan dönemi kapsamış oluruz. Böyle olunca örneğin 30/06/2010 kesit tarihine bakmaya gerek kalmamaktadır. O dönem çıkarıldığında müşterinin sonraki performansı hakkında bilgi kaybımız olmaz.

Bu 3 aşamalı eleme işleminden sonra geri kalan canlı, yapılandırma ve temerrütler gözlemleriyle model kurulmuştur. Elemeler yapıldıktan sonra veri setinin dağılımı şöyledir:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Segment Adı** | **Gözlem Sayısı** | **Elenen Gözlem Sayısı** | **Canlı Müşteri Sayısı** | **Temerrüt Eden Müşteri Sayısı** | **Yapılandırma Yapılan Müşteri Sayısı** |
| **İşletme** | **39,606** | **23,462** | **37,699** | **990** | **917** |

## 3.4 Veri Analizi

Kullanılan veri, ekransal kontroller (veri setinde raporlanan bilgi ile ilgili ekrandaki bilginin aynı olup olmadığının teyidi) ve çapraz sorgu kontrolleri (bazı mantıksal kurallar çerçevesinde verinin teyidi, örneğin son 3 ayda gecikmeli ödenen taksit sayısı son 6 ayda gecikmeli ödenen taksit sayısından büyük olamaz vb.) gibi kontrol aşamalarından geçmiş ve doğruluğu teyit edilmiş bir veridir. Bu yüzden verinin doğruluğu hususunda ilave bir çalışma yapılmamıştır.

Kayıp değerler hususunda, hangi örneklem büyüklüğü için ne kadar kayıp değerin tolere edilebileceğine ilişkin belirlenmiş bir ölçüt yoktur. Çözümlerden birisi, % 30 gibi belli bir oranın üzerinde kayıp değeri olan değişkenler modelden çıkartılmasıdır. Diğer bir çözüm, kayıp değer yerine ortalama değeri yazmaktır. Üçüncü bir yöntem, sürekli değişkenler kategorik hale getirilirken kayıp değerler için ayrı bir kategori oluşturulmasıdır. Bu çalışmada kullanılan yaklaşım şöyledir: Kayıp değer yüzdesi belli bir oranın üstündeyse değişken çıkartılmıştır.

Veri kontrolleri yapılırken uç değerler açısından da kontrol yapılmıştır. Veri girişi yapanların hatalarından kaynaklı hatalar tespit edildiğinde kaynak tablolarda düzeltme yapılmıştır. Her bir değişkendeki uç değerler, değişken özelinde analiz edilmiş ve ilgili iş birimleriyle görüşülüp bilgi alınmıştır. Uç değerlerin etkisini azaltmanın bir yolu, sürekli değişkenleri kategorik hale getirirken uç değerler için ayrı bir kategori oluşturulmasıdır. Böylece, ilgili gözlemi hariç tutmak zorunda kalınmaz ve bilgi kaybı minimize edilebilir.

## 3.5 Modellemeye Konu Bağımsız Değişkenler

Alpar (2003: 192)’ın belirttiği üzere araştırmada kullanılacak bağımsız değişkenler tespit edilirken karşılaşılan önemli bir sorun, belirleme hatası yani “ilgisiz değişkenlerin modele alınması, ilgili değişkenlerin ise model dışında bırakılmasıdır”. Aydın (2016: 253) öncelikle “ki-kare, iki ortalama arası farkın önemlilik testi veya Mann-Whitney U gibi testler aracılığıyla bağımlı değişkenle ilişkisiz olan bağımsız değişkenler”in tespit edilip modelleme çalışmasının kapsamından çıkartılmasını önerir. Modellemede temel amaç modellemeye konu ilişkiyi en az sayıda değişkenle açıklamaktır. Bu yüzden ilgisiz değişkenlerin öncelikle elenmesi gerekmektedir. İlgili bir değişken modellemeye alınmadığında ise modelin açıklayıcılığı azalacaktır.

