T.C. İSTANBUL TİCARET ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

GİRİŞİM ŞİRKETLERİNİN FİNANSAL BAŞARISIZLIKLARININ YAPAY SİNİR AĞLARI VE LOJİSTİK REGRESYON ANALİZİ İLE TAHMİN EDİLMESİ

Kubilay ERİŞLİK

Danışman
Prof. Dr. Özlem DENİZ BAŞAR

YÜKSEK LİSANS TEZİ
İSTATİSTİK ANABİLİM DALI
İSTANBUL - 2020

KABUL VE ONAY SAYFASI

Kubilay ERİŞLİK tarafından hazırlanan "Girişim Şirketlerinin Finansal Başarısızlıklarının Yapay Sinir Ağları ve Lojistik Regresyon Analizi ile Tahmin Edilmesi " adlı tez çalışması 14/07/2020 tarihinde aşağıdaki jüri üyeleri önünde başarı ile savunularak, İstanbul Ticaret Üniversitesi Fen Bilimleri Enstitüsü **İstatistik Anabilim Dalı**'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Danışman	**Prof. Dr. Özlem Deniz BAŞAR** İstanbul Ticaret Üniversitesi
Jüri Üyesi	**Prof. Dr. Münevver TURANLI** İstanbul Ticaret Üniversitesi
Jüri Üyesi	**Doç. Dr. Serpil KILIÇ DEPREN** Yıldız Teknik Üniversitesi

Onay Tarihi : 24/07/2020

İstanbul Ticaret Üniversitesi, Fen Bilimleri Enstitüsünün 24.07.2020 tarih ve 2020/288 numaralı Yönetim Kurulu Kararının 1. maddesi gereğince, ders yüklerini ve tez yükümlülüğünü yerine getirdiği belirlenen "Kubilay ERİŞLİK" (TC:18523693456) adlı öğrencinin mezun olmasına oy birliği ile karar verilmiştir.

Prof. Dr. Necip ŞİMŞEK
Enstitü Müdürü

AKADEMİK VE ETİK KURALLARA UYGUNLUK BEYANI

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

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

beyan ederim.

24.07.2020



Kubilay ERİŞLİK

İÇİNDEKİLER

	Sayfa
İÇİNDEKİLER	i
ÖZET	iii
ABSTRACT	iv
TEŞEKKÜR	v
ŞEKİLLER DİZİNİ	vi
ÇİZELGELER DİZİNİ	vii
SİMGELER VE KISALTMALAR DİZİNİ	ix
1. GİRİŞ	1
2. LİTERATÜR ÖZETİ	3
3. YAPAY SİNİR AĞLARI	6
3.1. Yapay Sinir Ağları'nın Tanımı ve Tarihçesi	9
3.1.1. Camelot Çağı (1890 - 1969)	10
3.1.2. Karanlık Çağ (1969 - 1982)	11
3.1.3. Rönesans Çağı (1982 - 1986)	11
3.1.4. Bağlantıcılık Çağı (1988 - 1994)	12
3.1.5. Yapay Zeka Çağı (1998 -)	12
3.2. Yapay Sinir Ağları'nın Çalışma Prensipleri	12
3.3. Yapay Sinir Ağları'nın Yapısı ve Elemanları	13
3.3.1. Katmanlar	14
3.3.2. Bağlantılar	14
3.4. Yapay Sinir Ağları'nın Çeşitleri	15
3.4.1. Öğrenme kurallarına göre yapay sinir ağları	15
3.4.1.1. Denetimli öğrenme	15
3.4.1.2. Denetimsiz öğrenme	16
3.4.1.3. Pekiştirmeli öğrenme	16
3.4.2. Ağ yapılarına göre yapay sinir ağları	17
3.4.2.1. İleri beslemeli ağlar	17
3.4.2.1.1. Geri yayılım ağı	18
3.4.2.2. Geri beslemeli ağlar	19
3.5. Yapay Sinir Ağlarının Uygulama Alanları	20
3.6. Yapay Sinir Ağlarının Tahmin Amacı Kullanımı	20
4. LOJİSTİK REGRESYON ANALİZİ	21
4.1. Lojistik Regresyon Analizinin Tarihçesi	23
4.2. Lojistik Regresyon Modelleri	24
4.2.1. İkili lojistik regresyon modeli	25
4.2.2. Sıralı lojistik regresyon modeli	28
4.2.3. İsimsel lojistik regresyon modeli	28
4.3. Çok Değişkenli Lojistik Regresyon Analizi	29
4.4. Uyum İyiliği Testleri	31
4.4.1. Pearson Ki-Kare testi	31
4.4.2. Hosmer – Lemeshow testi	31
4.4.3. Sınıflandırma tablosu	32

5. UYGULAMA 33
5.1. Araştırmanın Metodolojisi 33
5.1.1. Araştırmanın amacı ve önemi 33
5.1.2. Araştırmanın kapsamı ve kısıtları 34
5.1.2.1. Girişim şirketleri 34
5.1.2.2. Girişim şirketlerinin yatırım aşamaları 35
5.1.2.3. Girişim şirketlerinin başarısız olması 36
5.1.2.4. Araştırmaya dahil edilen sektörlerin belirlenmesi 37
5.2. Araştırmanın Yöntemi 39
5.2.1. Araştırmanın verileri 40
5.3. Bulguların Değerlendirilmesi 42
5.3.1. Tanımlayıcı istatistikler 43
5.3.2. İkili lojistik regresyon analizinin uygulanması 49
5.3.3. Yapay sinir ağlarının uygulanması 53
5.3.4. Bulguların karşılaştırılması 60
6. SONUÇ VE ÖNERİLER 63
KAYNAKLAR 67
EKLER 71
EK A. Lojistik Regresyon Analizi Kodları 71
EK A. Yapay Sinir Ağları Kodları 73
ÖZGEÇMİŞ 74

ÖZET

Yüksek Lisans Tezi

GİRİŞİM ŞİRKETLERİNİN FİNANSAL BAŞARISIZLIKLARININ YAPAY SİNİR AĞLARI VE LOJİSTİK REGRESYON ANALİZİ İLE TAHMİN EDİLMESİ

Kubilay ERİŞLİK

İstanbul Ticaret Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Prof. Dr. Özlem DENİZ BAŞAR

2020, 74 sayfa

Girişim şirketleri üzerine yapılan bu araştırmanın amacı 2010 – 2015 yılları arasında melek, çekirdek ve girişim evrelerinde yatırım almış olan girişim şirketlerinin günümüzde başarılı ya da başarısız olduğunun tespit etmektir. Ayrıca yatırım aşamalarında alınan yatırımlar ve sektörler ile oluşturulan lojistik regresyon analizi modeli ve yapay sinir ağları modeli ile girişim şirketlerinin başarılı bir şekilde sınıflandırılıp sınıflandırılamayacağı ve bu modellerin yeni kurulan girişim şirketleri için ve yatırım alacak girişim şirketleri için ön analiz olarak kullanabilir olup olmadığı üzerinde durulmuştur.

Çalışma kapsamında bir takım istatistiksel analizler yapılarak değişkenler arasındaki ilişkiler ortaya konulmuştur. Bununla birlikte girişim şirketlerinin finansal açıdan başarılı olup olmayacağının tahmin etme çalışmaları gerçekleştirilmiştir. Tahminleme yapmak için "İkili Lojistik Regresyon Analizi" ve "İleri Beslemeli Geri Yayılımlı Yapay Sinir Ağları" olmak üzere iki farklı metot kullanılmıştır ve 1675 girişim şirketi verileri üzerinde tahmin işlemi gerçekleştirilmiştir. Çalışma neticesinde ileri beslemeli geri yayılımlı yapay sinir ağı yöntemi ile daha başarılı bir sonuç elde edilmiştir.

Anahtar Kelimeler: Finansal başarısızlık, girişim şirketleri, lojistik regresyon analizi, yapay sinir ağları, yatırım aşamaları.

ABSTRACT

M.Sc. Thesis

PREDICTION OF FINANCIAL FAILURES OF STARTUPS BY ARTIFICIAL NEURAL NETWORKS AND LOGISTIC REGRESSION ANALYSIS

Kubilay ERİŞLİK

İstanbul Commerce University
Graduate School of Applied and Natural Sciences
Department of Statistics

Supervisor: Prof. Dr. Özlem DENİZ BAŞAR

2020, 74 pages

The aim of this research on startup companies is to determine whether the startup companies that received investment in angel, seed and venture phases between 2010 and 2015 are successful or unsuccessful today. In addition, it is emphasized whether the startup companies can be successfully classified with the logistic regression analysis model and artificial neural networks model created by the investments in the companies in their sectors and investment stages and whether these models can be used as a preliminary analysis for newly established startup companies.

Within the scope of the study, statistical analyzes were used and the relationships between the variables were determined. In addition, pretiction were made on whether startup companies would be financially successful or not. Two different methods were used to make prediction, namely "Binary Logistic Regression Analysis" and "Feed Forward Back Propagation Artificial Neural Networks" and estimation was performed on 1675 venture company data. As a result of the study, a more successful result was obtained with the feed forward back propagation artificial neural network method.

Keywords: Artificial neural networks, financial failure, investment stages, logistic regression analysis, startups.

TEŞEKKÜR

Bu araştırma için beni yönlendiren, karşılaştığım zorlukları bilgi ve tecrübesi ile aşmamda yardımcı olan değerli Danışman Hocam Prof. Dr. Özlem DENİZ BAŞAR'a teşekkürlerimi sunarım.

Hem eğitim hayatımda hem de iş hayatımda bana her konuda destek olan değerli hocalarım Prof. Dr. Münevver TURANLI, Prof. Dr. Ünal Halit ÖZDEN ve Doç. Dr. Seda BAĞDATLI KALKAN'a teşekkür ederim.

Araştırmanın yürütülmesinde maddi ve manevi yardımlarını gördüğüm Akif ÇİÇEK olmak üzere tüm Kuramsoft Yazılım Bilgisayar Güvenlik Ltd. Şti. personeline teşekkür ederim.

