

BANKACILIK SEKTÖRÜNDE SORUNLU KREDİ OLUŐUMUNUN TESPİTİNE BİLİMSSEL YAKLAŐIM*

SCIENTIFIC APPROACH TO THE DETERMINATION OF PROBLEM LOANS IN BANKING INDUSTRY

Mustafa CERAN**
Serpil ERGÜN BÜLBÜL***

Öz

Dijitalleşmenin iş dünyasındaki önemi son zamanlarda artmaktadır. Yapay zekâ tabanlı uygulamaların kullanımı yaygınlaşmakta ve sektör içi rekabette başarının sağlanmasında bu uygulamalar ön plana çıkmaktadır. Bankacılık sektörünün de sorunlu kredi oluşumuna engel olmak adına, söz konusu trende uyararak, kredi analizi ve tahsisi süreçlerini dijital yapılarla kuvvetlendirmesi gerekmektedir. Bu sayede, olumsuz sonuçların meydana gelmesi öncesinde bertaraf edilmesi mümkün hale gelecektir. Bu çalışmada; kredilerin sorunlu hale gelmesi, bu kredileri kullanan işletmelerin finansal başarısızlıklarıyla ilişkilendirilmekte ve sorunlu kredi oluşumunun tespitine ilişkin uygulanabilecek bilimsel çalışmalar değerlendirilmektedir. Sonuçta, “yapay sinir ağlarının” sorunlu kredi oluşumunun tespitinde başarılı bir bilimsel yöntem olacağı ve bu yöntemin işletmelerin finansal verilerinin bir işleme sürecinden geçirilerek gerçekleştirilmesiyle daha başarılı çıktılar üreteceği öne sürülmektedir.

Anahtar Kelimeler: Finansal Başarısızlık, Kredi Analizi ve Tahsisi, Sorunlu Krediler, Yapay Sinir Ağı Modelleri.

JEL Sınıflandırması: C45, G21, G32, M41

Abstract

The significance of digitalization in the business world has been rising, recently. The usage of artificial intelligence based applications has become prevalent and these applications have come into prominence for providing success in intra-industry competition. Banking industry should fortify loan analysis and assignment processes by digital frameworks, in order to prevent problem loans, keeping up with this trend, as well. Thus, it will be possible to get rid of negative results before they happen. In this study,

* Bu çalışma ilk yazarın doktora tezinden türetilmiştir.

** Marmara Üniversitesi, Bankacılık ve Sigortacılık Enstitüsü, Bankacılık Anabilim Dalı. mustafa.ceran@windowslive.com, Orcid: 0000-0001-9496-7957

*** Marmara Üniversitesi, Bankacılık ve Sigortacılık Enstitüsü, Aktüerya Anabilim Dalı. serpilbulbul@marmara.edu.tr., Orcid: 0000-0002-6636-7329

problem loans are being associated with the financial failures of the companies that utilize these loans and scientific studies, which can be applied on determination of problem loans, are being evaluated. As a result, it is set forth that “artificial neural networks” shall be a prosperous scientific method in determination of problem loans and this method will generate more successful outputs realizing it after passing the financial data of the companies through a sifting process.

Keywords: Financial Failure, Loan Analysis and Assignment, Problem Loans, Artificial Neural Network Models.

JEL Classification: C45, G21, G32, M41

1. Giriş

Sorunlu kredilerin yönetimi ve bu kredilerden doğan kayıpların minimize edilmesi bankacılık sektöründe önemli bir yönetim tarzıdır. Bununla beraber; sorunlu olmaya aday kredilerin önceden tespit edilerek tahsisinin engellenmesi veya tahsisi yapıldıktan sonra sorunlu hale gelmesi muhtemel olan kredilere dair belirtilerin yakın izlenerek potansiyel sorunlu kredilerin tespit edilmesi de başarılı bir banka yönetimi için elzemdir. Dolayısıyla; potansiyel sorunlu kredilerin engellenmesi amacıyla bankalarca uygulanacak kredi analizi ve tahsisi politikaları, risk izleme uygulamaları ve risk tasfiye yöntemleri, ileride yaşanacak muhtemel olumsuz sonuçların bertaraf edilmesi ve bu sayede başarılı bir aktif yönetimi sağlanması adına bankalar için önem arz etmektedir.

İşletmelerin yaşadığı finansal başarısızlıklar literatürde çok sayıda esere konu olmuş olup bu doğrultuda çok sayıda analiz ve değerlendirme çalışması yapılmıştır. Kredi tahsis edilen işletmelerin finansal başarısızlıklarının bankalar için potansiyel sorunlu kredi oluşumuna yol açabileceği düşünüldüğünde, finansal başarısızlık üzerine literatürde yer alan çeşitli önemli çalışmaların mercek altına alınması yerinde olacaktır. Bu çalışmaların sonucunda elde edilen çıkarımlar, işletmelerin finansal başarısızlıkları ve muhtemel sorunlu kredi oluşumları üzerine önemli yönlendirmeler sağlayacaktır.

Finansal başarısızlık ve sorunlu kredi kavramlarının birbiriyle ilişkilendirildiği bu çalışmanın amacı; bankaların sorunlu kredi oluşumunu önceden tespit edebilmeleri adına kullanabilecekleri bilimsel yöntemlerden yapay sinir ağlarının başarısını ve uygulama sistematiğini değerlendirmektir.

2. Finansal Başarısızlık Üzerine Literatür Taraması

Finansal başarısızlığa yönelik çalışmalar 1930’lu yıllardan bu yana literatürde yer almaktadır. İlk zamanlarda yapılan çalışmalarda özellikle işletmelerin finansal oranları üzerinden değerlendirmeler yapılmış ve bu oranların işletmelerin iflas etmesinde önemli bir tahmin potansiyeline sahip olduğu çıkarımı yapılmıştır (Altman, 1968:590). Günümüze değin çok sayıda çalışmaya konu olmuş olan finansal başarısızlığa yönelik yapılan araştırmaları 1960lı yıllardan bu yana kronolojik sırayla değerlendirmeye almak yararlı olacaktır.

1960 – 1980 arasında yapılan araştırmalarda ön plana çıkartılabilecek çalışmalar; 1966 yılında Meir Tamari, yine 1966 yılında William Beaver ve özellikle 1968 yılında Edward Altman tarafından yapılan çalışmalardır. Tamari (1966), İsrail’de 28 başarısız işletme üzerine bir çalışma gerçekleştirmiş ve 1956-1960 yılları arasındaki verileri kullanmıştır. Bu çalışmanın en önemli katkısı, altı farklı orana farklı ağırlıkların verildiği bir risk endeksi üzerinden geliştirilerek çok boyutlu analize doğru bir adım olmasından ileri gelmektedir (Aktaş, 1991:41).

Beaver’ın çalışmasında (1966) ise 38 farklı endüstride faaliyet gösteren 79 sorunlu ve 79 sorunsuz işletme tespit edilmiş ve otuz adet finansal oran kullanılarak iflastan 5 yıl öncesine kadar analiz genişletilmiştir. Sorunlu ve sorunsuz işletmeler aynı endüstriden ve aktif büyüklüğü birbirine en yakın olan işletmelerden seçilmiştir. Bu çalışmada, öncelikle sorunlu ve sorunsuz işletmelerin finansal oranlarının ortalama değerleri birbiriyle kıyaslanmış ve her bir finansal oran için kritik noktalar saptanarak bu kritik noktalar doğrultusunda örnekte yer alan işletmelerin sorunlu hale gelip gelmeyeceği araştırılmıştır. Sonuç olarak sorunlu ve sorunsuz işletmeleri ayırmakta tahmin gücü yüksek olan altı adet değişken öne sürülmüştür. Bu değişkenler;

- 1) Nakit Akışı / Toplam Borçlar
- 2) Net Kâr / Toplam Aktifler
- 3) Toplam Borçlar / Toplam Aktifler
- 4) İşletme Sermayesi / Toplam Aktifler
- 5) Cari Oran (Dönen Varlıklar / Kısa Vadeli Borçlar)
- 6) (Koruyucu Aktifler – Kısa Vadeli Borçlar) / Operasyonel Fon Harcamaları

olarak belirlenmiştir. Bu oranlardan nakit akışı / toplam borçlar oranı, finansal başarısızlığı 1 yıl öncesinde %87 oranında doğrulukla tahmin etmiştir.

Altman’ın çalışmasında ise (1968); kendisinden önceki çalışmalarda genellikle, kârlılık, likidite ve borç ödeme gücünü ölçmeye yönelik oranların en önemli göstergeler olarak öne sürüldüğü, ancak bu oranlara dair bir önem sırasının net olmadığı, çünkü hemen hemen tüm çalışmalarda farklı bir oranın en önemli gösterge olarak aktarıldığı belirtilmiştir. Altman analizinde her biri 33 işletmeden ve biri sorunlu diğeri sorunsuz işletmelerden oluşan iki grup üzerine yoğunlaşmıştır. Sorunlu işletmeler iflas eden işletmelerden, sorunsuz olanlar ise varlığını sürdüren işletmelerden seçilmiştir. Bu gruplar değerlendirilirken; likidite, kârlılık, kaldıraç, borç ödeme gücü ve faaliyet oranları olmak üzere beş ana kategoride toplanan yirmi iki adet oran dikkate alınmıştır. Altman çoklu diskriminant analizini kullanarak bir model oluşturmuş ve Altman’ın oluşturduğu bu model; tüm örnekleme, 1 yıl öncesinde %95 oranında doğru sınıflandırma gerçekleştirmiştir. Bu modelin içerisinde değişken olarak şu oranlar yer almıştır:

- 1) İşletme Sermayesi / Toplam Aktifler
- 2) Dağıtılmamış Geçmiş Yıl Karları / Toplam Aktifler

- 3) Faiz ve Vergi Öncesi Kar / Toplam Aktifler
- 4) İşletmenin Piyasa Değeri / Toplam Borçlar
- 5) Satışlar / Toplam Aktifler

1980li yıllarda yapılan çalışmalar arasından ön plana çekilebilecek bir araştırma Ohlson'a aittir. Ohlson gerçekleştirdiği araştırmasında (1980), logit modelini kullanarak 1970-1976 yılları arasında dair bir bilgi setini araştırmasına konu etmiş ve bu bilgi setinde önceki çalışmalardan farklı olarak 105 iflas etmiş işletme ve 2.058 iflas etmemiş işletme yer almıştır. Çalışmasında dokuz adet değişkeni dikkate alan Ohlson; bu çalışmanın neticesinde, işletmelerin başarısızlığında 1 yıl öncesinden etkili olduğu istatistiksel anlamda tespit edilen 4 temel faktör öne sürmüştür:

- 1) Büyüklük
- 2) Finansal yapı
- 3) Performans
- 4) Likidite

Tahmin edilen süre bazında 3 farklı model oluşturan Ohlson'un Model 1 adını verdiği 1 yıl öncesinden yapılan tahminlerin doğruluk oranı %96,12 seviyesindedir.

Ülkemizde finansal başarısızlığının tahminine yönelik ilk çalışmalardan biri Erkut Göktan tarafından yapılan ve diskriminant analizinin kullanıldığı çalışmadır (1981). 25 başarılı ve 14 başarısız işletmenin 1976-1980 yılları arasındaki üçer yıllık finansal tablolarından üretilen on dokuz adet finansal oranın kullanıldığı çalışmada, iflastan 1 yıl öncesinin %92,9 oranında doğrulukla tahmin edildiği bir model çıktısı oluşturulmuştur.

Altman 1968 yılındaki çalışmasından sonra sunduğu kitabında (1983) ise; 53 iflas etmiş ve 58 iflas etmemiş işletmeyi içeren ve imalat sektörünün ağırlığının %50 olduğu yeni bir model çıkarımında bulunmuştur. Yedi değişkenli bir model üreten Altman'ın bu modelinde ise şu değişkenler yer almıştır:

- 1) Faiz ve Vergi Öncesi Kar / Toplam Aktifler
- 2) Faiz ve Vergi Öncesi Kar hesabının istikrarlılığı
- 3) Faiz Karşılama Oranı (Faiz ve Vergi Öncesi Kar / Faiz Giderleri)
- 4) Dağıtılmayan Geçmiş Yıl Karları / Toplam Aktifler
- 5) Cari Oran (Dönen Varlıklar / Kısa Vadeli Borçlar)
- 6) Özsermaye / Toplam Pasifler
- 7) İşletme Büyüklüğü (Maddi Varlıklar Toplamı)

Bu modelin doğruluğu %92,8 olarak hesaplanmış ve farklı sektörlerdeki uygulamalarda da başarılı sonuçlar görülmüştür.

Yine ülkemizde Targan Ünal tarafından sunulan eserde (1986), Türk İmalat Sektörü'ne ait veriler kullanılarak bir erken uyarı modeli oluşturulmuştur. Araştırmada; 1979-1984 yılları arasında gıda üzerine faaliyet gösteren işletmeler incelenmiş ve veriler Türkiye Sermaye Piyasası Kurulu'ndan temin edilmiştir. Araştırmanın birinci kısmında, başarılı ve başarısız işletmelerin faaliyetlerini açıklayan oran gruplarını tespit etmek amacıyla çeşitli çok değişkenli analiz yöntemleri kullanılırken, ikinci kısımda işletme gruplarının ayırımında etkili olabilecek finansal oranların belirlenmesi için başarısızlıktan bir yıl öncesinin verileri üzerinde birbirlerini destekleyici olarak; kümeleme, asal faktörler ve Q faktör analizleri uygulanmıştır. Araştırmanın son kısmında da birinci ve ikinci kısımda yapılan analizler bir arada değerlendirilmiş ve erken uyarı modelinde yer alacak finansal oranlar belirlenmiştir. Bu oranlar üzerinden uygulanan diskriminant analizinden elde edilen fonksiyon, yani erken uyarı modeli, başarısızlık tarihinin bir yıl öncesinden işletme başarısızlıklarını %97 olasılıkla tahmin edebilmiştir.

Türkiye'de 1990lı yıllarda da finansal başarısızlık üzerine önemli çalışmalar yapılmıştır. Örneğin; Ramazan Aktaş tarafından yapılan çalışmada (1991), çoklu diskriminant analizi, çoklu regresyon modeli, logit ve probit modelleri kullanılarak araştırma şekillendirilmiştir. 25 başarısız işletme ile 35 başarılı işletmenin incelemeye alındığı çalışmada işletmeler Türkiye Sermaye Piyasası Kanunu'na tabi olan 300 anonim şirketten seçilmiştir. İşletmelerde yapılan ayırım 3 yıl peşpeşe zarar etme ve iflas ya da finansal darboğazdan dolayı faaliyetlerini durdurma noktaları etkili olmuştur. Yirmi üç adet finansal orandan yapılan çıkışla önce on beş adet finansal orana indirgeme yapılmış, ayrıca dört finansal oranla ek bir değerlendirme yapılmıştır. Çalışmanın sonuçlarına göre; probit ve logit modeller değişken sayısı azalsa da son 1 yıllık tahmin gücü anlamında kuvvetli çıkmıştır.

Tablo 1: İstatistiki Yöntemlerin Başarısızlığı Tahmin Etme Olasılıkları

1 Yıl Öncesinden Tahmin Gücü	11 Değişkenli Modelin Tahmin Gücü	4 Değişkenli Modelin Tahmin Gücü
Doğrusal Diskriminant	%90,0	%86,7
Kuadratik Diskriminant	%85,0	%81,7
Çoklu Regresyon Modeli	%85,0	%88,3
Probit Model	%90,0	%90,1
Logit Model	%90,0	%90,1

Kaynak: Aktaş (1991: 176,177).

1999 yılında Serpil Ergün Bülbül tarafından yazılan bir doktora tezinde, İMKB'ye kayıtlı olan 33 başarılı ve 14 başarısız şirketin verileri kullanılarak, finansal başarısızlıkların öngörülmesinde doğrusal diskriminant analizi ve lojistik regresyon yöntemi dışında çok değişkenli istatistik yöntemlere alternatif olarak ilk kez Cox regresyon modeli kullanılmış ve ayrıca tahmin

modelinde kullanılan değişkenlerin önem sıralaması yapılmıştır. Elde edilen modellerin tahmin gücü oldukça yüksektir.

1990lı yıllarda; uluslararası literatürde görülen çeşitli araştırmalarda, yapay sinir ağları ile oluşturulan modeller diğer istatistiki yöntemlerle oluşturulan modellerle mukayese edilmiştir. Örneğin; yapılan bir çalışmada (Salchenberger, Çınar ve Lash, 1992), 100 başarılı ve 100 başarısız işletme üzerinden yapılan analizin sonucu olarak; finansal başarısızlığın tahmininde, yapay sinir ağları modelinin logit modelinden daha az varsayım gerektirmekle beraber daha yüksek bir sınıflandırma derecesiyle öngörüler sunduğu savunulmuştur. Başka bir çalışmada (Coats ve Fant, 1993) ise; 94 başarısız, 188 başarılı işletme analize tabi tutulmuş ve yapay sinir ağları ile geliştirilen modellerin özellikle finansal başarısızlığı son 2 yıl içerisinde öngörü gücünün, diskriminant yöntemiyle geliştirilen modellerin öngörü gücünden daha kuvvetli olduğu saptanmıştır. Yine başka bir çalışmada (Wilson ve Sharda, 1994), 64 başarılı ve 65 başarısız firma diskriminant ve yapay sinir ağları yöntemiyle deney ve kontrol grupları olarak ikiye ayrılmış ve analize tabi tutulmuştur. Doğru sınıflandırma noktasında yapay sinir ağları modellerinin diskriminant modellerine göre daha başarılı olduğu sonucu bu çalışmada da teyit edilmiştir.

2000li yıllarda yapılan araştırmalarda, hem yapay sinir ağları ile yapılan çalışmaların çokluğu hem de hibrit analizlerin yapıldığı çalışmalar dikkat çekmekte olup farklı yöntemlerin birleştirildiği analizler eşliğinde kredi skorlaması ve derecelendirmesine yönelik öngörüler sunulmaktadır.

Ramazan Aktaş, Mete Doğanay ve Birol Yıldız tarafından yapılan bir çalışmada (2003); 53'ü başarılı 53'ü başarısız olmak üzere 106 işletme analize tabi tutulmuştur. Çok boyutlu istatistiksel yöntemlerden çoklu regresyon modeli, diskriminant analizi ve logit modeli ile deney grubu (70 işletme) üzerinden finansal başarısızlık modelleri oluşturulmuş ve bu modeller kontrol grubu (36 işletme) üzerinde uygulanarak geçerlilik testi gerçekleştirilmiştir. Bu test sonuçlarına göre, 1 yıl öncesinden finansal başarısızlığı öngörmeye en iyi yöntemin çoklu regresyon modeli olduğu saptanmıştır. Bununla birlikte; aynı işlemler yapay sinir ağı modeli için de uygulanmış ve sonuç olarak yapay sinir ağı modelinin, 1 yıl öncesinde, çoklu regresyon modelinden de daha iyi bir öngörüye sahip olduğu belirlenmiştir. Yapay sinir ağları modeli; deney grubunda %95,71, kontrol grubunda ise %86,11 doğru öngörü gücüne sahip görülmüştür. Bu çalışmanın bir sonucu da tüm modellerde istatistiki anlamda önemli olduğu tespit edilen dört adet orandır:

- 1) Likidite Oranı ((Dönen Varlıklar – Stoklar) / Kısa Vadeli Borçlar))
- 2) Kısa Vadeli Borçlar / Özsermaye
- 3) Uzun Vadeli Borçlar / Toplam Aktifler
- 4) Dönem Karı / Özsermaye

Yasemin Keskin Benli tarafından yapılan başka bir çalışmada (2005) ise; lojistik regresyon ve yapay sinir ağı karşılaştırması yapılarak bu sefer bankalar araştırmaya konu edilmiş ve bankaların finansal başarısızlıklarına yönelik bir değerlendirme gerçekleştirilmiştir. Çalışmada 38 tane özel sermayeli banka yer almış, bunlardan 17 tanesi TMSF'ye devrolan bankalar arasından seçilmiş

ve finansal başarısızlığa konu edilmiştir. Çalışmada bankalara dair on iki adet finansal oran kullanılmıştır. Sonuçta; lojistik regresyon modeline göre yapay sinir ağı modeli daha doğru bir sınıflandırma yapmıştır.

Tablo 2: Lojistik Regresyon – Yapay Sinir Ağları Karşılaştırması

Lojistik Regresyon		
Başarısız	Başarılı	Genel
%76,5	%90,5	%84,2
Yapay Sinir Ağları		
Başarısız	Başarılı	Genel
%82,4	%90,5	%87,0

Kaynak: Belli (2005:42).

Hibrit çalışmalara örnek olarak ise; Hsieh tarafından yapılan bir araştırmada (2005), kümeleme ve yapay sinir ağları yöntemlerinin birleştirilmesiyle oluşturulan melez bir modellemenin kredi skorlamasında etkili sonuçlar üretebileceği öne sürülmüşken, Tsai ve Chen tarafından yapılan başka bir çalışmada (2010) da lojistik regresyon ve yapay sinir ağlarının kombine edilmesi sonucu elde edilen modelin kredi derecelendirmesindeki tahmin gücünün etkili bir seviyede gerçekleştiği tespit edilmiştir.

Finansal başarısızlık konusu, yakın dönemde de akademik çevreden ilgi görmeye devam etmektedir. Jacobs tarafından Güney Afrika'da yapılan bir çalışmada (2007), Altman'ın modellemelerine atıf yapılmış ve farklı sektörlerde faaliyet gösteren 20 tanesi tasfiye olmuş 20 tanesi faal olan toplam 40 tane firma üzerinden yapılan değerlendirmede; Altman modelinin, borsada işlem görmeyen bu firmalardaki doğru sınıflandırma gücünün en az %75 olduğu tespit edilmiştir.

Ülkemizde Mehmet Yazıcı tarafından hazırlanan bir doktora tezinde (2007), KOBİ niteliğindeki firmalar dikkate alınarak yıllık ciro hacmi 25 Milyon TL'ye kadar olan 42 tane başarısız, 58 tane de başarılı firma, 1993-2004 yıllarına ait veri setiyle, ticari bir bankanın kayıt sisteminden temin edilerek analize tabi tutulmuştur. Başarısızlık kriteri olarak bankaya kredi geri ödeme gecikme süresi 90 günün üzerine çıkmış olma durumu baz alınmış ve elli dört adet değişken dikkate alınarak diskriminant, lojistik regresyon ve yapay sinir ağları yöntemleriyle değerlendirme yapılmıştır. Yapılan analiz sonucunda; diskriminant analizi deney grubunda %85, test grubunda %34 doğru sınıflandırma yüzdesine sahip olması sebebiyle öngörü gücü zayıf olarak değerlendirilmiştir. Lojistik regresyon analizinde ise diskriminant analizinin de altında sağlıklı bir sonuç üretildiği tespit edilmiştir. Yapay sinir ağlarında ise deney grubu verilerinin doğru sınıflandırma yüzdesi %100 olarak sağlanabilmiş ve test grubundaki doğru sınıflandırma yüzdesi ise %96 olarak hesaplanmıştır. Sonuç olarak, yapay sinir ağları ile oluşturulan modelin, sağladığı doğru sınıflandırmayla ön plana çıktığı ve alternatif bir yöntem olarak dikkate alınmasının yerinde olacağı kanaati savunulmuştur.

Emre Ergin tarafından hazırlanan başka bir doktora çalışmasında (2009), İMKB (yeni adıyla Borsa İstanbul) üzerinden 2004-2006 arası finansal veriler kullanılarak işletmelerde finansal başarısızlık olasılığına dair erken uyarı araştırması yapılmıştır. Araştırmada son 3 yıldaki birikimli karı negatif olan, İMKB'de işlem sırası kalıcı olarak kapatılan, özsermayesi negatif olan ve gözaltı pazarına alınıp 1 yılı aşkın süredir burada kalan işletmeler sorunlu olarak sınıflandırılmıştır. Bu kapsamda, 308 toplam işletme içerisinde 93 başarısız işletme tespit edilmiştir. İşletmelerin sektörleri ise gıda, içki ve tütün sanayi, dokuma, giyim eşyası ve deri, kağıt ve orman ürünleri, kimya, petrol ve plastik ürünleri, taş ve toprağa dayalı sanayi, metal ana sanayi ve metal eşya, makine yapımı olarak çeşitlendirilmiştir. Araştırmada, pazar verilerine dayalı BSM modeli ile muhasebe verilerine dayalı Z-skor modeli uygulanmıştır. Sonuçta; modellerin genel başarıları açısından bakıldığında, muhasebeye dayalı verilerin daha üstün olduğu ortaya çıkmıştır.

Serkan Terzi tarafından yapılan çalışmada (2011) ise; gıda sektörü üzerine bir araştırma yapılmış ve işletmelerin finansal başarısızlıklarının incelenmesi amacıyla on dokuz adet finansal oran araştırmaya dâhil edilmiştir. Yapılan tekli ve çoklu istatistiksel analizlerle bu oranların önemli olanlarının sayısı altı adet olarak belirlenmiştir. Araştırmada 2009 ve 2010 yıllarında İMKB'de yer alan 22 işletmenin denetlenmiş finansal tabloları kullanılmış ve finansal başarısızlık tespitinde önceki kısımlarda açıklanmış olan, Altman'ın 1968 yılında sunduğu finansal oranlar uygulanmıştır. Diskriminant uygulamasının yapıldığı çalışmada modele sadece iki değişken dâhil edilmiştir. Bu değişkenler özkaynaklar / toplam borçlar ile faiz ve vergi öncesi kar / toplam aktifler oranlarıdır. Sonuçta modelin sınıflandırma başarıları %90,9 olarak saptanmıştır.

Yapay sinir ağlarının tahmin gücünü bireysel krediler kapsamında değerlendiren bir çalışmada (Budak ve Erpolat, 2012); Türk Bankacılık Sektörü'nde faaliyet gösteren bir bankadan yaklaşık bir yıl önce bireysel kredi kullanmış olan 1.639 kişi rastgele olarak seçilmiş ve bu kişiler kredilerini düzenli ödeyenler ve ödemeyenler olarak iki gruba ayrılmıştır. İstatistiki yöntemlerle analize tabi tutulan bu örneklem üzerinde yapılan çalışmanın sonucunda; yapay sinir ağları modelinin doğru sınıflandırma oranı %70,3 olarak tespit edilirken, lojistik regresyon modelinin doğru sınıflandırma oranının %65,1 olduğu görülmüştür. Benzer bir amaca yönelik, Ferdi Sönmez tarafından yayınlanan başka bir çalışmada (2015) ise; bankalara kredi talebinde bulunan bireysel müşterilerin değerlendirilmesine yönelik olarak yapay sinir ağlarını ve karar ağaçlarını dikkate alan bir çalışma gerçekleştirilmiş ve yapay sinir ağları metodolojisinin tahmin gücünün daha yüksek çıkması sebebiyle bu yöntemi merkeze alan bir yazılım modeli sunulmuştur. Bu modelin genelleştirme ve sınıflandırmada iyi rol oynaması, bu çalışmanın en önemli çıkarımı olmuştur.

Emin Zeytinoğlu ve Yasemin Deniz Akarım tarafından yapılan araştırmada (2013) ise; Serkan Terzi'nin daha önce bahsedilen 2011 yılında gıda sektörüne yönelik gerçekleştirdiği çalışmasına benzer şekilde, Altman'ın 1968 yılındaki yaklaşımı kullanılarak İMKB'de 2009-2011 yılları arasında faaliyet gösteren 115 işletme üzerinden başarısız olanlar tespit edilmiştir. 2009 yılında 75 başarılı 40 başarısız, 2010 yılında 50 başarılı 65 başarısız ve 2011 yılında 79 başarılı 36 başarısız işletme saptanarak bu işletmeler yirmi adet finansal oran kullanılarak diskriminant analizine tabi tutulmuştur. Oluşturulan modellerin 2009, 2010 ve 2011 yıllarındaki başarıları sırasıyla %88,7,

%90,4 ve %92,2 olarak hesaplanmıştır. Araştırmanın ilginç bir sonucu ise her 3 dönemde de önemli olduğu görülen iki adet oranın tespitidir. Bu oranlar; Özsermaye / Toplam Aktifler ile Net İşletme Sermayesi / Toplam Aktifler oranlarıdır.