Kullanılan veriyi sağlayan banka kapsamlı bir veri ambarına sahiptir. Bu veri ambarından davranışsal modellemeye uygun, doluluk oranı yüksek, iş mantığı açısından kabul edilebilir olan ve ön testlerden geçen 58 değişken belirlenmiş ve bu değişkenlerle regresyon analizi yapılmıştır. Böylece ilgili tüm bağımsız değişkenlerin modele dâhil edilmesi ve ilgisiz değişkenlerin elenmesi sağlanmıştır.

# 4. UYGULAMA VE İSTATİSTİKİ YORUMLAR

Çalışmada Ek 1’de ayrıntıları verilen değişken seti kullanılmış ve öncelikle çoklu bağıntı açısından kontroller yapılmıştır.

## 4.1 Çoklu Bağlantı Açısından Kontrol

Çoklu bağlantı sorununun çözümü için faktör analizi yöntemiyle modeldeki değişkenler birleştirilip elde edilen faktörlerle model kurulabilir. Ya da bu soruna neden olan değişkenler modelden çıkartılabilir. Bu sorunun tespiti için aşağıdaki yöntemler kullanılabilir (Çoklu vd, 2014:35-38):

Öncelikle, değişkenler arasındaki basit korelasyon incelenir. Değişkenler arasında güçlü korelasyon olması çoklu bağlantı sorununa işaret eder.

Daha sonra, VIF (Variance Inflation Factor = Varyans artış faktörü) incelenir. Genel kural olarak VIF’ın 10’a eşit veya daha büyük olması durumunda çoklu bağlantı sorunu olduğu anlaşılmış olur. Ardından, tolerans değeri incelenir. Bu değerin 0.1’den büyük olması durumunda, değişkenler arasında çoklu bağıntı sorununun olmadığını söyleyebiliriz. Daha sonra, Condition Index incelenir. Bu değerin 10 ile 30 arasında olması orta düzeyde, 30’dan büyük olması durumunda yüksek düzeyde bağıntı problemi olduğu anlaşılır.

Çoklu bağıntı sorununu çözebilmek için birbiri ile yüksek korelasyon olan değişken çiftlerinde birisi çıkartıldı. Ayrıca, yukardaki kriterlere takılan değişkenler de elendi ve kalan değişkenlerle nihai analiz yapıldı. Modellemeye konu değişkenler şöyledir:

|  |  |
| --- | --- |
| Değişken | Kod |
| CANLI\_TEMERRUT\_YAPILANDIRMA | Bağımlı Değişken |
| LastLimitDayCount | X1 |
| G\_LastMonthOverdueDayCount | X2 |
| Last6MonthOverdueDayCount | X3 |
| L12MUtilHasRiskCashPCount | X4 |
| Last12MonthsAvgCashLimit | X5 |
| TCMBChkUnpaidCountLast1YearG | X6 |
| G\_WorkingBankCount | X7 |
| AccuralAmount | X8 |
| TtlCashLmtUtlztntRMMemzucT0 | X9 |
| FactoringtoTotalRiskT0 | X10 |
| SharingRate1 | X11 |
| Fourth\_BO3 | X12 |

## 4.2 Modelin Oluşturulması

[Bu bölüm geliştirme aşamasındadır.]

Model R’da nnet paketinde yer alan multinom fonksiyonu ile oluşturuldu.

Öncesinde, aşağıdaki kodla bağımlı değişkeni faktöre çevirdik.

> mn$Y <- factor(mn$Y)

Modeli oluşturmadan önce bağımlı değişkenimizdeki bir kategoriyi referans kategorisi olarak seçmek ve ardından model kodunu çalıştırmamız gerekir.