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

Kubilay ERİŞLİK
İSTANBUL, 2020

ŞEKİLLER

Sayfa

Şekil 3.1. Çok katmanlı bir yapay sinir ağı 14
Şekil 3.2. Hopfield Ağı 19
Şekil 4.1. Lojistik fonksiyonu 26
Şekil 5.1. Alınan melek yatırım miktarı ve başarı durumuna ait kutu grafiği..... 49
Şekil 5.2. Alınan çekirdek yatırım miktarı ve başarı durumuna ait kutu grafiği..... 50
Şekil 5.3. Alınan risk yatırım miktarı ve başarı durumuna ait kutu grafiği .. 50
Şekil 5.4. %70 - %30 için en iyi yapay sinir ağı yapısı grafiği 56
Şekil 5.5. %80 - %20 için en iyi yapay sinir ağı yapısı grafiği 57

ÇİZELGELER

Sayfa

Çizelge 2.1. Literatür özeti 3
Çizelge 3.1. Geleneksel yöntemler ve yapay sinir ağlarının bazı özellikleri 8
Çizelge 4.1. Bağımlı ve bağımsız değişkenlerin türüne göre analiz yöntemleri 22
Çizelge 5.1. Tahmini örneklem büyüklüğü 41
Çizelge 5.2. Bağımlı değişkene ait bilgiler 42
Çizelge 5.3. Bağımsız değişkenlere ait bilgiler 42
Çizelge 5.4. Şirketlerin yatırım aşamasında aldıkları yatırımlara ilişkin istatistikler 43
Çizelge 5.5. Bulut hizmetlerinde bulunan şirketlerin yatırım aşamasında aldıkları yatırımlara ilişkin tanımlayıcı istatistikler 43
Çizelge 5.6. Büyük veri ve makine öğrenmesinde bulunan şirketlerin yatırım aşamasında aldıkları yatırımlara ilişkin tanımlayıcı istatistikler 44
Çizelge 5.7. E-Ticaret sektöründe bulunan şirketlerin yatırım aşamasında aldıkları yatırımlara ilişkin tanımlayıcı istatistikler 45
Çizelge 5.8. Mobil uygulama sektöründe bulunan şirketlerin yatırım aşamasında aldıkları yatırımlara ilişkin temel istatistikler 45
Çizelge 5.9. Sosyal medya sektöründe bulunan şirketlerin yatırım aşamasında aldıkları yatırımlara ilişkin tanımlayıcı istatistikler 46
Çizelge 5.10. Örneklemde bulunan girişim şirketlerinin başarı durumları 46
Çizelge 5.11. Bulut hizmetleri sektöründe bulunan girişim şirketlerinin başarı durumları 47
Çizelge 5.12. Büyük veri ve makine öğrenmesi sektöründe bulunan girişim şirketlerinin başarı durumları 47
Çizelge 5.13. E-Ticaret sektöründe bulunan girişim şirketlerinin başarı durumları 47
Çizelge 5.14. Mobil uygulama sektöründe bulunan girişim şirketlerinin başarı durumları 48
Çizelge 5.15. Sosyal medya sektöründe bulunan girişim şirketlerinin başarı durumları 48
Çizelge 5.16. İkili lojistik regresyon modeline ait katsayılar 51
Çizelge 5.17. İkili lojistik regresyon modeline ait katsayılar 52
Çizelge 5.18. Modellere ait AIC değerleri 52
Çizelge 5.19. Model 1'e ait sınıflandırma tablosu 52
Çizelge 5.20. Model 2'ye ait sınıflandırma tablosu 53
Çizelge 5.21. Modellere göre eğitim setinin hata kareler ortalamaları 54
Çizelge 5.22. %70 - %30 için en iyi yapay sinir ağı yapısı 55
Çizelge 5.23. %80 - %20 için en iyi yapay sinir ağı yapısı 56
Çizelge 5.24. %70 - %30 Eğitim seti performans sonuçları 57

Çizelge 5.25. %70 - %30 Test seti performans sonuçları 58
Çizelge 5.26. %80 - %20 Eğitim seti performans sonuçları 59
Çizelge 5.27. %80 - %20 Test seti performans sonuçları 59
Çizelge 5.28. Lojistik regresyon analizi performans tablosu 61
Çizelge 5.29. Yapay sinir ağları performans sonuçları..... 61

SİMGELER VE KISALTMALAR

LRA : Lojistik Regresyon Analizi
YSA : Yapay Sinir Ağları
$: Amerikan Doları

1. GİRİŞ

Ekonomik, sosyal, kültürel ve siyasal yaşamın lokomotifi olan girişimcilik kavramı, sanayi toplumundan bilgi toplumuna geçiş ile birlikte daha büyük önem kazanmaya başlamış, orta çağdan günümüze kadar pek çok teorisyenin tanımlamaya çalıştığı bir kavram olmuştur. Girişimcilik, hem işletme, hem de iktisat literatüründe en fazla ilgi çeken konulardan birisi olarak araştırılmaya devam etmektedir (Semra Güney, 2019).

Bu çalışmanın amacı; girişim şirketlerinin yatırım aşamalarında aldıkları yatırımlar ile şirketlerin başarılı olup olmayacağı arasındaki ilişkiyi tespit etmek, girişim şirketlerinin bulunduğu sektörlerin başarıya etkisi incelenmektir. Ayrıca yatırım aşamalarında alınan yatırımlar ve sektörler ile oluşturulan lojistik regresyon analizi modeli ve yapay sinir ağları modeli ile girişim şirketlerinin başarılı bir şekilde sınıflandırılıp sınıflandırılamayacağı ve bu modellerin yeni kurulan girişim şirketleri ve yatırım alacak girişim şirketleri için ön analiz olarak kullanabilir olup olmadığı üzerinde durulmuştur.

Girişimcilik bireysel ve toplumsal anlamda bir refah yarattığı için pek çok araştırmanın ilgi odağı olmuştur. Girişimcilik alanında yapılan araştırmalar şu tür sorulara yanıt aramıştır: Neden bazı insanlar yeni fırsatları görme ya da yaratma yeteneğine sahipken diğerlerinin bu tür bir özelliği yoktur? Neden bazı kişiler iş fikirlerini ya da hayallerini gerçek bir işletmeye dönüştürürken diğerleri bunu başaramaz. Neden bazı girişimciler başarılı iken bazıları başarısızdır (Baron, 1998; Semra Güney, 2019).

Bu çalışmanın amacı da yukarıda belirtilen sorulardan bazılarına girişim şirketlerinin aldıkları yatırımlar ve bulundukları sektörler ile bu sorulara yanıt aramaktır.

Çalışmanın ikinci bölümünde; yapay sinir ağları, literatür, kavramsal ve kuramsal açıdan incelenmiştir. Yapay sinir ağları tarihçesi ve yöntemleri detaylı bir şekilde

açıklanmıştır. Çalışmada kullanılan yapay sinir ağı modeli ayrıntılı bir şekilde ele alınmıştır.

Çalışmanın üçüncü bölümünde; lojistik regresyon analizi yöntemi detaylı bir şekilde açıklanmıştır. Bu bölümde lojistik regresyon analizi kuramsal açıdan incelenmiştir. Çalışmada kullanılacak ikili lojistik regresyon analizi ayrıntılı bir şekilde ele alınmıştır. Performans karşılaştırması için kullanılacak yöntemler kısaca anlatılmıştır.

Çalışmanın dördüncü bölümünde; araştırmanın amacı, araştırmanın metodolojisi ve araştırmada kullanılan veri gibi konular ele alınmıştır. Girişim şirketleri, girişim şirketlerinin başarısız olma durumu ve yatırım aşamaları gibi kavramlar detaylıca açıklanmıştır. Çalışmada yapılan analizlere ait bulgular değerlendirilmiştir. 2010 – 2015 yılları arasında yatırım aşamalarının tamamında yatırım alan ve bulut hizmetleri, büyük veri ve makine öğrenmesi, e-ticaret, mobil uygulama ve sosyal medya sektöründe bulunan 1675 girişim şirketi analize dahil edilmiştir. Araştırmada elde edilen veriler R programı kullanılarak analiz edilmiştir.

2010 – 2015 yılları arasında belirtilen sektörlerde bulunan girişim şirketlerine ait yatırım aşamalarında aldıkları yatırım miktarı Crunchbase web site üzerinden alınmıştır ve analize dahil edilmiştir. Elde edilen verilere ileri beslemeli geri yayımlı yapay sinir ağı modelleri uygulanmış ve en iyi model seçilerek çalışmada kullanılmıştır. Elde edilen veriler aynı zamanda ikili lojistik regresyon analizi ile analiz edilmiştir. İkili lojistik regresyon modeli ile elde edilen sınıflandırma işlemi ve yapay sinir ağları modeli ile elde edilen sınıflandırma işlemi bulguları karşılaştırılmıştır.

Çalışmanın beşinci bölümünde; elde edilen bulgular yorumlanmış ve tahminleme işlemine ait sonuçlar incelenmiştir. Çalışmada kullanılan yöntemler kendi aralarında kıyaslanmış ve hangisinin neden daha iyi sonuç verdiği açıklanmıştır.

2. LİTERATÜR ÖZETİ

Lojistik regresyon analizine ait ilk çalışmalar 19. yüzyılın ortalarında başlamıştır. Ancak 1950'li yıllardan sonra daha sık literatür çalışmalarında kullanılmaya başlanmıştır. Yapay sinir ağlarına ait çalışmalar ise 1900'lü yılların başlarında başlamasına rağmen 1990 yılından sonra teknolojik gelişmeler sayesinde popüler hale gelerek literatür çalışmaları gerçekleştirilmiştir. 1990'lı yıllardan sonra kategorik verilerin analiz edilmesinde ve sınıflandırılmasında iki yöntemde sıklıkla kullanılmıştır. Literatür incelendiğinde sıklıkla yöntemlerin aynı veriler üzerindeki başarılarının karşılaştırıldığı tespit edilmiştir. Genellikle her iki yöntem de sağlık ve ekonomi alanlarında tercih edilmiştir. Lojistik regresyon analizinin ve yapay sinir ağlarının kullanıldığı literatür çalışmaları Çizelge 2.1'de tablolaştırılarak verilmiştir.

Çizelge 2.1. Literatür özeti

Yazar	Tarih	Çalışmanın Adı	Açıklama
King, M.Y., Tam, K.Y.	1992	Managerial applications of neural networks: The case of bank failure predictions	Çalışmada, banka varsayılan verileri kullanılarak yapay sinir ağları, lojistik regresyon analizi KNN ve ID3 yöntemleri karşılaştırılmıştır. Çalışmada yapay sinir ağlarının modelleme amaçlı kullanılırlığı tartışılmaktadır.
Dasgupta, C.G., Dispensa, G.S., Ghose, S.	1994	Comparing the predictive performance of a neural network model with some traditional market response models	Çalışmada, lojistik regresyon modeli, diskriminant analizi ve geri yayılımlı yapay sinir ağı modellerinin performansları karşılaştırılmış ve finansal risk alma ve tüketici segmentlerini belirlemede kullanılmıştır.
Tu, J.V.	1996	Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes.	Çalışmada, yapay sinir ağları ve lojistik regresyon analizi tıpta ikili sonuç öngörücü modeller olarak kullanılmış ve avantajarı ve dezavantajları belirtilmiştir.
Dreiseiti, S., Ohno-Machado, L.	2002	Logistic regression and artificial neural network classi fication models: A methodology review	Derleme çalışmada, lojistik regresyon analizi ve yapay sinir ağları modellerinin farklılıkları ve benzerlikleri teknik açıdan özetlenmiştir.

