# Çoklu Lojistik (Multinominal Logistic) Regresyon Analizi ile Temerrüt Tahminlerine Yapılandırmanın Dâhil Edilmesi

Celil Taşkın, İstanbul Ticaret Üniversitesi, İstatistik Tezli Yüksek Lisans Öğrencisi

celiltaskin@gmail.com

**Özet**

Anahtar Kelimeler: Çoklu lojistik regresyon, yapılandırma, kredi derecelendirme, temerrüt olasılığı

**Abstract**

Keywords: Multinominal Logistic Regression, Re-structuring, rating, probability of default

**1. GİRİŞ**

Bir kredi müşterisinin mevcut kaynaklarıyla kredi sözleşmesi şartları doğrultusunda borcunu ödemeyeceği belli olunca bankaların önünde iki ana alternatif bulunmaktadır. Ya kredinin temerrüt ettiği kabul edilip yasal takip başlatılacaktır ya da yapılan analizler ve görüşmelerden sonra sözleşme şartlarında değişiklik yapılırsa müşterinin borcunu ödeyebileceğine dair kanaat oluşursa kredi yeniden yapılandırılacaktır. Bankalar genelde esnek bir yeniden yapılandırma politikasına sahip olduklarından bu yeniden yapılandırmaya dair kanaat, konjektürel şartlardan sıklıkla etkilenebilmektedir. Bu yapılandırma kararları bankaların temerrüt tahmin modelleriyle ilgili ciddi bir sıkıntıya yol açmaktadır. Şöyle ki banka yapılandırma hususunda yanlış bir karar verdiğinde yani yapılandırılmaması gereken bir krediyi yapılandırdığında temerrüt edecek bir kredi temerrüt etmemiş olacak ve dolayısıyla modelleme çalışmalarına temerrüt vakası olarak girmesi gereken bir vaka canlı kredi vakası olarak tahmin modelinde yer alacaktır. Diğer bir ifadeyle, benzer özelliklere sahip ve dışsal bir müdahale olmadığı durumda temerrüt edecek iki krediden birisi yapılandırıldığı için temerrüt etmezken diğeri temerrüt edecektir. Temerrüt tahmin modellerinde bu hususta bir çözüm önerisi getirilmesi gerekmektedir.

Tahmin modellerine yapılandırmanın dâhil edilmesiyle sübjektif yapılandırma kararların temerrüt tahminlerini olumsuz etkilemesinin önüne geçilebilir. Ayrıca, 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.

**2. LİTERATÜR TARAMASI**

**Kaynakçada tespit edilen kaynaklar taranarak bu alan doldurulacaktır. Aşağıda konuyla ilgili daha önce hazırlamış olduğum notlar yer almaktadır. Bunlarla ilgili akademik referanslar ilave edilecektir.**

Lojistik Regresyon (LR), analizinin temelinde, bireylerin hangi grubun üyesi olduğunu kestirmede kullanılacak bir regresyon denklemi oluşturmak vardır. Bu analizde temel amaç, bağımlı (yordanan) ve bağımsız (yordayıcı) değişkenler arasındaki ilişkiyi, en az değişken ile en iyi uyuma sahip olacak biçimde tanımlayabilen, kabul edilebilir bir model kurmaktır. Kategorik bağımlı değişkenin değerini tahmin etmek, diğer bir ifadeyle iki ya da daha fazla gruba ilişkin “üyelik” tahminidir. Buna göre analizin amaçlarından birisi, sınıflandırmadır. Diğer bir amacı ise bağımlı ve bağımsız değişkenler arasındaki ilişkileri araştırmaktır.

Lojistik Regresyon modelinin temeli, üstünlük oranına (odds ratio) dayanır. Bir olayın meydana gelme olasılığının, meydana gelmeme olasılığına oranlanmasıyla üstünlük oranı elde edilir. Model kestiriminde en küçük kareler (ordinary least square) yöntemi yerine, en çok olabilirlik (maximum likelihood) yöntemi kullanılır. Bunun sebebi lojistik dönüştürme gerektiren doğrusal olmayan doğasıdır. LR, sapmaların karesini en az yapmaya çalışmaz, bir olayın olma ihtimalini en çok yapmakla ilgilenir.