> mn$Y2 <- relevel(mn$Y, ref = "0")

Ardından multinom fonksiyonunun çalıştırdık.

test <- multinom(Y2 ~ X1 + X2 + X3, data = mn)

# weights: 15 (8 variable)

initial value 30.761144

iter 10 value 28.696762

final value 28.526486

converged

Bu kodla birlikte modelimize test adını vermiş olduk. Oluşturduğumuz modeli “summary” kodu ile görebiliriz.

summary(test)

Call:

multinom(formula = Y2 ~ X1 + X2 + X3, data = mn)

Coefficients:

 (Intercept) X1 X2 X3

1 -4.294991 -0.03491795 0.5723493 0.6301150

2 -2.137939 -0.15431016 0.3290166 0.3451514

Std. Errors:

 (Intercept) X1 X2 X3

1 3.659792 0.1612273 0.4431742 0.5783516

2 3.850371 0.1820062 0.4927616 0.6262413

Residual Deviance: 57.05297

AIC: 73.05297

Summary kodu ile elde ettiğimiz bilgileri şöyle yorumlayabiliriz:

*Call bölümü*: Bizim yazdığımız koddur.

*Katsayılar ve Standart Hatalar bölümleri*: Katsayıları ve standart hataları içerir. Tablo p değerini içermez. P değerlerini Wald istatistikleriyle şöyle üretebiliriz:

> z <- summary(test)$coefficients/summary(test)$standard.errors

> z

 (Intercept) X1 X2 X3

1 -1.1735613 -0.2165759 1.2914769 1.0895016

2 -0.5552552 -0.8478290 0.6676994 0.5511476

2-tailed z test sonuçları şöyledir:

> p <- (1 - pnorm(abs(z), 0, 1)) \* 2

> p

 (Intercept) X1 X2 X3

1 0.2405708 0.8285388 0.1965384 0.2759328

2 0.5787201 0.3965332 0.5043255 0.5815325

Akaiki (AIC= Akaiki Information Criteria), model uyumunu ölçmek için kullanılan bir uyum ölçütüdür. AIC = -2logL + 2r formülüyle hesaplanmaktadır. Bir modeli tek başına değerlendirirken kullanılmaz. Kurulan farklı modellerin karşılaştırılmasında kullanılır. En küçük AIC değerine sahip model, en uygun model olarak kabul edilir.

Bu formüldeki LogL değeri, MLR modelinde elde edilen log-olabilirlik değerini, r değişken (parametre) sayısını ifade etmektedir.

Katsayıları incelediğimizde, referans kategorimiz olan “0”a kıyaslanarak bağımlı değişkenin diğer iki kategorisi için katsayıları görürüz. Bu katsayıları şöyle yorumlayabiliriz:

X3 değişkenindeki bir birimlik artış, referans kategorisi olan 0 ile kıyaslanarak 1 nolu bağımlı değişken kategorisinde olmanın log odds u 0.63 kez artar.

X2 değişkenindeki bir birimlik artış, referans kategorisi olan 0 ile kıyaslanarak 1 nolu bağımlı değişken kategorisinde olmanın log odds u 0.57 kez artar.

X1 değişkenindeki bir birimlik artış, referans kategorisi olan 0 ile kıyaslanarak 1 nolu bağımlı değişken kategorisinde olmanın log odds u -0.3 kez azalır.

X3 değişkenindeki bir birimlik artış, referans kategorisi olan 0 ile kıyaslanarak 2 nolu bağımlı değişken kategorisinde olmanın log odds u 0.34 kez artar.

X2 değişkenindeki bir birimlik artış, referans kategorisi olan 0 ile kıyaslanarak 2 nolu bağımlı değişken kategorisinde olmanın log odds u 0.32 kez artar.

X1 değişkenindeki bir birimlik artış, referans kategorisi olan 0 ile kıyaslanarak 2 nolu bağımlı değişken kategorisinde olmanın log odds u -0.15 kez azalır.

## 4.3 Katsayıların Odds a Dönüştürülmesi

Bağımlı değişkenin kategorilerinden herhangi birisinin seçilme ihtimalini referans kategorisinin seçilme ihtimaline bölerek nispî riski buluruz. Bu nispî risk, odds ratio olarak da ifade edilir. Bu nispî risk oranlarını görmek için katsayıların üstü alınır.

Katsayıların yorumlanması nispeten zordur. Çünkü logaritmik bir yapıdadırlar. Katsayıların üstü alınırsa odds ratio olarak yorumlanabilirler.