Nurcan Öcal tarafından hazırlanan ve Ünal Aysal Tez Değerlendirme Yarışması'nın 2015 yılı doktora tezleri sıralamasında birincilik kazanarak 2016 yılında İktisadi Araştırmalar Vakfı'na yayınlanan doktora tezinde (2015) ise; imalat sektöründe faaliyet gösteren bir firmanın derecelendirme notunun belirlenmesi sürecinde lojistik regresyon ve karar ağacı algoritmaları kullanılarak modeller oluşturulmuştur. Bu çalışmada, Borsa İstanbul pay piyasasında işlem gören 206 tane imalat sektörü firmasının 2007-2013 dönemine ait kamuya açıklanmış finansal verileri dikkate alınarak otuz beş adet finansal oran değerlendirmeye alınmıştır. İyi ve kötü firma ayrımı üzerinden yapılan analizde kötü firma tanımında sayısal olmayan ölçütler kullanılmış ve sonuç olarak karar ağaçlarından C5 algoritması ile oluşturulan modelin kötü firmaları doğru sınıflandırmada en başarılı alternatifi sunduğu öne sürülmüştür.

Finansal başarısızlık öngörüsüne yönelik literatürde yer alan çalışmalar değerlendirildiğinde, yapay sinir ağları ile gerçekleştirilen analizlerin başarısı dikkat çekicidir. Yukarıda detayları aktarılan çalışmalarda olduğu gibi; bankacılık sektörüne yönelik yapılan diğer bazı çalışmalarda da yapay sinir ağlarının, doğrusal diskriminant analizine ve doğrusal olmayan logit ve probit analizlerine kıyasla en az onlar kadar ya da onlardan daha başarılı tahminler ürettiği, literatürde konu edilmiştir (McNelis, 2005:57). Sorunlu kredilerin temelde, işletmelerin finansal başarısızlıkları sonucunda meydana geldiği düşünüldüğünde ve literatürde yer alan çalışmaların sonuçları dikkate alındığında; uygulamada, bankacılık sektörünün yapay sinir ağlarını kullanarak sorunlu kredilerine yönelik modelleme oluşturmasının başarılı bir kredi yönetimi için yararlı olacağı düşünülmektedir.

3. Aktarma – Arındırma İşlemleri ve Finansal Tablo Manipülasyonlarının Giderilmesi

İşletmelerin mali yapılarının sağlıklı bir şekilde analiz edilebilmesi ve kredi tahsis kararlarının doğruluk kalitesinin artırılması adına; bankalara sunulan finansal tabloların sunulduğu şekilden öte, duru ve gerçeğe uygun bir şekilde değerlendirilmesi gerekmektedir. Bununla beraber, sorunlu kredi yönetimine yönelik kurulacak modellerin etkinliğinin ve verimliliğinin artırılması adına finansal verilerin bir işleme sürecinden geçirilerek modellenmesi yerinde olacaktır.

Bankalarca işletme mali verilerinde yapılması yerinde olacak aktarma – arındırma işlemleri Tablo 3'te detaylarıyla sunulmaktadır. Bu başlığın sonunda; Tablo 3'te detaylandırılan aktarma – arındırma işlemleri neticesinde oluşturulan ve analize uygun bir işletme bilançosu formatı (Şekil 1) ile bu bilançoyla eş zamanlı, analize uygun bir gelir tablosu formatı (Şekil 2) da sunulacaktır. Bu formatlarda; aktarma – arındırma işlemlerini tamamlayıcı nitelikte olarak, analizin kolaylığını ve etkinliğini artıracak hesap birleştirmeleri de yer almakta ve finansal tablo formatını özet, net ve özgün bir hale getirmektedir.

Tablo 3: Aktarma – Arındırma İşlemleri ve Finansal Tablo Manipülasyonlarının Giderilmesi

HESAP	TESPİT	SEBEP	UYGULAMA
Kasa	100.000 TL'nin üzerinde bakiye görülmesi (aktif büyüklüğü 5.000.000 TL'nin altında olan işletmelerde 50.000 TL'nin üzeri)	Fiziki olarak kasada bulunması makul görülmeyen tutar (kuyumculuk, döviz bürosu işletmeciliği ve perakende market zincirleri dışında)	100.000 TL'nin üzerinde görülen tutarın kasa hesabından silinmesi; yoruma göre ¹ , kaynağı makul görülüyorsa ² aktifte düzenleyici hesaba aktarılması ³ , aksi durumda özkaynaklara negatif olarak eklenmesi ⁴
Bankalar	Hesabın ters bakiye vermesi	Kredili mevduat hesapları kapaması yapılmaması	İlgili tutarların kısa vadeli banka borçlarına aktarılması
Verilen Çekler ve Ödeme Emirleri	Hesapta bakiye görülmesi	Aktif hesaplar arasında eksi çalışan, ters bir hesap olması	Pasif hesaplara, ticari borçlar altına alınması
Menkul Kıymetler	İflas etmiş, piyasa değerini kaybetmiş bir şirkete ait olduğunun görülmesi	Likidite özelliğini kaybetmiş olması	Yoruma göre; gelecekte düzelmesi muhtemel görülen şirketler için aktifte düzenleyici hesapta takip edilmesi, aksi durumda özkaynaklara negatif olarak eklenmesi
Alacak Hesapları	Vadesi geçmiş, avukata verilmiş, takipteki çekler ve senetler ile önceki dönemlerle kıyaslandığında hareketsiz olan alacakların görülmesi	Likidite özelliğini kaybetmiş olması	Özkaynaklara negatif eklenmesi (işletmenin sorunlu işletmelerden temin ettiği bir maddi teminatı varsa ve bu belgelenebilirse, ilgili tutarların aktifte düzenleyici hesap altına alınması)
Şüpheli Ticari Alacaklar	Bu hesabın tamamına karşılık ayrılmaması	İlgili alacakların tahsilat kabiliyetinin zayıf olması	Özkaynaklara negatif eklenmesi (işletmenin sorunlu işletmelerden temin ettiği bir maddi teminatı varsa ve bu belgelenebilirse, ilgili tutarların aktifte düzenleyici hesap altına alınması)
Ortaklardan Alacaklar	Hesapta bakiye görülmesi	Dönen varlık niteliği taşımama ihtimali olması	Yoruma göre; kaynağı makul görülüyorsa aktifte düzenleyici hesaba aktarılması, geri dönüşü olmadığı düşünülen bir tutar ise özkaynaklara negatif eklenmesi

- 1 Bu tabloda bahsi geçen “yoruma göre” ifadesi, kredi analizi ve tahsisi yapan kişilerin inisiyatif alanında olan durumları ve bu durumlarda analistlerin işletmeyi yerinde ziyaret, hesap hareketlerini detaylı inceleme gibi elde ettiği ek gözlemler ve tespitler sayesinde alabileceği aksiyonların farklılaşabileceğini ifade etmektedir.
- 2 Bu tabloda bahsi geçen “kaynağı makul görülme” durumu, işletme tarafından sunulabilecek bir fatura, tapu vb. belge ile desteklenecek durumları içermektedir.
- 3 Bu tabloda bahsi geçen “aktifte düzenleyici hesap” kavramı; yapılacak tespitler sonrasında, esasen dönen varlık niteliği taşımayan hesapların aktarılacağı ve aktif içerisinde dönen varlıklar ile duran varlıklar arasında konumlandırılacak “düzenleyici aktif” anlamına gelmektedir. Bu hesabın bilanço içerisinde konumlandırılacağı yer, ilerleyen kısımlarda belirtilmektedir.
- 4 Bu tabloda bahsi geçen “özkaynaklara negatif ekleme” hususu, özkaynaklar içerisinde yer alabilecek “indirimler” adlı bir hesap üzerinden yapılabilecektir. Bu hesabın bilanço içerisinde konumlandırılacağı yer, ilerleyen kısımlarda belirtilmektedir.

İştiraklerden / Bağlı Ortaklıklardan Alacaklar (İlişkili Taraflardan Alacaklar)	Hesaplarda bakiye görülmesi	Dönen varlık niteliği taşımama ihtimali olması ve ticaret ya da ticaret dışı kapsamda olması	1) İlgili hesaplarda mizan hareketliliği düşüğe bu tutarların Mali Duran Varlıklar hesabına aktarılması, ya da faaliyeti göstermeyen, batık yatırımlara aitse bunların özkaynaklara negatif eklenmesi 2) İlgili hesaplarda mizan hareketliliği var ve bu alacaklar ticari faaliyetlerle ilgiliyse vadesine göre ticari alacaklar hesabına, ticaret kapsamı dışındaysa vadesine göre diğer dönen (1 yıla kadar vadeli) ya da diğer duran varlıklara (1 yıldan uzun vadeli) aktarılması
Diğer Ticari Alacaklar, Diğer Çeşitli Alacaklar	Mizan analizinde hareketsiz alacaklar görülmesi ya da işletme ziyaretinde yapılan tespitler	Dönen varlık niteliği taşımama ihtimali olması	“Diğer” başlığı altında mizan hareketliliği düşük bir alacak var ise özkaynaklara negatif eklenmesi (işletmenin ilgili alacağa yönelik temin ettiği bir maddi teminatı varsa ve bu belgelenebilirse, ilgili tutarların aktifte düzenleyici hesap altına alınması)
Stoklar	Mizan analizinde hareketsiz stoklar görülmesi, şişik stoklar ya da işletme ziyaretinde yapılan tespitler	Satış kabiliyeti kalmayan stokların likidite özelliğini kaybetmiş olması	Özkaynaklara negatif eklenmesi
Stoklar, Ortaklara Borçlar	Net Satışlar çok artmazken, stoklarda yüksek bakiye görülmesi ve buna paralel ortaklara borçlar hesabının olması	Kayıtdışı satışlar sebebiyle gerçek olmayan stokların ve ortaklara borçların yer alması	Yoruma göre; kayıtdışı satış oranında stokların ve ortaklara borçların karşılıklı düşülmesi, kalan ortaklara borçların özkaynaklara pozitif eklenmesi ⁵
Stoklar, Duran Varlıklar, Finansman Giderleri	Yıllık kredinin maliyetiyle kıyaslandığında oldukça düşük finansman gideri işlenmiş olması ⁶	İşletmenin finansman giderlerini stoklar ve duran varlıklar üzerinde aktifleştirerek fiktif bir aktif yapısı ve kârlılık göstermesi	İşletmenin Kurumlar Vergisi Beyannamesi Gelir Tablosu Dipnotlarında belirtmesi halinde burdaki tutarın, belirtmemesi halinde mizanda tespit edilen tutarın, hiç tespit edilemiyorsa bankalardan kullandığı yıllık kredi hacminin yıllık kredi maliyeti ile çarpılması sonucu çıkan tutar ile yıllık finansman gideri arasında kalan tutarın aktiflerden (yoruma göre; tamamının stoklardan ya da stoklar ve duran varlıklardan karma bir şekilde) düşülerek finansman giderlerine eklenmesi
Banka Borçları	Memzuç verileriyle bilanço verilerinin karşılaştırılması sonrasında işletmenin olduğundan daha az banka borcu göstermesi ⁷ ya da toplamın aynı olması ancak vade uyumsuzluğu olması	İşletmenin gerçek banka borçluluğunun daha yüksek olması ya da kısa vadeli borçluluğunun daha yüksek olması	Öncelikle memzuç ve bilanço verilerinin kısa vadeli kısımları karşılaştırılarak kısa vadeli banka borçları, daha sonra uzun vadeli kısımları (12 ay üzeri borçlar) karşılaştırılarak tüm kısımları eşitlenir. Eğer memzuç verilerinde fazlalık varsa, kısa vadeli kısımlar eşitlendikten sonra bu fazlalık uzun vadeli banka borçlarına eklenir; bu eklemenin karşılığı bilanço da bir hesapta (örneğin ortaklara borçlarda olabilir) kayıtlıysa burdan silinir, değilse yoruma göre; bu fark aktifte dönen varlıklara veya düzenleyici hesaba eklenir ya da özkaynaklara negatif eklenir ⁸ .

5 Bu tabloda bahsi geçen “özkaynaklara pozitif ekleme” hususu, özkaynaklar içerisinde konumlandırılabilen “ihtiyatlar” adlı bir hesap üzerinden yapılabilmektedir. Bu hesabın konumu ilerleyen kısımlarda belirtilmektedir.

6 Bazı durumlarda işletmeler, finansman gideri olarak işlemesi gereken tutarları aktifleştirmeden satışların maliyeti hesabına da aktarabileceğinden, burada belirtilen aktarma – arındırma işleminin yapılmadan önce bu durumun

Ticari Borçlar, Stoklar ve Verilen Sipariş Avansları	Memzuç verilerinde akreditif riski bulunmakla beraber bu akreditife konu malların bilanço karşılığının olmaması	İşletmenin gerçek aktif toplamının, gerçek dönen varlık ve borçluluk yapılarının görülmemesi	Memzuç dönemlerinde akreditif riski içerisinde görülen ve bilanço'ya gerektiği şekilde yansıtılmayan tutarlar ticari borçlara ve karşılığında, ilgili akreditiflere konu mallar işletmeye ulaşmışa stoklara, henüz ulaşmamışa verilen sipariş avanslarına eklenir. ⁹
İştiraklere / Bağlı Ortaklıklara Borçlar (İlişkili Taraflara Borçlar)	Hesaplarda bakiye görülmesi	Grup içi fonlama kapsamında olması ve ticaret ya da ticaret dışı kapsamda olması	1) Bu borçların ticari faaliyetlerle ilgili olması halinde ve mizan hareketliliği varsa vadesine göre ticari borçlar hesabına aktarılması, 2) Ticaret kapsamı dışındaysa veya mizan hareketliliği kısıtlıysa vadesine göre diğer kısa ya da diğer uzun vadeli borçlara aktarılması
Ortaklara Borçlar	Hesapta bakiye görülmesi	Borç niteliği taşımaması	Yoruma göre; ortakların gerçek finansmanı ise özkaynaklara pozitif eklenmesi, aksi durumda düzenleyici pasif ¹⁰ altında izlenmesi
Alıcılar, Satıcılar	Hesapların ters bakiye vermesi	Ön ödeme ya da tahsilat yapan işletmelerin hesap kapaması yapmamış olması	Alıcılarda ters bakiye veren tutarların alınan sipariş avanslarına, satıcılarda ters bakiye veren tutarların verilen sipariş avanslarına aktarılması
Çeşitli Dönen Varlıklar, Çeşitli Uzun Vadeli Borçlar	Dönen varlıklar ve uzun vadeli borçlar altında izlenen ancak mizan analizinde ya da ziyarette farklı yapıda olduğu tespit edilen hesaplar	İşletmenin kısa vadeli borçluluğunun düşük, likidite yapısının olduğundan güçlü gösterilmesi	1) Dönen varlıklar altında izlenen ancak cari hesap niteliği taşımayan varlıkların aktifte düzenleyici hesaba aktarılması 2) Uzun vadeli borçlarda takip edilen ancak esasen kısa vadeli olan borçların uzun vadeden silinip kısa vadeli aynı hesap başlıkları altına alınması
Diğer Faaliyetlerden Olağan Gelirler, Diğer Olağandışı Gelir ve Kârlar	Duran varlıkların ya da stokların değerlemeye tabi tutulması ve bunun sonucunda oluşan değer artışlarının gelir tablosu hesapları üzerinden kaydedilmesi	İşletmenin kârlılığının reel olmayan bir şekilde yüksek gösterilmesi	Kapsamı tamamen yeniden değerlendirme ile ilgili olan tutarların ilgili gelir tablosu hesaplarından silinmesi ve özkaynaklara pozitif eklenmesi
700 Koduyla Başlayan Maliyet Hesapları	Geçici dönem mali verilerinin incelenmesi ve mizanda gerekli kapamaların yapılmadığının görülmesi	Bilanço – gelir tablosu dengesinin sağlanması	İlgili maliyetlerin mali tablolarındaki karşılığının tespit edilerek düzenleme yapılması (örneğin, satılmış ancak maliyet kapaması yapılmayan stokların maliyet hesaplarında görülmesi akabinde stokların eksiltilmesi)

netleştirilmesi gerekmektedir

- 7 İşletmelerin banka borçları ile memzuç verileri arasındaki farkın sebebi; işletmelerin yurtdışı kaynaklı kredi kullanıyor olması sebebiyle bu kredilerin bilançoda görülürken memzuç verilerinde gayrinakdi olarak görülmesi ya da hiç görülmemesi, ithalatlarının finansmanında açılan akreditiflerin bankalardan fonlanırken yani nakdi kredi de kullanılırken bunun yine memzuç verilerinde gayrinakdi olarak görülmesi, diğer yandan, yurtdışından kullanılan kredilerin

Not: İnşaat sektöründe faaliyet gösteren firmalar için yukarıda belirtilen aktarma – arındırma işlemleri farklı yorumlar gerektirebileceğinden aşağıdaki noktalara dikkat edilmelidir:

- 1) Yap-Sat tarzı projeler geliştiren firmalarda henüz teslim edilmemiş işlerle alakalı olan ve giderleştirilemeyen maliyetler olarak nitelendirilebilecek; kasa hesabındaki yüksek bakiye, ortaklardan alacaklar hesabındaki şişiklik ya da diğer çeşitli alacaklarda görülen tutarların inşaat projeleri ile ilgili harcamalar kapsamında olması halinde bu tutarlar stoklar hesabı altında takip edilebilir ya da yoruma göre bu hesaplarla ilgili olarak yukarıda belirtilen uygulamalara tabi tutulabilir. Taahhüt projeleri geliştiren firmalarda, bu kapsamdaki tutarların ise yıllara yaygın inşaat ve onarım maliyetlerinde takip edilmesi ya da yoruma göre bu hesaplarla ilgili olarak yukarıda belirtilen uygulamalara tabi tutulması yerinde olacaktır.
- 2) 1) no'lu maddede belirtilen durumdan farklı olarak projeler teslim edilmiş ise; ilgili tutarların hem yap-sat hem de taahhüt firmalarında, projeler aynı bilanço döneminde teslim edilmiş satışların maliyeti hesabına aktarılması, önceki yıllarda teslim edilmiş özkaynaklara negatif eklenmesi yerinde olacaktır. İlgili tutarlar, yoruma göre, bağlı oldukları hesaplara ilişkin yukarıdaki tabloda belirtilen uygulamalara da tabi tutulabilir.
- 3) Yap-Sat tarzı projeler geliştiren firmaların ortaklara borçlar hesabında takip ettiği ve kayda alınmamış olan satış tutarları; ilgili işler bitmemişse alınan avanslara, aynı bilanço döneminde bitmiş net satışlara aktarılmalı (ilgili stoklar da satışların maliyetine aktarılmalı), önceki dönemlerde bitmiş stoklarla karşılıklı kayıt dışılık elimine edikten sonra özkaynaklara pozitif eklenmelidir. Aynı durum taahhüt projeleri geliştiren firmalar kapsamında da, projeler bitmemişse yıllara yaygın inşaat ve onarım hakedişlerine, aynı yıl bitmiş net satışlara aktarım yapılmalı (ilgili maliyetler de satışların maliyetine aktarılmalı), önceki dönemlerde bitmiş aktifteki kayıt dışılıkla beraber elimine edildikten sonra özkaynaklara pozitif eklenmelidir.
- 4) Taahhüt projeleri geliştiren firmaların projelerine ilişkin aldıkları avanslar ve gelecek aylara ve yıllara ait gelirleri, bir ön hakediş niteliği içermeleri sebebiyle yıllara yaygın inşaat ve onarım hakedişlerine; gelecek aylara ve yıllara ait giderler, personel ve iş avansları, verilen sipariş avanslarında bulunan ya da aktifte bulunması gereken hesap altında olmayan taşeronlara verilen avanslar ve inşaat projelerine fiziken dâhil olmuş varlıklara ilişkin proje maliyetleri ise yıllara yaygın inşaat ve onarım maliyetleri hesaplarına aktarılmalıdır. Yap-sat tarzı proje geliştiren firmalarda görülecek benzeri durumlarda ise aktifte stoklar, pasifte alınan avanslar hesabına aktarım yapılmalıdır.
- 5) Taahhüt projeleri geliştiren firmaların ara dönem ya da geçici yıl sonu mali verilerinde; bilanço kesinleşmemişken ve mizan üzerinden analizler yapılırken, 700 koduyla başlayan maliyet hesaplarında yapılmamış kapamalara ilişkin tutarlar; projeler bitmemişse stoklara (bu harcamalara dair varlıklar projeye fiziken eklenmediyse) veya yıllara yaygın inşaat ve onarım maliyetlerine (bu harcamalara dair varlıklar projeye fiziken eklendiyse), projeler bitmiş satışların maliyetine aktarılmalıdır. Yap-sat tarzı proje geliştiren firmalar için bu maddede belirtilen duruma dair tutarlar; projeler bitmemişse stoklara, projeler bitmiş satışların maliyetine aktarılmalıdır.

bankalar tarafından nakdi krediler içerisinde hem aracılık edilen / garanti verilen krediler arasında hem de diğer nakdi kredilerde bildirilmesi yoluyla memzuç verilerinin bilançodan daha yüksek görülmesi yüzünden de oluşabilmektedir. Bu tabloda anlatılan aktarma – arındırma işlemleri; bu durumların

- 8 Memzuçta yer alan reeskont tutarı kısa vadeli banka borcuna ve karşılığında düzenleyici aktife eklenebilir. Ayrıca; bankaların memzuç bildirimlerinde yaşanabilecek hatalardan dolayı, memzuçtaki orta vadeli kredilerinin yarısının kısa vadeli, diğer yarısının da uzun vadeli olduğunu baz almak şeklinde bir uygulama da tercihe açıktır.
- 9 Söz konusu mallar alınıp satıldıysa, satışlarda gösterilme durumu da incelenerek satışların maliyetine eklenebilir.
- 10 Bu tabloda bahsi geçen “düzenleyici pasif” kavramı; yapılacak tespitler sonrasında, esasen tam olarak tanımlanması yapılamayan tutarların aktarılacağı bir hesaptır. Bu hesabın konumu ilerleyen kısımlarda belirtilmektedir.

Şekil 1. Analiz Formatına Uygun İşletme Bilançosu

AKTİF	PASİF
CARİ VARLIKLAR	KISA VADELİ BORÇLAR
Nakit ve Nakit Benzeri Değerler	Banka Borcu
Ticari Alacaklar	Ticari Borçlar
Stoklar	Diğer Kısa Vadeli Borçlar
Verilen Sipariş Avansları	
Diğer Cari Aktifler	
	UZUN VADELİ BORÇLAR
	Banka Borcu
	Diğer Uzun Vadeli Borçlar
DÜZENLEYİCİ VARLIKLAR	DÜZENLEYİCİ KAYNAKLAR
Geçici Aktif ¹¹	Geçici Pasif
Düzenleyici Aktif	Düzenleyici Pasif
Yıllara Yaygın İnşaat ve Onarım Maliyetleri	Yıllara Yaygın İnşaat ve Onarım Hakedişleri
DURAN VARLIKLAR	ÖZKAYNAKLAR
Mali Duran Varlıklar	Ödenmiş Sermaye
Arsa ve Binalar	Net Dönem Karı
Tesis, Makine ve Cihazlar	Net Dönem Zararı (-)
Yapılmakta Olan Yatırımlar	İhtiyatlar
Diğer Duran Varlıklar	Geçmiş Yıl Zararları (-)
Birikmiş Amortismanlar (-)	İndirimler (-)

Şekil 2. Analiz Formatına Uygun İşletme Gelir Tablosu 1

Net Satışlar (1)
Satışların Maliyeti (2)
Faaliyet Giderleri (3)
Faaliyet Kârı (1-2-3)
FAVÖK (1-2-3+8)
Finansman Giderleri (4)
Net Kambiyo Karı (5)
Net Diğer Gelirler (6)
Vergi Öncesi Kar (1-2-3-4+5+6)
Vergi (7)
Net Dönem Karı (veya Zararı) (1-2-3-4+5+6-7)
Dönemin Amortisman Gideri (8)
Amortisman Öncesi Net Kâr (veya Zarar) (1-2-3-4+5+6-7+8)

11 Geçici aktif ve geçici pasif; aktifte ve pasifte yer almakla beraber gelir tablosundaki gelir ve giderlere dönüşecek geçici hesapları ifade etmekte olup, bu hesapların dönen/duran varlıklar ve kısa vadeli/uzun vadeli borçlar harici takip edilmesi, bilançonun verimli analizi açısından anlamlı olacaktır. Bu hesaplar altında toplanacak alt kalemler ilerleyen kısımda belirtilecektir.

Gerekli aktarma – arındırma işlemlerinin yapılması akabinde oluşturulacak Şekil 1’de belirtilen bilançoda¹² ve Şekil 2’de belirtilen gelir tablosunda, kredi analizinin etkinliğini artıracak hesap birleştirmeleri de yapılmış olup bu sayede özgün ve net bir finansal tablo yapısı elde edilmiştir. Aktarma – arındırma sürecini tamamlayan birleştirmeler şunlardır:

AKTİF

Nakit ve Nakit Benzeri Değerler: Kasa, Bankalar, Diğer Hazır Değerler¹³, Menkul Kıymetler.

Ticari Alacaklar: Dönen Varlıklar içerisindeki Alıcılar, Alınan Çekler, Alacak Senetleri (Net), Diğer Ticari Alacaklar.

Stoklar: İlk Madde ve Malzeme, Yarı-mamuller, Mamuller, Ticari Mallar, Diğer Stoklar, Stok Değer Düşüklüğü Karşılığı (-).

Verilen Sipariş Avansları: Aynı isimle Verilen Sipariş Avansları.

Geçici Aktif: Gelecek Aylara Ait Giderler, İş Avansları, Personel Avansları ve Kısa ve Uzun Vadeli Alınan Avanslara ilişkin Aktifte yer alan bağlı tutarlar¹⁴.

Düzenleyici Aktif: Aktarma – arındırma işlemleri sonrası oluşan düzenleyici hesap, bu işlemler sonrası kalan Uzun Vadeli Ticari Alacaklar ve Uzun Vadeli Diğer Alacaklar ve Sayım ve Tesellüm Noksanları.

Yıllara Yaygın İnşaat ve Onarım Maliyetleri: Aynı isimle Yıllara Yaygın İnşaat ve Onarım Maliyetleri.

Mali Duran Varlıklar: Aynı isimle Mali Duran Varlıklar ve aktarma – arındırma işlemi ile gelen hesaplar.

Diğer Cari Aktifler: Yukarıda belirtilen dönen varlıklar dışında kalan dönen varlıklar.

Arsa ve Binalar: Arazi ve Arsalar, Yeraltı ve Yerüstü Düzenleri, Binalar.

Tesis, Makine ve Cihazlar: Tesis, Makine ve Cihazlar, Taşıtlar, Demirbaşlar.

Yapılmakta Olan Yatırımlar: Yapılmakta Olan Yatırımlar, Verilen Avanslar (Duran Varlık).

Diğer Duran Varlıklar: Yukarıda belirtilen duran varlıklar dışında kalan tüm duran varlıklar (birikmiş amortismanlar ve itfa ve tükenme payları hariç).

Birikmiş Amortismanlar: Tüm duran varlıklar için ayrılmış olan birikmiş amortismanlar ve itfa ve tükenme payları.

PASİF

Banka Borcu (Kısa Vadeli): Kısa Vadeli Mali Borçlar.