## 2.1 Doğrusal Regresyon ve Lojistik Regresyon

Doğrusal regresyon analizi ve lojistik regresyon analizi arasındaki temel fark, LR’da bağımlı değişkenin ikili veya çoklu olması gerekir. DR’da ise tahmin edilecek bağımlı değişken süreklidir. DR’da hem bağımlı hem de bağımsız değişkenlerin sürekli olması gerekir. İkinci fark ise, DR’da bağımlı değişken değeri, LR’da ise bağımlı değişkenin alabileceği değerlerden birinin gerçekleşme olasılığı tahmin edilir. Üçüncü fark ise, DR’da bağımsız ve bağımlı değişkenlerin çoklu normal dağılım şartını sağlaması gerekir. LR’da bu şart aranmaz.

## 2.2 Diskriminant Analizi ve Lojistik Regresyon

Bağımlı ve bağımsız değişken ayrımının yapıldığı çok değişkenli bir modelde bağımlı değişken nominal ölçekli bir değişken olduğunda en küçük kareler tekniğiyle elde edilen tahminler yetersiz kalmaktadır. Çünkü ekk –en küçük kareler- tekniği bağımlı değişkenin normal dağılıma uyduğunu varsaymaktadır.

Bağımlı değişken nominal ölçekli ise ekk ya alternatif olarak kullanılabilen teknikler arasında diskriminant ve lojistik regresyon yer almaktadır. Fakat diskriminant analizinin grupları en iyi biçimde ayırabilmesi için bağımsız değişkenlerin normal dağılıma sahip olması ve bağımsız değişkenlerin kovaryanslarının her grup düzeyinde eşit olması gerekmektedir. Bu nedenle diskriminant analizinde bağımsız değişkenler arasında nominal veya ordinal ölçekli değişkenlerin kullanılması durumunda bu varsayımlar sağlanamamaktadır. Bu noktada lojistik regresyon gündeme gelmektedir. Çünkü lojistik regresyon analizinde bu iki varsayım aranmamaktadır.

## 2.3 Lojistik Regresyon Çeşitleri

LR kendi içinde bağımlı değişkene göre 3’e ayrılır. a) İkili LR (binary) b) Sıralı LR (ordinal) ve c) Çok Kategorili İsimsel (multi-nominal) LR.

İkili LR: Bağımlı değişken 0 – 1 gibi (yani temerrüt etmiş – etmemiş gibi) iki değer alır.

Sıralı LR: Bağımlı değişken 2 den fazla kategoriden oluşur. Örneğin, beğeniyorum, az beğeniyorum, beğenmiyorum, hiç beğenmiyorum gibi opsiyonlar vardır. Bir hiyerarşi vardır.

**Çoklu LR**: Bağımsız değişken en az 3 kategori olmalıdır. Kategoriler arasında üstünlük-altlık yani bir hiyerarşi yoktur.

## 2.4 Lojistik Regresyonun Yaygın Kullanım Nedenleri

1) Bağımlı değişken kategorik (kesikli, süreksiz) olmakla birlikte bağımsız değişkenler sürekli, kategorik yada ikilem olabilir. Bağımsız değişkenlerin sürekli yada süreksiz olmasına yönelik hiçbir kısıtlama getirmemektedir. 2) Kolaylıkla yorumlanabilmekte ve matematiksel olarak kullanımı kolay olan fonksiyonlar üretmektedir. 3) Lojistik modele dayalı analizleri yapmayı sağlayan çok sayıda bilgisayar paket programı (SPSS ve SAS gibi) vardır. 4) Negatif tahmin ihtimalleri üretmez. Tüm tahmin değerleri pozitiftir ve 0 ile 1 arasındadır. 5) Bağımsız ve bağımlı değişken arasındaki ilişkinin doğrusal olmasını gerektirmez. Doğrusal olmayan ilişkiyi koruyarak, ilişkinin formunu doğrusal hale getiren logaritmik dönüştürmeler yapar.

## 2.5 Lojistik Regresyonun Temel Varsayımları

1) Örneklem büyüklüğü hususunda genel bir kabul şöyledir: Bağımlı değişken ikiden fazla kategoriden oluşuyorsa her bir bağımsız değişkende en az 50 kişilik bir grup büyüklüğüne gerek vardır. 2) Çoklu bağlantı (multicollinearity): Bağımsız değişkenler arasındaki yüksek korelasyona oldukça duyarlıdır. Aralarında çoklu bağıntı olan değişkenlerin olması durumunda değişkenlerin azaltılması yani sorunu çözmeye yetecek kadar değişkenin çıkarılması gerekir. 3) Uç Değerler: 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. 4) Uygun Tüm Bağımsız Değişkenler Modele Dahil Edilmelidir: Bazı değişkenlerin modele dahil edilmemesi modelin yetersizliğine neden olabilir.