Odd ratio ları şöyle buluruz:

> exp(coef(test))

 (Intercept) X1 X2 X3

1 0.0136367 0.9656847 1.772426 1.877827

2 0.1178976 0.8570062 1.389601 1.412204

Bulduğumuz odds ratio ları örneğin şöyle yorumlayabiliriz:

X3 teki bir birimlik artışın odds ratio su, referans kategorisine kıyasla 1 nolu kategori için 1.87 dir.

# 5. SONUÇ VE ÖNERİLER

Çalışmaya veri sağlayan banka, çalışma sonucunda geliştirilen regresyon modelini kullanarak modelin çalıştırıldığı raporlama tarihinden sonraki bir yıllık süreçte aktif kalitesine dair öngörülere sahip olacaktır. Diğer bankalar kendi içsel verileriyle benzer modeller kurarak aktif kalitesinin yönetiminde kullanabilirler. Bir bankanın aktif kalitesi değerlendirilirken yapılandırılan kredi tutarı önem kazanmaktadır. Çünkü yapılandırılan kredilerden takibe gitme oranı oldukça yüksektir. Yapılandırılan kredi tutarı bankanın yakın zamanda ilave ne kadar daha kredisinin batacağını öngörmek için kullanışlı bir bilgidir. Banka batan kredileri için özel karşılık ayıracağı için karlılığı düşecektir. Bu yüzden yeniden yapılandırılacak olan müşterilerin önceden tespit edilmesi önem arz etmektedir. Bankalar yapılandırma hususunda daha erken aksiyon alabilirler ve böylece krediyi teminatlandırma ihtimalleri artabilir veya ilave kredi riskine girmeyebilirler. Müşteri açısından bakıldığında, daha erken bir dönemde finansal borçları hususunda daha uygulanabilir bir plan yapma imkânı doğabilir.

Bu tahmin modeli sayesinde finansal varlık kalitesi hakkında büyük mevduat sahiplerine ve sukuk (kira sertifikası) alıcılarına sonraki bir yılda beklenen temerrüt ve yapılandırma adetleri ve tutarları hakkında ikna edici bilgiler verilebilir. Bankalar kendi aralarında işlem yaparken birbirlerine limit açarlar. Aktif kalitesi limit tahsisinde bakılan önemli hususlardan birisidir. Bankalar diğer bankalar nezdindeki limitlerini arttırmak istediklerinde bu tahmin modelinin çıktılarını kullanabilirler.

Son olarak, yapılandırmayı içeren bu tahmin modeli banka tarafından kredi risk yönetiminde kullanılabilir.

# 6. KAYNAKÇA

Alpar, R. 2003. Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş 1. Değiştirilmiş ve Genişletilmiş 2. Baskı. Ankara: Nobel Yayın Dağıtım.

Aydın, N. 2016. Türkiye de Hanehalkı İş Bulma Yöntemleri İle İşgücü Yapısı Arasındaki İlişkinin Multinominal Regresyon Yöntemiyle Analizi. Dumlupınar Üniversitesi Sosyal Bilimler Dergisi. ICEBSS Özel Sayısı. 248-266.

Bayram, N. 2011. Multinominal Lojistik Regresyon Analizinin İstihdamdaki İşgücüne Uygulanması. İktisat Fakültesi Mecmuası. 54(2). 60-75.

Çokluk, Ö. 2010. Lojistik Regresyon Analizi: Kavram ve Uygulama. Kuram ve Uygulamada Eğitim Bilimleri. 10(3). 1357-1407.

Çokluk, Ö., Şekercioğlu, G., Büyüköztürk, Ş. 2014. Sosyal Bilimler İçin Çok Değişkenli İstatistik: SPSS ve LISREL Uygulamaları. 3.Baskı. Ankara: Pegem Akademi.

Harrell, F., E. 2001. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression and Survival Analysis New York: Springer.

Hosmer, D., W., Lemeshow, S. 2000. Applied Logistic Regression. Second Edition. New York: John Wiley & Sons, Inc.

Kleinbaum, D., G., Klein, M. 2002. Logistic regression : a self-learning text. New York : Springer.