Ticari Borçlar: Kısa Vadeli Borçlar içerisindeki Satıcılar, Borç Senetleri (Net), Diğer Ticari Borçlar.

12 Alt hesap gruplarında (-) işareti ile belirtilen hesaplar aktif ve pasif toplamına negatif, diğer hesaplar ise pozitif katılmaktadır.

13 Diğer Hazır Değerler altında yer alan ve bir ayın üzerinde vadeye sahip olan kredi kartı alacaklarının ticari alacaklarda takip edilmesi daha doğru olacaktır.

14 Önceden belirtildiği gibi; inşaat taahhüt firmalarında projelere ilişkin hesaplar, yıllara yaygın inşaat ve onarım maliyetleri hesabına; inşaat yap-sat firmalarında projelere ilişkin hesaplar, stoklar hesabına alınmalıdır. Alınan avanslara ilişkin aktifte yer alan bağlı tutarlar aktarmaya tabi tutulurken; söz gelimi toplam 10 Milyon TL tutarında alınan avans aktifte hangi hesaplara yansıtıldıysa, söz konusu hesaplardan toplam 10 Milyon TL olacak şekilde geçici aktife aktarım yapılır.

Geçici Pasif: Kısa ve Uzun Vadeli Alınan Avanslar, Gelecek Aylara ve Yıllara Ait Gelirler¹⁵.

Düzenleyici Pasif: Aktarma – arındırma işlemleri sonrası oluşan düzenleyici hesap ve Sayım ve Tesellüm Fazlaları.

Yıllara Yaygın İnşaat ve Onarım Hakedişleri: Aynı isimle Yıllara Yaygın İnşaat ve Onarım Hakedişleri.

Diğer Kısa Vadeli Borçlar: Yukarıda belirtilen kısa vadeli borçlar dışında kalan kısa vadeli borçlar. Banka Borcu (Uzun Vadeli): Uzun Vadeli Mali Borçlar.

Diğer Uzun Vadeli Borçlar: Yukarıda belirtilen uzun vadeli borçlar dışında kalan uzun vadeli borçlar.

Ödenmiş Sermaye: Ödenmiş Sermaye toplamı.

Net Dönem Kârı (veya Zararı): Gelir tablosunda oluşan Net Dönem Kârı (veya Zararı) olup konsolidasyona tabi işletme gruplarında kâr / zarar eden işletmelerin bir arada olması halinde her iki hesapta da bakiye bulunması mümkündür.

İhtiyatlar: Aktarma – arındırma işlemleri sonrası oluşan hesap ile Sermaye Yedekleri, Kâr Yedekleri, Geçmiş Yıl Kârları.

Geçmiş Yıl Zararları (-): Aynı isimle Geçmiş Yıl Zararları.

İndirimler (-): Aktarma – arındırma işlemleri sonrası oluşan hesap.

GELİR TABLOSU

Net Satışlar: Brüt Satışlar ile Satış İndirimleri arasındaki fark.

Satışların Maliyeti: Aynı isimle Satışların Maliyeti.

Faaliyet Giderleri: Aynı isimle Faaliyet Giderleri.

Net Diğer Gelirler: Kambiyo Kârları ve Zararları ile Faiz Gelirleri hariç olmak üzere Diğer Faaliyetlerden Olağan Gelir ve Kârlar, Olağandışı Gelir ve Kârlar, Diğer Faaliyetlerden Olağan Gider ve Zararlar, Olağandışı Gider ve Zararlar hesaplarının toplamının neti.

Net Kambiyo Kârı: Kambiyo Kârları ve Zararları hesaplarının arasındaki fark.

Finansman Giderleri: Finansman Giderleri ile Faiz Gelirleri hesaplarının neti (aralarındaki fark).

Vergi: Dönem Kârı Vergi ve Diğer Yasal Yükümlülük Karşılıkları (Kurumlar Vergisi).

Dönemin Amortisman Gideri: Bağımsız Denetim Raporunda veya Kurumlar Vergisi Beyannamesinde yazan dönemin tüm amortisman giderleri ile itfa ve tükenme payları, eğer bunlar temin edilemiyorsa mizandaki 7 ile başlayan hesaplarda bulunan amortisman giderlerinin toplamı, bu da mümkün değilse son 2 yılın birikmiş amortisman farkı baz alınabilir (son 2 yılın birikmiş amortisman farkının da negatif çıkması durumunda ilgili dönem için amortisman gideri 0 kabul edilebilir; bu hesaplama yapılırken, farklı duran varlıklar için ayrılan amortismanlar solo bazda değerlendirilebilir).

Gerek sunulan verilerin doğası itibarıyla, gerek dönem sonlarında (özellikle 3, 6 ve 9 aylık ara dönemlerde) işletmelerin hesap kapamalarını henüz yapmamış olması sebebiyle ve gerekse manipülatif yöntemlere tabi tutulabilmeleri sebebiyle, işletme mali verilerinin kredi analizine uygun hale getirilebilmesi için yukarıda belirtilen düzenlemelerin yapılması yerinde olmaktadır. Bu

15 Önceden belirtildiği gibi; inşaat taahhüt firmalarında projelere ilişkin hesaplar, yıllara yaygın inşaat ve onarım hakedişleri hesabına; inşaat yap-sat firmalarında projelere ilişkin hesaplar, alınan avanslar hesabına alınmalıdır.

düzenlemeler; çeşitli hesapların farklı hesaplara aktarılmasını, manipülasyonların arındırılmasını ve bu doğrultuda basite indirgenmiş, özet ve analize uygun bir finansal tablolar bütünlüğüne kavuşulmasını sağlayacaktır. Bu bütünlük üzerinden yapılacak yapay sinir ağları modellemesi de işletmelerin kredibilitesinin daha net bir şekilde görülmesine imkân oluşturacaktır.

4. Sonuç

Bankacılık sektörünün ana fonksiyonunu oluşturan ve aktif toplamının önemli bir kısmını teşkil eden kredilerin sağlıklı yönetimi, banka yönetimlerinin sürdürülebilir bir finansal dengeyi tesis edebilmeleri adına ciddiye arz etmektedir. Diğer yandan; bankacılık mevzuatında yapılan son güncellemeler, Uluslararası Finansal Raporlama Standartları'na uyum süreçleri ve dijitalleşmenin getirdiği etkiler, kullanılan krediler sorunlu hale gelmeden önce alınacak tedbirlerin önemini artırmakta ve kredilerin dinamik yönetimini oldukça değerli hale getirmektedir.

Bankalarca işletmelere kullanılan kredilerin sorunlu hale gelmesini önceden tespit etmeye yönelik bir çalışma gerçekleştirilirken, sorunlu kredilerin temelinde yatan sebebin söz konusu işletmelerin finansal başarısızlıklarıyla ilgili olduğunu öngörmek yerinde olacaktır. Literatürde yapılan çalışmalar dikkate alındığında, yapay sinir ağları üzerine gerçekleştirilen çalışmaların finansal başarısızlığın tahmininde başarılı sonuçlar üreterek ön plana çıktığı görülmektedir.

Yapay sinir ağları ile gerçekleştirilen modelleme uygulamalarının başarısının artırılabilmesi adına; söz konusu uygulamaların gerçekleştirilmesi öncesinde, analize tabi tutulacak işletmelerin finansal verilerinin bir işleme sürecinden geçirilerek (bu çalışmada belirtildiği şekilde) daha net, özgün ve manipülasyonlardan arındırılmış bir yapıyla uygulamaya dâhil edilmesi fayda sağlayacaktır. Modellemenin çıktı kalitesinin temelinde yer alan verilerin bu çalışmada sunulan işleme süreci sonrasında rafine edileceği öngörülmekte ve bu doğrultuda isabetli bir modelleme çıktısına ulaşılacağı ileri sürülmektedir.

KAYNAKÇA

- Aktaş, R. (1991). Endüstri İşletmeleri İçin Mali Başarısızlık Tahmini – Çok Boyutlu Model Uygulaması. *Yayınlanmamış Doktora Tezi*, Ankara Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı.
- Aktaş, R., Doğanay, M. ve Yıldız, B. (2003). Mali Başarısızlığın Öngörülmesi: İstatistiksel Yöntemler ve Yapay Sinir Ağı Karşılaştırması. *Ankara Üniversitesi SBF Dergisi*, Cilt 58, No.4: 1-24.
- Altman, E. (1968). Financial Ratios, Discriminant Analysis And The Prediction of Corporate Bankruptcy. *The Journal of Finance*, XXIII (4): 589-609.
- Altman, E. (1983). *Corporate Financial Distress: A Complete Guide to Predicting, Avoiding and Dealing With Bankruptcy*. New York: John Wiley & Sons.
- Beaver, W. (1966). Financial Ratios as Predictors of Failure. *Journal of Accounting Research*, 4(3): 71-111.
- Bülbül, S.E. (1999). İstatistiksel Başarısızlık Zamanı Modelleri ve Finansal Analizlerde Uygulaması. *Yayınlanmamış Doktora Tezi*, Marmara Üniversitesi, Sosyal Bilimler Enstitüsü, Ekonometri Anabilim Dalı, İstatistik Bilim Dalı.

- Belli, Y.K. (2005). Bankalarda Mali Başarısızlığın Öngörülmesi Lojistik Regresyon ve Yapay Sinir Ağı Karşılaştırması. *Gazi Üniversitesi Endüstriyel Sanatlar Eğitim Fakültesi Dergisi*, Sayı16: 31-46.
- Budak, H. ve Erpolat, S. (2012). Kredi Riski Tahmininde Yapay Sinir Ağları ve Lojistik Regresyon Analizi Karşılaştırması. *Online Academic Journal of Information Technology*, 3(9): 23-30.
- Coats, P.K. ve Fant, L.F. (1993). Recognizing Financial Distress Patterns Using A Neural Network Toll. *Financial Management*, 22(3): 142-155.
- Ergin, E. (2009). İşletmelerde Finansal Başarısızlık Olasılığının Erken Tanısı: İMKB Uygulaması. *Yayınlanmamış Doktora Tezi*, Kocaeli Üniversitesi İşletme Anabilim Dalı Muhasebe Finansman Programı.
- Hsieh, N. (2005). Hybrid Mining Approach In The Design Of Credit Scoring Models. *Expert Systems With Applications*, 28: 655-665.
- Jacobs, J.J. (2007). The Application Of Failure Prediction Models On Non-Listed Companies. *Tshwane University of Technology Faculty of Management Sciences Business School*: 65-71.
- Mcnelis, P.D. (2005). *Neural Networks in Finance: Gaining Predictive Edge In The Market*. USA: Elsevier Academic Press.
- Ohlson, J. (1980). Financial Ratios And The Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 18(1): 109-131.
- Öcal, N. (2016). Firma Derecelendirmesi ve Borsa İstanbul'da İmalat Sektörü İçin Model Önerisi, *İktisadi Araştırmalar Vakfı İktisadi İşletmesi*, Yayın No.26: 153-204.
- Salchenberger, L.M., Cinar, E.M. ve Lash N.A. (1992). Neural Networks: A New Tool For Predicting Thrift Failures. *Decision Sciences*, 23(4): 899-916.
- Sönmez, F. (2015). Kredi Skorunun Belirlenmesinde Yapay Sinir Ağları ve Karar Ağaçlarının Kullanımı: Bir Model Önerisi, *ABMYO Dergisi*, Sayı 40: 1-22.
- Tamari, M. (1966). Financial Ratios as a Means of Forecasting Bankruptcy. *Management International Review*, 6(4): 15-21.
- Terzi, S. (2011). Finansal Rasyolar Yardımıyla Finansal Başarısızlık Tahmini: Gıda Sektöründe Ampirik Bir Araştırma. *Çukurova Üniversitesi İİBF Dergisi*, 15(1): 1-18.
- Tsai, C. ve Chen M. (2010). Credit Rating By Hybrid Machine Learning Techniques. *Applied Soft Computing*, 10: 374-380.
- Ünal, T. (1986). Firma Başarısızlıklarının Belirlenmesinde Bir Erken Uyarı Modeli, *Yayınlanmamış Doktora Tezi*, İstanbul Teknik Üniversitesi.
- books?id=s7zonQEACAAJ&dq=irma+ba%259Far%C4%B1s%C4%B1zl%C4%B1klar%C4%B1n%C4%B1n+belirlenmesinde+bir+erken+uyar%C4%B1+modeli&hl=tr&sa=X&ei=RwuIVb_LNISWsgG4zoLYCg&ved=0CBsQ6AEwAA, (22.06.2015).
- Wilson, R.L. ve Sharda, R. (1994). Bankruptcy Prediction Using Neural Networks. *Decision Support Systems*, 11(5): 545-557.
- Yazıcı, M. (2007). Bankalarda KOBİ Kredilerini Değerlendirmeye İlişkin Bir Yaklaşım: Yapay Sinir Ağları. *Yayınlanmamış Doktora Tezi*, Kadir Has Üniversitesi Sosyal Bilimler Enstitüsü Finans ve Bankacılık Doktora Programı.
- Zeytinoğlu, E. ve Akarım, Y.D. (2013). Financial Failure Prediction Using Financial Ratios: An Empirical Application On Istanbul Stock Exchange. *Journal of Applied Finance & Banking*, 3(3): 107-116.

DOES TURKEY SUFFER FROM MIDDLE INCOME TRAP OR IS IT STUCK IN PERMANENT LOW GROWTH? *

TÜRKİYE ORTA GELİR TUZAĞINDA MI YOKSA SÜREKLİ DÜŞÜK BÜYÜMENİN KISKACINDA MI?

Fatma DOĞRUEL**
A. Suut DOĞRUEL***

Abstract

Explaining the disparities in growth rates across both developing and developed countries is one of the vital questions in the theory of economic growth. As an answer to this question the term middle-income trap was introduced about a decade ago and it quickly became popular in politics, media and academics. The aim of the paper is to seek to what extent this term is capable to explain the growth performance of the Turkish economy. Following a brief discussion on middle-income trap, the paper focuses on the Turkish economic growth. It is concluded that the major problem of Turkey is persistent slow long-run growth rather than the growth slowdown following fast growth as suggested by the literature.

Keywords: Growth Slowdown, Middle Income Trap, Turkey.

JEL Classification: O11, O40, O47

Öz

Ekonomik büyüme teorisindeki temel sorulardan biri hem gelişmiş hem de gelişmekte olan ülkeler arasındaki büyüme oranlarındaki farklılıkların açıklanmasına dönüktür. Yaklaşık on yıl önce bu sorunun yanıtlanmasında, orta gelir tuzağı kavramı ortaya çıkarılmıştır ve siyaset biliminde, medyada ve akademiye hızlı bir şekilde popülerlik kazanmıştır. Bu çalışmanın amacı, bu kavramın ne ölçüde Türkiye ekonomisinin büyüme performansını açıklayabildiğinin araştırılmasıdır. Orta gelir tuzağı üzerine kısa bir tartışmanın yürütülmesinin ardından çalışma, Türkiye'nin ekonomik büyümesi üzerine odaklanmaktadır. Çalışmada ulaşılan temel sonuç, Türkiye'nin ilgili literatürün savunduğu hızlı büyümeyi takiben yavaşladığı görüşünden daha çok sürekli olarak yavaş seyreden bir uzun dönem büyümesine sahip olduğuna yöneliktir.

* The authors would like to thank the anonymous reviewers for their evaluations. The authors would also like to thank the Marmara University Scientific Research Committee for its financial support (Project No: SOS-D-120.514.0195).

** Marmara University, E-mail: fatma.dogrueel@marmara.edu.tr, Orcid: 0000-0003-1113-4621

*** Marmara University, E-mail: suut.dogrueel@marmara.edu.tr

Anahtar Kelimeler: Büyümedeki Yavaşlama, Orta Gelir Tuzağı, Türkiye.

JEL Sınıflandırması: O11, 040, 047

I. Introduction

The disparities in growth rates across both developing and developed countries are one of the vital questions in the theory of economic growth. The cross-country empirical growth studies constitute a vast literature on this issue. However, as Elhanan Helpman states “... *the mystery of economic growth has not been solved*” (Helpman, 2004: 2). The term of middle-income trap raised under this circumstance and became popular among politicians, academia, and media. Then Indermit S. Gill and Homi Kharas wrote a review paper on the literature developed over ten years and they assert that “... *these papers do not use a common definition. Instead, the term has been loosely used to describe situations where a growth slow-down results from bad policies in middle-income countries that prove difficult to change in the short-run (hence, “trap”)*” (Gill and Kharas, 2015: 6).¹

Although Indermit S. Gill and Homi Kharas emphasize the fuzziness of the term, an empirical literature rapidly developed. Growth slowdown in the middle-income countries is the main issue in these studies. Based on these debates, the main question of this paper is to ask whether Turkey holds the dynamics of middle-income trap or has a persistent low growth rate. The paper examines the growth path of Turkey starting from the beginning of the Republican Period and checks whether there is a growth slowdown by adopting the conditions of Eichengreen et al. (2012; 2014). The long-term growth path displays a very close movement to a linear trend. In contrast to China and South Korea there was no high growth rates during the period that we observe. Therefore, we cannot say that Turkish economy suffered from middle income trap. The reason behind staying 50 years in the lower middle-income group is low growth rate and it seems that Turkey will wait long time in the upper middle-income group to enter into high-income group.

This paper is organized in the following sections: Section 2 describes the term of middle-income trap. Section 3 investigates the growth performance of Turkey. The last section concludes the paper.

2. Middle-Income Trap: The Fuzziness of the Term

This section seeks to clarify the term by briefly reviewing the literature around middle-income trap. The term of middle-income trap was first used in a World Bank publication which was prepared by a World Bank team (Gill and Kharas, 2007: xiii). The geographic focus of the book was East Asia. The book has discussed the remarkable growth successes and export performances of this region countries. The authors called these developments as “*an economic renaissance*” (Gill and Kharas, 2007: 2). Later the term is widely adopted in the development literature and by

1 Gill and Kharas (2015: 6) also state that “*Academics also became interested in the subject. By May 2015, a search of Google Scholar returned over 3,000 articles including the term “middle-income trap” and close to 300 articles with the term in the title.*”

politicians and media as well. The term is still used to explain growth and development problems facing middle-income countries. Although a vast literature followed the contribution of Gill and Kharas (2007) more than a decade, there are skeptical remarks on the term. For example, Felipe et al. (2012: 44) state that “*The problem with the debate of what prevents these [middle-income] countries from becoming high-income economies is that it is not clear what the trap refers to, as there is no accepted definition.*” And, Han and Wei (2017) pose a question “*Is this a bubble, or is it a conceptual breakthrough?*” By considering this fuzziness first, we discuss how Gill and Kharas (2007) introduced the term. Then we discuss the growth slowdown and the transition from one income group to another.

Gill and Kharas (2007) did not define a term as the middle-income trap. Instead, they focused on the growth of East Asian Countries based on the fundamentals of modern growth theory. They defined what should be the transformation of these countries while they are middle-income countries. The transformation that they define was related to a suitable growth strategy. Their approach shows how a trap may arise and whether it could be inevitable. Kharas and Kohli (2011) take one step forward and try to define a transition strategy for middle-income countries on how they may avoid from middle income trap: they highlight first, the importance of transition from diversification to specialization in production structure; second, the role of education and innovation in total factor-productivity growth; and third, decentralization in the economic decision mechanism. However, the strategy does not correspond to an exact definition. These highlights recall some fundamentals from growth theory. We should remember what Helpman (2004) says on economic growth: He emphasizes accumulation of physical and human capital, total factor productivity, and economic and political institutions as the factors on the growth performance of a country. The debates on middle-income trap are also linked economic development theory. For example, productivity changes are related well-known literature on structural change in development economics. The origin of the productivity and growth nexus is based on sectoral shifts in an economy during economic development. Sectoral shifts represent structural change *a la* Kuznets: As an economy develops first the share of agriculture decreases in favor of manufacturing and then the share of industry declines and the share of service increases (Kuznets Facts).² Therefore, Gill and Kharas (2007) lead a new discussion on growth problems of middle income countries rather than provide an empirical definition for middle-income trap.³

The literature predominantly deals with the empirical side of this issue rather than theoretical side. The debates widely arise from empirical studies in which the observations mainly focus on growth slowdown. The growth slowdown was not only experienced in 2000s. Ben-David and Patell (1998) state that there were growth slowdowns in major continental countries and Japan in the

2 “Characteristics of modern economic growth, Item 4” (Kuznets, 1966: 492). See also Kongsamut et al. (2001).

3 After ten years, in their review paper they state that:

“We introduced the term “middle-income trap” while writing a report to assess economic developments in East Asia since the crisis of the 1990s. We did so with modesty, because we had not rigorously established its prevalence. (...) To our surprise, the phrase “middle-income trap” immediately became popular among policy makers and development specialists (...) As government leaders repeatedly referred to the term, first academics and then the mainstream media started to adopt the term” (Gill and Kharas, 2015: 5).

early 1970s while the US, Canada and the UK were exceptions.⁴ Besides, it is a phenomenon not only in developed countries, but also developing countries have experienced growth slowdowns. Ben-David and Patell (1998) indicate the years 1978 through 1983 for growth slowdown for developing countries. We may also mention about Eichengreen et al. (2012; 2014) for the growth slowdown issue. Growth slowdown is matter for fast growing countries. Eichengreen et al. (2012) focus on the timing of the slowdown in fast-growing economies and the effects of characteristics and conditions of countries on timing: They consider fast-growing countries and exclude low income ones, and the data starts from 1957 to the current year available. In Eichengreen et al. (2012: 46) growth slowdown was determined by the following conditions:

“The first condition requires that the 7-year average growth rate of per capita GDP is 3.5 percent or greater prior to the slowdown (earlier growth was fast). The second one identifies a growth slowdown with a decline in the 7-year average growth rate of per capital GDP by at least by 2 percentage points (the slowdown is nonnegligible). The third condition limits slowdowns to cases in which per capita GDP is greater than US\$ 10,000 in 2005 constant international PPP prices.”

As a condition, they also add that the reduction in the growth slowdown should be continuous for seven years (Eichengreen et al., 2014). The empirical analyses reveal that growth slowdown may typically appear in two different income level: the first is \$10,000–11,000 and the second is around \$15,000–16,000 (Eichengreen et al., 2014).⁵

Aiyar et al. (2018) is another empirical study which focus on growth slowdown. Although Aiyar et al. (2018: 2) is an empirical study, the authors are critical about the earlier contributions on this topic and say that “[they] propose a clear identification procedure for growth slowdowns, one that takes theory seriously rather than simply relying on structural breaks in the time series patterns of economic growth.” By emphasizing the fundamentals of growth theory, Aiyar et al. (2018) present institutions, demographic structure (high dependency ratio), macroeconomic environment and policies, economic structure (sectoral composition and sectoral shifts), trade structure, and other factors like wars and civil conflicts as key factors of continuous growth slowdown. Similar factors (as independent variables) also take place in the empirical analyses of (Eichengreen et al., 2012 and 2014) with some variations. Thus, it is possible to say the leading empirical studies rest, in general, on the existing growth theory. However, their frameworks/approach is restricted with the existing discourse of globalization. The variables employed to explain growth slowdown are related to the old paradigm (Washington Consensus).

The growth slowdown is an important fact to discuss existence of middle-income trap. As we discussed above, empirical studies meticulously investigate the patterns of growth slowdown for developed and developing countries. Their methods help to identify the countries in the

4 They indicate the following European countries: France, Germany, and Italy (Ben-David and Patell, 1998).

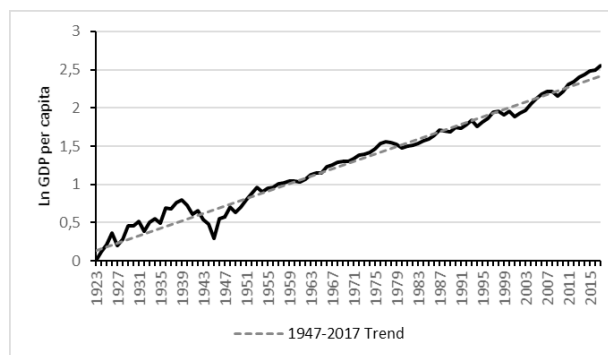
5 Income levels are defined in terms of constant 2005 PPP \$. Although they identified two different income level, in their first analysis there exists only one income level around \$15,000–16,000 at which growth slowdowns typically happen (their data source for the first and second analyses respectively are Penn World Tables 6.3 and Penn World Tables 7.1) (Eichengreen et al., 2012; 2014).

trap with few exceptions. However, there is another approach to identify the countries which are capturing or escaping middle income trap. This approach classifies countries by per capita income and investigate the pattern of transition from one income group to another. Felipe et al. (2012), first classify countries by their GDP per capita income in 1990 PPP dollars (the time span is 1950-2010). The middle-income countries are classified in two groups: between \$2,000 and \$7,250 lower-middle-income and \$7,250 and \$11,750 upper-middle-income countries. One of the important contributions of Felipe et al. (2012) is calculation of “*threshold number of years for a country to be in the middle-income trap*”. Im and Rosenblatt (2013) use transition matrices to identify the countries in middle-income trap and the transition of countries from one income group to another. Bulman et al. (2014) again use income classification for their analyses. Although most of the studies do not include Turkey in the analyses, Bulman et al. (2014) focus on Turkey as a counterpart of non-escapees which are Mexico, Malaysia, and Brazil: these countries displayed relative stagnation in their growth path for the entire period considered. The data they use is taken from Pen Word Table 7.

3. Do Growth Slowdown and Middle-Income Trap Exist in Turkey?

Considering the discussions presented above the main question to answer is “Is there any growth slowdown in Turkey in the long-run?” Figure-1 displays log of GDP per capita during 1923-2017 setting the 1923 value as equals to one in Turkey. Excluding the turbulent years following the foundation of the Republic and the Second World War, log of GDP per capita moves around a linear trend line during 1947-2017. Downward deviations from the long-run trend correspond to the major economic crises such as 1977-1980, 2001 and 2008-2009. Following 1 or 2-year recovery period, Turkish economy turned to its long-run growth trend. Persistent stay above the long-run trend correspond to the years where comprehensive development planning was implemented and the years after 2008-2009 crisis.

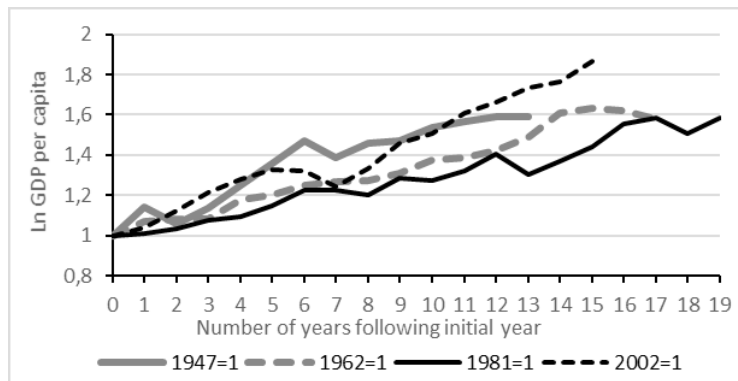
Figure 1: GDP per capita in Turkey (Ln, 1923-2017)



Source: TURKSTAT database.

To separate the post war period into the sub-periods may provide a closer look at the Turkish economic growth. The period following 1947 can be defined as the entrance to the Western Block of divided world economy. 1963-1979 is the comprehensive development planning period. 1981-2000 is transition from inward looking to export orientation and liberalization period. The period starting 2002 following 2001 crisis can be marked as the intensification of the integration of the Turkish economy with the global financial markets. GDP growth for each sub-period displayed in Figure-2, by setting initial year's GDP per capita is equal to 1. Although the growth performances slightly vary across sub-periods, there is no significant downward or upward breaks in the long-run growth trend.

Figure 2: GDP per capita in Turkey by Periods (Ln)



Source: TURKSTAT database.