## 2.6 Lojistik Regresyonun Zayıf Yönü

LR’nun yansız ve sapmasız istatistikler ortaya koyabilmesi için büyük örneklemler gerekir. Küçük örneklemler için ideal değildir.

**3. UYGULAMA**

Bu çalışma ile bir bankanın verilerinden hareketle bir yıllık süre zarfında kredilerin batma, yapılandırılma veya canlı kalma ihtimallerini hesaplamaya yönelik regresyon modelleri kurulacaktır. Bir kesit tarihi örneğin 30.06.2014 baz alınarak sonraki bir yıldaki gerçekleşmelere bakılacaktır.

Bağımlı değişken 3 kategoriden oluşacaktır: a) Sonraki bir yılda temerrüt etmiş (yani batmış), b) sonraki bir yılda yapılandırma yapılmış (müşteri borcunu ödeyemediği için kredisinin vadesi değiştirilmiş) ve sonraki bir yılda kredilerini sorunsuzca ödemeye devam etmiş.

## 3.1 Tanımların Netleştirilmesi

Temerrüt tanımı tüm araştırmayı etkileyeceği için etraflıca tanımlanmalıdır. Tanımı şöyle olacaktır:

Müşteri, kesit tarihinde riski devam eden projelerinde veya performans dönemi içinde yeni kullandığı projelerinden herhangi bir tanesinde aşağıdaki durumlardan birisi gerçekleşirse,

(a) ödemesini 90 günden fazla geciktirirse veya (b) geç gün sayısı 90 gün olmadan takip hesaplarına virmanlanırsa veya (c) Temerrüt Etmiş Yapılandırma (Müşteriye ilk yapılandırma işlemi yapıldıktan sonra ikinci defa yapılandırma işlemi yapılıyorsa ve ilk yapılandırma işleminden karlı risk bazında % 10’dan az tahsilat yapılmışsa temerrüt olarak kabul edilecektir.) veya (d) İflas veya iflas erteleme ile ilgili bir karar varsa müşteri temerrüt etmiş sayılır. Temerrüt tarihi olarak hangisi önce gerçekleşmişse o kabul edilecektir.

## 3.2 Modelin Kurulacağı Tarihsel Dönem

Aşağıdaki 12 kesit alınarak 5 yıllık bir data üzerinden model kurulması sağlanmış olacaktır:

|  |  |  |  |
| --- | --- | --- | --- |
| 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 |

Her bir kesit tarihinden sonraki bir yıl içindeki performans değerlendirilecek. Yani, 12 ayrı performans dönemi olacak. Bu kesitler birleştirilerek analizler yapılacaktır.

## 3.3 Modellemeye Konu Müşteriler

Kesit tarihinde gecikmesi olmayan canlı müşteriler ve 90 güne kadar gecikmesi olan idari takip müşterileri üzerinden temerrüt tahmini yapılacaktır. Yasal takip hesaplarında izlenen müşteriler ve terkin müşterileri kapsam dışındadır.

Geliştirilecek olan modeller davranışsal modeller olacağı için müşterilerin bankada belli bir geçmişinin olması gerekmektedir. Genel uygulama olarak, banka ile 6 aydır çalışan müşteriler eski müşteri kabul edildi. Davranış modellerinde kullanılacak 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.

## 3.4 İstatistiki Analizler

**Kaynakçada tespit edilen kaynaklar taranarak bu alandaki referanslar doldurulacaktır. Aşağıda konuyla ilgili daha önce hazırlamış olduğum notlar yer almaktadır. Bunlarla ilgili akademik referanslar ilave edilecektir.**

## 3.4.1 Kayıp Değerler Açısından Kontrol

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. Fakat en ideal çözüm, kayıp değer kestirimi için regresyon yaklaşımından faydalanılmasıdır. Kayıp değerler içeren değişken bağımlı değişken olur. Birkaç bağımsız değişken, bağımlı değişkenin değerini tahmin etmede kullanılabilir.

Bizim modelde kullanacağımız yaklaşım ise şöyle olacaktır: Kayıp değer yüzdesi belli bir oranın üstündeyse değişkeni çıkartacağız. Sürekli değişkenleri kategorik hale getirirken kayıp değerler için de bir kategori oluşturacağız.