Baesens, B., Suykens, J., Stepanova, M., Van Gestel, T., Vanthienen, J. Viaene, S.,	2003	Benchmarking state-of-the-art classification algorithms for credit scoring	Çalışmada, lojistik regresyon analizi, diskriminant analizi, KNN, yapay sinir ağları ve karar ağaçları sekiz gerçek kredi puanlama veri seti üzerinde kullanılmış ve performansları karşılaştırılmıştır.
Liou, F.	2008	Fraudulent financial reporting detection and business failure prediction models: A comparison	Çalışmada, lojistik regresyon analizi, yapay sinir ağaçları ve sınıflandırma ağaçları uygulanarak hileli finansal raporlamalar tespit edilmiş ve iş başarısızlıkları sınıflandırılması tahmin edilmiştir.
Kumar, U.A Paliwal, M.,	2009	Neural networks and statistical techniques: A review of applications	Çalışmada, yapay sinir ağları ve tahmin, sınıflandırma alanında kullanılan istatistiksel tekniklerin karşılaştırıldığı çalışmalar incelenmiştir.
Çelik, M.	2010	Bankaların Finansal Başarısızlıklarının Geleneksel ve Yeni Yöntemlerle Öngörüsü	Çalışmada, bankaların finansal başarısızlıklarının öngörülebilmesi için diskriminant analizi ve yapay sinir ağları ile erken uyarı modelleri oluşturulmuştur.
Perols, J.	2011	Financial statement fraud detection: An analysis of statistical and machine learning algorithms	Çalışmada, finansal tablo sahtekarlığının tespiti için altı farklı istatistik yöntemi ve makine öğrenmesi yöntemleri karşılaştırılmıştır.
Budak, H., Erpolat. S.	2012	Kredi Riski Tahmininde Yapay Sinir Ağları ve Lojistik Regresyon Analizi Karşılaştırılması	Çalışmada, bankaların kredi risklerini öngörmelerine yardımcı olması amacıyla, kredi talep eden müşterilerin ödeme alışkanlıklarının düzenli olup olmayacağının tahmin edilmesi için yapay sinir ağları ve lojistik regresyon analizi kullanılmıştır.
Altunöz, U.	2013	Bankaların Finansal Başarısızlıklarının Diskriminant Analizi ve Yapay Sinir Ağları Çerçevisinde Tahmini	Çalışmada, banka başarısızlıklarının önceden tespitinde kullanılan dikriminant analizi ve yapay sinir ağları modeli kullanılmıştır.
Altınırmak, S., Karamaşa, Ç.	2014	Comparison Of Machine Learning Techniques For Analyzing Banks' Financial Distress	Çalışmada, bankalarda meydana gelecek finansal başarısızlıkların önceden tahmin edilebilmesi için destek vektör makineleri ve yapay sinir ağları erken uyarı sistemi olarak kullanılmıştır.
Kaygın, C. Y., Tazegül, A., Yazakan, H.	2016	Estimation Capability of Financial Failures and Successes of Enterprises Using Data Mining and Logistic Regression Analysis	Çalışmada, Borsa İstanbul'da 2010-2013 döneminde kesintisiz olarak işlem gören 143 imalat sanayi şirketinin yıllık bilanço ve gelir tablosu verilerinden yararlanılarak işletmelerin finansal başarılı ve başarısız olma durumları veri madenciliği ve lojistik regresyon ile tahmin edilmeye çalışılmıştır.

Kuzu, S.	2017	Yapay Sinir Ağı Modeli Yöntemiyle İşletme Başarısının Tahmin Edilmesi: 2008 Küresel Finansal Kriz Örneği	Çalışmada, Borsa İstanbul da (BİST) İmalat sektörlerinde faaliyet gösteren çeşitli endüstri işletmelerinin çeşitli finansal göstergeleri temel alarak yapay sinir ağları modeli oluşturulmuştur.
Ayan, T., Değirmenci, N.	2018	Firma Finansal Başarısızlık Öngörüsü İçin Bir Lojistik Regresyon Modeli	Çalışmada, Borsa İstanbul Sanayi Endeksinde faaliyet gösteren firmaların finansal başarısızlıklarının göstergesi olabilecek oranları belirlemek ve bu oranlar yardımıyla başarısızlığı öngörebilmek için çeşitli lojistik regresyon modelleri geliştirmektir.

3. YAPAY SİNİR AĞLARI

Günümüzde toplanan verilerin artması ile büyük veri kavramı daha sık kullanılmaya başlandı. Elde edilen bu veriler ile istatistiksel analiz gerçekleştirmek zorlaştı. Bu nedenden dolayı daha yaygın olarak yapay zeka, makine öğrenmesi ve yapay sinir ağları gibi kavramlar kullanılmaya başlandı (Russel, 2019).

Makine öğrenmesi modern toplumda web aramalarından, sosyal alanda içerik filtrelemelerine, e-ticaret web sitelerinde size önerilen ürünler gibi aklınıza gelebilecek birçok alanda makine öğrenmesi kullanılmaktadır. Makine öğrenmesi mevcut verilerden bir model veya algoritma oluşturarak yeni verilerin tahmin edilmesi işlemidir. Yapay sinir ağları, makine öğrenmesinin eğitim kısmında kullanılan yöntemlerden bir tanesidir. Yapay sinir ağları, insan beyninin özelliklerinden olan öğrenme yolu ile yeni bilgiler üretebilme gibi kabiliyetini herhangi bir yardım olmaksızın otomatik olarak ortaya çıkartma niyetiyle geliştirilen bilgisayar sistemleridir (Öztemel, 2012).

Yapay sinir ağları birçok çalışmada ve medyada insan beyni gibi çalışan bir bilgisayar sistemi olarak gösterilmektedir. Ancak yapay sinir ağları beyin sistemi ile karşılaştırıldığında asıl fark ortaya çıkmaktadır. Günümüzde oluşturulan en gelişmiş yapay sinir ağlarında en fazla birkaç milyon nöron bulunmaktadır. Ancak bu sayı insan beyni için 85 milyar nörondur. Bir diğer önemli fark ise insan beynindeki nöronlar yakınlarındaki birkaç nörona bağlanırken, yapay sinir ağlarında bir katmandaki nöron farklı bir katmandaki nöronlara bağlanabilir. Yapay sinir ağlarında bilgi çoğunlukla tek taraflı olarak iletilir.

Yapay sinir ağları, psikoloji, matematik, nörobilim, fizik, mühendislik, bilgisayar bilimi, felsefe ve dil bilimi gibi farklı disiplinlere ait çalışmalarda kullanılmaktadır.

Yapay sinir ağları bazı ayırt edici özellikleri sayesinde tahmin etme işlemlerinde kullanılan diğer yöntemlere göre daha değerlidir. Bu özelliklerden birincisi,

geleneksel model tabanlı yöntemlerin aksine yapay sinir ağları varsayıma değil veriye dayalı kendinden uyarlamalı bir yöntemdir. Yapay sinir ağları örneklerden öğrenirler ve tanımlaması zor olsa da veriler arasındaki ince fonksiyonel ilişkileri yakalarlar. Bu anlamda yapay sinir ağları çok değişkenli doğrusal olmayan yöntemlerden biri olarak ele alınabilir (Cheng ve Titterington, 1994; Hornik vd., 1989; Ripley, 1994).

Yapay sinir ağlarının ikinci özelliği veride bulunan eksik gözlemleri doğru bir şekilde tamamlayabilmesidir. Aynı zamanda bu sayede geçmiş gözlemlerden yararlanarak gelecek tahmini yapılmasını mümkün kılar. Yapay sinir ağlarının üçüncü özelliği geleneksel istatistiksel yöntemlerden daha genel ve esnek işlevsel biçimlere sahip olmasıdır. Geleneksel istatistiksel tahmin modellerinde, girdiler ile çıktılar arasında temel bir ilişki olduğu varsayılır. Gerçek hayattaki problemlerin karmaşıklığında çoğunlukla bu varsayımlar sağlanamaz. Bundan dolayı yapay sinir ağları tahmin işlemleri için iyi bir alternatif olabilirler.

Yapay sinir ağlarının yakın gelecekte çok daha fazla başarıya sahip olacağı düşünülmektedir, çünkü çok daha az mühendislik ve hesaplama kullanılmaktadır. Günümüzde toplanan veri miktarının artış hızı düşünüldüğünde yapay sinir ağlarının ve öğrenme algoritmalarının ilerlemesinin de benzer oranda olacağıdır. Bu sayede yapay sinir ağları diğer analizlere göre daha doğru ve hızlı sonuçlar vereceklerdir.

Yapay sinir ağları model seçimi ve sınıflandırması, işlev tahmini, en uygun değeri bulma ve veri sınıflandırması gibi işlerde başarılıdır. Geleneksel yöntemler ise özellikle model seçme işinde verimsizdir ve sadece algoritmaya dayalı hesaplama işlemleri ile kesin aritmetik işlemlerde hızlıdır (Elmas, 2011).

Geleneksel yöntemler açıkça tanımlanmış kurallara ve denklemlere dayanır. Yapay sinir ağları ise sisteme giren verilere göre en uygun kuralı veya fonksiyonu kullanır. Yapay sinir ağlarının geleneksel yöntemlere göre en çok eleştiri aldığı nokta, kural tabanlı sistemlerin aksine bulduğu sonuçların doğruluğunu kanıtlayabilecek önemli açıklamaları gerçekleştirememesidir (Swingler, 2001).

Geleneksel yöntemler ve yapay sinir ağları herhangi bir sorunu çözmek için farklı yaklaşımları kullanırlar. Geleneksel yöntemler belirli bir problemi çözmek için hazır algoritmaları kullanmaktadır. Geleneksel yöntemler hangi adımları izleyeceği önceden biliniyorsa faydalı olur ve problemi çözebilir. Yapay sinir ağları ise daha önceden karşılaştığı veriler ile kendi kendini eğiterek problemleri çözmeye çalışır. Buradaki en önemli husus yapay sinir ağlarını eğitmek için kullanılan veri setidir. Veri seti ne kadar probleme benzer durumları içeriyorsa o kadar iyi eğitilir ve daha doğru sonuçlar verir (Russel, 2019).

Yapay sinir ağları, geleneksel yöntemlerden çok farklı özellikler gösterirler Çizelge 3.1'de geleneksel yöntemler ve yapay sinir ağları bazı özellikleri verilmiştir.

Çizelge 3.1. Geleneksel yöntemler ve yapay sinir ağlarının bazı özellikleri (Elmas, 2011)

Geleneksel Yöntemler	**Yapay Sinir Ağları**
Tümdengelimli usavurma: Çıkış üretmek için giriş bilgilerine bilinen kurallar uygulanır.	Tümevarımlı usavurma: Giriş ve çıkış bilgileri (eğitilen örnekler) verilir, kuralları kullanıcı koyar.
Hesaplama merkezi, eş zamanlı ve ardışıktır.	Hesaplama toplu, eş zamansız ve öğrenmeden sonra paraleldir.
Bellek paketlenmiş, hazır bilgi depolanmış ve yer adreslenebilir.	Bellek ayrılmıştır, dahilidir ve içerik adreslenebilir.
Hata toleransı yoktur.	Eğer bilgi, gürültülü veya kısmi ise kurallar bilinmiyorsa ya da karışıksa hata toleransı uygulanabilir.
Hızlıdır.	Yavaştır.
Bilgiler ve Algoritmalar kesindir.	Yapay sinir sistemleri deneyimden yararlanır.

Çok katmanlı yapay sinir ağları, örüntü sınıflandırma, örüntü işleme ve fonksiyon yaklaşımı için kullanılabilir. Bir Gauss veya sigmoid fonksiyonu gibi sürekli farklılaşabilen bir aktivasyon fonksiyonu ekleyerek, ağın istenilen doğruluk derecesinde pratik olarak öğrenme gerçekleştirmesi sağlanabilir (White, 1989).

Çok katmanlı yapay sinir ağları çoğunlukla satır sütun formatında oluşan büyük tabloları analiz etmede kullanılır. Yapay sinir ağlarında sınıflandırma işlemi için

girdi ilk önce doğrusal olmayan bir dönüşüm formülü kullanılarak değiştirilir. Dönüştürülen bu girdi verileri doğrusal olarak ayrılabileceği bir alana yerleştirilir. Bu işlemler genellikle ağın gizli katmanlarından birinde gerçekleşir.

3.1. Yapay Sinir Ağları'nın Tanımı ve Tarihçesi

Yapay sinir ağları, sinir hücrelerinin birbirine çeşitli şekillerde bağlanmasından oluşur. Hücrelerin bağlantı şekillerine, öğrenme kurallarına ve transfer fonksiyonlarına göre birçok yapay sinir ağı yapısı geliştirilmiştir (Arıkan Kargı, 2015). Elmas'a göre yapay sinir ağları bir algoritmaya ihtiyaç duymadan programcının geleneksel yeteneklerini gerektirmeyen, kendi kendine öğrenme düzenekledir. Bu ağlar öğrenmenin yanı sıra, ezberleme ve bilgiler arasında ilişkiler oluşturma yeteneğine sahiptir.