These observations reveal that the Turkish economy did not experienced permanent growth slowdown during last 70-year period. Variations across sub-period partly can be explained by the internal factors related with the economic growth. However, growth performance of an economy cannot be isolated from performance of the world economy. Turkey raised from lower middle-income group to upper middle-income group in 2005 (Table-2). Table-1 compares the Turkish average annual GDP per capita growth rates with lower middle-income group average for the period before the year 2005 and with upper middle-income group average for the period after that year. Both inward looking and outward oriented development periods (1961-1976 and 1981-2004), Turkish average growth rates are more than one point above the average of the lower middle-income group. Despite this relative success of Turkey stayed 50 years in the lower middle-income group. Two success stories in economic growth (China and South Korea) stayed much shorter period in lower middle-income stage (Table-2). Considering the high growth rate score of these countries and quite stable growth trend of Turkey it is possible to conclude that the long stay of Turkey in lower middle-income stage is the result of its relatively low growth rate rather than lower middle-income trap.

Table 1: Average Annual Per capita Growth in Turkey and Income Groups

	Turkey	Lower middle income	Upper middle income
1961-1976	3.36	2.09	
1981-2004	2.61	1.70	
2005-2017	4.35		4.75

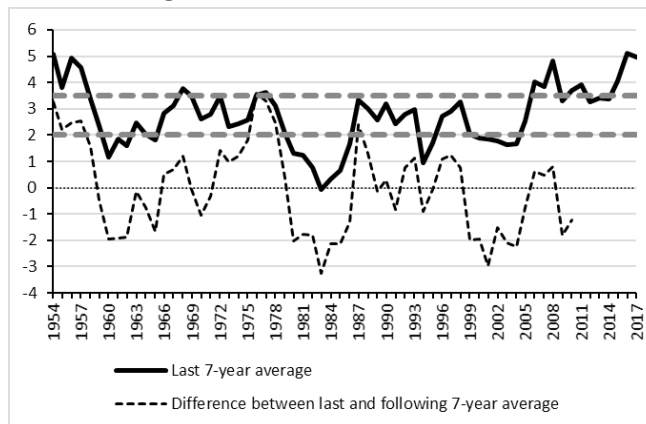
Source: The World Bank, World Development Indicators database.

Table 2: From Lower Middle Income to Upper Middle Income

Country	Year country turned Lower Middle Income	Year country turned Upper Middle Income	Number of years as Lower Middle Income
China	1992	2009	17
South Korea	1969	1988	19
Turkey	1953 and 1955	2005	50

Source: Felipe, Abdon and Kumar (2012, Table 3).

Growth slowdown following a fast growth epoch is shown as the main indicator of the middle-income trap. Using the conditions defined by Eichengreen et al. (2012: 46) for growth slowdown, 7-year average growth rate and the average growth rates difference between two successive 7-year period are displayed in Figure-3. 7-year average annual per capita GDP growth rate exceeds 3.5 percent level (upper horizontal dashed line) only in the decade following the Second World War and last two decades. In the first half of the 1970's, 7-year average growth rate touched to 3.5 percent level. During the first decade following the War and the first half of the 1970's second condition is satisfied: average growth rates difference between two successive 7-year period is above 2 percent level (lower horizontal dashed line). However, in these periods PPP GDP per capita was not above US\$ 10,000. These results show that three conditions did not simultaneously satisfied after the War. Therefore, it is not possible to conclude that the Turkish economy suffered from middle income trap. Considering the comparative growth rates presented in Table-1 and linear long rung growth trend, we may expect that Turkey will wait long time in the upper middle-income group to enter into high-income group.

Figure 3: Growth Slow-Down Measure

Source: TURKSTAT database.

4. Conclusion

The term of middle-income trap became popular during the last decade. Politicians, academia and media enthusiastically embraced the term. The term became a buzzword to explain the development problems that middle-income countries faced. However, the story of the first appearance and spreading of the term displays awkwardness. Indermit S. Gill and Homi Kharas are the pioneer of this term and it was first used in Gill and Kharas (2007). The authors explain their aim as just to prepare a report on East Asia to assess the economic developments of the period afterwards of 1990s economic crisis (Gill and Kharas, 2015). Although there was no clear definition for middle-income trap, a vast literature emerged on this famous term. The empirical literature developed by two ways: the first focuses on the growth slowdown after a high growth period. The second focuses on the classification on income groups and the transition pattern from one income group to another. Low income countries are excluded, and emphasis is given on middle-income countries.

The paper examines Turkey as a case. The empirical analysis on Turkey suggests a slow but steady growth. We observe that Turkey is stuck in the middle-income in a very long time due to slow growth on a flat linear trend rather than slowing after a rapid growth. Therefore, this is not a middle-income trap that literature suggested. It is a smooth slow growth performance. This may be explained partly as a unique growth pattern. However, Filipe et al. (2012) display many countries who are stuck in low-income or middle-income levels.

References

- Aiyar, S., Duval, R., Puy, D., Wu, Y. and Zhang, L. (2018). Growth slowdowns and the middle-income trap. *Japan and the World Economy*, 48 (December): 22-37.
- Ben-David, D. and Papell, D. (1998). Slowdowns and Meltdowns: Postwar Growth Evidence from 74 Countries. *Review of Economics and Statistics*, 80: 561-571.
- Bulman, D., Eden, M. and Nguyen, H. (2014). Transitioning from Low-Income Growth to High-Income Growth: Is there a Middle-income Trap? *WB Policy Research Working Paper No. 7104*. World Bank, Washington, DC.
- Eichengreen, B., Park, D. and Shin, K. (2012). When Fast-Growing Economies Slow Down: International Evidence and Implications for China. *Asian Economic Papers*, 11(1): 42-87.
- Eichengreen, B., Park, D. and Shin, K. (2014). Growth slowdowns redux. *Japan and the World Economy*, 32(November): 65-84.
- Felipe, J, Abdon, A. and Kumar, U. (2012) Tracking the Middle-income Trap: What Is It, Who Is in It, and Why? *The Levy Economics Institute Working Paper*.
- Gill, I. and Kharas, H. with Deepak Bhattasali et al. (2007). *An East Asian Renaissance: Ideas for Economic Growth*. World Bank, Washington, DC.
- Gill, I. S., and Kharas, H. (2015). The Middle-Income Trap Turns Ten. *Policy Research Working Paper 7403*. World Bank, Washington, DC.
- Han, X. and Wei, S-J. (2017). Re-examining the Middle-Income Trap Hypothesis (MITH): What to Reject and What to Revive? *NBER Working Paper No. 23126*, February.

- Helpman, E. (2004). *The Mystery of Economic Growth*. Belknap Press of Harvard University Press, Cambridge.
- Im, F., and Rosenblatt, D. (2013). Middle-income traps: A conceptual and empirical survey. *World Bank Policy Research Working Paper* No. 6594, World Bank, Washington, DC.
- Kharas, H., and Kohli, H. (2011). What Is the Middle Income Trap, Why do Countries Fall into It, and How Can It Be Avoided? *Global Journal of Emerging Market Economies*, 3(3): 281–289.
- Kongsamut, P., Rebelo, S. and Xie, D. (2001). Beyond Balanced Growth. *The Review of Economic Studies*, 68(4): 869-882.
- Kuznets, S. (1966). *Modern economic growth: rate, structure and spread*, Yale University Press, London.

DİJİTAL REKLAM VERİLERİNDEN YARARLANARAK POTANSİYEL KONUT ALICILARININ RASTGELE ORMAN YÖNTEMİYLE SINIFLANDIRILMASI

CLASSIFICATION OF POTENTIAL RESIDENTIAL BUYERS BY USING RANDOM FOREST METHOD TAKING ADVANTAGE OF DIGITAL ADVERTISING DATA

Haydar EKELİK*
Dilek ALTAŞ**

Öz

Günümüzde internet ağlarının yaygınlaşması ve internete erişimin bir ihtiyaç haline gelmesi internet sitelerinde ve diğer dijital platformlardaki reklamların kullanılmasını yaygınlaştırmıştır. Dijital reklamcılık olarak adlandırılan bu süreç firmalar, markalar ve diğer kuruluşlar için insanlara ulaşma ve reklam amaçları doğrultusunda hedeflerini gerçekleştirmelerinde gerekli bir reklam aracı olmuştur. En önemli özelliği ölçülebilir olan dijital reklamcılık, firmalara çok geniş veriler (istatistikler) sağlamaktadır. Firmalar bu verileri kullanıp dijital reklamların değerlendirmesini yaparak gelecek reklam planları için ön görüye sahip olurlar. Bu çalışmanın amacı bir inşaat firmasının dijital reklam kampanyasından elde edilen kullanıcı verilerini kullanarak bir sınıflandırma yapmaktır. Kullanıcıların satış ofisine gelip gelmediklerinin kaydının tutulduğu veriler analiz edilerek bir sınıflandırıcı oluşturulmuştur. Bundan sonraki süreçte reklamlarla elde edilen kullanıcı verileri bu sınıflandırıcı kullanılarak sınıflandırılabilir. Böylece kullanıcıların satış ofisine gelip gelmemeleriyle ilgili bir ön bilgi elde edilir. Firma bu ön bilgi sayesinde satış ve pazarlama hedeflerini daha doğru bir şekilde belirleyebilir. Çalışmanın amacı doğrultusunda bağımlı değişken olarak kullanıcıların satış ofisine gelip gelmemesi, bağımsız değişken olarak ise dijital reklamlar sayesinde kullanıcın iletişim bilgilerini hangi gün firma çalışanlarına gönderdiği, kullanıcının cinsiyeti, reklamı hangi sitede görüp siteye geldiği, reklamı hangi reklam alanında (doğal, 300*250 görsel boyutlu vb.) gördüğü, hangi cihazdan (bilgisayar veya telefon) gördüğü, kullanıcının daha önce ilgili firmada kayıtlı olup olmaması ve bu formu hangi amaçla doldurduğu (yatırım, ev sahibi olma vb.) olmak üzere toplamda 7 adet bağımsız değişken kullanılmıştır. Uygulamada R programından yararlanılmış ve verileri analiz etmek için bir topluluk

* İstanbul Üniversitesi İktisat Fakültesi Yöneylem Anabilim Dalı. haydar.ekelik@istanbul.edu.tr,
Orcid: 0000-0002-0661-4164

** Marmara Üniversitesi İktisat Fakültesi İstatistik Anabilim Dalı. dilekaltas@marmara.edu.tr,
Orcid: 0000-0001-5103-9018

öğrenme algoritması olan Rastgele Orman Yöntemi kullanılmıştır. Temelinde karar ağaçları olan bu yöntem diğer sınıflandırma algoritmalarına göre daha iyi sonuçlar vermiştir.

Anahtar Kelimeler: E-ticarette Tahmin Analizi, Kitle Analizi, Rastgele Orman

JEL Sınıflandırması: L81, C45, M3

Abstract

The widespread use of internet networks and the need to access the internet has become widespread in the use of internet sites and other digital platforms. This process, which is called digital advertising, has become an indispensable advertising tool for companies, brands and other organizations to reach their goals and to realize them in accordance with advertising purposes. Digital advertising, the most important feature of which is measurable, gives companies a very large data (statistics). Firms use this data to evaluate the digital advertising and have a look to the future advertising plans. The purpose of this study is to make a classification by using user data from a construction company's digital advertising campaign. A classifier is created by analyzing the data that the users are kept in the sales office and the user data obtained with the subsequent ads can be classified using this classifier. Thus, a preliminary information can be obtained about whether the users come to the sales office. Through this preliminary information, the company will determine its targets in sales and marketing more accurately. Through the purpose of the study, as dependent variable, whether the users came to the sales office, as the independent variable, the user sent which days the contact information to the employees of the company, the gender of the user, the site in which the advertisement is seen, which advertisement area (native, 300*250 visual size etc.), where ads were seen by the user (computer or telephone), whether the user had previously been registered in the relevant company and for which purpose he filled in this form (investment, host, etc.) a total of 7 independent variables were used belirlenmişti. R program was used and the Random Forests Method, a community learning algorithm, was employed to analyze the data. This method, which is based on decision trees, yields better results than other classification algorithms.

Keywords: Predictive Analysis in E-commerce, Mass (Audience) Analysis, Random Forest

JEL Classification: L81, C45, M3

I.Giriş

Son dönemlerde, internet ağlarının yaygınlaşması ve internete erişimin bir ihtiyaç haline gelmesi ile birlikte internet sitelerinde ve diğer dijital platformlardaki reklamların kullanılması yaygınlaşmıştır. Dijital reklamcılık olarak adlandırılan bu süreç, internet teknolojilerinin geldiği son noktada, insanların bilgiye en hızlı ulaştığı, ilk aramayı yaptığı internete bağlı cihazlar üzerinde, markaların ya da ürünlerin tanıtımının yapılmasıdır. Teknolojik gelişmeler ve internete olan erişimin artması sonucunda dijital reklamcılık, reklam sektörü için yeni ve gerekli bir hale gelmiştir(Barfoot, Burtenshaw ve Mahon 2006; Akt: Yürekli,2016). Dijital reklamcılık, günümüzde çevrimiçi platformların (internet site ve mobil uygulamaları) ve içeriklerin en yüksek payını finanse eden baskın bir faaliyet haline gelmiştir(McStay,2010 Yürekli, 2016). Markalar ya da ürünler bu teknolojiyi kullanarak reklam kampanyalarında daha ulaşılabilir ve daha görünür olmaktadır. Günümüzde şirketler, web sitelerini ziyaret ettikten sonra tüketicilerin davranışlarındaki değişikliklere bağlı olarak yapılan araştırmaların sonucunda web sitelerine trafik

sağlamak için çeşitli reklam türleri kullanarak dijital reklamcılık yatırımlarını genişletmektedir. Çoğu şirket, son dönemlerde çevrimdışı reklamlara (televizyon, gazete, radyo vb.) ek olarak farklı türde çevrimiçi reklamlara yatırım yapmaktadır. Şirketler, müşterilerine dijital reklamlar sayesinde kişisel veya belirli bir hedef kitleye özgü reklamları gösterebilirler.

Bu çalışmada bir inşaat firmasının, belirli bir dönemde yapmış olduğu dijital reklam yayınlarından elde edilen veriler (kullanıcı verileri) analiz edilmiştir. İnşaat firmasının reklamlarını dijital platformlarda görmüş olan kullanıcılar bu reklamlar sayesinde firma sitesine gelmiş ve burada konut alımıyla ilgili olarak iletişim bilgilerini firmaya bir form aracılığıyla göndermişlerdir. Bu kullanıcıların (müşteriler) inşaat firmasının satış ofisine gelip gelmedikleri kayıt edilmiştir. Elde edilen bu bilgiler sayesinde kullanıcılar satış ofisine gelip gelmeme durumuna bağlı olarak veri madenciliğinde çokça kullanılan ve karar ağaçları için bir topluluk öğrenme algoritması olan Rastgele Orman (Random Forest) yöntemiyle sınıflandırılmıştır. Böylelikle elde edilen yeni bir kullanıcı bilgisi, geliştirilen model sayesinde bu kullanıcının satış ofisine gelip gelmemesi tahmin edilecektir. Böylelikle firma hangi kullanıcıların satış ofisine geleceğini önceden öngörerek her bir kullanıcıyı arayarak harcadığı zamanı en aza indirecek, hedef kitle analizini gerçekleştirecek ve elindeki kullanıcı verilerini daha etkin şekilde kullanmış olacaktır.

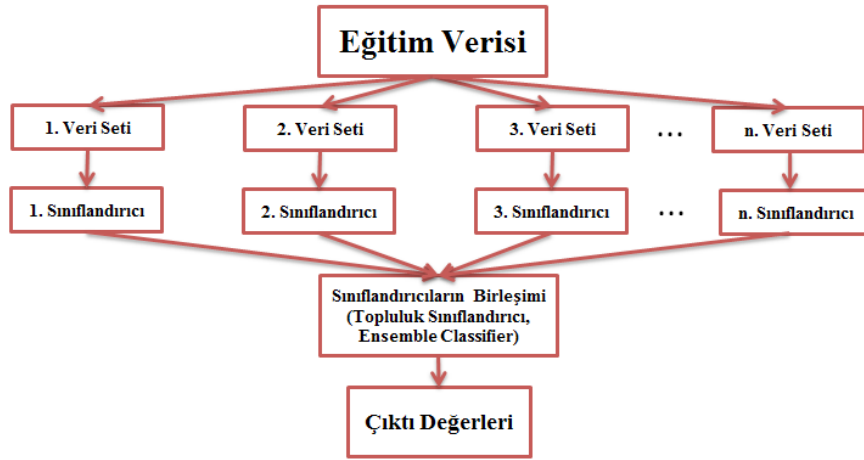
2. Yöntem

2.1. Karar Ormanları (Topluluk Öğrenme Algoritmaları)

Topluluk öğrenme yöntemlerinin ana fikri, her biri aynı problemi çözen, daha doğru ve güvenilir tahminler veya kararlarla tek bir model kullanarak elde edilenden daha iyi bir model oluşturmaktır. Çoklu modelleri birleştirerek tahmin modelini oluşturma fikri uzun süredir araştırılmaktadır. Topluluk öğrenme yöntemi, önemli bir karar vermeden önce çevresine danışan, farklı görüşler alan bir insanın nihai kararını verme sürecine benzer(Polikar, 2006).

Topluluk yöntemlerinin tahmin performansının iyileştirilmesinde kullanılabileceği bilinmektedir. İstatistik, makine öğrenimi, örüntü tanıma ve veri madenciliği gibi çeşitli disiplinlerden araştırmacılar, topluluk yönteminin kullanımını dikkate almışlardır.

Literatürde, “topluluk yöntemleri” terimi genellikle aynı temel modelin küçük değişikliklerine sahip model koleksiyonları olarak yer almaktadır. Ayrıca, literatürde “çoklu sınıflayıcı sistemler” olarak da bilinmektedir (Rokach ve Maimon, 2014).

Şekil 2.1. Karar Ormanları Bağımsız Sınıflandırıcı Çalışma Prensipleri

2.1.1. Torbalama (Bagging)

1996 yılında Breiman tarafından geliştirilen yöntem olan Bootstrap toplaması (bootstrap aggregating) veya Torbalama (bagging), istatistiksel öğrenme yönteminin hata varyansını azaltmak için amaçlanan genel bir yöntemdir. Torbalama tahmini, bir tahmincinin birden çok sürümünü oluşturur ve bunları toplu bir tahmin elde etmek için kullanan bir yöntemdir. Torbalama, sayısal bir sonuç tahmin ederken tahminlerin ortalamasını alır ve bir sınıf tahmin ederken oylamaya göre tahmin yapar (Breiman, 1996). Yöntem, öğrenilmiş sınıflandırıcıların çeşitli çıktılarını tek bir tahmin halinde birleştirerek, geliştirilmiş bir bileşik sınıflandırıcı oluşturarak doğruluğu arttırmayı amaçlamaktadır. Yeni bir örneği sınıflandırmada, her bir sınıflandırıcı bilinmeyen örnek için sınıf tahminini oluşturur. Sonuç olarak, torbalama, orijinal tekli verilerden oluşturulan tek modelden daha iyi performans gösteren birleşik bir model üretmektedir. Breiman 1996 yılındaki makalesinde, özellikle kararsız sınıflandırıcılar için bunun doğru olduğunu belirtmektedir. Çünkü torbalama kararsızlıkları ortadan kaldırabilir. Bu bağlamda, öğrenme setini bozmak, yapılandırılmış sınıflandırıcıda önemli değişikliklere neden olabiliyorsa, kararsız olarak kabul edilir. Bu yöntem karar ağaçları için yararlıdır ve sıklıkla kullanılır (Rokach ve Maimon, 2014).

Torbalama algoritmasını kısaca özetlemek gerekirse; Bir bootstrap örneği ile iadeli olarak eğitim setinden eşit oranda m tane örnek içeren örneklemeler üretilir. T bootstrap örneklemeleri B_1, B_2, \dots, B_T üretilir ve her bir bootstrap örneği B_i için bir C_i sınıflandırıcısı oluşturulur. Son sınıflandırıcı olan C^* , C_1, C_2, \dots, C_T sınıflandırıcılarının en çok tahmin ettiği sınıfı baz alarak elde edilen sınıflandırıcıdır (Bauer ve Kohavi, 1999).

2.1.2. Rastgele Orman (Random Forest)

Rastgele Orman (Random Forest) 2001 yılında Leo Breiman tarafından geliştirilmiştir. Rastgele orman, torbalama yöntemi ve Tim Kam Ho¹ tarafından önerilen rastgele alt uzay yöntemlerinin birleşiminden oluşmaktadır. Rastgele alt uzay yönteminde en uygun dallara bölünmeyi sağlayacak değişken tüm değişkenler arasından rastgele seçilmiş az sayıda değişken tarafından belirlenir (Uzbaş, 2017). Rastgele ormanlar, ağaç tipi sınıflandırıcılar topluluğudur, ormandaki her ağaç bağımsız olarak örneklenen bir rasgele vektörün değerlerine ve aynı dağılıma bağlıdır (Breiman, 1999). Torbalama yönteminin gelişmiş bir şekli olarak kabul edilebilir. Bu yöntemi Breiman makalesinde şöyle açıklamaktadır. k 'ncü ağaç için, geçmiş rastgele vektörlerden Q_1, \dots, Q_{k-1} bağımsız ancak aynı dağılımla sahip Q_k rasgele vektörü oluşturulur; ve bir ağaç, x 'in giriş vektörü olduğu $h(x, Q_k)$ sınıflandırıcısının sonucu olarak eğitim veri seti ve Q_k kullanılarak oluşturulur (Breiman, 2001). Örneğin, torbalamada, rasgele vektör Q , N eğitim veri setindeki gözlem sayısı olmak üzere, kutularda rastgele atılan N gözlemlerinden elde edilen N kutudaki sayımlar olarak oluşturulur. Rastgele bölünme seçiminde Q , 1 ile k arasında bir dizi bağımsız rastgele tam sayıdan oluşur. Q 'nun yapısı ve boyutsallığı, ağaç yapımındaki kullanımına bağlıdır (Breiman, 2001).

Rastgele orman yönteminde, torbalama ve rastgele değişken (özellik) seçimi birlikte kullanılır. Her yeni eğitim seti, orjinal eğitim setinden iadeli olarak (bootstrap yöntemiyle) çekilir. Ardından rastgele değişken (özellik) seçimi kullanılarak yeni eğitim setinde bir ağaç yetiştirilir. Yetiştirilen ağaçlarda budama yapılmaz (Breiman, 2001). Yapılan çalışmalar, özellik seçimi ölçütlerinin değil, budama yöntemlerinin seçiminin ağaç tabanlı sınıflandırıcıların performansını etkilediğini göstermektedir (Mingers, 1989). Torbalamanın kullanılmamasının iki nedeni vardır. Birincisi, rastgele değişken seçimi (özellikler) kullanıldığında torbalama kullanımının doğruluğu arttırdığı görülmektedir. İkincisi, torba dışı hataların (Out-of-bag (OOB)) hesaplanmasıdır (Breiman, 2001). Budamanın olmaması rastgele ormanları diğer karar ağacı yöntemlerinden daha avantajlı hale getirmektedir. Torbalama ve rastgele orman yöntemi arasındaki temel fark m boyutlu değişkenlerin seçimidir. Örneğin, toplamda p değişken varsa ve $m=p$ olarak oluşturulursa, o zaman yöntem torbalama'ya eşit olur. $m = \sqrt{p}$, yi kullanan rastgele ormanlar hem test hatasında hem de torbalama üzerindeki OOB hatasında bir azalışa yol açar (James vd. 2014).

Rasgele orman karar ağaçları için tanımlanmış olmasına rağmen, bu yaklaşım tüm sınıflandırıcılar için geçerlidir. Rastgele orman yönteminin önemli bir avantajı, çok sayıda girdi değişkenlerini ele almasıdır (Skurichina ve Duin, 2002). Rastgele ormanın bir başka önemli özelliği de hızlı olmasıdır.

1 Ayrıntılı bilgi için bkz. Ho (1998).

3.Verilerin Analizi

Uygulamada bir inşaat firmasının, belirli bir dönemde yapmış olduğu dijital reklam yayınlarından elde edilen veriler (kullanıcı verileri) R programının randomForest kütüphanesi kullanılarak analiz edilmiştir. İnşaat firmasının reklamlarını dijital platformlarda görmüş olan kullanıcılar bu reklamlar sayesinde firma sitesine gelmiş ve burada konut alımıyla ilgili olarak iletişim bilgilerini firmaya bir form aracılığıyla göndermişlerdir. Firma çalışanları kullanıcıların iletişim bilgileri sayesinde onlara ulaşmış ve konutlar hakkında daha ayrıntılı bilgi vermek için uygun bir zamanda satış ofislerine davet etmişlerdir. Tüm bu süreç yani kullanıcının reklamı görmesi, ilgili reklam sayesinde firma sitesine gelmesi, form doldurması ve satış ofisine gelmesi kayıt altına alınmıştır. Uygulama sayesinde bu veri seti analiz edilmiş ve kullanıcılar satış ofisine gelen ve gelmeyen olarak oluşturulan model tarafından tahmin edilmiştir. Amaç bundan sonraki süreçte oluşturulan model yardımıyla yeni kullanıcıların satış ofisine gelip gelmeyeceğinin tahmin edilmesidir. Böylelikle firma hangi kullanıcıların satış ofisine geleceğini önceden öngörerek her bir kullanıcıyı arayarak harcadığı zamanı en aza indirecek, hedef kitle analizini gerçekleştirecek ve elindeki kullanıcı verilerini daha etkin şekilde kullanmış olacaktır.

3.1.Uygulama Kapsamı ve Veri Yapısı

Uygulamada bağımlı değişken olarak kullanıcıların satış ofisine gelip gelmemesi, bağımsız değişken olarak ise kullanıcının iletişim bilgilerini hangi gün firma çalışanlarına gönderdiği, kullanıcının cinsiyeti, reklamı hangi internet sitesinde (mecra) görüp firma sitene geldiği, reklamı hangi reklam alanında (doğal, 300*250 görsel boyutu vb.) gördüğü, hangi cihazdan (bilgisayar veya mobil) gördüğü, kullanıcının daha önce ilgili firmada kayıtlı olup olmaması ve bu formu hangi amaçla doldurduğu (yatırım, ev sahibi olma vb.) olmak üzere toplamda 7 adet bağımsız değişken kullanılmıştır.

Uygulamada kullanılan Rastgele Orman yöntemiyle veri seti analiz edilmiş ve kullanıcılar satış ofisine gelen ve gelmeyen olarak model tarafından tahmin edilmiştir.

Bağımlı değişkene göre eğitim ve test verileri;

Tablo 3.1. Sınıf sayıları

Eğitim Veri Seti;	Satış Ofisi	0	1	Toplam
			397	44

Test Veri Seti;	Satış Ofisi	0	1	Toplam
			81	14

İlk başta 1.492 kullanıcı verisi elde edilmiş ancak veri temizleme aşamasında geriye 536 kullanıcı verisi kalmıştır. Bu temizleme işlemi kullanıcıların hatalı iletişim bilgisi vermesi, ulaşılacakları, formu yanlışlıkla doldurmaları ve konut alımıyla ilgilenmedikleri bilgisini vermeleri üzere analizden çıkarılmıştır. Analiz bu 536 kullanıcı verisiyle yapılmıştır. Bu kullanıcı verisinin %80'i eğitim, %20'si test verisi olarak kullanılmıştır. Bağımlı değişken olan satış ofisine gelme durumunu gösteren değişken, gelenler 1, gelmeyenler 0 sembolleriyle gösterilmiştir.

Uygulamada kullanılan bağımsız değişkenlerin hepsi kategoriktir.

Gün değişkeni; 1'den 7'ye kadar olan rakamlarla gösterilmiştir. 1 pazartesi gününe karşılık gelmekte olup sırayla devam ederek 7 Pazar gününe karşılık gelmektedir. Kullanılan R programında gun diye kodlanmıştır.