## 3.4.2 Uç Değerler Açısından Kontrol

Veri Girişinde yapılan hatalar: Uç değerler, veri girişiyle alakalıysa kontrol edilerek düzeltilmeli veya örneklemden çıkarılmalıdır.

Deneğin örneklemin alındığı evrenin üyesi olmaması: Bu durumda örneklemden çıkarılması gerekir.

Deneğin örneklemin geri kalanından farklı olması: En kolayı örneklemden çıkarılmasıdır. Fakat modele tanıtılmadıklarından dolayı model uygulamaya alındığında model bu örneklerle karşılaştığında yanlış sonuçlar verecektir. İdeal olan, uç değerlerin verilerde yer aldığı bir versiyon ile uç değerlerin yer almadığı bir versiyonun birlikte çalışılıp sonuçların karşılaştırılmasıdır.

## 3.4.3 Çoklu Bağlantı Açısından Kontrol

İlk denenmesi gereken alternatiflerden birisi örneklem büyüklüğünü arttırmak gerekir. Değişken de eklenebilir.

Eğer bu ikisi yapılamıyorsa, çözümü için faktör analizi yöntemiyle modeldeki değişkenleri birleştirilip elde edilen faktörlerle model kurulabilir. Ya da bu soruna neden olan değişkenler modelden çıkartılmalıdır.

Bu kontrol SPSS’de şöyle yapılır: Analyze 🡪 Regression 🡪 Linear… kısa yoluyla Linear Regression penceresi açılır. Bağımlı ve bağımsız değişkenler ilgili kutucuklara atılır. Statistics segmesinde aşağıdaki kutucuklar işaretlenir: Estimates, Model fit, R squared change, Part and partial correlations ve Collinearity diagnostics.

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.

Tolerans değeri incelenir. Bu değerin 0.1’den büyük olması durumunda, değişkenler arasında çoklu bağıntı sorununun olmadığını söyleyebiliriz.

Condition Index incelenir. Bu değerin 10 ile 30 arasında olması orta düzeyde, 30’dan büyük olması durumunda yüksek düzeyde bağıntı problemi olduğu anlaşılır.

## 3.4.4 Veri Gruplama

Sürekli değişkenler, veri gruplama yöntemiyle bağımlı değişken baz alınarak gruplandı.

Bu kontrol SPSS’de şöyle yapılır: Transform 🡪 Optimal Binning penceresi açılır. Bağımlı ve bağımsız değişkenler ilgili kutucuklara atılır. Save segmesinde aşağıdaki kutucuklar işaretlenir: Create variables that contain binned data values. Output segmesinde aşağıdaki kutucuklar işaretlenir: Endpoints for bins, Descriptive statistics for variables that are binned ve Model entropy for variables that are binned.

## 3.5 Veri Analizi

Orijinal veride sorun olduğu için test amaçlı küçük bir örneklem alınmıştır. O yüzden anlamlı istatistiksel testler yapılamamıştır. Veri ile ilgili sorun giderildiğinde araştırma nihayete erdirilecektir. Gerçek bir banka verisi kullanılacak olup veri temini hususunda banka yetkililerinden sözlü onay alınmıştır. Bankanın modelleme için veri hazırlığı bittiğinde çalışmaya başlanabilecektir.

Modellemedeki amaç bir regresyon modelinin geliştirilmesidir.

**4. SONUÇ VE ÖNERİLER**

Çalışma sonucunda geliştirilen regresyon modelleri ile banka gelecek bir yıllık süreçte aktif kalitesine dair öngörülere sahip olacaktır.

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 durum katılım bankaları için şu açıdan kritiktir: Katılım bankaları mevduat sahiplerine sabit bir kar payı garanti etmedikleri için banka zarar ettiğinde mevduat sahipleri de bu zarara ortak olmak zorundadırlar. Katılım bankası ancak kar ederse, mevduat sahiplerine kar payı dağıtabilecektir.

Küçük yatırımcılar her ne kadar yapılandırılan kredi tutarlarına bakmaksızın mevduatlarını bankalarda tutsalar da büyük yatırımcılar bu tutarlara dikkat etmektedirler.

Bu tahmin modeli, varlık kalitesi hakkında büyük mevduat sahiplerine ve sukuk (kira sertifikası) alıcılarına bilgi verirken ve diğer bankalar nezdinde bankanın kendisi için limit tahsisi süreçlerinde kullanabileceği değerli bilgiler sunabilir. Ayrıca, yapılandırmaya yönelik bu tahmin modelleri banka tarafından kredi risk yönetiminde kullanılabilir.