Yapay sinir ağları, sinir ağları, bağlantıcılık, bağdaşır sistem, uyarlanabilir ağlar, nöral bilgisayarlar ve paralel dağıtımlı işleme olarak da adlandırılmaktadır. Yapay sinir ağları teorisi, psikoloji, matematik nörobilim, fizik, mühendislik, bilgisayar bilimi, felsefe ve dilbilimi gibi birçok disiplinden türetilmiştir (Simpson, 1990).

Swingler'a göre yapay sinir ağları, gerçek dünya sisteminde ağırlık adı verilen parametrelerin ayarlanması ile oluşturulan istatistiksel modellerdir. Eberhart ve Shi'ye göre yapay sinir ağları, girdileri kabul edip onlardan çıktı üreten bir kara kutu olarak tanımlanmıştır.

İlk yapay sinir ağ modeli 1943 yılında, bir sinir hekimi olan Warren McCulloch ile bir matematikçi olan Walter Pitts tarafından gerçekleştirilmiştir (Elmas, 2011).
1959 yılında, Bernard Widrow ve Marcian Hoff (Stanford Üniversitesinde) ADALINE ve MADALINE diye adlandırdıkları ağ modellerini oluşturmuşlardır. MADALINE telefon hatlarında oluşan yankıları yok eden bir uyarlanabilir süzgeç olarak kullanılmış ve gerçek dünya sorunlarına uygulanmış olan ilk yapay sinir ağıdır ve hala kullanımda bulunmaktadır (Elmas, 2011).

Yapay sinir ağları tarihsel olarak beş çağa ayrılmıştır. İlk çağ 1890 yılında William James ile başlamış ve bu döneme Camelot Çağı adı verilmiştir. Camelot Çağı 1969 yılına Misky ve Papert'ın yayınladığı Algılayıcılar (Perceptrons) adlı kitapla sona ermiştir. Camelot çağından sonra 1696 yılı ile 1982 yılları arasında Karanlık Çağ yaşanmıştır. Hopfield ile yapay sinir ağları Rönesans Çağına girmiştir ancak Rönesans Çağı Rumelhart ve McClellan'ın yayınladığı kitap ile 1986 yılında son bulmuştur. Cowan ve Sharp'ın 1988 yılında yayınladıkları makale ile Bağlantıcılık Çağı başlamıştır. 1998 yılında gerçekleştirilen IEEE Dünya Yapay Zeka Kongresi ile Yapay Zeka Çağı başlamış ve günümüzde hala devam etmektedir (Eberhart ve Shi, 2007).

3.1.1. Camelot Çağı (1890 - 1969)

Yapay sinir ağlarının ilk adımı ünlü psikolog William James tarafından atılmıştır. James beyinin yapısı ve işlevi ile ilgili, ilişkisel öğrenme ve çağrışımsal bellek gibi birtakım gerçekleri yayınlayan ilk kişi oldu. James sinirsel bir faaliyetin, girdilerin toplamının bir fonksiyonu olduğunu öne sürdü ve ara bağlantılar arasında ağırlıkların bulunduğunu dile getirdi (Eberhart ve Shi, 2007).

Yarım asırdan fazla bir süre sonra 1943 yılında McCulloch ve Pitts, 1940'larda biyolojik yapılar hakkında bilinenlere dayanarak nöronal sistem modelleri ile ilgili teoremleri türettikleri sinir ağı makalelerini yayınladılar. David O Hebb 1949 yılında yayınladığı Davranış Organizasyonu (The Organization of Behavior) kitabında sinaptik ağları güncelleme yöntemini tanımlamış ve Camelot Çağına önemli bir katkı sağlamıştır (Eberhart ve Shi, 2007).

1958 yılında Frank Rosenblatt yapay sinir ağları için bir dönüm noktası olan algılayıcıları tanımlamıştır. Algılayıcılar bilgisayarda simüle edildiği için muhtemelen geçerli olan ilk yapay sinir ağı uygulamasıdır. 1960 yılında Bernard Widrow ve Marcian Hoff Uyarlamalı Anahtarlama Devreleri (Adaptive Switching Circuits) başlıklı makale yayınlamıştır. Bu makale özellikle mühendislik açısından yapay sinir ağları teknolojileri için en önemli belgelerden biri olmuştur (Eberhart ve Shi, 2007).

3.1.2. Karanlık Çağ (1969 - 1982)

1969 yılında Marvin Minsky ve Seymour Papert yapay sinir ağları topluluğuna bomba gibi düşecek Algılayıcılar adlı bir kitap yayınladı. Algılayıcılar gizli katmanı bulunmayan giriş ve çıkış katmanlarının bulunduğu basit bir yapay sinir ağı modelidir. Aynı yıllarda Stephen Grossberg ve Gail Carpenter Uyarlanabilir Rezonans Teorisi adlı ağ yapısını geliştirdi (Eberhart ve Shi, 2007).

1972 yılında dünyanın iki farklı kıtasında bulunan iki isim aynı ağ modelini geliştirdi. Teuvo Kohonen modeline Çağrışımsal Bellek adını verirken James Anderson ise Etkileşimli Bellek adını vermiştir. İki ağ yapısı da birbirine yakın algoritmaları ve transfer fonksiyonlarını kullanarak benzer sonuçları vermektedir. Aynı yıllarda Kohonen İşleme Elemanı adlı bir makale yayınladı. Bu makalede Kohone, McCulloch-Pitts ve Widrow-Hoff gibi ikili değerlere sahip bir model kullanmak yerine doğrusal ve sürekli değerler kullandı. 1979 yılında Kunihiko Fukushima Neocognitron adını verdiği bir ağ yapısı ve modeli geliştirdi. Neocognitron görsel örüntü tanıma mekanizması için geliştirilmiş biyolojik olasılıkları dayanan ilk model olarak literatüre geçmiştir (Eberhart ve Shi, 2007).

3.1.3. Rönesans Çağı (1982 – 1986)

1982 – 1986 yılları arasında yapay sinir ağları araştırmalarının durumu önemli bir ölçüde artıran birçok yayın ortaya çıkmıştır. Bu dönemde grup olarak ya da bireysel olarak birçok araştırmacı yapay sinir ağlarına katkı sağlamıştır. Bu araştırmacılara Rönesans Adamları denmektedir (Eberhart ve Shi, 2007).

3.1.4. Bağlantıcılık Çağı (1988 – 1994)

Yapay sinir ağlarının geliştirilmesi ve kişisel bilgisayarlarda sinir ağlarının uygulanabilmesi ile Cowan ve Sharp'ın Bağlantıcılık olarak adlandırdığı çağa girildi. 1987 yılında ilk defa yapay sinir ağları ile ilgili Uluslararası Yapay Sinir Ağları Konferansı olarak adlandırılan bir konferans düzenlendi. Bu konferans

sayesinde yapay sinir ağları ile ilgili birçok paradigma ve yüzlerce varyasyon literatüre kazandırıldı. 1987 yılından sonra yapay sinir ağlarının uygulama alanları genişledi. Biyoloji ve psikolojinin yanı sıra biyomedikal dalga formu sınıflandırması, müzik kompozisyonu ve emtia fiyatlarının tahmini gibi çeşitli alanlarda yapay sinir ağı kullanılmaya başlandı (Eberhart ve Shi, 2007).

1994 yılında ilk defa IEEE Dünya Yapay Zekâ Kongresi düzenlendi. İlk kez yapay sinir ağları, evrimsel hesaplamalar ve bulanık mantık üzerine büyük bir konferans ile bilim insanları bir araya geldi.

3.1.5. Yapay Zekâ Çağı (1998 –)

İkinci IEEE Dünya Yapay Zekâ Kongresi 1998 yılında düzenlendi. Yapay sinir ağları, evrimsel hesaplamalar ve bulanık mantık konuları arasındaki sınır tamamen ortadan kaldırılmıştır. Bu olay da Yapay Zekâ Çağının başlamasına vesile olmuştur.

3.2. YSA'nın Çalışma Prensipleri

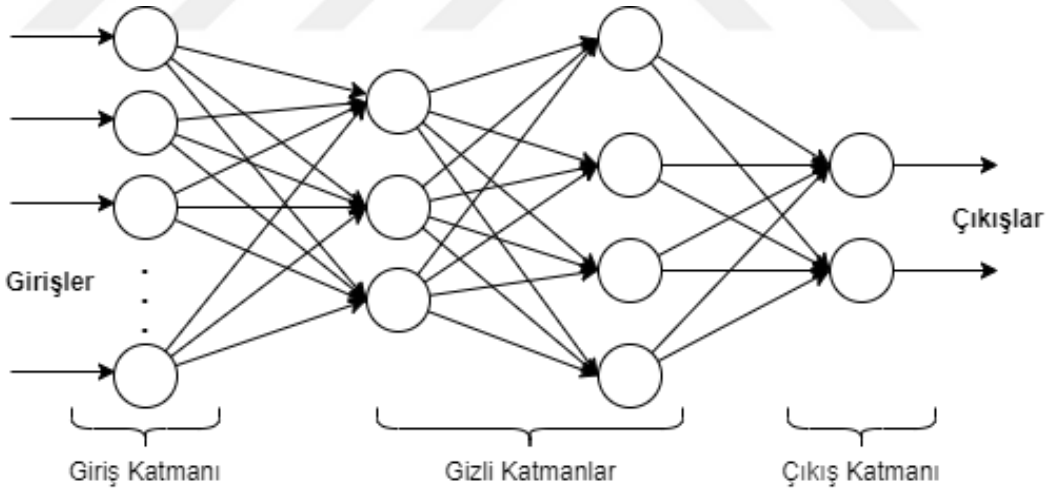
Bir yapay sinir ağı projesine ait yaşam döngüsü aşağıda belirtildiği gibidir;

1. Görev Tanımı ve Tasarımı: Nihai sistemden beklenilenlerin kesin olarak belirlenmesi,
2. Fizibilite: Problemin yapay sinir ağları çözümü için uygun olup olmadığının belirlenmesi,
3. Verilerin Kodlanması: Verinin ağa aktarılabilmesi için uygun olarak verilerinin kodlanma işleminin gerçekleştirilmesi,
4. Ağın Tasarımı: Doğru ve genelleştirilmiş bir modelin sağlanması,
5. Verinin Toplanması: Gerekli veri miktarının ölçülmesi ve bu ölçüm sonucuna göre verilerin toplanması,
6. Verinin Kontrol Edilmesi: Verilerin problemin çözülmesine izin verip vermeyeceğinin belirlenmesi,
7. Eğitim ve Test: Veriden en iyi ağın oluşturulması,

8. Hata Analizi: Hata türlerini ve nedenlerini belirleme ve ağ çıkışına güven sınırlarını atama,
9. Ağ analizi: Oluşturulan yapay sinir ağının analiz edilerek açıklanması ve yorumlanması,
10. Sistem Uygulaması: Yapay sinir ağının son kez kullanılması ve izlenmesi (Swingler, 2001).

3.3. Yapay Sinir Ağlarının Yapısı ve Elemanları

Yapay sinir ağları sadece girdi ve çıktı katmanlarından oluşabileceği gibi girdi katmanı, gizli (ara) katmanlar ve çıktı katmanından da oluşabilmektedir. Sadece girdi ve çıktı katmanından oluşan yapay sinir ağları karmaşık problemleri çözmede başarılı değillerdir. Bu nedenle karmaşık problemlerin çözülebilmesi için ağlarda en az bir adet gizli katman bulunmalıdır. Şekil 3.1.'de iki gizli katmanlı toplam dört katmandan oluşan bir yapay sinir ağı görülmektedir.