Cinsiyet değişkeni; 0 ve 1 rakamlarıyla gösterilmiştir. 0 erkeklere 1 kadınlara karşılık gelmektedir. Kullanılan R programında cins diye kodlanmıştır.

Cihaz değişkeni; 0 ve 1 rakamlarıyla gösterilmiştir. 0 mobil(cep telefonu) 1 bilgisayara karşılık gelmektedir. Kullanılan R programında cih diye kodlanmıştır.

Mecra değişkeni; 1'den 23'e kadar olan rakamlarla gösterilmiştir. 1. mecra, 2. mecra olarak 23'e kadar numaralandırılmıştır. Kullanılan R programında mec diye kodlanmıştır.

Kreatif değişkeni; 1'den 21'e kadar olan rakamlarla gösterilmiştir. 1. kreatif, 2. kreatif olarak 21'e kadar numaralandırılmıştır. Kullanılan R programında kre diye kodlanmıştır.

Kayıt değişkeni; 0 ve 1 rakamlarıyla gösterilmiştir. 0 kayıtlı olmayan kullanıcıya, 1 kayıtlı kullanıcıya karşılık gelmektedir. Kullanılan R programında kay diye kodlanmıştır.

Görüş değişkeni; 1'den 6'ya kadar olan rakamlarla gösterilmiştir. Kullanılan R programında grs diye kodlanmıştır.

1; 1+1 konutlarla ilgilenen kullanıcı

2; 2+1 konutlarla ilgilenen kullanıcı

3; 3+1 konutlarla ilgilenen kullanıcı

4; detaylı bilgi almak isteyen kullanıcı

5; yatırım yapmak isteyen kullanıcı

6; diğerler amaçlar için, olarak ifade edilmiştir.

Bağımsız değişkenlerin 0 ve 1 arasında olması ve değişken normlarının eşit olması için gün, mecra, kreatif ve görüş değişkenlerine aşağıdaki denklemde gösterildiği gibi normalizasyon dönüşümü yapılmıştır(Han, Kamber ve Pei, 2012).

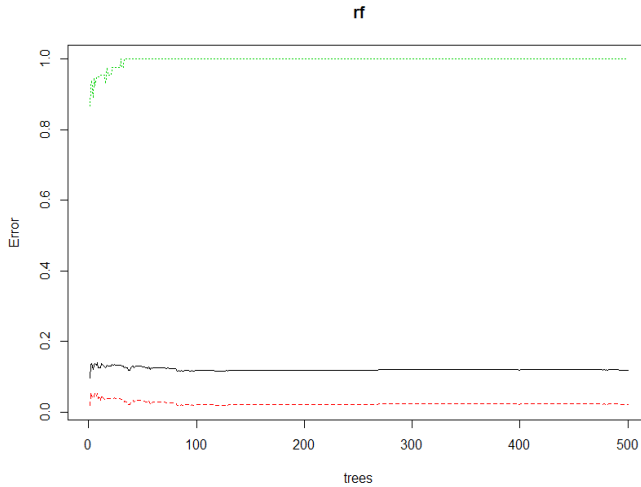
$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (3.1)$$

3.2.Elde Edilen Bulgular

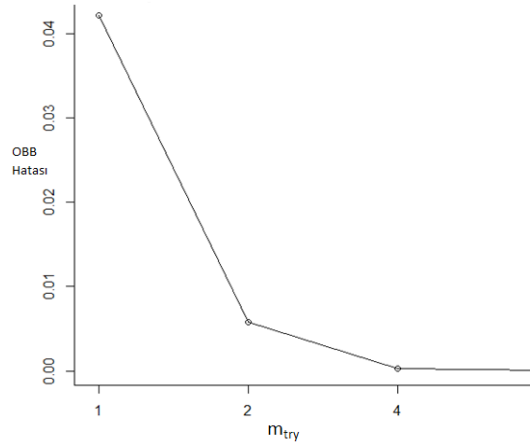
Uygulama analiz sonuçları, R (R Core Team,2017) programında randomForest(Liaw ve Wiener,2002) paketi kullanılarak elde edilmiştir.

Öncelikle ağaç sayısı ve ağaç oluşumda her bölünmede dikkate alınan değişkenlerin sayısı keyfi olarak belirlenir ve yapılan analiz sonucunda bu değerler için optimum değerler bulunarak bu parametreler yeniden belirlenir. Keyfi olarak ormanda oluşan ağaç sayısının 500 seçildiği ve ağaç oluşumda her bölünmede dikkate alınan değişkenlerin yine keyfi olarak 2 olarak seçildiği yöntemin çıktıları şekil 3.1 ve şekil 3.2'de gösterilmiştir.

Şekil 3.1. Ağaç sayısı grafiği



Grafik OOB hatasının ağaç sayısına bağlı olarak grafiğini çizmektedir. Ağaç sayısının artıkça bu hatanın sabitleştiğini görülmektedir. Yeni oluşturulacak parametre değerinde ağaç sayısı 200 olarak alınabilir ya da 500 olarak kalabilir. Ağaç sayısının fazla olması OOB'de bir azalışa yol açmasa da ormanın daha kararlı olmasını sağlar (James ve diğerleri, 2014).

Şekil 3.2. Ağaç oluşumda her bölünmede dikkate alınan değişken sayısı

OBB hatasını en aza indirecek her bölünmede dikkate alınan değişken sayısının grafiğine bakarak bu parametre değerinin 4 olması gerektiği söylenebilir.

Kullanıcı tabanlı parametrelerin optimal değerinin ne olacağı belirlendikten sonra analiz tekrardan yapılır ve aşağıdaki sonuçlar elde edilir.

Tablo 3.2. Eğitim verisi için sınıflandırma tablosu

Tahmin	Gerçek	
	0	1
0	395	17
1	2	27

Tablo 3.3. Eğitim verisi için sınıflandırma tablosuna ait değerlendirme ölçütleri

Doğruluk=	95,69%
Yanlış Sınıflandırma Oranı=	4,31%
Kesinlik=	93,10%
Doğru Pozitif Oran=	61,36%
Yanlış Pozitif Oran=	4,55%
Doğru Negatif Oran=	99,50%
Yanlış Negatif Oran=	38,64%
Kappa=	71,73%

Kappa istatistiği; Cohen'in kappa katsayısı olarak bilinir ve iki değerleyici arasındaki karşılaştırmalı uyuşmanın güvenilirliğini ölçen bir istatistik yöntemidir. Eğer $Pr(a)$ iki değerleyici için örtüşen uyuşmaların toplama olan oranı ve $Pr(e)$ ise bu uyuşmanın şans eseri ortaya çıkma olasılığı ise, Cohen'in kappa katsayısı bulunması için kullanılacak formül şu olur: Aşağıdaki formül yardımıyla hesaplanır. Kappa değerinin 1'e yakın olması iki değerleyicinin o derece iyi bir şekilde uyuşmakta olduğunu göstermektedir (Cohen,1960).

$$\kappa = \frac{\Pr(\alpha) - \Pr(e)}{1 - \Pr(e)} \quad (3.2)$$

Tablo 3.3'de görüldüğü gibi eğitim verisinde doğruluk %95 gibi çok yüksek bir değer çıkmıştır. Doğru pozitif oranın %61,36 olduğu görülmektedir. Bu değerler karar ormanının eğitim verisini iyi anlamda öğrendiğini göstermektedir. Kappa istatistiğinin %70 olması da modelin yüksek oranda gerçek değerlerle uyuştuğunun bir göstergesidir.

Tablo 3.4. Test verisi için sınıflandırma tablosu

Tahmin	Gerçek	
	0	1
0	77	13
1	4	1

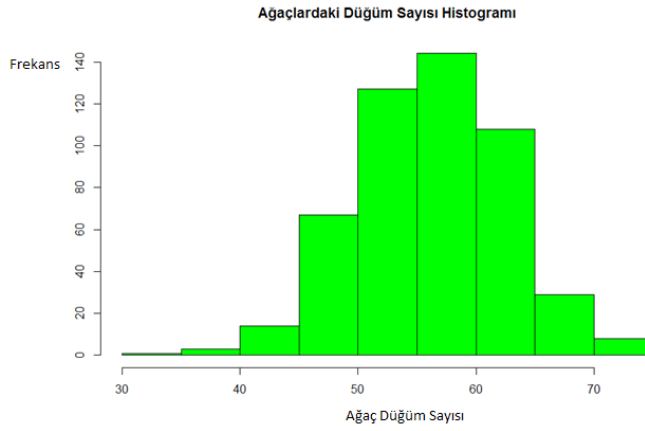
Tablo 3.5. Eğitim verisi için sınıflandırma tablosuna ait değerlendirme ölçütleri

Doğruluk=	82,11%
Yanlış Sınıflandırma Oranı=	17,89%
Kesinlik=	20,00%
Doğru Pozitif Oran=	7,14%
Yanlış Pozitif Oran=	28,57%
Doğru Negatif Oran=	95,06%
Yanlış Negatif Oran=	92,86%
Kappa=	0,05%

Tablo 3.5'de görüldüğü gibi test verisinin doğruluğunun yüksek %82 gibi yüksek bir değer olmasına rağmen doğru pozitif oranının %7 olması veri yapısından kaynaklanmaktadır. Bu durumu Brieman ve arkadaşları bir makalesinde şöyle açıklamaktadır. Random forest sınıflandırıcısı, çoğunluk sınıfını azınlık sınıfına göre daha doğru sınıflandırma eğilimi gösterir. Bu nedenle

azınlık sınıfını iyi bir şekilde sınıflandıramaz. Bu durumda ağırlıklandırılmış Rf (Random Forest) algoritması kullanılır. Ancak yapılan çalışmalar dengesiz veriyi öğrenmek için kullanılan ağırlıklı Rf ile dengeli Rf (uygulamada kullanılan) arasında net bir kazananın olmadığını ortaya çıkarmıştır (Chen vd., 2003).

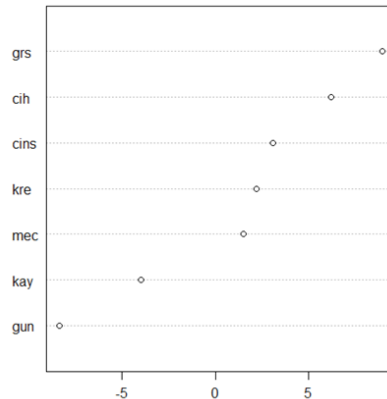
Şekil 3.3. Ağaçlardaki düğüm sayısı histogramı



Toplamda 501 tane ağaçtan oluşan ormanda düğüm sayılarının histogramı yukarıda verilmiştir. Ağaçlardaki düğüm sayısı yukarıdaki grafikten görüldüğü gibi ortalama olarak 50-60 arasında değişmektedir.

Ortalama Düşüş Doğruluğu

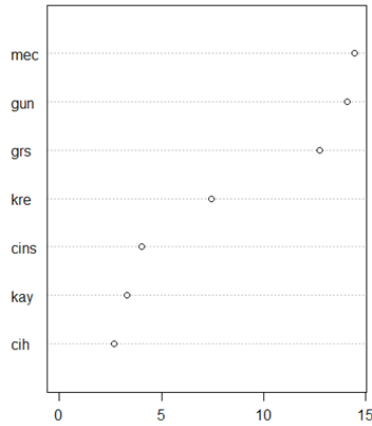
Şekil 3.4. Değişkenlerin doğruluğa olan katkıları



Şekil herhangi bir değişken olmadığında model performansının ne kadar kötü olduğunu test eder. Ağaçları yaparken herhangi bir değişkenini kaldırırsak, doğrulukta ortalama azalma ne olur sorusunun cevabını veren grafikdir. Örneğin grs (görüş) değişkeninin değeri çok yüksek olduğundan doğruluğun hesaplanmasında yüksek bir öneme sahip olduğu söylenir.

Gini Değeri Ortalama Düşüşü

Şekil 3.5. Dğümlerin değişkenlere göre saflık değerleri



Bu grafik, herhangi bir değişken olmadan ağacın sonunda düğümlerin ne kadar saflıkta olduğunu ölçer. Yani değişkenlerden biri modelden çıktığında Gini katsayısında ne kadar bir azalma olacağını cevabını verir. Örneğin mec (mecra) değişkeni değeri yüksek olduğu için diğer değişkenlere göre Gini indeksinde yüksek katkısı olduğu söylenir.

Tablo 3.6. Değişkenlerin doğruluk ve gini katkıları için sayısal değerler

Değişkenler	Ortalama Doğruluk	Gini
gun (gün)	-8,36	14,09
cins (cinsiyet)	3,11	4,03
cih (cihaz)	6,22	2,67
mec (mecra)	1,49	14,46
kre (kreatif)	2,2	7,43
grs (görüşme)	8,97	12,75
kay (kayıt)	-4	3,31

Bu tablo yukarıdaki verilen grafiklerde yer alan değişkenlerin sayısal değerlerini göstermektedir.

Tablo 3.7. Değişken kullanım sayıları

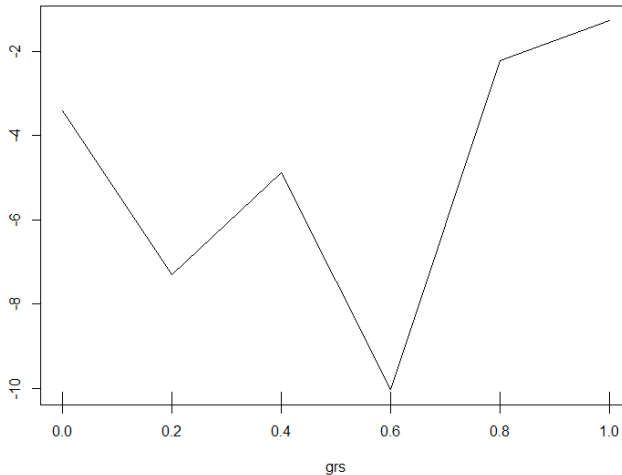
Değişkenler	Kullanım Sayıları
gun (gün)	7.760
cins (cinsiyet)	2.215
cih (cihaz)	873
mec (mecra)	7.119
kre (kreatif)	4.120
grs (görüşme)	3.534
kay (kayıt)	2.275

Bu tablo, değişkenlerin ağaç oluşumdaki kullanım sayılarını göstermektedir. Örneğin en fazla kullanılan değişken gun(gün) değişkenidir. Bu değişkenin ortalama doğruluğa katkısı - 8,36 iken gini indeksine katkısı ise 14,09 olarak hesaplanmıştır.

Kısmi Bağımlılık grafikleri;

Bu grafikler bir değişkenin, sınıf olasılığı üzerindeki marjinal etkisini vermektedir.

Şekil 3.6. Görüş değişkeninin kısmi bağımlılığı



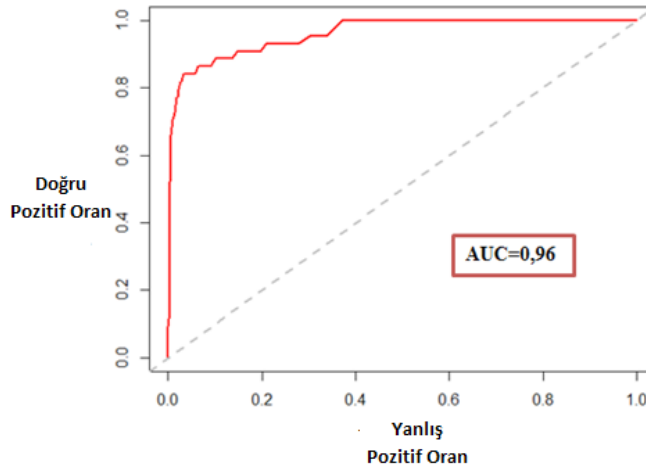
Şekil 3.6 yorumlandığında grs (görüş) değişkeni 0,6 değerinden büyük olduğunda 1 sınıfını 0,6 değerinden küçük olmasına göre daha kuvvetli tahmin etmektedir.

Diğer değişkenlerin kısmi bağımlılıkları da o değişkenlere göre çizilen grafiklere bakılarak yapılabilir. Bu grafikler ekler kısmında verilmiştir.

ROC Eğrileri

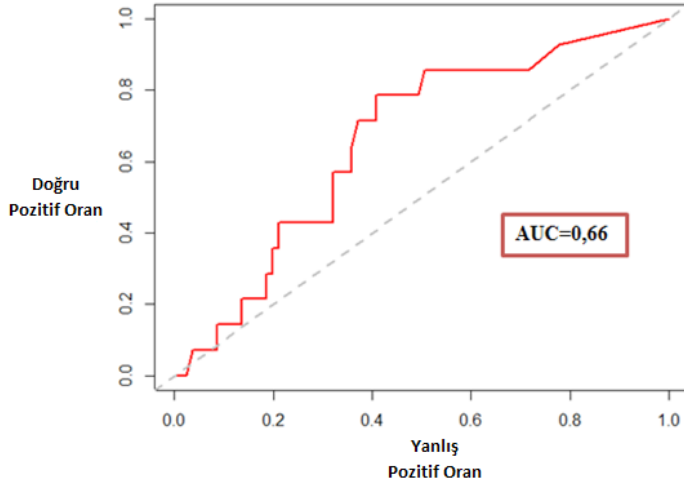
Kullanılan random forest algoritması gözlemleri aldığı oylara göre sınıf atamaları yapmaktadır. Bu oylar her bir gözlem için hesaplanır ve OOB örnek ağacı sınıflandırması için de hesaplanır. Bu oylar kısaca bir olasılığı temsil eder ve bu nedenle ROC grafiği çizilip AUC değeri hesaplanabilir.²

Şekil 3.7. Eğitim verisi için ROC eğrisi



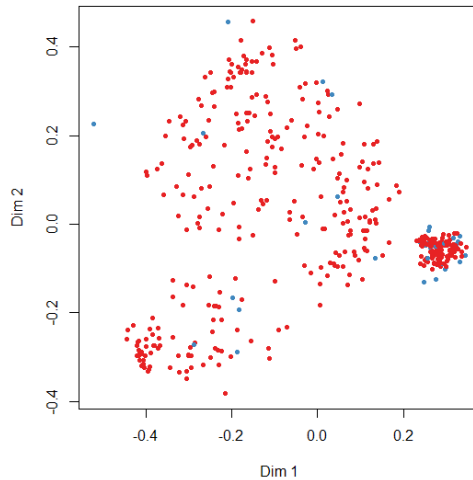
R programındaki ROCR paketi kullanılarak çizilen ROC eğrisi Şekil 3.7'de gösterilmektedir. Yine aynı paket kullanılarak hesaplanan AUC (area under curve) değeri **0,96** olarak hesaplanmıştır. Bu değer sınıflandırma performansının çok iyi olduğunu göstermektedir.

2 www.stats.stackexchange.com/questions/188616/how-can-we-calculate-roc-auc-for-classification-algorithm-such-as-random-forest,2017 Erişim tarihi:12.11.2018

Şekil 3.8. Test verisi için ROC eğrisi

Test verisi için hesaplanan ROC eğrisi Şekil 3.8'de gösterilmiştir. Bu eğri altında kalan AUC değeri de 0,66 olarak bulunmuştur. Bu değer 0,5'nin üzerinde olduğu için sınıflandırma performansının iyi olduğu söylenebilir.

Çok Boyutlu Ölçekleme Grafiği

Şekil 3.9. Eğitim verisindeki geliş değişkenin çok boyutlu ölçekleme grafiği

Veride 0 sınıfına ait olan gözlemler çoğunlukta olduğu için kırmızı renkli olanlar 0 sınıfını, mavi renkli olanlar ise 1 sınıfını göstermektedir.

4.Sonuç

Çalışmada bir inşaat firmasının dijital reklam yayınları sonucunda elde edilen kullanıcı verileri analiz edilmiştir. Veri analizinde Rastgele Orman Yöntemi kullanılarak sınıflandırma modeli oluşturulmuştur. Kullanıcıların satış ofisine gelip gelmemesinin bağımlı değişken olduğu sınıflandırma modelinde, kullanıcının iletişim bilgilerini hangi gün firma çalışanlarına gönderdiği, kullanıcının cinsiyeti, reklamı hangi internet sitesinde görüp firma sitesine geldiği, reklamı hangi reklam alanında gördüğü, hangi cihazdan (bilgisayar veya telefon) gördüğü, kullanıcının daha önce ilgili firmada kayıtlı olup olmaması ve bu formu hangi amaçla doldurduğu (yatırım, ev sahibi olma vb.) olmak üzere toplamda 7 adet bağımsız değişken kullanılmıştır.

Bağımsız değişkenlerin normlarını eşitlemek için tüm bağımsız değişkenleri 0-1 aralığına dönüştüren normalizasyon dönüşümü yapılmıştır. Oluşturulan sınıflandırma modelinde eğitim verisinde sınıflandırma doğruluğu %95 seviyesinde yüksek bir oranda olurken test verisinde bu oran %82 seviyelerinde gerçekleşmiştir. Bu tarz çalışmalarda test verisi dikkate alınarak değerlendirme yapılmaktadır. Test verisi içinde doğruluk iyi bir seviyede gerçekleşmiştir.

Değişken bazında en fazla görüş değişkeni olmak üzere, cihaz değişkeninin de doğruluğu artırıcı bir etkisi olduğu ortaya çıkmıştır. Yani kullanıcıların (müşterilerin) satış ofisine gelmelerindeki en önemli değişken (tahmin doğruluğunu arttıran değişken) görüş değişkeni ve cihaz değişkeni olduğu görülmektedir. Bu değişkenler tahmin doğruluğuna en fazla katkıyı veren değişkenlerdir. Bu sonuç bundan sonraki form doldurma sayfalarında kullanıcıların konut alma amaçlarını yansıtacak bir seçeneğin olması gerektiğinin ön bilgisini vermektedir. Mecra ve gün değişkenlerinin gini indeksine en fazla katkıyı sağlayan değişkenler olduğu görülmüştür. Mecra ve gün değişkenleri ormandaki karar ağaçlarında en fazla kullanılan değişkenler olmuşlardır. Ayrıca oluşturulan karar ağaçları ortalama 50-60 düğümde oluşmaktadır.

Eğitim verisinde, 0 etiketine sahip olan sınıfın doğru sınıflandırılmasını gösteren doğru negatif oran (duyarlılık) değeri %99 seviyelerinde, 1 etiketine sahip olan sınıfın doğru sınıflandırılmasını gösteren doğru pozitif oran (özgüllük) değeri %61 gibi iyi bir seviyede gerçekleşmiştir. Test verisinde doğru negatif oran (duyarlılık) değeri %95 seviyesinde ve doğru pozitif oran (özgüllük) %7 seviyelerinde olmuştur. Test verisinde özgüllük değerini küçük ve duyarlılık değerinin yüksek çıkma nedeni azınlık sınıf varlığından kaynaklanmaktadır. Yani 0 olarak etiketlenen sınıftaki gözlem sayısı, 1 olarak etiketlenen sınıftaki gözlem sayısından çok fazladır. Bu tip verilerde çoğunluk sınıfı iyi bir seviyede tahmin edilirken azınlık sınıfın tahmin doğruluğu düşük seviyelerde gerçekleşmektedir.

Kullanılan Random Forest algoritmasında gözlemlerin aldıkları oylara göre hangi sınıfta oldukları belirlenir. Bu oylar bir olasılığı temsil etmektedir. Böylece ROC grafiği çizilip AUC değeri

hesaplanabilir. Çizilen ROC eğrilerinde eğitim verisi için eğri sol üst köşeye yakın olup eğri altında kalan alan ise 0.96'dır. Eğri altında kalan alanın 1'e yakın olması yapılan sınıflandırmanın iyi seviyelerde olduğunu göstermektedir. Test verisi için çizilen ROC eğrisi sol üst köşeye fazla yakın olmamasına rağmen eğri altında kalan alan 0.66 seviyelerinde olmuştur. Buradan da eğri altında kalan alan 0.5'in üzerinde olduğu için yapılan sınıflandırmanın rastlantısal olmadığı sonucuna varılır. Hem eğitim verisi hem de test verisi değerlerine bakarak yapılan sınıflandırmanın iyi bir seviyede olduğu söylenebilir.

Oluşturulan sınıflandırma modeli ile veri tabanına gelen yeni bir kullanıcı bu modele göre belirli hata oranlarıyla sınıflandırılabilir. Böylece yeni kullanıcılar hakkında önceden bir ön bilgiye sahip olunur. Bu ön bilgi sayesinde firma her bir kullanıcıyı arayarak harcadığı zaman kaybını minimize edebilir ve müşteri portföyünü bu bilgiler ışığında değerlendirme imkânına kavuşur. Değişken bazında en önemli değişkenlerin mecra, gün ve görüş değişkenleri olduğu görülmüştür. Bu önemli değişkenlerin kısmi grafiklerine bakılarak yeni hedef kitleler belirlenebilir ve bu hedef kitlelere göre dijital reklam stratejileri oluşturulabilir. Görüş değişkeninin önemli bir değişken olarak bulunması sonucunda bundan sonraki form doldurma sayfalarında kullanıcıların konut alma amaçlarını yansıtacak bir seçeneğin olmasını gerektiğinin ön bilgisini de vermektedir. Bir başka kullanım amacı ise, şu anki mevcut durumda satış ofisine gelmeyen ancak model tarafından satış ofisine geldi olarak tahmin edilen kullanıcılar da belirlenip bu kullanıcılara özel hedeflemeler ve satış ofisine gelmelerini sağlayacak farklı pazarlama stratejileri de geliştirilebilir.

Referanslar

- Bauer, E. ve Kohavi, R. (1999). An Empirical Comparison of Voting Classification Algorithms : Bagging , Boosting , and Variants. *Machine Learning*, 36: 105–139.
- Breiman, L. (1996). Bagging Predictors. *Machine Learning*, 24: 123–140.
- Breiman, L. (1999). Random Forests – Random Features, Tecnic Report 567, Statistic Department, University of California, Berkeley, 1–29. (<https://www.stat.berkeley.edu/~breiman/random-forests.pdf>, erişim tarihi:08.10.2018).
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1): 5–32. <https://doi.org/10.1023/A:101.093.3404324>
- Chen, C., Liaw, A. ve Breiman, L. (1999). Using Random Forest to Learn Imbalanced Data, 1–12.
- Cohen, J. (1960). A coefficient of agreement for nominal scales, *Educational and Psychological Measurement* 20(1): 37-46
- Han, J., Kamber, M. ve Pei, J. (2012). *Data Mining Concepts and Techniques Third Edition*. Simon Fraser University: Morgan Kaufman Publishers.
- Ho, T. K. (1998). The Random Subspace Method for Constructing Decision Forests, *IEEE Trans. Pattern Anal. Mach. Intell.* 20: 832-844.
- James, G., Witten, D., Hastie, T. ve Tibshirani, R. (2014). *An Introduction to Statistical Learning with Applications in R*. Performance Evaluation, 64. <https://doi.org/10.1016/j.peva.2007.06.006>
- Liaw, A. ve Wiener, M. (2002). Classification and Regression by randomForest. *R News* 2(3),18-22.
- Mingers, J. (1989). An empirical comparison of selection measures for decision tree induction. *Machine Learning*, 3: 319–342. <https://doi.org/10.1007/BF00116837>

- Polikar, R. (2006). Ensemble based systems in decision making. IEEE Circuits and Systems Magazine, 6(3): 21–44. <https://doi.org/10.1109/MCAS.2006.168.8199>.
- R Core Team (2017). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL <http://www.R-project.org/>.
- Rokach, L. ve Maimon, O. (2014). Data Mining With Decision Tree Theory and Applications, 2.Edition, World Scientific Publishing.
- Skurichina, M. ve Duin, R. P. W. (2002). Bagging, boosting and the random subspace method for linear classifiers. Pattern Analysis and Applications, 5(2), 121–135. <https://doi.org/10.1007/s100440.200011>.
- Uzbaş, B.(2017), Sayısal Dental Modellerden Otomatik Cinsiyet Tespiti, Konya (Doktora Tezi).
- Yürekli,K..(2016). Dijital Reklamcılıkta Reklam Ajansı – Reklam Veren İlişkisinin Analizi, İstanbul, (Yüksek Lisans Tezi)

SİGARAYI BIRAKMA VE OBEZİTE ARASINDAKİ İLİŐKİ: TÜRKİYE ÖRNEĐİ

GIVING UP SMOKING AND OBESITY: THE CASE OF TURKEY

Deniz KARAOĐLAN*

Öz

Obezitenin başlıca sebepleri arasında hareketsizlik, masa başı çalışma, yüksek seviyede hazır ve kalorili yemek tüketimi olarak gösterilmekle beraber literatürde deđişik savlar da ortaya atılmaktadır. Sigara tüketimindeki düşüőle beraber obezitenin yaygınlığının artması, literatürde birçok kez ortaya konulan bir savdır. 2014 Türkiye Sağlık Arařtırması (TSA) veri seti kullanılarak, çift deđişkenli probit modelleri uygulanarak yapılan bu çalışmanın sonucunda kişinin eski sigara tüketicisi ve obez olma olasılığı arasında anlamlı pozitif bir ilişki bulunmuştur. Bu bağlamda sigara bırakmayı destekleyici politikalar ile birlikte eş zamanlı olarak obeziteyi azaltılacak politikaların uygulanması önerilmektedir.