**5. KAYNAKÇA**

Akbulut, Y. 2010. Sosyal bilimlerde SPSS uygulamaları: Sık kullanılan istatiksel analizler ve açıklamalı SPSS çözümleri. İdeal Kültür Yayıncılık, 200, İstanbul.

Aydın, D., 2014. Uygulamalı regresyon analizi: kavramlar ve R hesaplamaları. Nobel Akademik Yayıncılık, 560, Ankara.

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. Pegem Akademi, 414, Ankara.

Dietmar Zellner, Frieder Keller & Günter E. Zellner (2004) Variable Selection in Logistic Regression Models, Communications in Statistics - Simulation and Computation, 33:3, 787-805, DOI: 10.1081/SAC-200033363.

Effects of Missing Data in Credit Risk Scoring. A Comparative Analysis of Methods to Achieve Robustness in the Absence of Sufficient Data Author(s): R. Florez-Lopez Source: The Journal of the Operational Research Society, Vol. 61, No. 3, Consumer Credit Risk Modelling; Transportation, Logistics and the Environment (Mar., 2010), pp. 486-501.

Mertler, C. A., Vannata, R. A., 2013. Advanced and multivariate statistical methods: practical application and interpretation. Pyrczak Publishing, 368, California.

Petrucci, Carrie 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.

Upadhyay, A., Bandyopadhyay, G., Dutta, A., 2012. Forecasting Stock Performance in Indian Market using Multinomial Logistic Regression. Journal of Business Studies Quarterly, 3(3), 16-39.

David W. Hosmer, Stanley Lemeshow. Applied Logistic Regression, Second Edition. Published Online: 28 JAN 2005

J. Scott Long, 1997. *Regression Models for Categorical and Limited Dependent Variables. Advanced Quantitative Techniques in the Social Sciences Number 7.* Sage Publications: Thousand Oaks, CA.

Allison, P. D. (1995). *Survival analysis using SAS*: *A practical guide*. Cary, NC: SAS Institute.

Carrie J. Petrucci (2009) A Primer for Social Worker Researchers on How to Conduct a Multinomial Logistic Regression, Journal of Social Service Research, 35:2, 193-205, DOI: 10.1080/01488370802678983

ELVİRA TURE, BANKALARDA KREDİLERİN YÖNETİMİ VE SORUNLU KREDİLERİN YENİDEN YAPILANDIRILMASI SÜRECİ, Yüksek Lisans Tezi, 2015, MARMARA ÜNİVERSİTESİ BANKACILIK VE SİGORTACILIK ENSTİTÜSÜ

Başak TANINMIŞ YÜCEMEMİŞ ve İnanç Asım SÖZER, BANKALARDA TAKİPTEKİ KREDİLER: TÜRK BANKACILIK SEKTÖRÜNDE TAKİPTEKİ KREDİLERİN TAHMİNİNE YÖNELİK BİR MODEL UYGULAMASI

Abdalla M. EL-HABIL, An Application on Multinomial Logistic Regression Model, Pak.j.stat.oper.res. Vol.VIII No.2 2012 pp271-291

Arun Upadhyay, Gautam Bandyopadhyay ve Avijan Dutta, Forecasting Stock Performance in Indian Market using Multinomial Logistic Regression. Journal of Business Studies Quarterly
2012, Vol. 3, No. 3, pp. 16-39

Sadi ELASAN, MULTİNOMİYAL LOGİSTİK REGRESYON ANALİZİ VE UYGULAMASI, Yüksek Lisans Tezi, 2010, YÜZÜNCÜ YIL ÜNİVERSİTESİ SAĞLIK BİLİMLERİ ENSTİTÜSÜ

Tika Arundina , Mohd. Azmi Omar, Mira Kartiwi, The predictive accuracy of Sukuk ratings; Multinomial Logistic and Neural Network inferences. [http://dx.doi.org/10.1016/j.pac***fi***n.2015.03.002](http://dx.doi.org/10.1016/j.pacfin.2015.03.002)

Anass BAYAGA, MULTINOMIAL LOGISTIC REGRESSION: USAGE AND APPLICATION IN RISK ANALYSIS. Journal of Applied Quantitative Methods, Vol.5 No.2 Summer 2010.