Petrucci, C., J. 2009. A Primer for Social Worker Researchers on How to Conduct a Multinomial Logistic Regression. Journal of Social Service Research. 35(2). 193-205.

Siddiqi, N. 2006. Credit Risk Scorecards Developing and Implementing Intelligent Credit Scoring. New Jersey: John Wiley & Sons, Inc.

# EK 1: Değişken Listesi

|  |  |
| --- | --- |
| Değişken | Tanımı |
| CANLI\_TEMERRUT\_YAPILANDIRMA | CANLI\_TEMERRUT\_YAPILANDIRMA |
| LastLimitDayCount | Son Limit Açılma Tarihi (Gün) |
| CashProductCount | Nakdi Ürün Limiti Adedi |
| G\_LastMonthOverdueDayCount | Son 1 Ay Gecikme Gün Sayısı |
| G\_Last12MonthOfNegotiationCount | Son 12 Ayda Gruptaki Müşterilerin Görüşme Sayısı |
| CollateralTypeCount | Teminat Türü Sayısı (İpotek, Araç, Nakit Bloke,Gks Hariç) |
| ChqCltrlAvgAmtLastYr | Son 1 Yıl Teminat Alınan Çeklerin Ortalaması |
| BdCqCtlAmntLtYrRtOvAllCqCtlLtYr | Son Bir Yıl İçinde Karşılıksız Çek Adedi / Son Bir Yıl İçinde Teminata Alınan Toplam Çek Adedi |
| ChequeRiskAmount | Çek Riski |
| AvgDelayDayCountL12M | Önceki 12 Ay Ortalama Gecikme Gün Sayisi |
| Last3MonthOverdueDayCount | Son 3 Ay-Gecikme Gün Sayısı |
| Last6MonthOverdueDayCount | Son 6 Ay-Gecikme Gün Sayısı |
| Plus7ArrearAvgAmtL6M | önceki son 6 ayda 7 gün ve üzeri gecikmeye girmiş toplam tutar / önceki son 6 ayda 7 gün ve üzeri gecikme sayısı |
| Last6MonthOfNegotiationCount | Son 6 ay Aranma adedi (aktif görüşme sayısı)  |
| First6MonthInstallmentCount | İlk 6 Ay Toplam Taksit Adedi |
| BFirst6And12MInsCount | İlk 6-12 Ay Arası Toplam Taksit Adedi |
| L12MUtilHasRiskCashPCount | Son 12 Ayda Yapılan Projelerden Riski Devam Eden Nakdi Proje Sayısı |
| Last24MonthCashLoanProjectCount | Son 24 Ayda Yapılan Nakdi Proje Sayısı |
| Last24MonthLoanProjectCount | Son 24 Ayda Yapılan Toplam Proje Sayısı |
| L612MonthAvgNonCashRisk | önceki ilk 6-12 aydaki ortalama gayrinakdi riski |
| L6MonthAvgRisk | Son 6 Aydaki Ortalama Total Riski |
| CashLimitUtil | Nakdi Limit Kullanım Oranı |
| L6MAvgTotalLimitUtil | Önceki Son 6 Aydaki Ortalama Total Limiti Kullanım Oranı (%) |
| L12MMaxTotalLimitUtilRate | Önceki son 12 aydaki maksimum total limiti kullanım oranı (%) |
| Last12MonthsAvgCashLimit | Önceki son 12 aydaki ortalama nakdi limiti |
| Last12MonthsMaximumCashLimit | Önceki Son 12 Aydaki Maksimum Nakdi Limiti |
| LastMonthAvgCurrentBalance |  Son 1 Ay Cari Hesaplarında Ortalama Bakiye |
| Last3MonthAvgCurrentBalance | Cari Hesaplarında Son 3 Ay Ortalama Bakiye |
| OfficialCorrespondenceCountG | Grup Ve Grup Firma Ortakları Resmi Yazışma Adedi  |