Anahtar Kelimeler: Obezite, Sigarayı Bırakma, Türkiye, Çift Deđişkenli Probit Modelleri

JEL Sınıflaması: I12, I18, C21

Abstract

Although the main reasons of obesity are considered to be sedentary life-style and eating high-calorie, prepared convenience food, other reasons are discussed in the literature. One of the reason is the decrease in the smoking rates leads to rise in the obesity rates. In this paper, we test the relationship between obesity and giving up smoking by applying bivariate probit analysis. We find that giving up smoking and obesity are positively and significantly correlated with each other. Therefore, we propose health policies which aim help to reduce both smoking and obesity simultaneously.

Keywords: Obesity, Giving up Smoking, Turkey, Bivariate Probit Models

JEL Classification: I12, I18, C21

* Bahçeşehir Üniversitesi, Ekonomi Bölümü. E-posta: hanifedeniz.karaoglan @eas.bau.edu.tr,
Orcid: 0000-0002-8000-8613

I. Giriş

Obezite tüm dünyada ve Türkiye’de giderek yaygınlaşmakta olan bir sağlık sorunudur. Obezite kardiyovasküler hastalıkların, yüksek şekerin ve buna bağlı sorunların en önemli sebebi olarak kabul edilmektedir (Stewart v.d, 2009).Ekonomik Kalkınma ve İş Birliği örgütünün 2016 verilerine göre Türkiye’deki yetişkin kadınların yüzde 24.5’i ve erkeklerin yüzde 15.3’ü obez olarak kabul edilmektedir. Dünya Bankası’nın 2008 Türkiye raporuna göre Türkiye’de obezitenin ve aşırı kiloluluk oranının artmasındaki başlıca sebepleri arasında çok şekerli, yağlı ve kalorili besinlerin aşırı tüketimi, hareketsizlik, masa başında çok vakit geçirme gösterilmektedir.

Dünyadaki obezite oranlarının artışına sebep olarak da benzer unsurlar öne sürülmüştür. Ancak obezitenin en sık görüldüğü yer olan Amerika Birleşik Devletleri (ABD) için yapılan çalışmalarda obezitenin artışına sebep olarak sigara tüketimindeki düşüş de gösterilmektedir. Sigara tüketiminin kalori yakmayı ve kilo vermeyi hızlandırdığı düşünülmektedir, ayrıca sigarayı bırakmanın kişileri daha fazla yemeye sevk ettikleri ve dolayısıyla kilo aldırıldığı bazı yazarlarca üzerinde sıklıkla tartışılan bir konudur (Chou vd. (2004), Rashad ve Grossman (2004), Rashad (2006)). Sigara tüketimi ve obezite arasındaki ilişki daha önce bir kaç defa ABD için incelenirken (Gruber ve Frakes (2005), ve Rashad (2006)), bildiğimiz kadarıyla gelişmekte olan ülkeler için daha önce hiç bakılmamıştır.

Bu çalışma gelişmekte olan bir ülke, Türkiye için bu alanda yapılan ilk çalışma olacaktır. Türkiye Sağlık Araştırması 2008, 2010 ve 2012 veri setleri ile yapılan bir başka çalışmada hem amprik hem de betimleyici analizler genel olarak kişilerin Vücut Kitle İndeksi (VKİ) seviyelerinin yıllar itibariyle arttığını göstermektedir (Karaoğlu ve Tansel, 2018). Türkiye’de son yıllarda sigara kullanımını sınırlandırıcı birçok atılım yapılmıştır. Örneğin, “Kamu Spotu” adı altında sigaranın zararlarını anlatan reklamlar televizyonlarda sık sık verilmektedir. 2010 yılı itibari ise de açık alanlarda sigarayı kullanma yasağı getirilmiştir. Ayrıca, Türkiye’de sigara ve alkol üzerine konulan vergiler de çok yüksektir. Ancak aynı anda Türkiye’de obezitenin görünürlüğünün yıllar itibariyle artması istatistiki sonuçlarla kanıtlanan bir gerçektir. Bu çalışmanın temel amacı sigarayı bırakma ve obezite arasındaki ilişkiyi çift değişkenli probit (bivariate probit) modelleri kullanarak incelemektir. Ekonometrik analizlerde kullanılan diğer açıklayıcı değişkenler (kişinin yaşı, cinsiyeti, medeni durumu, eğitim seviyesi, hanehalkı gelir durumu, istihdam durumu) ve bağımlı değişkenler arasındaki ilişki çalışmanın yan bulgularıdır.

Analizlerde Türkiye İstatistik Kurumu tarafından hazırlanan 2014 Türkiye Sağlık Araştırması (TSA) veri seti kullanılmıştır. TSA veri seti kişinin demografik özellikleri, sağlık durumu ve sağlık davranışları ile ilgili sorular içermesi bakımından zengin bir veri setidir. Analizler 25 yaş üstü yetişkin kişiler için uygulanmıştır. Uygulanan bu kısıtın sebebi kişinin sağlık davranışlarını açıklayan önemli değişkenlerden biri olan kişinin eğitim seviyesinin Türkiye’de ortalama 25 yaşından sonra netlik kazanmasıdır. (Tansel ve Karaoğlu, 2014). Çift değişkenli probit modellerinin sonucunda sigarayı bırakma ve obezite arasında anlamlı bir pozitif ilişki bulunmuştur. Bu sonuç, Chou vd. (2004), Rashad ve Grossman (2004) ile Rashad (2006)’ın

sigara içmenin kalori yakımını hızlandırması veya sigarayı bırakan kişilerin yemeğe daha fazla yönelmesi hipotezlerini destekler niteliktedir. Bu bağlamda, sonuçları en az sigara içmek kadar ağır olan obezitenin önlenebilirliği açısından sigarayı bıraktırma ve obezitenin tehlikelerini vurgulayan kampanyaların eş zamanlı olarak uygulanması önem taşımaktadır.

Bu çalışmanın, Türkiye için, kişinin sağlık durumunu etkileyen bu iki sağlık davranışının birbirleriyle olan ilişkisini inceleyen ilk çalışma olması dolayısıyla, sağlık ekonomisi literatürüne önemli bir katkı yapması düşünülmektedir. Tansel ve Karaoğlan (2014) TSA veri setinin daha önceki versiyonlarını kullanarak (2008, 2010 ve 2012), yetişkinlerde sağlık davranışlarını belirleyici etkenleri incelemişlerdir. Tansel ve Karaoğlan (2016) ise eğitim ve sağlık arasındaki nedensellik ilişkisine bakmıştır. Son olarak, Karaoğlan ve Tansel (2018) yetişkin kişilerin VKİ'lerini etkileyen demografik ve sosyoekonomik faktörleri kantil regresyon metodları kullanarak incelemiştir. Ancak, bu çalışmaların hiçbirinde iki farklı sağlık davranışı arasındaki ilişki incelenmemiştir. Sigarayı bırakmayı etkileyen sosyoekonomik ve demografik özellikler de bu çalışmaların hiçbirinin incelediği konulardan değildir. Dolayısıyla, bu çalışma yukarıdaki çalışmalardan bakış açısı, uygulanan metodlar ve asıl amacı bakımından çok farklıdır.

Makalenin ikinci bölümünde sigara ve obezite arasındaki ilişkiyi inceleyen teorik ve ampirik makalelerin yöntemleri ve bulguları ayrıntılı olarak ele alınmıştır. Üçüncü bölümde TSA veri seti hakkında bilgi verilmiş ve kullanılan bağı ve bağımsız değişkenler için yapılan frekans analizleri yorumlanmıştır. Çalışmanın dördüncü bölümünde kullanılan ekonometrik metodoloji ayrıntılı olarak ele alınmıştır. Beşinci bölümde analizlerden elde edilen bulgular açıklanmış ve sonuçlar yorumlanmıştır. Sonuç bölümünde ise bulguların kısa bir özeti verildikten sonra çalışmanın sınırlılıkları tartışılmıştır.

2.Literatür Taraması

Obezite sağlık ekonomisi literatüründe sigaradan sonra en fazla incelenen ikinci sağlık problemidir. Hatta son on yılda obezite üzerine yapılan çalışmalar sigara konusunda yapılan çalışmaları geride bırakmıştır. Bunun nedeni obezitenin ABD başta olmak üzere tüm dünyada insan sağlığını hızla tehdit eden bir problem haline gelmesidir. Stewart ve diğerleri (2009) ABD için 18 yaşındaki insanların yaşam beklentisinin ne olduğuna dair tahminlemeler yapmıştır. Bu tahminlemelerdeki amaç son 30 yılda sigara tüketimindeki düşüşün yaşam süresinin nasıl değiştirdiğini görmektir. Stewart ve arkadaşlarının yaptığı tahminlemeler ve projeksiyonlar sonucunda obezitenin hızla arttığı ve artacağını, insanların sigarayı bırakmalarının olumlu etkilerinin obezitenin olumsuz etkileri üzerine geçemediğini bulmuşlardır. Dolayısıyla, kişilerin yaşam süresinde bir artış olmayacağı sonucuna varmışlardır.

Chou ve diğerleri (2004) ABD için obezitedeki artış sebeplerini şöyle sıralandırmışlardır: Birincisi, kadınların iş gücüne girmesiyle evde daha az zaman geçirmeleri ve hazır yemek tüketimine yönelmeleri; ikincisi sigara fiyatlarındaki artış sonucunda sigara tüketiminin azalması ve kişilerin daha çok kalorili yemekler tüketmesi; üçüncüsü “fast-food” tarzı ucuz ve yüksek

kalorili yiyecekler satan yemek yerlerinin çoğalması ve kişilerin bunları çok tercih etmeleri. Chou ve arkadaşları bu hipotezlerini En Küçük Kareler (EKK) yöntemleri ile test etmişler ve obezitenin sigara tüketimiyle hiç bir ilişkisinin olmadığını fakat kadınların iş gücündeki artışın obezitedeki artışı doğrudan ve anlamlı olarak etkilediğini öne sürmüşlerdir. Ancak, bir başka çalışmada Cutler vd. (2003) Amerika ve gelişmiş bazı Avrupa ülkeleri için (İngiltere, Almanya, Fransa) ile Avusturalya ve Kanada için kadın iş gücündeki artış ve obezite arasında anlamlı bir ilişki olmadığını göstermişlerdir. Cutler ve arkadaşlarına göre obezitedeki artışta başlıca sebep teknolojiye gelişmelerle beraber mikrodalga, paket yemek, dondurulmuş gıda gibi ürünlerin yaygın bir şekilde zaman içinde kullanılmaya başlanmasıdır. Örneğin Almanya ve Fransa gibi yemeklerde geleneksel mutfaklarını kullanmaktan vazgeçmeyen ülkelerde obezite oranlarının ABD, Kanada ve Avusturalya'ya göre daha düşük olduğunu gözlemlemişlerdir. Benzer bir şekilde Phillopson ve Posner (2003) de obezitedeki artışı teknolojiye gelişmelere bağlamış ve insanların bir süre sonra yemeğe bağımlı hale geldiklerini vurgulamışlardır.

Rashad ve Grossman (2004), Chou ve arkadaşlarının aksine sigaranın kilo kontrolünde önemli bir faktör olduğunu söylemişler ve sigara tüketimindeki düşüşün obeziteyi arttırdığını savunmuşlardır. Ancak Rashad ve Grossman bu çalışmalarında amprik yöntemler kullanmamışlar, literatüre ve betimleyici istatistiklere dayanarak önermeler yapmışlardır. Phillopson ve Posner (2008) de gene betimleyici istatistiklere dayanarak sigara tüketimindeki düşüşün ve alkol tüketimindeki artışın obezite oranlarını arttırdığını savunmuşlardır.

Sigara tüketimi ve obezite arasındaki amprik ilişki Gruber ve Frakes (2005) tarafından ABD için incelenmiştir. Gruber ve Frakes yukarıda önerilen tüm hipotezlerin aksine sigara tüketimindeki düşüş ve obezite oranlarındaki artış arasında pozitif bir ilişki olmadığı, dahası negatif bir ilişki olduğu; yani kişilerin sigarayı bırakmaları ile birlikte VKİ'lerinin daha da normal seviyelere geldiğini bulmuşlardır. Rashad (2006) ise gene ABD için sigara tüketimi ve obeziteyi incelerken bu iki sağlık davranışı arasındaki içsellik problemini ele almış ve araç değişken (AD) metodları kullanarak sigara tüketimi ve obezite arasındaki ilişkiyi incelemiştir. Rashad AD metodu kullandığı zaman aradaki olası negatif ilişkinin giderek anlamını yitirdiğini görmüştür. Dolayısıyla genetik özelliklerin ve küçüklükten beri yerleşen sağlık ile ilgili davranışlarının kişinin aşırı kilolu veya obez olmasında daha etkili oldukları sonucuna varmıştır.

3.TSA Veri Seti ve Frekans Analizleri

Türkiye Sağlık Araştırması (TSA) Anketi Türkiye İstatistik Kurumu (TÜİK) tarafından 2008 yılından itibaren iki yılda bir uygulanmaktadır. Gidilen hanelerde kişilere sağlık durumları ve sağlığı belirleyici davranışlarının yanında yaş, medeni durum, eğitim seviyesi, gelir seviyesi ve istihdam durumları hakkında çeşitli sorular sorulmaktadır. Anket formunda 0-6 yaş, 7-14 yaş ve 15 yaş ve Üstü olmak üzere 3 farklı soru formu bulunmaktadır. 0-6 yaş ve 7-14 yaş arası çocukların soru formları ebeveynleri tarafından doldurulmaktadır. Bu çalışmada 25 yaş ve üstündeki kişilerin sigarayı bırakma ve aşırı kilolu/obez olma olasılıkları arasındaki ilişki

incelenmiştir. Çalışma TSA'nın en güncel veri seti olan 2014 veri seti kullanılarak yapılmıştır. Veri setinde 25 yaşüstü toplam 15,724 tane gözlem vardır. Bu gözlemlerin 8,597 tanesi kadın, 7,124 tanesi ise erkektir. Ancak analizlerde kullanabildiğimiz toplam gözlem sayısı 8,515'tir.

Çalışmada kullanılan birinci bağımlı değişken kişinin önceden sigara tüketicisi olması, fakat anketin yapıldığı dönemde sigarayı bırakmış olmasıdır. Bu bağılı değişken için, bir kukla değişken oluşturulmuştur. Kukla değişkeni yaratırken kullanılan soru aşağıdaki gibidir:

“Tütün mamülü kullanıyor musunuz? (Bir kez dahi olsa)”

Bu soruya verilen cevaplar sırasıyla şöyledir: *“Evet, her gün/evet, ara sıra/Hiçbir Zaman/Bıraktım.”*

Eğer soruya verilen cevap *“Bıraktım”* ise bu kişi için bağımlı kukla değişken 1 olarak hesaplanmış, *“Evet, her gün”* veya *“evet, ara sıra”* cevabını veren kişiler içinse 0 olarak tanımlanmıştır. Hiçbir zaman sigara kullanmadığını söyleyen kişiler ise örneklemden çıkarılmıştır. Çünkü bu çalışmanın amacı sigarayı bırakmanın obezite artışını desteklediği hipotezini test etmektir. Veri setinde yapılan bu düzenlemeden sonra toplamda 8,515 kişinin eski veya yeni sigara içicisi olduğu gözlenmiştir. Bu kişilerin % 61.31'i hala sigara içmeye devam ettiklerini söylerken, % 38.69'u sigarayı bıraktıklarını açıklamıştır. 2014 TSA veri setine göre kadınlarda ve erkeklerde sigarayı bırakma olasılığı hemen hemen aynıdır. Kadınlarda bu oran % 39.46 iken erkeklerde %38.29'dur. Yapılan olasılık testlerinde bu iki oranın anlamlı olarak farklı olduğu tespit edilmiştir. Yani kadınların sigarayı bırakma eğilimleri erkeklerden daha fazladır.

Çalışmanın ikinci bağılı değişkeni ise kişinin obez olup olmamasıdır. Bu bağılı değişken de eğer kişi obez veya aşırı kilolu ise 1, değilse 0 değerini alan bir kukla değişken olarak tanımlanmıştır. Bu kukla değişkeni tanımlanmadan önce kişinin Vücut Kitle Endeksi (VKİ) hesaplanmıştır. VKİ veri setinde var olan kişinin kendi rapor ettiği boy ve kilo bilgileri kullanılarak aşağıdaki gibi hesaplanmaktadır:

$$VKİ = \frac{\text{Kişinin Kilosu (kg)}}{(\text{Kişinin boyu (m)})^2}$$

Dünya Sağlık Örgütü'nün tanımlarına göre eğer kişinin VKİ'si 30 ve daha fazla ise kişi obez, 25 ve daha fazla ise kişi fazla kilolu olarak tanımlanmaktadır. Bu tanıma göre kişinin VKİ'si 30 veya daha fazlaysa çalışmada kullandığımız 2. bağılı kukla değişken 1, eğer kişinin VKİ'si 30'un altında ise bağılı değişken 0 olarak tanımlanmıştır. 2014 TSA veri setine göre Türkiye'de yaşayan kişilerin % 68.03'ü obez veya aşırı kiloludur. 2014 TSA veri setine göre sadece obezite oranı ise % 21.67'dir. Erkeklerde obezite oranı % 16.93 iken, kadınlarda bu oran %25.69'dur. Yani Türkiye'de obezitenin kadınlarda görünürlüğü erkeklere göre daha fazladır.

Analizlerde kullanılan diğer bağımsız değişkenler kişinin cinsiyeti, yaşı, eğitim seviyesi, medeni durumu, hanehalkı gelir grubu, istihdam durumu ve yaşadığı bölgedir. Analizlerde kişinin cinsiyeti kadınsa 0, erkekse 1 olarak kodlanmıştır. Kişinin yaşı veri setinde rapor ettiği sürekli değişkendir. Kişinin eğitim seviyesi için 4 farklı kukla değişken tanımlanmıştır. “İlkokul ve

daha az” değişkeni kişi ilkokul mezunu ise veya okuma-yazma bilip de diploması yok ise veya okuma yazma bilmiyorsa 1, diğerleri için 0 olarak tanımlanmıştır. “Ortaokul” değişkeni eğer kişi ilkokuldan sonra 3 yıllık herhangi bir okulu bitirdiyse veya 8 yıllık ilköğretim mezunu ise 1’e, değil ise sifıra eşittir. “Lise” değişkeni, eğer kişi lise veya mesleki lise mezunu ise 1, değil ise 0 olarak tanımlanmıştır. Son olarak “Üniversite ve daha fazla” değişkeni eğer kişi yüksekokul veya 4 yıllık fakülte mezunu veya master/doktora derecelerine sahip ise 1, değil ise 0 olarak hesaplanmıştır.

Kişinin rapor ettiği medeni durumuna göre ise 4 farklı medeni hal kukla değişkeni tanımlanmıştır. “Evli” değişkeni, kişi evli ise 1, değil ise sıfır değerini almaktadır. “Bekar” değişkeni kişi hiç evlenmemişse 1, aksi halde 0’a eşittir. “Dul/boşanmış” değişkeni kişi eşini kaybetmişse veya eşinden ayrıldıysa 1, değilse 0 olarak tanımlanmıştır.

Kişinin hanehalkı gelir grubu 1 ile 5 arasında değişen sıralı bir değişkendir. Hanehalkı gelir grubu aşağıdaki sorunun cevabına göre belirlenmiştir:

“Hananizin ortalama aylık geliri aşağıdaki gelir gruplarından hangisinde yer almaktadır?”

Soruya verilen cevaplar şu şekildedir: 1. 0-1080 TL. 2. 1081-1550 TL. 3. 1551-2170 TL.

4. 2171-3180 TL. 5. 3181+ TL.

Kişinin verdiği cevaba göre gelir endeksi değişkeni 1 ile 5 arasında değerler alan sıralı bir değişken olarak tanımlanmış, daha sonra bu değişkenlere göre beş farklı kukla değişken tanımlanmıştır. (Örneğin, gelir grubu 1 kukla değişkeni kişinin gelir endeksi 1’e eşitse 1, değilse 0 değerini almaktadır.)

Kişinin istihdam durumu ise aşağıdaki sorulara verilen cevaplara göre tanımlanmıştır. Kişinin istihdam durumunu belirleyici ilk soru şu şekildedir:

“Son bir hafta içerisinde aynı (mal) veya nakdi (para) bir gelir elde etmek amacıyla (ev kadını, öğrenci veya emekli de olsanız) bir saat bile olsa bir işte çalıştınız mı?”

Soruya verilen cevaplar: “Çalıştı”. “Çalışmadı fakat işle ilgisi devam ediyor”. “Çalışmadı” şeklindedir.

Kişi bu soruya “Çalışmadı” cevabını vermediği sürece istihdamda kabul edilmiş ve “istihdam” kukla değişkeni yaratılmıştır. Kişi istihdamda ise kukla değişken 1, değil ise 0 değerini almaktadır.

“Çalışmadı” cevabı veren kişiler içinse aşağıdaki soru sorulmuştur:

“Çalışmama nedeniniz nedir?”

Soruya verilen cevaplar: “İş bulamama/işsiz olup iş arıyor”, “Mevsimlik çalışıyor”, “Eğitim ve öğretime devam ediyor”, “Ev işleri ile meşgul”, “Emekli veya işi bırakmış”, “Özürlü veya hasta”, “Yaşlı”, “İrad Sahibi”, “Ailevi ve kişisel nedenler”, “Diğer” şeklindedir.

Bu soruya verilen cevap “İş bulamama/işsiz olup iş arıyor” ise kişi “işsiz” olarak tanımlanmış, “işsiz” kukla değişkeni ise kişi işsiz ise 1, değil ise 0 olarak hesaplanmıştır. Eğer kişi mevsimlik çalıştığını rapor etmişse “Mevsimlik çalışan” olarak tanımlanmıştır. “Mevsimlik çalışan” kukla değişkeni ise bu kişiler için 1, diğer kişiler için sıfır olarak verilmiştir. Kişi bu soruya bu iki seçenek dışında bir cevap veriyse kişi “işgücü dışında” olarak tanımlanmış ve “işgücü dışında” kukla değişkeni bu kişiler için 1, diğer kişiler için 0 olarak girilmiştir.

Tablo 1’de analizlerde kullanılan açıklayıcı değişkenlere göre obezite ve önceden sigara içip daha sonra bırakmış olmanın 2014 TSA veri setine göre frekans analizleri verilmektedir¹.

Tablo 1. Bağımsız Değişkenlere Göre Frekans Analizleri

Değişken	(1) Obezite (VKİ>=30)	(2) Sigarayı Bırakma
Cinsiyet		
Erkek	16.93	36.27
Kadın	25.69	39.84
Ortalama Yaş	50.43	49.45
Eğitim Seviyesi		
İlkokul ve daha az	29.50	42.16
Ortaokul	10.80	28.27
Lise	12.77	29.45
Üniversite ve daha fazla	13.19	40.84
Medeni Durum		
Evli	25.70	39.77
Bekar	5.08	24.40
Dul veya boşanmış	30.79	43.51
Hanehalkı Gelir Grubu		
Gelir Grubu 1	22.27	37.55
Gelir Grubu 2	22.38	38.49
Gelir Grubu 3	22.96	37.31
Gelir Grubu 4	19.51	34.56
Gelir Grubu 5	20.58	39.59
İstihdam Durumu		
İstihdamda	17.41	30.38
İşsiz	9.37	23.13
Mevsimlik İşçi	23.74	25.26
İşgücünde Değil	25.56	47.72

Kaynak: Yazarın TSA 2014 veri seti kullanarak yaptığı hesaplamalar.

Frekans analizi sonuçlarının açıkça gösterdiği gibi obezite en fazla ilkökul veya daha düşük eğitim seviyesine sahip olan kişilerde görülmektedir (Bu gruptaki kişilerin %29.50’si obezdir).

1 Süreli değişken olan “yaş” değişkeni için ilgili gruptaki ortalama yaş verilmiştir. Örneğin obez kişilerin yaş ortalaması 50.43 olarak hesaplanmıştır. Tüm sıklık oranları ağırlıklı ile birlikte hesaplanmıştır.

Sigarayı bırakma sıklığı en fazla olan kişiler de gene bu gruba veya en üst eğitim seviyesi grubuna dahildirler. İş gücüne dahil olmayan kişilerde hem obezite hem de sigara bırakma oranlarının diğer istihdam gruplarına göre daha yüksek olduğu gözlenmiştir. Gelir gruplarında anlamlı değişiklikler görülmemekle birlikte en alt 3 gelir grubunda obezite oranlarının daha yüksek olduğu saptanmıştır. Frekans analizleri medeni duruma göre bakıldığı zaman ise dul veya boşanmış kişilerde hem obezite hem sigarayı bırakma oranlarının evli veya bekar kişilere göre daha yüksektir. Frekans analizlerinde her iki bağlı değişken için oranların aynı anda aynı gruplarda en yüksek ve en düşük çıkması obezite ve sigarayı bırakma arasında yüksek bir korelasyon olması ihtimalinin çok yüksek olduğunu düşündürmektedir.

4. Model

Sigarayı bırakma ve obezite arasındaki ilişkiyi kişinin diğer sosyo-ekonomik özellikleri kontrol edilerek test edilebilmesi için çift değişkenli probit (bivariate probit) modeli kullanılmıştır. Çift değişkenli probit modelleri iki bağlı değişken arasındaki ilişkiyi test ederken aynı zamanda bu iki değişkeni aynı anda açıklayan değişkenleri kontrol eden bir modeldir. Çift değişkenli probit modelleri literatürde benzer çalışmalarda sıklıkla kullanılmıştır. Örneğin, Morris (2007) çalışmasında İngiltere’de obezite ve istihdam durumu arasındaki ilişkiyi çift değişkenli probit modeli kullanarak incelemiş, ve iki değişken arasında güçlü bir negatif ilişki bulmuştur. Türkiye için yapılan başka bir çalışmada Başlevent ve Onaran (2003) kadınlarda ek çalışan etkisinin üstünlüğünü 1988 ve 1994 yılları için çift değişkenli probit modeli kullanarak incelemiş ve kriz yılı olan 1994’de hanehalkı reisinin işini kaybetmesi ve kadının iş gücüne girmesi arasında güçlü bir pozitif ilişki bulmuştur.