Şekil 3.1. Çok katmanlı bir yapay sinir ağı (Elmas, 2011)

Şekil 3.1.'de gösterilen ağ yapısında, giriş katmanı, bir veya daha fazla ara (gizli) katman ve çıkış katmanı bulunmaktadır. Bir katmandaki her bir nöron, bir sonraki katmanın bütün nöronlarına bağlanmaktadır. Aynı katmandaki nöronlar arasında bağlantı bulunmamaktadır.

3.3.1. Katmanlar

Yapay sinir ağlarının yapısı basitçe ele alınırsa, katmanlar ve katmanlar arasındaki bağlantılar olarak adlandırılabilir. Nöronların bir kısmına gerçek dünyadan veri girişi sağlanır. Bu nöronların bulunduğu katmana giriş katmanı denir. Giriş katmanındaki nöronlar, gerçek dünyadan aldıkları verileri, ağın yapısına göre gizli katmanlara ya da çıkış katmanına iletilir. Eğer ağda gizli katmanlar mevcutsa, giriş katmanından aldığı verileri ağırlıklandırarak bir sonraki gizli katmana veya çıkış katmanına iletilir. Çıkış katmanında bulunan nöronlar giriş katmanından veya gizli katmanlardan elde ettiği veriler ile çıktı oluşturur ve oluşturdukları çıktıları gerçek dünyaya iletilir.

3.3.2. Bağlantılar

Bir nöronun girişi, diğer bir nörona giriş olarak yollarla bağlanır. Nöron birçok nörondan giriş alır ancak fakat diğer nöronlara aktarılan tek bir çıkış üretir. Katmanlar arasında kullanılan değişik türde bağlantılar vardır. Bu bağlantılardan bazıları tam bağlantı, kısmi bağlantı, ileri besleme, çift yönlü, hiyerarşik ve rezonans bağlantıdır (Elmas, 2011).

3.4. Yapay Sinir Ağlarının Çeşitleri

Yapay sinir ağları çeşitleri, öğrenme kurallarına ve yapılarına göre iki farklı şekilde çeşitlenmektedir. Öğrenme kurallarına göre yapay sinir ağları denetimli, denetimsiz ve pekiştirmeli ağlar olmak üzere üçe ayrılmaktadır. Yapı açısından bakıldığında ise yapay sinir ağları ileri ve geri beslemeli ağlar olmak üzere ikiye ayrılmaktadır.

3.4.1. Öğrenme kurallarına göre yapay sinir ağları

Yapay sinir ağlarının eğitimi kullanılan eğitim kuralları genellikle üç başlık altında toplanmaktadır. Bunlardan birincisi denetimli öğrenmedir. Denetimli öğrenmede bir öğretmene ihtiyaç vardır. Öğretmen, veriler ile ağa alıştırmalar

yaptırır ve ağ sonuçlarını değerlendirerek öğrenmenin ne zaman biteceğine kadar verir. İkinci öğrenme şekli ise denetimsiz öğrenmedir. Denetimsiz öğrenmede ağ verilerden öğrenme işlemini kendi kendine gerçekleştirir. Öğrenme işleminin ne zaman tamamlanacağına ise ağ karar verir. Diğer bir öğrenme kuralı ise pekiştirmeli öğrenme yöntemidir. Bu yöntemde ağ doğrudan gerçek ağ çıkışını vermez, ağ çıkışının iyi veya kötü olarak değerlendirilmesini sağlar (Elmas, 2011).

3.4.1.1. Denetimli öğrenme

Denetimli öğrenme algoritmalarında, öngörülen çıktı ile belirli bir çıktının karşılaştırılması yapılır. Yapılan bu karşılaştırmaya göre tüm parametreler karakterize edilir. Bir sinir ağının parametreleri onun ağırlıklarıdır. Denetimli öğrenmede arzu edilen ağ çıktısının elde edilebilmesi için, çıkış hatasının düşürülmesi ve ağırlıkların uyarlanabilir hale getirilmesi gerekir. Denetimli öğrenmede her bir giriş değer için istenen çıkış sisteme tanıtılır ve yapay sinir ağları giriş-çıkış ilişkisini gerçekleştirene kadar ağırlıkları aşama aşama ayarlar. Denetimli öğrenme harici bir öğretmen veya genel bilgi kullanan bir süreçtir. Denetimli öğrenme algoritmaları Hebbian, rekabetçi ve hata korelasyonu algoritmalarından yararlanır. Denetimli öğrenme algoritmalarında, eğitimin ne zaman tamamlanacağına, eğitim ve performans bilgilerinin sağlanması için her bir ilişkilendirmenin ne kadar ve ne sıklıkta sunulacağına denetleyici karar verir (Eberhart ve Shi, 2007).

Denetimli öğrenme iki alt kategoriye ayrılır. Bu alt kategoriler Yapısal ve Geçicidir. Yapısal uyum her bir örüntü çifti için mümkün olan en iyi girdi – çıktı ilişkisini bulmakla ilgilidir. Geçici uyum ise nihai sonuçlara ulaşmak için gerekli olan bir dizi örüntüyü yakalamakla ilgilidir (Eberhart ve Shi, 2007).

3.4.1.2. Denetimsiz öğrenme

Denetimsiz öğrenmede bir danışman veya öğretmen, sinir ağına girişin hangi veri parçası sınıfına ait olduğunu veya ağın nerede iyi sonuç vereceğini söylemez. Ağ

veriyi üyeleri birbirinin benzeri olan öbeklere yol gösterilmeksizin ayırır (Elmas, 2011).

Denetimsiz öğrenme bazı yerlerde kendi kendini örgütleme olarak da adlandırılır. Denetimsiz öğrenmede tüm uyarlama süreci boyunca yalnızca yerel bilgilere yer verilir. Denetimsiz öğrenme algoritmaları verilerin kümelenmesini sağlar, sunulan verileri düzenler ve ortaya çıkan kolektif özellikleri keşfeder (Eberhart ve Shi, 2007).

3.4.1.3. Pekiştirmeli öğrenme

Pekiştirmeli öğrenme algoritmaları denetimli ve denetimsiz algoritmaların bir arada kullanıldığı algoritmalardır. Bazı durumlarda denetimli ve denetimsiz algoritmalar tek başlarına doğru sonuçları elde etmek için yeterli olmazlar. Bundan dolayı iki veya daha fazla algoritma analizde bir arada kullanılır. Bu tür öğrenmelere de pekiştirmeli öğrenme denir. Pekiştirmeli öğrenme algoritmalarında bir ödül ve ceza programı uygulanmaktadır. Pekiştirmeli öğrenme, dinamik bir yapıya sahiptir ve deneme yanılma etkileşimleri yolu ile öğrenme işlemini gerçekleştirir. Deneme işleminin sonucunda sonuç hatalı çıkarsa bir ceza uygulanır (Kaelbling vd., 1996).

3.4.2. Ağ yapılarına göre yapay sinir ağları

Yapay sinir ağlarının düğümleri ve bağlantıları çok değişik biçimlerde bir araya getirilebilir. Ağlar bu düğüm ve bağlantı mimarilerine göre farklı isimler alırlar (Elmas, 2011).

Yapay sinir ağları, nöronlar arasında oluşturulan bağlatıların bilgi taşıma yönlerine göre ikiye ayrılmaktadır. Yapay sinir ağlarında, ileri beslemeli ve geri beslemeli ağlar olmak üzere iki temel ağ yapısı bulunmaktadır.

3.4.2.1. İleri beslemeli ağlar

Reel değerli n boyutlu girdi özel vektörleri şu şekilde ifade edilir; j gizli katman nöronu, i girdisini $w_{ij}(i = 1,2,\cdots,n,\ j = 1,2,\cdots,\mathrm{m},)$ ağırlığına göre alır. J birimi x girdi işaretinin ve w_{ij} ağırlıklarının bir işlevini hesaplayıp, sonucu sonraki tüm komşu nöronlara iletilir. Bu işlem çıkış katmanındaki nöron tarafından da yapıldıktan sonra tamamlanır. Bu ağlar, çok katmanlı ileri beslemeli ağlar olarak adlandırılırlar (Elmas, 2011).

İleri beslemeli ağlar, makine öğrenmesi için oldukça önemli bir yere sahiptir. Evrişimsel ağlar, görüntüden nesne tanıma için özelleştirilmiş bir ileri beslemeli yapay sinir ağı türüdür. Buna ek olarak birçok doğal dil işleme uygulamalarında kullanılan yinelemeli ağlar, ileri beslemeli yapay sinir ağlarının özel alt bir türüdür. İleri beslemeli ağların bir başka önemli alt türü de geri yayılımlı ileri beslemeli yapay sinir ağlarıdır.

3.4.2.1.1. Geri yayılım ağı

Yapay sinir ağı belirlenen görevi yerine getirebilmek için eğitime ihtiyaç duyar. Bu eğitimde hedef çıktı ile gerçekleşen çıktı arasında meydan gelen hata miktarında bir azalmanın olması sağlanır. Bunun için yapay sinir ağı ağırlıkların türevini hesaplar. Ağ, yaptığı hesaplamaların sonuçlarına bağlı olarak ağırlıkları artırır veya azaltır. Bu sayede hedef çıktı ile gerçekleşen çıktı arasındaki hata en aza indirgenir. Geri yayılım algoritması da bu işlemin gerçekleşmesi için kullanılan algoritmalardan bir tanedir (Russel, 2019).

Geri yayılım ağı, 1970'lerin başında geliştirilmiş, en popüler, en etkili ve karmaşık, tanımlanmamış problemlere doğrusal olmayan çözümler getirebilen bir ağ çeşitidir. Geri yayılım ağları çok katmanlı ağlarda kullanılan delta kuralı için geliştirilmiş bir algoritmadır ve hiyerarşik yapıdadır (Elmas, 2011).

Geri yayılım ağı, birbirinden bağımsız olarak geniş disiplinlerden birçok kişi tarafından türetilmiştir. Çok katmanlı yapay sinir ağlarına yönelik ilk gradyan iniş

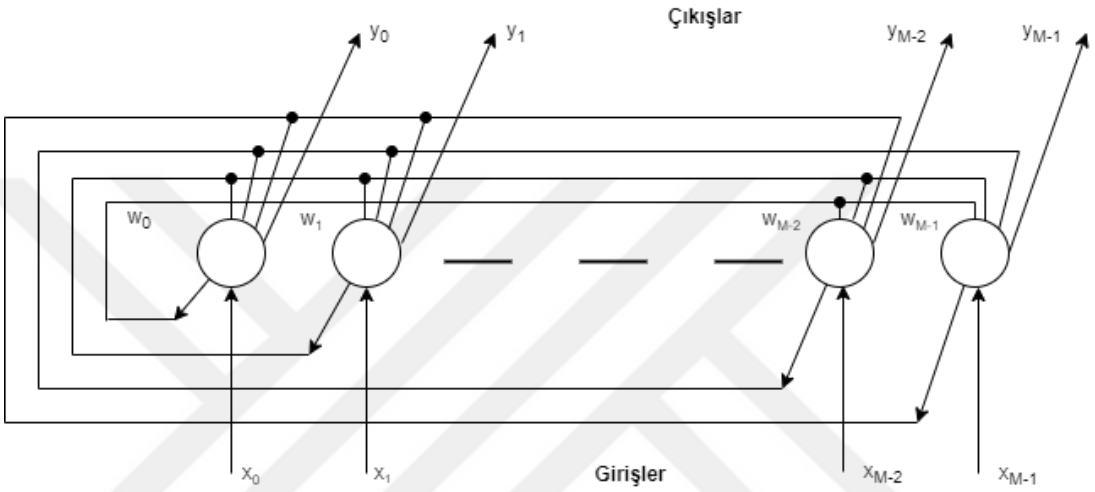
yaklaşımı 1967 yılında Amari tarafından geliştirilmiştir. 1969 yılında Bryson ve Ho doğrusal olmayan problemler için geri yayılım algoritmasına benzer bir algoritma geliştirmiştir. Webros 1974 yılında geri yayılım ağlarını ve birkaç varyantını keşfetmiş ve keşfettiği bu yöntemlere dinamik geri besleme adını vermiştir. 1982 yılında Parker geri yayılım ağlarını tekrardan bulmuş ve öğrenme mantığı olarak adlandırmıştır. 1986 yılında Rumelhar, Hilton ve Williams geri yayılım algoritmasının tüm gücünü ve potansiyelini ortaya çıkartmıştır. Bilim topluluğu bu yapay sinir ağı paradigmasından oldukça etkilenmiştir (Simpson, 1990).