| TCMBChkUnpaidCountLast1YearG | Grup ve Grup Firma Ortakları TCMB Karşılıksız Çek Firma/Şahıs Adedi(Son 1 Yıl) |
| TCMBChkUnpaidCountLast5YearG | Son 5 Yıl Grup Ve Grup Firma Ortakları TCMB Karşılıksız Çek Firma/Şahıs Adedi  |
| FirmAge | Firma Yaşı (Gün) |
| GroupFirmCount | Grup Firma Sayısı |
| OpenAccountDayCount | Hesap Açılma Tarihi\_Müşteri (Gün) |
| G\_WorkingBankCount | Grup Çalışılan Banka Sayısı |
| IsExistsAccrual | Faiz Tahakkuku Var Mı? (1,0) |
| AccuralAmount | Memzuç Faiz Tahakkuk Tutarı |
| WorkingBankCount | Çalışılan Banka Sayısı |
| TtlCashLmtUtlztntRMMemzucT0 | Toplam Nakdi Limit Kullanımı |
| TtlNonCashRiskPrcntgtRMMemzucT0 | Memzuç Toplam Gayrinakdi Risk / Memzuç Toplam Risk  |
| MidTermCashtoTotalRiskT0 | Orta Vade Nakdi Risk / Toplam Risk  |
| ShortTermCashtoTotalLimitT0 | Kısa Vadeli Nakdi Risk / Toplam Limit Oranı (Memzuç T0) |
| ShortTermCashtoTotalCashLimitT0 | Memzuç Kısa Vade Nakdi Risk / Memzuç Toplam Nakdi Limit |
| TtlLimitUtilizationatRMMemzucT6 | Toplam Limit Kullanımı= 6 Dönem Öncesinin Memzuç Toplam Risk / 6 Dönem Öncesinin Memzuç Toplam Limit |
| ChangingFactoringRiskT0T12 | Memzuç Faktoring Riski / Memzuç 12 Dönem Öncesinin Faktoring Riski  |
| FactoringtoTotalRiskT0 | Faktoring Riski/ Toplam Risk |
| ChanginginLongTermCashRiskT0T3 | Memzuç Uzun Vade Nakdi Risk / Memzuç 3 Dönem Öncesi Uzun Vade Nakdi Risk  |
| ChanginginTotalNonCashLimitT0T6 | GN Limit Değişimi=Toplam GayriNakdi Limit / 6 Dönem Öncesinin Toplam GayriNakdi Limit |
| ChanginginTotalRiskT0T6 | Memzuç Toplam Risk / Memzuç 6 Dönem Öncesi Toplam Risk |
| ChangingIntAccuralT0T3 | Memzuç Tahakkuk Tutarı / 3 Dönem Öncesi Memzuç Tahakkuk Tutarı |
| SharingRate1 | Ortalık Payı Yüksek Olan Ortağın Ortaklık Payı |
| AvailblApplicntCountThreeMonth1 | Son 3 Ay Ortalık Payı Yüksek Olan Ortağın -Toplam başvuru sayisi  |
| FourthCon\_KA15 | (Faaliyet Kar Zararı+Amortisman Giderleri)/(Kısa Vadeli Yabancı Kaynaklar-Ortaklara Borçlar-Alınan Avanslar) |
| FourthCon\_AK13 | Alacak Devir Süresi |
| Fourth\_BO0 | Özkaynaklar/Toplam Borçlar |
| Fourth\_BO3 | (Özkaynaklar+Ortaklara Borçlar)/ (Toplam Borçlar-Ortaklara Borçlar-Hakedişler-Alınan Avanslar) |
| Fourth\_KA63 | Vök/Toplam Banka Borçları |
| Fourth\_LI2 | Hazır Değerler/Kısa Vadeli Yabancı Kaynaklar |
| Fourth\_NA9 | (Esas Faaliyetlerden Kaynaklanan Nakit-Dağıtılan Temettüler)/Toplam Borçlar |

1. Sorumlu Yazar, İstanbul Ticaret Üniversitesi, Yrd. Doç. Dr., sbagdatli@ticaret.edu.tr [↑](#footnote-ref-1)
2. İstanbul Ticaret Üniversitesi, İstatistik Tezli Yüksek Lisans Öğrencisi [↑](#footnote-ref-2)