Çalışmada kullanılan çift değişkenli probit modeli matematiksel olarak aşağıdaki şekilde yazılabilir:

$$y_1^* = \beta_1'x + u_1, \quad y_1 = 1 \text{ eğer } y_1^* > 0, \\ \text{diğer durumda } y_1 = 0 \quad (1)$$

$$y_2^* = \beta_2'x + u_2, \quad y_2 = 1 \text{ eğer } y_2^* > 0, \\ \text{diğer durumda } y_2 = 0 \quad (2)$$

Yukarıdaki modelde tahmin edilen y_1^* ve y_2^* değişkenlerinin işaretleri y_1 ve y_2 nin alacağı değerleri belirler. (1) numaralı denklemde y_1 bağımlı değişkeni kişinin obez olma olasılığını gösterirken, y_2 bağımlı değişkeni kişinin sigarayı bırakma olasılığını göstermektedir. (1) ve (2) numaralı denklemlerde x vektörü kişinin analizlerde kontrol edilen demografik ve sosyoekonomik göstergeleridir. Son olarak u_1 ve u_2 hata terimlerini göstermektedir.

Eş zamanlı olarak tahmin edilen (1) ve (2) numaralı denklemlerden elde edilen ρ (rho) katsayısı sigarayı bırakma ve obezite arasındaki korelasyonu göstermektedir. Çalışmanın ana sorusu ρ

katsayısının anlamlı olup olmadığıdır. Yani, sıfır (null) ve alternatif (alternative) hipotezler şu şekilde yazılabilir:

$$H_0: \rho = 0 \quad (3)$$

$$H_1: \rho \neq 0 \quad (4)$$

Sıfır hipotezini red edip etmeme kararı hesaplanan Wald Test istatistiğine göre verilecektir. Eğer sıfır hipotezi reddediliyorsa, obezite ve sigarayı bırakma arasında ilişki vardır. Korelasyon katsayısının işareti ilişkinin yönünü vermektedir (pozitif veya negatif).

Obezite ve sigarayı bırakma arasındaki korelasyon katsayısının anlamlı çıkması aynı zamanda bu iki değişken arasında içsellik olduğu anlamına gelmektedir (Wooldridge, 2002). Bu çalışmanın amacı içsellik problemini ele almak değil, iki değişken arasındaki ilişkiyi bulup; kişinin sigarayı bırakmasının aşırı kilo alımına neden olup olmadığını tespit etmektir. Şunu da belirtmek gerekir ki analizlerde kontrol edilen değişkenler bir nedensellik ilişkisi vermemektedir. Örneğin, kişinin eğitim seviyesi ve Vücut Kitle İndeksi (VKİ) arasında içsellik problemi olabilir (Tansel ve Karaoğlan, 2016). Bu değişkenler kontrol amaçlı analizlere eklenmiş olup, bağlı değişkenleri ne yönde etkiledikleri hakkında fikir vermektedir.

5.Ampirik Bulgular

Tablo 2 obezite ve sigarayı bırakma arasındaki ilişkiyi test eden çift değişkenli probit modellerin sonuçlarını vermektedir. Tablo 2'de 5 farklı ana sütun vardır. Her ana sütunun altında 2 tane alt-sütun vardır ve bu alt sütunların birincisi kontrol değişkenlerinin kişinin obez olma olasılığını nasıl açıkladıklarını gösterirken, ikincisi değişkenlerin kişinin sigarayı bırakma olasılığını nasıl açıkladığını göstermektedir. En son satırlarda ise toplam gözlem sayısı ile birlikte korelasyon katsayısı ve Wald test istatistiği sonuçları verilmiştir. Açıklayıcı değişkenler aralarındaki muhtemel içsellik problemlerinden dolayı, sonuçların daha güçlü ve tutarlı olması için analizlere teker teker dahil edilmiştir. (1) numaralı sütunda kişinin tamamen dışsal değişkenleri (yaş, cinsiyet) kontrol edilmiştir. (2) numaralı sütunda kişinin eğitim seviyesini gösteren kukla değişkenler regresyona eklenmiştir. (3) numaralı sütunda kişinin medeni durumunu gösteren kukla değişkenler analizlere dahil edilmiştir. (4) numaralı sütunda hanehalkı gelir endeksini gösteren kukla değişkenler ve (5) numaralı sütunda kişinin istihdam durumu probit analizlerine kontrol değişkeni olarak eklenmiştir.

Tablo 2, sigarayı bırakma ve obezite arasında anlamlı ve pozitif bir korelasyon olduğunu göstermektedir. Kişinin sadece dışsal faktörleri kontrol edildiği zaman aradaki ilişkinin büyüklüğü 0.17 bulunmuştur. Diğer kontrol değişkenleri eklendiğinde korelasyonun büyüklüğünün azalsa da anlamlılığını yitirmediğini ve ilişki yönünün değişmediği gözlenmiştir. Yani gelişmiş ülkeler için önerilen hipotez Türkiye için de geçerli gibi görülmektedir: Kişinin sigarayı bırakması kilo alma olasılığını arttırmaktadır.

Çift değişkenli probit sonuçlarına göre, her iki bağımlı değişkenle açıklayıcı değişken arasındaki ilişki Tablo 1'de verilen frekans analizleriyle tutarlı görülmektedir. Örneğin, ilkokul mezunu veya daha düşük eğitim seviyesine sahip olan kişiler de obezitenin anlamlı bir şekilde daha fazla görüldüğü gözlenmiştir. Bekar kişilerde evli kişilere göre obezite görülme ve sigarayı bırakma olasılıklarının daha az olduğu görülmektedir. Gelir grubunun obez olma olasılığını anlamlı bir şekilde açıklamadığı fakat özellikle alt gelir grubu için sigarayı bırakma olasılıklarını azalttığı görülmüştür. İstihdamda ve işsiz olan kişilerin ise iş gücünde olmayan kişilere göre obez olma ve sigarayı bırakma olasılıklarının daha az olduğu görülmektedir. Kişinin yaşla birlikte obez olma olasılığının artması çalışmanın bir başka önemli sonucudur. Son olarak kadınların obez olma ve sigarayı bırakma olasılıklarının erkeklere göre anlamlı bir şekilde fazla olduğu gözlenmiştir.

6.Sonuç

Son yıllarda tüm Dünya ile birlikte Türkiye'de de obezite oranlarında artış görülmüştür. Gelişmiş ülkeler için yapılan bazı çalışmalar sigara tüketimindeki düşüşün obeziteyi arttırdığı yönündedir. Bu çalışmada Türkiye Sağlık Araştırması (TSA) veri seti kullanarak sigarayı bırakma ve obezite arasındaki ilişki çift değişkenli probit teknikleri kullanılarak test edilmiştir. 25 yaş ve üstü kişilerin demografik ve sosyo-ekonomik özellikleri kontrol edilerek yapılan analizlerin sonucunda iki değişken arasında anlamlı pozitif bir ilişki saptanmıştır. Dolayısıyla bu sonuç Chou vd. (2004), Rashad ve Grossman (2004) ve Rashad (2006)'ın sonuçlarını destekler niteliktedir. Yani, sigarayı bırakmanın kişiyi daha fazla yemeğe yönelttiği ve kalori yakımını yavaşlattığı söylenebilir.

Sigara tüketimi de obezite de insan sağlığı için oldukça zararlıdır. Dolayısıyla, obeziteyi düşürmek için sigara tüketiminin arttırılması veya sigara fiyatlarının düşürülmesi yanlış bir politika önerisi olacaktır. Ortaya çıkan sonuçlar bağlamında sigarayı bırakma ve obezitenin tehlikelerini vurgulayan kampanyaların eş zamanlı olarak uygulanması önem taşımaktadır.

Çalışmanın sonuçlarının beklenildiği gibi olmasına ve sağlık davranışları ve sosyo-ekonomik faktörlerin ilişkisi ile ilgili anlamlı ve tutarlı sonuçlar vermesine rağmen, politika önerisi yapmak için yeterli bir çalışma olmadığı kabul edilmektedir. Çalışmada kullanılan veri seti kesit bir veri setidir (Cross-sectional data set). Türkiye için bu alanda henüz bir panel veri seti bulunmamaktadır. Örneğin, panel veri seti ile daha önce sigara içen bir kişinin, sigarayı bırakmış olması halinde Vücut Kitle İndeksi (VKİ)'nin nasıl etkilendiğini görebilmek bu tip araştırmanın sonucunu daha kuvvetli kılacaktır.

Bu çalışmanın sonucunda obezite ve sigarayı bırakma arasında içsellik ilişkisi olduğu tespit edilmiştir. Morris (2007) benzer bir çalışmada kişinin istihdam durumu ve obezite arasındaki güçlü bir negatif ilişki bulduktan sonra, araç değişken (AD) yöntemleri kullanarak bu iki değişken arasındaki nedensellik ilişkisine bakmıştır. Bu çalışmada kullanılan veri setindeki kısıtlardan dolayı obezite ve sigarayı bırakma arasındaki nedensellik ilişkisini AD yöntemleri kullanarak incelemek mümkün olmamaktadır. Ancak, bu çalışmanın ileriki araştırmalar için sağlam bir alt yapı oluşturması beklenmektedir.

Referanslar

- Başlevant, C., ve Ö. Onaran (2003). "Are Married Women in Turkey More Likely to Become Added or Discouraged workers?." *Labour* 17 (3): 439-458.
- Chou, S. Y., M. Grossman, ve H. Saffer (2004) "An Economic Analysis of Adult Obesity: Results from the Behavioral Risk Factor Surveillance System." *Journal of Health Economics* 23 (3): 565-587.
- Cutler, D. M., E.L. Glaeser, ve J.M. Shapiro (2003) "Why Have Americans Become More Obese?," *The Journal of Economic Perspectives* 17 (3): 93-118.
- Gruber, J., ve M. Frakes (2006) "Does Falling Smoking Lead to Rising Obesity?," *Journal of Health Economics* 25 (2): 183-197.
- Karaoğlan, Deniz ve Aysit Tansel (2018) "Determinants of Body Mass Index in Turkey: A Quantile Regression Analysis from a Developing Country" *Boğaziçi Journal Review of Social, Economic and Administrative Studies*. 32 (2): doi no:10.21773/boun.32.2.1.
- Morris, S. (2007). The Impact of Obesity on Employment. *Labour Economics* 14 (3), 413-433.
- Philipson, T. J., ve R. A. Posner (2003) "The Long-Run Growth in Obesity as a Function of Technological Change." *Perspectives in Biology and Medicine* 46 (3): 87-107.
- Philipson, T. J., ve R. A. Posner (2008) "Is the Obesity Epidemic a Public Health Problem? A Review of Zoltan J. Acs and Alan Lyles's" Obesity, Business and Public Policy." *Journal of Economic Literature* 46 (4): 974-982.
- Rashad, I. ve M. Grossman (2004) "The Economics of Obesity." *Public Interest* 156: 104-113.
- Rashad, I. (2006) "Structural Estimation of Caloric Intake, Exercise, Smoking, and Obesity." *The Quarterly Review of Economics and Finance* 46 (2): 268-283.
- Stewart, S. T., D. M. Cutler, ve A. B. Rosen (2009) "Forecasting the Effects of Obesity and Smoking on US Life Expectancy." *New England Journal of Medicine* 361 (23): 2252-2260.
- Tansel, A., ve D. Karaoğlan, (2014) "Health Behaviors and Education in Turkey." Bonn, Germany: Institute for the Study of Labor (IZA) Discussion Paper No: 8262.
- Tansel, A., ve D. Karaoğlan, (2016) "The Causal Effect of Education on Health Behaviors: Evidence from Turkey" Ankara, Türkiye: Economic Research Center (ERC) Working Paper No: 16/06.

Ek I:**Tablo 2.** Çift Değişkenli Probit Sonuçları

	(1)		(2)		(3)		(4)		(5)	
Değişkenler	Obezite	Sigarayı Bırakma	Obezite	Sigarayı Bırakma	Obezite	Sigarayı Bırakma	Obezite	Sigarayı Bırakma	Obezite	Sigarayı Bırakma
Cinsiyet	-0.291*** (0.033)	-0.160*** (0.030)	-0.290*** (0.0327)	-0.146*** (0.031)	-0.292*** (0.033)	-0.176*** (0.0313)	-0.292*** (0.0337)	-0.173*** (0.0313)	-0.231*** (0.0390)	-0.0891** (0.0366)
Yaş	0.102*** (0.008)	-0.009 (0.007)	0.0977*** (0.00774)	-0.009 (0.007)	0.091*** (0.007)	-0.0192** (0.00752)	0.0911*** (0.00799)	-0.0206*** (0.007)	0.0915*** (0.0080)	-0.0194*** (0.0074)
Yaş^2	-0.001*** (7.44e-05)	0.0004*** (7.18e-05)	-0.0008*** (7.50e-05)	0.0003*** (7.25e-05)	-0.001*** (7.62e-05)	0.0004*** (7.49e-05)	-0.001*** (7.69e-05)	0.0004*** (7.50e-05)	-0.001*** (7.82e-05)	0.0004*** (7.43e-05)
Eğitim Seviyesi:										

İlkokul veya daha az	0.117*** (0.0414)	-0.132*** (0.0383)	0.0975** (0.041)	-0.160*** (0.0388)	0.0988** (0.0455)	-0.129*** (0.0417)	0.0882* (0.0455)	-0.145*** (0.0418)
Ortaokul	0.00725 (0.0588)	-0.254*** (0.0534)	-0.006 (0.0591)	-0.271*** (0.0537)	-0.00874 (0.0611)	-0.252*** (0.0554)	-0.0222 (0.0613)	-0.272*** (0.0556)
Lise	-0.0599 (0.0521)	-0.215*** (0.0465)	-0.0704 (0.0524)	-0.227*** (0.0467)	-0.0732 (0.0536)	-0.219*** (0.0475)	-0.0815 (0.0537)	-0.232*** (0.0477)
Medeni Durum:								
Bekar			-0.254*** (0.0743)	-0.259*** (0.0604)	-0.253*** (0.0744)	-0.257*** (0.0605)	-0.237*** (0.0746)	-0.240*** (0.0610)
Dul/ Boşanmış			-0.0707 (0.0564)	-0.321*** (0.0556)	-0.0673 (0.0566)	-0.307*** (0.0557)	-0.0490 (0.0569)	-0.284*** (0.0558)
Gelir Grubu:								
Gelir Grubu 1					-0.0284 (0.0520)	-0.153*** (0.0473)	-0.0338 (0.0527)	-0.165*** (0.0480)
Gelir Grubu 2					0.0170 (0.0550)	-0.0332 (0.0495)	0.00630 (0.0553)	-0.0500 (0.0499)
Gelir Grubu 3					0.00968 (0.0561)	-0.0461 (0.0508)	-0.0022 (0.0563)	-0.0629 (0.0511)
Gelir Grubu 4					-0.0172 (0.0542)	-0.123** (0.0491)	-0.0223 (0.0543)	-0.132*** (0.0491)
İstihdam Durumu								
İstihdamda							-0.115*** (0.0392)	-0.164*** (0.0374)
İşsiz							-0.236** (0.0941)	-0.277*** (0.0829)
Korelasyon Katsayısı	0.170*** (0.0201)	0.170*** (0.0202)	0.166*** (0.0202)	0.165*** (0.0202)	0.165*** (0.0202)	0.162*** (0.0202)	0.162*** (0.0202)	0.162*** (0.0202)
Wald Test İstatistiği	71.884	70.843	67.571	66.931	66.931	64.271	64.271	64.271
Toplam Gözlem	8,515	8,515	8,515	8,515	8,515	8,515	8,515	8,515

Not: (1) Kaynak: Yazarın 2014 TSA veri seti kullanarak yaptığı hesaplamalar. (2) Robust standard sapmalar parantez içinde gösterilmiştir. (3)*** p<0.01, ** p<0.05, * p<0.1

PREDICTION OF TRANSITION PROBABILITIES FROM UNEMPLOYMENT TO EMPLOYMENT FOR TURKEY VIA MACHINE LEARNING AND ECONOMETRICS: A COMPARATIVE STUDY

TÜRKİYE İÇİN MAKİNE ÖĞRENMESİ VE EKONOMETRİ YOLUYLA İŞSİZLİKTEN İSTİHDAMA GEÇİŞ OLASILIKLARININ TAHMİNİ: KARŞILAŞTIRMALI BİR ÇALIŞMA

Yasin KÜTÜK *
Bülent GÜLOĞLU **

Abstract

In this study, it is mainly aimed to predict transition probabilities of individuals who are previously unemployed and get employment or stay unemployed. In order to do that, Household Labor Force Surveys conducted in Turkey are merged and matched from 2004 to 2016. Information about individuals only consists of individual characteristics and qualifications since there should not be any informative clue about the present situation. To predict those probabilities, logistic regression analysis as econometric approach, a shallow neural network and machine learning classification algorithms are run in order to compare them. The results indicate that classification in machine learning is slightly better than logistic regression and shallow neural network. While XGBoost classifier and Random Forest get 67% accuracy, logistic regression can predict only 63% of an individual's transition and shallow neural network forecasts 51%.

Keywords: Employment, Transition Probability, Machine Learning, Classification

JEL Classification: J21, C38, C45

Öz

Bu çalışma içerisinde, esasen daha önce işsiz olan bireylerin istihdama geçiş veya işsizlikte kalma olasılıklarını tahmin etmek amaçlanmaktadır. Bu amaçla, Türkiye'de yapılan Hanehalkı İşgücü Anketleri

* Corresponding Author, Ph.D. Candidate at Istanbul Technical University and R.A at Altınbaş University, Department of Economics, SEASS, Social Sciences Campus, Büyükdere Caddesi, No: 147, PK.34349, Esentepe, Şişli / Istanbul, Turkey, Tel:(+90) 212.604.0100 (2071). yasinkutuk@itu.edu.tr | yasin.kutuk@altinbas.edu.tr, Orcid: 0000-0003-2304-8309

** Istanbul Technical University, Department of Economics, Management Faculty, Harbiye Mahallesi, 34367 Maçka, Beşiktaş / Istanbul, Turkey, Tel: (+90) 212.293.1300 (2029). guloglu@itu.edu.tr

2004 ile 2016 arasında birleştirilmiş ve eşleştirilmiştir. Veriler, bireylerle ilgili mevcut istihdam/işsizlik durumu hakkında herhangi ipucu içermeyecek şekilde, bireysel özellik ve niteliklerle oluşturulmuştur. Bu geçiş olasılıklarını tahmin etmek ve bunları karşılaştırmak amacıyla, ekonometrik yaklaşım olarak lojistik regresyon analizi, tek katmanlı yapay sinir ağı ve yapay öğrenme sınıflandırma algoritmaları uygulanmıştır. Sonuç olarak, yapay öğrenme algoritmalarının, lojistik regresyon ve tek katmanlı sinir ağından görece daha iyi olduğunu göstermektedir. XGBoost sınıflandırıcısı ve Rassal Orman Karar Ağaçları algoritmaları %67 doğruluk ile, lojistik regresyon bir bireyin geçiş olasılığını yalnızca %63 düzeyinde ve tek katmanlı yapay sinir ağı ise % 51'ini tahmin edebilmektedir.

Anahtar Kelimeler: İstihdam, Geçiş Olasılığı, Yapay Öğrenme, Sınıflama

JEL sınıflaması: J21, C38, C45

I. Introduction

Not only in strong, rich and wealthier economies, but in all economies, more or less, unemployment is a phenomenon that exists not only in theoretical textbooks and articles, but also in everyday life or society. As ILO determines, “*the definition of unemployment covers people who are: out of work, want a job, have actively sought work in the previous four weeks and are available to start work within the next fortnight; or out of work and have accepted a job that they are waiting to start in the next fortnight*”¹. This definition includes only those who are actively seeking jobs, but the duration of active job search is not determined. For this reason, this active job search may be short or long. Important factors determining this period vary from the economic environment involved to the characteristic and experiential characteristics of the person. Economic environments vary from country to country. While some countries that may be export-oriented and can provide current account surpluses can create job opportunities in line with growth rates due to rich economic environment, some countries based on labor-intensive sectors can provide less employment opportunities due not to chronically reaching high-valued technology intensive sectors. From the perspective of labor economics, the significant point is that how an economy can deal with unemployment with successful policies within its economic environment. History of unemployment says most of countries have not been passed in that issue. Historically, there has been a marked increase in unemployment rates, throughout the countries. This increase in unemployment rate is quite obvious in Table 1:

Table I. Unemployment Rates all over the World

Country	1950-73	1974-83	1984-93	1994-98
Belgium	3.0	8.2	8.8	9.7
Finland	1.7	4.7	6.9	14.2
France	2.0	5.7	10.0	12.1
Germany	2.5	4.1	6.2	9.0
Italy	5.5	7.2	9.3	11.9
Netherlands	2.2	7.3	7.3	5.9

1 <http://www.ilo.org/ilostat-files/Documents/description UR EN.pdf> Accessed: 10.11.2017.

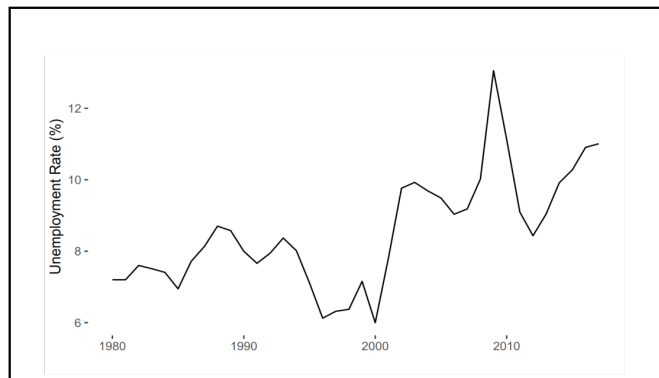
Norway	1.9	2.1	4.1	4.6
Sweden	1.8	2.3	3.4	9.2
United Kingdom	2.8	7.0	9.7	8.0
Ireland	n.a.	8.8	15.6	11.2
Spain	2.9	9.1	19.4	21.8
Average	2.6	6.0	9.2	10.7
Australia	2.1	5.9	8.5	8.6
Canada	4.7	8.1	9.7	9.4
United States	4.6	7.4	6.7	5.3
Average	3.8	7.1	8.3	7.8
Japan	1.6	2.1	2.3	3.4

Source: 1950-83 from Maddison et al. (1995) and updates from OECD, Labour Force Statistics accessed: 15.11.2017

There are many arguments for these fluctuations in history, but the reasons for all of them depend on several factors. Historically, despite the fact that the population has nearly doubled in all of the above countries, in all economies, except Japan, Italy and Netherlands, the unemployment ratio has tripled. When the average of each period is taken into account, the apparent increase is more pronounced. Moreover, all these countries, starting from the 1950s, have advanced to developed economies. Unemployment in developing countries presents worse scenarios than economically developed countries.

Turkey, the world's 17th economy, at the moment this study was carried out², is one of these developing economies. Having achieved significant growth rates for the last 15 years, Turkey did not present the same success in creating employment opportunities when the trend of unemployment rates is compared with others. After 2000, having experienced two crises, one in 2001 is national, the latter in 2008-2009 was international, unemployment rates has doubled regardless of getting back prior to 2000s' level. These increases have not been reverted back as these jumps can be easily seen in Figure 1:

Figure 1. Unemployment Rates in Turkey



² <https://www.worldatlas.com/articles/the-world-s-25-biggest-economies.html>

The main causes of unemployment in Turkey result from the inability to achieve the desired level of industrialization, the quality remained limited due to the excess weight of the share of rural communities, inability to access the level of investment required. These reasons can quickly increase the unemployment of young people and carry structural problems. One of the major differences between developed countries and Turkey; while the level of education is high in the unemployed, the level of education in the underemployed is low in Turkey. Another major difference, in contrast to developed countries, is the unemployment rate in Turkey is higher among educated people; the cause of this is that the jobs are far from being satisfactory and unsuitable for individuals looking for work for the first time.

In order to deal with unemployment, the private sector in employment, which was initiated in 2017, became the forefront of employing the public sector to support the social security premium that should be paid to the employer. In addition to this, the tax that will be deducted from the salary of the employee will be paid by the state. According to the explanation given at the beginning, the employer who wants to benefit from the support of “one sibling” will pay the first month’s tax, premium and salary of the new worker, according to the new incentives to be introduced in the coming period. Every two months from the second month, the expenses will be covered by the Unemployment Fund. Measures put into effect from the beginning of 2017 seem to be to provide temporary fixes by transferring private sector funds from the public sector rather than solving the problem permanently. The payment of the bill to the unemployment fund, which the public sector will undertake, brings many new debates.

The state of the economy is very important in the course of an individual’s employment or transition to it. As mentioned before, the other most important factors are all about personal abilities. Moreover, in economies where unemployment is high, individual developments gain more importance because of the increased competition. In this sense, individuals attach importance to their personal development, receive more training, establish and expand links with their business networks, and acquire new skills. Turkey’s labor market which is competitive in this perspective is a market where individual talents come to the fore. In that sense, numerous universities have opened to increase human capital, the number of individual vocational courses supported and subsidized by the government has been increased and the quality of pre-university education has been raised.

This study, at this point, examines individual skills, ingenuities and characteristics for the transition of individuals into employment. However, here it is necessary for individuals to be unemployed in accordance with the above ILO definition in the previous period, to move employment in the next period or to keep their situation. This is called as transition which can be calculated with probability. Since the calculation of this transitivity is important for policy makers or institutions, previous studies are not enough because both the estimation and prediction power are relatively low and the datasets are limited. As the second decade of the millennium brings renovated, strengthened and enriched tools to humankind in many areas, the most powerful one is data science. This study uses new estimation methods to increase predictability and forecast ability by

using machine learning techniques which use all the blessings of data science. Later, these results are compared to econometric ones.

In the following sections, first, a review of literature that covers job finding in the view of experimental studies in labor market, then some key figures are shown for Turkey. Later, the methods used in this study are presented with their data as well, and the results follow. Finally, the conclusion will summarize the whole study and provide some basic insight to the policy makers. Limitations and further studies conclude this study.

2.Literature Review

In the job market, all workers compete with each other. In a complete information and perfect competitive labor market, labor earns only marginal product according to classical theory. Again, a rational worker can invest as much as the cost of investment in his/her human capital (Becker, 1964) when the marginal investment of his/her human capital is positive. While it is positive, individual invests. Investment in human capital may be split into two main categories as offered: general and specific (Becker, 1964). General human capital arises if there is no firm-specific differentiation among workers and firms are not responsible for and do not tend to pay it to worker (Acemoglu and Shimer, 1999). Specific training is under the control of firms in order to increase workers' productivity while it helps to boost economic growth in a country by reducing quit rates and turnovers, increasing labor efficiency (Higashi, 2002; Parsons, 1972; Donaldson and Eaton, 1976). In the nature of specific investment, it is not certain that it accumulates during employment (Bougheas and Georgellis, 2004) leading to an increase in productivity and uncompetitive specializations. Wasmer (2006) and Cingano (2003) conclude specific investment in human capital and propose that firm-specific skills provide higher returns rather than others.

Empirical results show that the return on education is always positive, but researchers do not reach a common result on the subject of its amount. Dearden (1998), Ashenfelter et al. (1999) and Psacharopoulos (1994) find the effect of education around 5%, Mincer (1974), Krueger and Ashenfelter (1992) and Becker (1964) reach the effect of schooling on earnings around 11% level. In macroeconomics, Acemoglu (2002) conducts a study in less-developed regions and finds evidence that return on education has a multiplier effect on less developed countries. Blundell et al. (1999), Mincer and Ofek (1982) draw attention to the fact that gender differentiation on rate of return of education is in favor of male workers.