Geri yayılım ağlarında katman sayısı ve her katmandaki düğüm sayısı dikkatlice seçilmelidir. Bu sayıların ne olacağı hakkında kesin bir yöntem yoktur ancak takip edilecek bazı kurallar bulunmaktadır. Bu kurallar aşağıda belirtildiği gibidir.

- Girdi verisi ile istenilen çıktı arasındaki ilişkinin karmaşıklığı artınca, gizli katmanlardaki işleme elemanlarının sayısı da artmalıdır.
- Eğer ele alınan süreç birçok aşamalara ayrılabiliyorsa fazla sayıda gizli katman kullanılmalıdır. Eğer süreç aşamalara ayrılamıyorsa çok fazla sayıda gizli katman kullanılıyorsa ağda yalnızca ezberleme ortaya çıkar ve genel sonuçlara yol açar.
- Ağda kullanılan eğitim verisinin miktarı, gizli katmanlardaki işleme elemanlarının sayısı için üst bir sınır oluşturmaktadır. Bu üst sınırı bulmak için önce eğitim kümesindeki girdi ve çıktı çiftlerinin sayısı bulunur. Bulunan bu sayı ağdaki toplam giriş ve çıkış düğümlerinin sayısına bölünür. Çıkan sonuç beş ile on arasında bir dereceleme faktörüne bölünür. Gizli katmanların az miktarda işleme elemanına sahip olması önemli bir faktördür (Elmas, 2011).

3.4.2.2. Geri beslemeli ağlar

Geri beslemeli ağlar, genellikle danışmasız öğrenme algoritmalarının uygulandığı ağlarda kullanılmaktadır. Geri beslemeli ağlarda, isminden de anlaşılabileceği gibi bir tür geri besleme işlemi vardır. Şekil 3.2.'de örnek bir geri beslemeli yapay sinir ağı yapısı gösterilmiştir. Bu tür ağlarda bir nöronun çıkışı diğer her bir nöronun girişine bağlıdır.



Şekil 3.2. Hopfield Ağı (Elmas, 2011)

3.5. YSA'nın Uygulama Alanları

Yapay sinir ağları bilgi sınıflama ve bilgi yorumlamanın da içinde bulunduğu çok değişik problemlerin çözümünde kullanılmaktadır. Yapay sinir ağlarının kullanıldığı alanlar özet olarak şu şekilde sıralanabilir (Elmas, 2011);

- Denetim
- Sistem modelleme
- Ses tanıma
- El yazısı tanıma
- Parmak izi tanıma
- Elektrik işareti tanıma
- Meteorolojik yorumlama
- Otomatik araç denetimi

- Fizyolojik işaretleri (Kalp fonksiyonlarını gibi) izleme, tanıma ve yorumlama

3.6. YSA'nın Tahmin Amacı ile Kullanımı

Tahmin işlemlerinde uzun zamandır doğrusal istatistik yöntemleri kullanılmıştır. Doğrusal yöntemlerin anlaşılması, analiz edilmesi ve yorumlanması kolaydır. Ancak gerçek hayattaki problemler çoğunlukla doğrusal değildir. Birçok doğrusal olmayan tahmin modelleri vardır ancak doğrusal olmayan bir modelin belirli bir veri setine uygulanması oldukça zordur. Çünkü çok fazla sayıda olası doğrusal olmayan model vardır ve önceden belirlenmemiş doğrusal olmayan model verideki önemli özellikleri yakalamada yeterince başarılı olamayabilir. Veri odaklı olan yapay sinir ağları girdi ve çıktı değişkenleri arasındaki ilişkiler sayesinde önceden bilgi sahibi olmadan doğrusal olmayan modelleme yapabilir.

4. LOJİSTİK REGRESYON ANALİZİ

Lojistik regresyon analizinin kullanım amacı, bağımlı değişken ve bağımsız değişkenler arasındaki ilişkiyi, en az değişken ile en iyi uyuma sahip olacak biçimde tanımlayabilen, kabul edilebilir bir model kurmaktır (Atasoy, 2001). Örneğin kredi puanlamasında, lojistik regresyon analizi, bir ürünün krediye uygun olma olasılığını modellemek için kullanılır. Örneğin bir ürünün faturasının zamanında ödenme olasılığı, fatura büyüklüğü, kişinin yıllık geliri, kişinin mesleği, ipotek ve borç yükümlülükleri, kişinin geçmişte zamanında ödenen faturalarının yüzdesi ve başvuranın kredi geçmişi lojistik regresyon analizi ile analiz edilerek krediye uygun olup olmadığı kararlaştırılır.

Tatlıdil'e göre lojistik regresyon analizi çeşitli varsayım bozulmalarının bulunduğu durumlarda disktriminant analizi ve çapraz tablolara alternatif olabilecek bir yöntem olarak tanımlanmaktadır. Bağımlı değişkenin ikili veya ikiden çok düzey içeren kesikli bir değişken olması gerekmektedir (Tatlıdil, 2002).

Lojistik regresyon analizi disktriminant analizi ve çapraz tablolara göre daha esnek bir yapıya sahiptir. Lojistik regresyon analizinde bağımlı değişkenin dağılımı hakkında herhangi bir varsayım bulunmamaktadır. Lojistik regresyon analizinde tahmin edicilerin normal dağılım göstermesi, bağımlı değişkenle doğrusal ilişki içinde olması ya da grupların eşit varyansa sahip olması zorunluluğu bulunmamaktadır (Tabachnick vd., 2015).

Bağımlı ve bağımsız değişkenlerin türüne göre uygulanabilecek analiz yöntemleri Çizelge 4.1.'de verilmiştir.

Çizelge 4.1. Bağımlı ve bağımsız değişkenlerin türüne göre analiz yöntemleri (Osborne, 2017)

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="5">Bağımlı Değişken</td></tr>
<tr><td>Sürekli</td><td>İkili</td><td>Çok Kategorili</td><td>Sıralı Kategorili</td><td>Sayma Verisi</td></tr>
<tr><td rowspan="3">Bağımsız Değişken</td><td>Sürekli</td><td rowspan="2">En Küçük Kareler</td><td rowspan="2">İkili Lojistik Regresyon Analizi</td><td rowspan="2">İsimsel Lojistik Regresyon Analizi</td><td rowspan="3">Sıralı Lojistik Regresyon Analizi</td><td rowspan="2">En Küçük Kareler, Poisson Regresyon</td></tr>
<tr><td>Sürekli Kategorik Karışık</td></tr>
<tr><td>İkili / Kategorik</td><td>ANOVA</td><td>İkili Lojistik Regresyon Analizi Veya Log-Linear Modeller</td><td>İsimsel Lojistik Regresyon Analizi veya Log-Linear Modeller</td><td>Log-Linear Modeller</td></tr>
</table>

Lojistik regresyon modeli, yeterli sayıda ve iyi dağıtılmış bir örnek mevcut olması koşuluyla kategorik bağımlı değişkene sahip olan verileri ele almak için uygun bir yöntemdir (Domínguez-Almendros vd., 2011).

Lojistik regresyon analizinde en iyi modelin tanımı, çalışmanın türüne ve amacına bağlıdır. Modelin genellikle iki tür hedefi vardır; tahmin ve açıklama. Tahmin hedefi olan bir modelde, bağımlı değişkeni en iyi açıklayan, en az sayıda değişkeni içeren bir model oluşturulur. Açıklayıcı hedefi olan bir modelde "neden" değişkeni ile bir "etki" değişkeni arasındaki nedensellik ilişkisi açıklanmaya çalışılır (Domínguez-Almendros vd., 2011).

Lojistik regresyon analizini yaparken amacı dikkate almak önemlidir. Çünkü iki hedef farklı modelleme stratejileri kullanmaktadır. Tahmin modellerinde, istatistiksel olarak anlamlı katsayıları olan bir değişken mevcut ise, denkleme dahil edilir. Çünkü bu durumda daha güvenilir bir tahmin işlemi gerçekleştirilir. Açıklayıcı modellerde ise, anlamlı katsayıya sahip olan bağımsız değişken, oluşturulan modelde ilgilenilen neden ve etki değişkenleri arasındaki ilişkiyi etkilemiyorsa modele dahil edilmez (Domínguez-Almendros vd., 2011).

Lojistik regresyon analizinde kullanılan bağımsız değişkenlerden bazıları, cinsiyet, tedavi grubu, sektör, vb. nominal ölçekli değişkenler ise aralıklı ölçekli değişken gibi modele dahil edilmesi uygun değildir. Bu nominal değişkenlerin çeşitli seviyelerini temsil etmek için, modele kukla değişkenler eklenmelidir. Örneğin yapılan bir çalışmada sektör bağımsız değişkeninde 5 adet kategori bulunmaktadır. Bu beş kategori modele ayrı ayrı değişkenler olarak girilmelidir. Şirket hangi sektörde bulunuyorsa sadece o sektöre ait olan bağımsız değişken 1 değerini alırken diğer sektörler 0 olarak işaretlenmelidir.

4.1. Lojistik Regresyon Analizinin Tarihçesi

Lojistik regresyon analizi, 19. Yüzyılda popülasyonların büyümesinin ve oto katalitik kimyasal reaksiyonların seyrinin açıklaması için icat edilmiştir. İlk yıllarda Amerika Birleşik Devletleri gibi genç bir ülkede nüfusun orantısız artmasını engellemiştir. Daha sonrasında Quetelet üstsel büyümenin gelişi güzel dışlanmasının imkânsız değerlere yol açacağını fark etmiştir. Quetelet ve öğrencisi Verhulst üstsel büyümeye karşı artan direnci temsil edecek bir terim ekleyerek ve daha sonrasında Φ formlarını deneyerek sorunu çözüme ulaştırmıştır. Verhulst 1828 v e1847 yılları arasında üç adet makale yayınlayarak analizin adını lojistik olarak adlandırmıştır (Cramer, 2003).

1920 yılında Pearl ve Reed Amerika Birleşik Devletleri'nde nüfus artışı üzerine gerçekleştirdikleri bir çalışmada lojistik fonksiyonu tekrardan keşfetmişlerdir. Verhulst'un gerçekleştirdiği çalışmalardan habersiz olarak lojistik fonksiyonu ve eğrisine ulaştılar. Pearl ve arkadaşları 1920 – 1940 yılları arasında meyve sineklerinden, kavunların büyümesine, Kuzey Afrika'daki Fransız kolonilerinin nüfusuna kadar yaşayan hemen hemen her canlıya lojistik fonksiyonu uygulamışlardır (Cramer, 2003).