Tansel and Tasci (2004) uses 2001-2001 HLFS on the Turkey data to estimate unemployment transition probabilities, which may be the first application querying labor market transitions in Turkey. Proportional risk models, log-logistic and log-normal models were used in the study. The results indicate that urban or married individuals are more likely to be unemployed than other groups, but women are less likely to be unemployed than men. At the educational level, the probability of unemployment of college or undergraduate graduates is higher than other education groups, and the probability of becoming unemployed decreases with age. After that,

Tasci and Tansel (2005) estimated long-term unemployment durations with two-stage probit models using TURKSTAT's 2000-2001 Household Labor Force Surveys. Living in cities, for both types of unemployment, is reducing unemployment while women are increasing the possibility of long-term unemployment. Geographically, long-term unemployment decreases as we move from west to east. At the educational level, the probability of long-term unemployment decreases as the number of primary school graduates increases, but it still has positive effect, while the effect of long-term unemployment on graduates is negative. The appearance of gender discrimination in the labor market in Turkey, due to undergo more difficult than expected employment of women transition from unemployment is examined and this employment are high influence under finding a job in their field (İlkkaracan, 2012). Tunali and Ercan (2003) also indicates that although vocational school education is expected to help the graduates ease their way into the labor market, statistics reveal that a significant majority continues with higher education. Tasci and Ozdemir (2006) find that the transition from long-term unemployment to employment may be damaged by living in rural areas, being female, older and having lower education.

3. The Data

The data are collected from Household Labor Force Surveys (HLFS) conducted in between 2000 and 2016 by the Turkish Statistical Institute (TURKSTAT). In total, 7,323,449 individuals were surveyed by means of a set of questionnaires consisting of extensive information that were classified under seven headings such as household characteristics, individual's characteristics, employment status, income, unemployment and inactivity, past work experience, labor status one year before the survey. Total number of variables is 110 with three of them derived from other ones. The data as a csv file costs 3.18 GB to an Ext4³ file-system, which is an example of a big-data that it is hard to deal with classical computational methods. Rather, the data is compressed under Ext4 system by using "gzip" compression with a level of 6 by means of R that saves the data as an ".rdata" in order to process it with ease. The size of the big data after compression is reduced by up to 15%.

However, since this research is restricted only with both previously unemployed individuals and surveys done in 2003-2000 are not compatible to catch individuals who are previously unemployed and who are employed or unemployed at the survey date, the data merged are filtered against them⁴ to focus only on those who are interested in this study. So, given that information, the data are downsized to 203,891 individuals who will be split as train set (188,208 individuals) and test set (15,683). Due to the inconsistency of the surveys, a map for variables is generated to match them. However, even if the same question is asked to individual in the next year, the answers may be changed by TURKSTAT. So, the matching process for questions is not enough, the answers must be compared in terms of both content encodings and means. Variables are categorized into four groups: characteristics, employment, unemployment and past information. Later, some of

3 Fourth extended file-system, generally used by Linux OSs.

4 Individuals who are not in labor force, institutional population and are not actively searching for a job with the definition of ILO are dropped.

them are processed to derive certain variables. For instance, it is asked when the individual had a job in his/her past.⁵ The answer may not give a clear opinion, so the difference in between survey year and the answer is used to get the period of inactivity.⁶ The tabulation matrix, Table 2, is given below by year:

Table 2. Tabulation of HLFS (2004-2016)

Year	t0	t1_Unemp	t1_Emp
2004	Unemp	13426	7722
2005	Unemp	15136	8139
2006	Unemp	14110	7559
2007	Unemp	13685	7287
2008	Unemp	14024	7404
2009	Unemp	18625	10244
2010	Unemp	20097	9833
2011	Unemp	18223	7977
2012	Unemp	15941	6841
2013	Unemp	14997	6756
2014	Unemp	15064	7067
2015	Unemp	15497	7339
2016	Unemp	15066	7383
		SUM	203,891

Note: t0 means previous term while t1 means current period of survey.

The data have so many NAs in both qualitative and quantitative features. To be able to run a ML classifier that is so sensitive to NAs, NAs in this data were replaced as follows:

- Qualitative features/variables are encoded as factors. In order not to drop NAs, they are regarded as another level which is unknown in each qualitative features.
- In quantitative features, NAs are filled with median's of that feature for the current survey. Results do not show significant changes.

4.Methodology

As it was previously stated, since the dependent variable (feature) is a dichotomous one that indicates current employment or unemployment, it should be analyzed with discrete (or qualitative) choice models in econometrics or supervised-classification problem in machine

⁵ JOBLEFT YEAR

⁶ INACTIVE=SURVEY YEAR – JOBLEFT YEAR

learning literature. Greene (2003) splits discrete choice models as binary, multinomial or ordered.⁷ So, in econometric terms, the problem can be regarded as a binary choice problem.

In machine learning literature, classification problems can be solved with numerous methodologies. But here, the models which are shallow are preferred. The reason why shallow models are chosen is that since the econometric methodology can be regarded a one-layered neural network with some inputs that are weighted with their coefficients, an activation function that is a sigmoid and one output that is probabilities which are estimated with marginal coefficients is produced consequently. The whole econometric process, therefore, can be called as a shallow neural network which consists of one perceptron. It also allows comparing easily both estimation methodologies. Hence, the algorithms used in this study are given below:

Table 3. Machine Learning Algorithms

Function F	Methods
Naïve Bayesian Classifier	Gaussian
	Bernoulli
Support Vector Machines	Support Vector Classification (SVC)
	Linear SVC
Decision Trees	Extra Trees Classifier
	CART
	Logistic Model Trees
	Gradient Boosted Trees (XGBoost)
Discriminant Analysis	Linear Discriminant Analysis
	Quadratic Discriminant Analysis
Stochastic Gradient Descent	SGD Classification
Neighbors Classifier	KNeighbors
Ensemble Methods	Random Forest
	Bootstrapped Aggregation (Bagging)
	Adaboost
Neural Networks	Shallow/Wide Artificial Neural Network

All modules except the neural network were run under Python thanks to the Scikit-Learn (Pedregosa et al., 2011), the marvelous package, while shallow neural network was developed under Keras (Chollet et al., 2015). All were developed with and written in Python.

In these machine learning algorithms, the data must be split into two parts⁸: Train and test data (Breiman and Spector, 1992). An incisive separation of the data makes it vulnerable in prediction. In order to prevent that problem, Stratified K-Fold with randomized but not duplicated is a well suited algorithm to split the data (James et al., 2013). Strata are determined by the years, in the

⁷ and event counts, but it is not appropriate the case of this study.

⁸ Separation of cross validation/development data is skipped.

data, it is 13. So at first, Stratified K-Fold algorithm splits the data into 13 strata. Then each stratum is shuffled. Later, shuffled observations are attained into a test set at 8% approximately without duplication. The shape of the train data is [188208,43], while test's is [15683,43].

The final data reached, after being filtered by several conditions stated above, total observations which survived are 203,891 who were unemployed in the previous year of the survey date. To deal with the computationally costly process of ML algorithms, all quantitative variables/features are normalized in 0 and 1 according to the survey year. The variables that are used are classified according to their types within a broad classification: quantitative and qualitative variables. NAs in quantitative variables are accepted as another category since all the level of that kind of variables has levels.⁹ Hence for the quantitative variables, this process is simplified. All NAs are filled with the median value for the current year. Since median is more robust to outliers and large datasets, it is preferred (Rousseeuw and Bassett Jr, 1990).

Afterwards, a well-known econometric approach, logistic regression, for discrete/binary choice observations is applied to estimate transition probabilities for individuals who were unemployed in the previous term and moves to employment or stayed unemployed again for the current survey period.

5.Results

Descriptive statistics for the variables that are used in this study are given in Appendices. Quantitative features are summarized in Table 5 while qualitative features are in Table 6. Even the variable that was used to estimate population projections, WEIGHT_HLFS is also utilized to estimate transition probabilities, as it conveys regional information to estimation. Again, the deflator is also used since it has a small but significant role to convey information about the state of the economic environment for Turkey to the estimation processes. Therefore, no other variables were derived and used in conjunction with the variables specified in the descriptive statistics.

Table 4. Classifier Results

Classifier	Classes	Precision	Recall	F1-score	Support	Confusion Matrix	Accuracy	AUC
Random Forest Train	0	0.76	0.75	0.75	93740	70549 23191	0.7554	0.8486
	1	0.75	0.76	0.76	94468	22852 71616		
	Avg / Total	0.76	0.76	0.76	188208			
Random Forest Test	0	0.66	0.66	0.66	7811	5146 2665	0.6609	0.7404
	1	0.66	0.66	0.66	7872	2653 5219		
	Avg / Total	0.66	0.66	0.66	15683			
XGBoost	0	0.73	0.73	0.73	93740	68501 25239		

⁹ or factors, categories etc.

Train	1	0.73	0.74	0.73	94468	24955	69513	0.7333	0.8243
	Avg / Total	0.73	0.73	0.73	188208				
XGBoost Test	0	0.67	0.66	0.67	7811	5176	2635	0.6704	0.7492
	1	0.67	0.68	0.67	7872	2534	5338		
	Avg / Total	0.67	0.67	0.67	15683				
Ada Boost Train	0	0.65	0.6	0.62	93740	56347	37393	0.6371	0.6948
	1	0.63	0.67	0.65	94468	30916	63552		
	Avg / Total	0.63	0.63	0.63	188208				
Ada Boost Test	0	0.64	0.57	0.6	7811	4661	3150	0.6360	0.6958
	1	0.62	0.68	0.65	7872	2559	5313		
	Avg / Total	0.63	0.63	0.63	15683				
Extra Tree Train	0	1	1	1	93740	93740	0	OVERFIT	OVERFIT
	1	1	1	1	94468	26	94442		
	Avg / Total	1	1	1	188208				
Extra Tree Test	0	0.62	0.69	0.65	7811	5409	2402	0.6349	0.7073
	1	0.65	0.58	0.61	7872	3324	4548		
	Avg / Total	0.64	0.63	0.63	15683				
Logistic Regression		0.6371	0.5999	0.6179	102340	40634	67638	0.6305	0.6300

Later, the algorithms stated in Table 3 are run to get transition probabilities for the same sample again. For the Support Vector Classifier (SVC), radial basis function is used as kernel with 1.0 penalty parameter. Linear Support Vector uses squared hinge loss function. All tree based classifiers, Extra Tree Classifier and CART, since the data uses two classes which contain impurities, criterion contains only Gini for splitting the classes. Learning rate α for the XGBoost classifier is determined as 0.1 while the shallow network has 0.01. Logistic Regression Trees, Stochastic Gradient Descent Classifier and Linear Support Vector have their regularization terms prepared with L2, the ridge regression is utilized for them not to emerge over-fitting. However, Extra Tree Classifier gives over-fitted results due to catching information about next term employment for individuals.

The results indicate that the ML algorithm is relatively better than an econometric one (see Table 4, the Logistic Regression). In order to keep comparability between econometric and machine learning solutions, the shallow neural network is prepared as one layered since other algorithms including logistic regression has one process. Top two algorithms which get close accuracy rates are XGBoost Gradient Descent Classifier and Random Forest Classifier, their accuracies are 67%, 66%. Both of them use decision trees which is so suitable for the classification problems. As the classes that are employed or unemployed at the survey date but are unemployed in previous

term are not imbalanced, namely the ratio between two classes do not exceed 95% in any year, these classifiers successfully distinguish the classes by using characteristic information about individuals.

AdaBoost Classifiers follows these top two, where accuracy rate is around 63.6%. Extra-Tree classifier comes from later than AdaBoost Classifiers and gets 63.5% which is quite similar to AdaBoost's. But these latter classifiers perform so close to econometric approach since logistic regression which has a determined cut-off probability gets nearly 63% accuracy (see Table 4). In order to predict t1 term's unemployment (0) or employment (1), the cut-off value is taken as 0.5. So, if the probability is greater than the cut-off, the label gets a 1 which indicates predicted t1's situation is employment, otherwise, it is 0, that is individual stays unemployed in survey date. While logistic regression is beaten by four machine learning classifiers, it can successfully beat the rest of classifiers that are listed in Table 3 and resulted in Appendix in Table 7.

When it is elaborated why machine learning estimation methods could beat econometric ones, the following reasons come to mind. As stated in the literature, Mincer's (1974) equation establishes a non-linear equation where experience is taken as squared as independent variable. Since, especially XGBoost and Random Forests are completely superior tools when there is a non-linearity between features (explanatory variables) in a classification problem, they can deal with the problem rather than linear solutions even if the non-linear features do not feed the model. Second, linear models are good at monotonic relationships. Non-linearity breaks the assumption about monotonicity, therefore, machine learning estimation methods work splendidly. Theoretically, tree-based models in principle can approximate functions regardless of shape, whereas linear models can only produce functions with a linear shape with respect to a chosen set of features. Third is related to cut-off point. In logistic regression, it is required to determine cut-off as 0.5 where the z-value turns 0 indicating the class is not determined. However, in machine learning algorithms, this cut-off value does not depend on any value. Owing to determine the classes the path of trees that reach 1 or 0, there is no cut-off value which should be optimal.

The last concerns the size of the data, by virtue of processing relatively big data, computational issues take more time than expected. XGBoost and Random Forest classifiers can run in approximately 0.0007% of the time logistic regression and can beat it with ease, machine learning estimation methods can be chosen if the task needs to be solved within a certain time.

6. Conclusion

In this study, the main aim is to predict how individuals who are unemployed in previous term pass to employment or stay at current state with his/her characteristics and qualifications. In order to do so, Turkish labor market is handled with care by using relatively bigger data. Side task is to look at Turkish labor market with recent estimation methods which provide more accuracy compared to previous ones.

In order to look at transition from unemployment from a broader angle, the time span was chosen as wide as possible. Household labor force surveys which were decided on to calculate labor related issues were utilized from the beginning 2000. However, since it is impossible to derive the state of previous terms, starting year is accepted as 2004, then the data are expanded to 13 years, to 2016. Most of the variables are matched as much as possible. Though, some of them can give a clue to the present state of the individual in labor market, these variables are excluded. Survived variables are only related with characteristics, previous term and present qualifications of individuals.

Previous year state of individuals is known, however, present state is taken as unknown, which can be unemployment (0) or employment (1). Since the dependent variable that should be predicted is a dichotomous one, the problem turns out to be a binary choice model. Binary choice models can be solved with logistic regression in econometrics and classification methods in machine learning literature. For machine learning literature, classifiers and a shallow neural network with a sigmoid activation function that can be used to compute transition probabilities of each individual are used to compare them with logistic regression. The family of classifiers is Naïve Bayesian, Support Vector Machines, Decision Trees, Discriminant, Stochastic Gradient Classifier and Ensemble techniques.

By comparing accuracies, specifically the XGBoost classifier which is in Decision Trees, and Random Forest that in Ensemble have nearly the same, 67% accuracy rate to predict present state of individuals in the Turkish labor market. However, logistic regression with the same variables has slightly lower accuracy, 63%; this exceeds half of ML algorithms in general. Since the results obtained belong to the test data, the variables and models can be used to estimate new data produced in the future to predict an individual's transition probability.

It is important to estimate the likelihood of individuals going to employment in the future. Policy-makers, for example, may find individuals who will be employed in the future by using data from the past. They can predict which areas these individuals are educated. Thus, they can develop incentives, subsidiaries for those areas. Moreover, they can invest in education to cover the gap in the human capital that is missing in these areas. Identify common problems in individuals who cannot be employed. For these, they can develop policies aimed at eliminating the problems of re-education, re-training and common problems aimed at increasing their human capital. Based on these estimations, economists can realize their adaptations to develop social policies. Anyone who wants to use human resources effectively may also benefit from these estimates.

References

- Acemoglu, D. (2002). Technical change, inequality, and the labor market. *Journal of economic literature*, 40(1):7-72.
- Acemoglu, D. and Shimer, R. (1999). Holdups and efficiency with search frictions. *International Economic Review*, 40(4):827-849.
- Ashenfelter, O., Harmon, C. and Oosterbeek, H. (1999). A review of estimates of the schooling/earnings relationship, with tests for publication bias. *Labour economics*, 6(4):453-470.

- Becker, G. (1964). *Human capital: a theoretical and empirical analysis, with special reference to education*. National bureau of economic research publications: General series. National Bureau of Economic Research; distributed by Columbia University Press.
- Blundell, R., Dearden, L., Meghir, C. and Sianesi, B. (1999). Human capital investment: the returns from education and training to the individual, the firm and the economy. *Fiscal studies*, 20(1):1–23.
- Bougheas, S. and Georgellis, Y. (2004). Early career mobility and earnings profiles of german apprentices: Theory and empirical evidence. *Labour*, 18(2):233–263.
- Breiman, L. and Spector, P. (1992). Submodel selection and evaluation in regression. The x-random case. *International statistical review/revue internationale de Statistique*, 291–319.
- Chollet, F. et al. (2015). Keras. <https://keras.io>.
- Cingano, F. (2003). Returns to specific skills in industrial districts. *Labour Economics*, 10(2):149–164.
- Dearden, L. (1998). Ability, families, education and earnings in Britain. Technical report, *IFS Working Papers*.
- Donaldson, D. and Eaton, B. C. (1976). Firm-specific human capital: a shared investment or optimal entrapment? *Canadian Journal of Economics*, 462–472.
- Greene, W. H. (2003). *Econometric analysis*. Pearson Education, India.
- Higashi, Y. (2002). Firm specific human capital and unemployment in a growing economy. *Japan and the world economy*, 14(1):35–44.
- İlkkaracan, İ. (2012). Why so few women in the labor market in turkey? *Feminist Economics*, 18(1):1–37.
- James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). *An introduction to statistical learning*, 112. Springer.
- Krueger, A. and Ashenfelter, O. (1992). Estimates of the economic return to schooling from a new sample of twins. Technical report, *National Bureau of Economic Research*.
- Mincer, J. (1974). Schooling, experience, and earnings. *Human Behavior & Social Institutions*, 2.
- Mincer, J. and Ofek, H. (1982). Interrupted work careers: Depreciation and restoration of human capital. *Journal of Human Resources*, 3–24.
- Parsons, D. O. (1972). Specific human capital: An application to quit rates and layoff rates. *Journal of political economy*, 80(6):1120–1143.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Psacharopoulos, G. (1994). Returns to investment in education: A global update. *World Development*, 22(9):1325–1343.
- Rousseeuw, P. J. and Bassett Jr, G. W. (1990). The mediant: A robust averaging method for large data sets. *Journal of the American Statistical Association*, 85(409):97–104.
- Tansel, A. and Tasci, H. M. (2004). Determinants of unemployment duration for men and women in Turkey, *IZA Discussion Paper* no. 1258.
- Tasci, H. M. and Ozdemir, A. R. (2006). Trends in long-term unemployment and determinants of incidence of long-term unemployment in Turkey, *Journal of Economic and Social Research*, 7(2): 1-33.
- Tasci, H. M. and Tansel, A. (2005). Unemployment and transitions in the turkish labor market: evidence from individual level data, *IZA Discussion Paper* no. 1663.

Tunali, İ. and Ercan, H. (2003). Background study on labour market and employment in Turkey. European Training Foundation, Torino (June 2003).

Wasmer, E. (2006). General versus specific skills in labor markets with search frictions and firing costs. *American Economic Review*, 96(3):811–831.

Appendices

Table 5. Descriptive Statistics of Quantitative Features

Variable/Feature	N	Mean	St. Dev.	Min	Max
Age	203,891	32.150	11.160	15	99
tr_coming_year	3,569	15.360	11.580	0	71
living_place	3,569	15.360	11.580	0	71
educ_year	203,891	7.545	3.947	0	15
experience	203,891	14.970	11.650	0	83
deflator	203,891	1.114	0.291	0.681	1.655
weight_hlfs	203,891	160.800	79.950	5.620	754.200
hh_population	203,891	4.616	2.234	1	31
hh_population_estimated	203,891	4.485	2.183	1	31
inactive	52,774	7.486	9.646	0	73

Table 6. Descriptive Statistics of Qualitative Features

gender	age.group	birth_loc	abroad_exp6
Min. :1.00	Min. : 4.00	Min. :1	Min. :1
1st Qu.:1.00	1st Qu.: 5.00	1st Qu.:1	1st Qu.:2
Median :1.00	Median : 7.00	Median :1	Median :2
3rd Qu.:1.00	3rd Qu.: 9.00	3rd Qu.:1	3rd Qu.:2
Max. :2.00	Max. :14.00	Max. :2	Max. :2
		NA's :70381	NA's :100874
relative_type	recent_school_grad_k	foet99_k	nuts1
Min. : 1.00	Min. :0.00	Min. : 1	Min. : 1.00
1st Qu.: 1.00	1st Qu.:2.00	1st Qu.: 6	1st Qu.: 3.00
Median : 3.00	Median :3.00	Median : 6	Median : 6.00
3rd Qu.: 3.00	3rd Qu.:4.00	3rd Qu.:12	3rd Qu.: 9.00
Max. :11.00	Max. :6.00	Max. :21	Max. :12.00
		NA's :171336	
nuts2	prev_residence	prev_living	spouse
Min. : 1.0	Min. :1	Min. :1	Min. : 1
1st Qu.: 5.0	1st Qu.:1	1st Qu.:1	1st Qu.: 2
Median :11.0	Median :1	Median :2	Median :99
3rd Qu.:19.0	3rd Qu.:1	3rd Qu.:3	3rd Qu.:99
Max. :26.0	Max. :2	Max. :3	Max. :99
	NA's :155343	NA's :160119	NA's :13426

Table 6. Descriptive Statistics of Qualitative Features (cont'd)

mother	father	literacy_level	recent_school_grad
Min. : 1 1st Qu.: 2 Median : 3 3rd Qu.:99 Max. :99 NA's :13426	Min. : 1 1st Qu.: 1 Median :99 3rd Qu.:99 Max. :99 NA's :13426	Min. :1 1st Qu.:1 Median :1 3rd Qu.:1 Max. :2 NA's :122214	Min. :1.00 1st Qu.:2.00 Median :2.00 3rd Qu.:2.00 Max. :2.00
contd_school	contd_school_year	course_attendance	course_aim
Min. : 1 1st Qu.: 4 Median : 4 3rd Qu.: 5 Max. :32 NA's :189243	Min. :1 1st Qu.:1 Median :2 3rd Qu.:3 Max. :9 NA's :191939	Min. :1.00 1st Qu.:2.00 Median :2.00 3rd Qu.:2.00 Max. :2.00	Min. :1 1st Qu.:1 Median :1 3rd Qu.:2 Max. :3 NA's :202449
marital_status	job_prev	jobleft_reason	nace2_job_prev
Min. :1 1st Qu.:1 Median :2 3rd Qu.:2 Max. :9 NA's :17532	Min. :1 1st Qu.:1 Median :1 3rd Qu.:2 Max. :2 NA's :112294	Min. : 1 1st Qu.: 1 Median : 3 3rd Qu.: 7 Max. :13 NA's :164341	Min. : 1 1st Qu.: 2 Median : 3 3rd Qu.: 4 Max. :11 NA's :164341
isco08_job_prev	jobprev_status	jobprevear_nace	jobprevear_status
Min. :1 1st Qu.:5 Median :7 3rd Qu.:9 Max. :9 NA's :164341	Min. :1 1st Qu.:1 Median :1 3rd Qu.:2 Max. :5 NA's :164341	Min. : 1 1st Qu.: 1 Median : 3 3rd Qu.: 6 Max. :11 NA's :151684	Min. : 1 1st Qu.: 1 Median : 2 3rd Qu.: 3 Max. :10 NA's :151812

Table 7. Classifier Results

Classifier	Classes	Precision	Recall	F 1 - score	Support	Confusion Matrix	Accuracy	AUC
Gaussian Train	0	0.54	0.48	0.51	93740	44687 49053	0.5401	0.5409
	1	0.54	0.6	0.57	94468	37512 56956		
	Avg/Total	0.54	0.54	0.54	188208			
Gaussian Test	0	0.54	0.47	0.5	7811	3668 4143	0.5357	0.5365
	1	0.53	0.6	0.57	7872	3139 4733		
	Avg/Total	0.54	0.54	0.53	15683			
Bernoulli Train	0	0.61	0.15	0.24	93740	14356 79384	0.5289	0.5340
	1	0.52	0.9	0.66	94468	9282 85186		
	Avg/Total	0.56	0.53	0.45	188208			
Bernoulli Test	0	0.6	0.15	0.25	7811	1208 6603	0.5268	0.5320
	1	0.52	0.9	0.66	7872	818 7054		
	Avg/Total	0.56	0.53	0.45	15683			
Logistic Train	0	0	0	0	93740	0 93740	0.5019	0.5395
	1	0.5	1	0.67	94468	0 94468		
	Avg/Total	0.25	0.5	0.34	188208			
Logistic Test	0	0	0	0	7811	0 7811	0.5019	0.5483
	1	0.5	1	0.67	7872	0 7872		
	Avg/Total	0.25	0.5	0.34	15683			
KNeighbors Train	0	0	0.7	0.82	93740	93740 0	0.7843	0.9082
	1	1	0.57	0.73	94468	40605 53863		
	Avg/Total	0.85	0.78	0.77	188208			
KNeighbors Test	0	0.54	0.78	0.64	7811	6061 1750	OVERFIT	OVERFIT
	1	0.6	0.34	0.43	7872	5201 2671		
	Avg/Total	0.57	0.56	0.53	15683			

Note: Avg. is the average of classifier.

Table 7. Classifier Results (cont'd)

Classifier	Classes	Precision	Recall	F1-score	Support	Confusion Matrix	Accuracy	AUC
Decision Tree	0	1	1	1	93740	93740 0	0.9999	1.000
Train	1	1	1	1	94468 26 94442			
	Avg/ Total	1	1	1	188208			
Decision Tree	0	0.62	0.62	0.62	7811	4830 2981	0.6207	0.6207
Test	1	0.62	0.62	0.62	7872 2968 4904			
	Avg/ Total	0.62	0.62	0.62	15683			
Linear Discriminant	0	0.5	1	0.66	93740	93740 0	0.4981	0.6429
Train	1	0	0	0	94468 94468 0			
	Avg/ Total	0.25	0.5	0.33	188208			
Linear Discriminant	0	0.5	1	0.66	7811	7811 0	0.4981	0.6474
Test	1	0	0	0	7872 7872 0			
	Avg/ Total	0.25	0.5	0.33	15683			
Quadratic Discriminant	0	0.61	0.65	0.63	93740	60771 32969	0.6187	0.6771
Train	1	0.63	0.59	0.61	94468 38795 55673			
	Avg/ Total	0.62	0.62	0.62	188208			
Quadratic Discriminant	0	0.6	0.64	0.62	7811	5010 2801	0.6109	0.6665
Test	1	0.62	0.58	0.6	7872 3301 4571			
	Avg/ Total	0.61	0.61	0.61	15683			
SGD	0	0	0	0	93740	0 93740	0.4981	0.5000
Train	1	0.5	1	0.67	94468 0 94468			
	Avg/ Total	0.25	0.5	0.34	188208			
SGD	0	0	0	0	7811	0 7811	0.4981	0.5000
Test	1	0.5	1	0.67	7872 0 7872			
	Avg/ Total	0.25	0.5	0.34	15683			

Table 7. Classifier Results (cont'd)

Classifier	Classes	Precision	Recall	F1-score	Support	Confusion Matrix	Accuracy	AUC
SVM Train	0	1	0.75	0.85	93740	69922 23818	0.8734	0.9704
	1	0.8	1	0.89	94468	11 94457		
	Avg/Total	0.9	0.87	0.87	188208			
SVM Test	0	0.73	0	0	7811	11 7800	0.5024	0.5046
	1	0.5	1	0.67	7872	4 78.680.000		
	Avg/Total	0.62	0.5	0.34	15683			
Linear SVC Train	0	0.54	0.48	0.51	93740	44686 49054	0.5399	0.5401
	1	0.54	0.6	0.57	94468	37524 56944		
	Avg/Total	0.54	0.54	0.54	188208			
Linear SVC Test	0	0.54	0.47	0.5	7811	3669 4142	0.5365	0.5407
	1	0.53	0.6	0.57	7872	3127 4745		
	Avg/Total	0.54	0.54	0.53	15683			
Neural Networks Train							0.5018	0.5000
Neural Networks Test							0.5019	0.5000