Yule 1925 yılında yayınladığı çalışması ile Pearl ve Reed'in yaptığı çalışmalardan önce Verhulst'un çalışmalar yaptığını ortaya çıkartmış ve kullanılan analizin adının lojistik olarak adlandırılmasını sağlamıştır.

Reed ve Berkson 1929 yılında yayınladıkları çalışmada otokatalitik reaksiyonların seyrinde lojistik fonksiyonun kullanımını göstermiş ve 1883 yılında Alman kimya profesörü Wilhelm Ostwald'ın çalışmalarından alıntıda bulunmuşlardır. Berkson ve Reed'in gerçekleştirdiği çalışma lojistik fonksiyonun yaygınlaşmasını ve farklı alanlarda kullanılmasını sağlamıştır (Cramer, 2003).

Probit modellerin 1933 yılında Gaddum tarafından icat edildiği söylenmektedir. Ancak 1971 yılında Finney'in yayınladığı çalışmada modelin kökenlerinin Alman bilgin Fechner'e (1801 – 1887) dayanmaktadır. Probit yönteminin kabulü 1950'lere kadar Bliss'in düzenli olarak yayınladığı makaleler sayesinde gerçekleşmiştir (Cramer, 2003).

1944 yılında Berkson "logit" terimini "probit" ile benzer şekilde birleştirerek lojistik kullanımını önermiştir. Berkson'un önerisi biyometrik kuruluşlar tarafından kabul görmemiştir. Logit modelin daha kötü bir model olduğu kabul edilmiştir. Çünkü probitin aksine tolerans seviyeleri normal dağılım ile ilişkili olmadığı dile getirilmiştir. Ancak logit modellerin hesaplama kolaylığı, lojit modellere, probit modeller üzerinden açık bir avantaj sağlamıştır (Cramer, 2003).

Hem probit hem de lojit modeller ekonomide, epidemiyolojide ve sosyal bilimler alanlarında yapılan çalışmalarda sıkça kullanılmıştır.

4.2. Lojistik Regresyon Modelleri

Lojistik regresyon analizi bağımlı değişkenin yapısına göre üç farklı formda değerlendirilmektedir. Bu modellerden "İkili Lojistik Regresyon Modeli", kategorik bağımlı değişkenin iki durumlu olduğu durumda kullanılmaktadır. "İsimsel Lojistik Regresyon Modeli" kategorik bağımlı değişkenin çok kategorili olduğu durumlarda kullanılırken; çok kategorili ve sıralı bir yapı söz konusu ise "Sıralı Lojistik Regresyon Modelleri" kullanılmaktadır (Nevruz, 2019). İsimsel lojistik regresyon analizinde ise bağımlı değişkenin aldığı değerler arasında üstünlük durumu bulunmamaktadır.

4.2.1. İkili lojistik regresyon modeli

Bağımlı değişkenin iki kategoriden oluştuğu durumlarda ikili lojistik regresyon modeli kullanılır. Kodlama yapılırken, sıklıkla riskin olmadığı durum için 0 ve riskin olduğu durum için 1 kodu kullanılır. Bağımsız değişkenlerin türü ile ilgili herhangi bir kısıtlama bulunmamaktadır (Alpar, 2017).

Bağımlı değişken ikili bir yapıda olduğunda çoklu doğrusal regresyon analizi ve hipotez testleri için gerekli varsayımlar ihlal edilmiş olur. Bu durumda değerlerin normal dağılıma sahip olduğunun söylenmesi çok mantıklı olmamaktadır. Bundan dolayı da disktriminant analizinden ayrılmaktadır. Çünkü ikili lojistik regresyon analizinde bağımsız değişkenlerin dağılım özelliklerine yönelik bir şart yokken diskriminant analizinde bağımsız değişkenlerin çoklu normal dağılımda bulunması istenir (Akgül, 2003).

Lojistik regresyonda bağımlı değişkenin değerinin kestirilmesi ile ilgilenilmez. Bunun yerine, bağımlı değişkenin 1 değerini alması olasılığı kestirilmeye çalışılır. Elde edilen sonuç bir olasılık değeri olduğu için sadece 0 ile 1 arasında değerler alabilir (Alpar, 2017).

X: bağımsız değişkene ait veri matrisi olsun. Doğrusal regresyon modelinde X = x olduğunda Y'nin beklenen değeri:

$$E(Y \mid X = x) = \beta_0 + \beta_1 x \qquad (4.1)$$

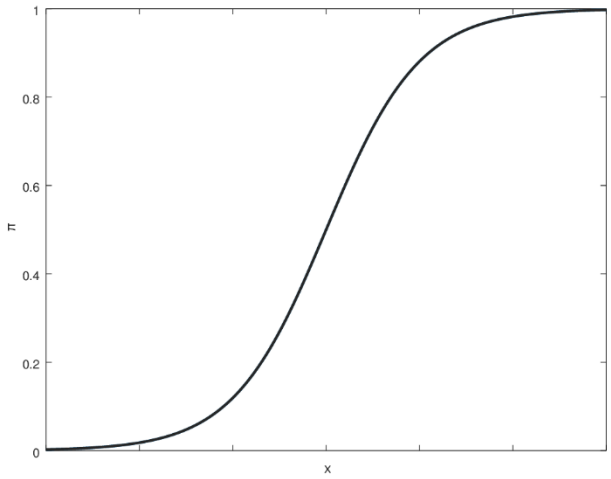
olarak elde edilir. Bu doğrusal fonksiyonun sonucu X'in değerine bağlı olarak $-\infty$ ile $+\infty$ aralığında değerler alabilir. Doğrusal model ile kestirilen değerler 0 ile 1 arasında sınırı olmadığından olasılığı modellemek için kullanılamaz. Ayrıca, doğrusal regresyonda katsayı kestirimi için kullanılan en küçük kareler yönteminin kullanılamamasının bir diğer nedeni de yanıt değişkeni olan Y'nin binomiyal rasgele değişken olmasıdır. Yani, Y'nin varyansı π 'nin bir fonksiyonudur $\left(\pi \times (1 - \pi)\right)$ ve X'in değlerine bağlı olarak değiştiğinden,

varyanların eşitliği varsayımını sağlamaz. Bu sorunu çözebilmek için ağırlıklandırılmış en küçük kareler yaklaşımı kullanılabilir; ancak bu yaklaşımla ilgili de bir sorun bulunmaktadır. Bu nedenlerle, olasılığın modellenmesi için Eşitlik (4.2)' de verilen lojistik regresyon modelinden yararlanır (Alpar, 2017).

$$\pi(x) = P(Y = 1 \mid X = x) = \frac{e^{(\beta_0+\beta_1 x)}}{1 + e^{(\beta_0+\beta_1 x)}} = \frac{1}{1 + e^{-(\beta_0+\beta_1 x)}} \qquad (4.2)$$

Olasılık π ile X bağımsız değişkenleri arasında ilişkili, sıklıkla lojistik yanıt/cevap fonksiyonu (S biçiminde bir eğri) ile gösterilir (Şekil 4.1). Şekil 4.1' deki lojistik fonksiyonunu tanımlayan lojistik regresyon modeli Eşitlik (4.2) ile verilir. Buna göre $X = x$ olduğunda $Y = 1$ olması olasılığı π'dir (Alpar, 2017).

İki kategorili sonuç değişkenlerinin analizi için çok çeşitli dağılım fonksiyonları önerilmiştir. Lojistik regresyonun kullanılmasının üç nedeni: fonksiyonun matematiksel açıdan çok esnek olması, kullanım kolaylığı ve sonuçların klinik açıdan anlaşılır bir şekilde yorumlanabilmesidir (Alpar, 2017).



Şekil 4.1. Lojistik fonksiyonu (Osborne, 2017)

Eşitlik (4.2) ile verilen ve doğrusal olmayan lojistik regresyon fonksiyonu logit dönüşümü uygulandığında doğrusallaştırılabilmektedir. Logit dönüşüm bir olayın odds'unun doğal logaritması alınarak yapılır. Bir olayın odds'u $P/(1-P)$ ya da $\pi/(1-\pi)$ ile verilir. Ancak bu oran 0 ile ∞ arasındabir değer alabilmektedir. Odds'ların doğal logaritması alındığında ise logit dönüşüm

yapılmış olur ve lojitlerin $-\infty$ ile $+\infty$ arasında bir değer alabilmektedir. Logit dönüşüm Eşitlik (4.3) ile tanımlanır.

$$\text{lojit } \pi(x) = g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) \tag{4.3}$$

Lojistik regresyon modeli, yanıt değişkeninin odds'u türünden aşağıdaki gibi belirtilebilir.

$$\frac{\pi(x)}{1-\pi(x)} = e^{(\beta_0+\beta_1 x)} \tag{4.4}$$

Odds'un doğal logaritması alındığında model doğrusal modele dönüşür.

$$g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \ln e^{(\beta_0+\beta_1 x)} = \beta_0 + \beta_1 x \tag{4.5}$$

Eşitlikten de görülebileceği gibi lojit dönüşüm β parametrelerinin doğrusal bir fonksiyonunu oluşturur (Alpar, 2017).

İki kategorili bir lojistik regresyon modelinin katsayılarının tahmin edilmesinde; en çok olabilirlik yöntemi, yeniden ağırlıklandırılmış en küçük kareler yöntemi, minimum lojit ki-kare yöntemi yaygın olarak kullanılan yöntemlerdendir. Belirtilen yöntemlerden en çok olabilirlik yöntemi en sık kullanılan yöntemdir. Bu fonksiyon, gözlenen verinin olasılığını, bilinmeyen parametrelerin bir fonksiyonu olarak belirtir. Bu fonksiyonu en büyük yapan değerler, bilinmeyen parametrelerin en çok olabilirlik kestiricileridir. Yani en çok olabilirlik yönteminde, bir olayın olması olasılığı en çok yapılmaya çalışılır (Alpar, 2017).

Modelde ki değişkenlerin önemliliği, Olabilirlik Oranı, Wald ya da Skor testlerinden biri kullanılarak gerçekleştirilir.

4.2.2. Sıralı lojistik regresyon modeli

Bağımlı değişken sıralı niteliksel değişken olduğu durumlarda kullanılır. Örneğin anketlerde bulunan likert ölçekleri (Kesinlikle katılıyorum, katılıyorum, kararsızım, katılmıyorum, kesinlikle katılmıyorum) veya eğitim durumu (İlkokul, orta okul, lise, lisans, lisans üstü) şeklindeki değişkenler sıralı nitelikli değişkenlerdir. Kategori sayısı 3 ya da daha fazla olan bağımlı değişkenin kategorileri arasında doğal olarak bir sıralama varsa sıralı lojistik regresyon modeli kullanılır (Alpar, 2017).

İkili lojistik regresyon modelinin bir uzantısı sıralı veya oransal olasılık modelidir. Bu model sıralı değişkenler kullanıldığı zaman daha doğru sonuçlar vermesi için özellikle tasarlanmıştır. Sıralı bir kategori yapısına sahip olan bir bağımlı değişkene isimsel lojistik regresyon analizi uygulanabilir ancak iyi bir sonuç elde edilemeyebilir.

Sıralı lojistik regresyon modeli belirli bir eşiğin altındaki tüm grupları eşiğin üzerindeki tüm gruplarla karşılaştıran bir dizi kümülatif ikili lojistik regresyon analizi olarak açıklanabilir (Osborne, 2017).

Sıralı lojistik regresyon modelinde örneğin bağımlı değişkenin 4 kategorisi varsa kategori 0 ile kategori 1,2 ve 3, kategori 0 ve 1 ile kategori 2 ve 3 kategori 0,1 ve 2 ile kategori 3 karşılaştırılır.

4.2.3. İsimsel lojistik regresyon modeli

İsimsel lojistik regresyon analizi kategori sayısı 3 ya da daha fazla olduğunda iki kategorili lojistik regresyon analizinin genişletilmiş biçimidir. Modelde parametre kestirimleri ikili lojistik regresyonda olduğu gibi en çok olabilirlik yöntemi ile elde edilir (Alpar, 2017).

İsimsel lojistik regresyon analizi, diskriminant analizine oldukça benzemektedir. Ancak diksriminant analizinin uygulanması için gerekli olan varsayımlar isimsel

lojistik regresyon analizinde bulunmamaktadır. Bundan dolayı uygulama alanı oldukça geniştir.

İsimsel lojistik regresyon analizi ve sıralı lojistik regresyon analizi her bir grubun birbiri ile karşılaştırıldığı bir dizi ikili lojistik regresyon analizidir. Öncelikle bir adet referans kategori seçilmelidir. Referans kategori seçilirken ilk veya son kategorinin referans kategori olarak alınması yorumlama işlemini kolaylaştırmaktadır. Referans olarak seçilen kategori ile diğer kategoriler birbiri ile karşılaştırılarak ikili lojistik regresyon analizi uygulanır. Kategori miktarı k ile gösterildiğinde k-1 tane lojit fonksiyon hesaplanır ve bu lojit fonksiyonlar birbirleri ile karşılaştırılarak yorumlanır (Osborne, 2017).

4.3. Çok Değişkenli Lojistik Regresyon Analizi

Diğer regresyon analizi yöntemlerinde olduğu gibi lojistik regresyon analizinde de birden fazla bağımsız değişkenin bağımlı değişken üzerindeki etkisi incelenebilir. Modelde iki ya da daha fazla bağımsız değişken olması durumunda çoklu lojistik regresyon çözümlemesinden yararlanılır. Çoklu lojistik regresyon çözümlemesi basit olarak tek değişkenli durumun genişletilmesi ile elde edilir. İlgilenilen değişkenlerin p olmak üzere bağımsız değişkenler vektörü $x = (x_1, x_2, \cdots, x_p)$ ile gösterilir. Buradan, çoklu lojistik regresyon denklemi aşağıdaki gibi elde edilir (Alpar, 2017).

$$\pi(x) = P(Y = 1 \mid x) = \frac{e^{(\beta_0+\beta_1 x_2+\beta_1 x_2+\cdots+\beta_p x_p)}}{1 + e^{(\beta_0+\beta_1 x_2+\beta_1 x_2+\cdots+\beta_p x_p)}} = \frac{1}{1 + e^{-(\beta_0+\beta_1 x_2+\beta_1 x_2+\cdots+\beta_p x_p)}} \tag{4.6}$$

Lojistik regresyon modeli, yanıt değişkeninin odds'u türünden aşağıdaki gibi belirtilebilir.

$$\frac{\pi(x)}{1-\pi(x)} = e^{(\beta_0+\beta_1 x_2+\beta_1 x_2+\cdots+\beta_p x_p)} \tag{4.7}$$

Odds'un doğal logaritması alınarak lojit dönüşüm yapılmış olur.

$$\text{lojit } \pi(x) = g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) \tag{4.8}$$

Odds'un doğal logaritması alındığında model doğrusal modele dönüşür.

$$g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x_2 + \beta_1 x_2 + \cdots + \beta_p x_p \tag{4.9}$$

Çoklu lojistik regresyon modeli kurulduktan sonra modeldeki değişkenlerin önemliliği test edilir. Olabilirlik oran testi modeldeki p adet β katsayısının sıfıra eşit olup olmadığını test eder (Alpar, 2017).

Lojistik regresyon modelinin temeli olabilirlik oranına dayanmaktadır. Olabilirlik oranı, bir olayın gerçekleşmesi ile söz konusu olayın gerçekleşmemesi olasılığını karşılaştırır. Böylece lojistik regresyon modeli, olabilirlik oranının doğal logaritması alınarak elde edilir. Elde edilen lojistik regresyon modelinin parametrelerini tahmin ederken en çok olabilirlik (maximum likelihood) yöntemi yaygın olarak kullanılmaktadır (Berenson ve Levine, 1996).

Çoklu lojistik regresyon analizi uygulamadan önce çalışmada kullanılan bağımsız değişkenlerin tek değişkenli analizlerle incelenerek ilgisiz olanların modele alınmaması önerilmektedir. Bu amaçla iki yaklaşımdan yararlanılır. Birinci yaklaşımda bağımlı değişken ile ilgili bağımsız değişkenler arasındaki bir ilişki olup olmadığı ki-kare, iki ortalama arasındaki farkın önemlilik testi, Mann-Whitney U testi gibi testlerden yararlanılarak incelenir (Alpar, 2017).

Diğer bir yaklaşımda ise her bir bağımsız değişkenle tek değişkenli lojistik regresyon analizi yapılması ve p değerlerinin incelenmesi gerekir. Her iki yaklaşım sonucunda p değeri 0,25'in altında bulunan değişkenlerin çok değişkenli çözümlemede dikkate alınması önerilmektedir (Alpar, 2017).

4.4. Uyum İyiliği Testleri

Lojistik regresyon analizinde elde edilen modelin sonuç değişkenini tanımlamakta ne kadar etkili olduğu modelin uygum iyiliği ile incelenir. Modelin veriye uyumunun iyi olup olmadığı çeşitli yöntemler ile incelenebilir: Pearson ki-kare testi, sapma istatistikleri, Hosmer-Lemeshow testi, sınıflandırma tablosu ve işlem karakteristiği eğrisi bu yöntemlerdir (Alpar, 2017).

4.4.1 Pearson Ki-Kare testi

Gözlenen değerler ve oluşturulan model yardımıyla elde edilen beklenen değerler karşılaştırılarak modelin uyumu değerlendirilir. Pearson ki-kare testinde gözlenen değer ve beklenen değerler arasındaki fark kullanılarak modelin, uyumu değerlendirilir.

$$\chi^2 = \sum_{i=1}^{J} r_i^2 \tag{4.10}$$

Burada J, x'in $\left(\mathrm{x} = \left(\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{p}\right)\right)$ gözlenen farklı değerlerini belirtmektedir. Pearson ki-kare istatistiğinde J-p-1 serbestlik derecesi ile ki-kare dağılımına uyduğu varsayılmaktadır (Alpar, 2017).

4.4.2 Hosmer-Lemeshow testi

Hosmer ve Lemeshow'un uyum iyiliğini incelemek için geliştirdiği yöntemde lojistik regresyon analizinden elde edilen kestirim olasılıkları kullanılır. Kestirilen olasılıklar küçükten büyüğe doğru sıraya dizilir ve bu olasılıklara göre k alt gruba bölünür. Sonrasında, her alt grupta gözlenen ve beklenen değerler hesaplanarak bilinen ki-kare testi uygulanır.

$$\hat{C} = \sum \frac{(G - B)^2}{B} \tag{4.11}$$

Bu istatistik k-2 serbestlik derece ile ki-kare dağılımına uyar (Alpar, 2017).

4.4.3 Sınıflandırma tablosu

Oluşturulan modelin uyum iyiliğini değerlendirmek için sıklıkla doğru sınıflama oranları kullanılmaktadır. Lojistik regresyon sonucunda elde edilen olasılık değerleri belirlenen kesim noktasına göre sınıflara her bir bireyin 0 ya da 1 değerlerinden hangisini alacağı kestirilir (Alpar, 2017).

Genellikle kesim noktası olarak 0,50 değeri alınır. Lojistik regresyon analizinden elde edilen olasılık değeri 0,50'nin üzerinde olduğunda 1. gruba, 0,50'nin altında olduğunda ise 0. gruba, atanır. Sonrasında kestirilen değerler ile gözlenen gerçek değerlerin çapraz tablosu oluşturulur. Çapraz tablo yardımı ile sonucu gerçekte pozitif olanların ne kadarının pozitif, negatif olanların ne kadarının negatif ve toplamda pozitif ve negatif sonuçların ne kadarının doğru sınıflandırıldığı hesaplanır. Modelin uyumu iyi olduğunda doğruluk değerinin yüksek olması beklenir (Alpar, 2017).

5. UYGULAMA

Bu bölüm araştırmanın metodolojisi, araştırmaya ait veriler, istatistiksel analizler ve bulguların değerlendirilmesi olmak üzere üç ana başlıktan oluşmaktadır. Araştırmanın metodolojisi kısmında; araştırmanın amacı ve önemi, kapsamı ve kısıtları ele alınmıştır. Araştırma verileri kısmında ise araştırmada kullanılan verilerin nereden alındığı, nasıl toplandığı ve verilere ait bazı istatistikler bulunmaktadır. Bulgular kısmında ise toplanan veriler yardımı ile yapay sinir ağları uygulaması ve ikili lojistik regresyon analizi uygulaması yapılmış ve elde edilen bulgular yorumlanmış, elde edilen bilgiler değerlendirilerek tablo ve şekiller halinde sunulmuştur.

5.1. Araştırmanın Metodolojisi

Araştırmanın metodolojisi kısmında araştırmanın amacı, araştırmanın kapsamı ve kısıtları ele alınmıştır. Girişim şirketleri, girişim şirketlerine ait yatırım aşamaları ve girişim şirketlerinin başarısız olması gibi konular bu kısımında açıklanmıştır.

Ayrıca, araştırmada kullanılan girişim şirketlerinin bulunduğu sektörler hakkında bilgi verilmiş ve neden bu sektörlerin tercih edildiği açıklanmıştır.

5.1.1. Araştırmanın amacı ve önemi

Bu çalışmanın amacı, 2010 – 2015 yılları arasında melek, çekirdek ve girişim evrelerinde yatırım almış olan girişim şirketlerinin günümüzde başarılı ya da başarısız olduğunun tespit etmektir. Alınan yatırımlar ile başarı veya başarısızlık arasındaki ilişkinin yapay sinir ağları ve ikili lojistik regresyon analizi ile tahmin edilerek sınıflandırılması incelenmiştir. Ayrıca "Yatırım şirketlerinin aldıkları yatırım miktarları ve sektörleri ile başarısız olmak durumları arasında ilişki nedir?" sorusu üzerine durulmuştur.

Aynı zamanda yapılan tahminleme ile başarı oranları hakkında bilgi verip vermediği incelemiştir. Seçilen sektörler arasında başarılı olan şirketlerin ve başarısız olan şirketlerin girişim evrelerinde aldıkları yatırım miktarları tespit edilmiş ve bu sektörlerde girişim şirketi açacak kişilere yatırımcılardan talep edecekleri yatırım miktarı hakkında yol haritası oluşturacaktır.

Elde edilen veriler kapsamında aşağıdaki hedefler belirlenmiştir:

1. Alınan yatırım miktarları ve sektörler ile girişim şirketlerinin başarısı arasındaki ilişkinin belirlenmesi
2. Başarılı olan şirketler ile alınan yatırım miktarlarına arasındaki ilişkinin belirlenmesi
3. Girişim şirketlerinin sektörleri ile başarılı olma arasındaki ilişkinin belirlenmesi

5.1.2. Araştırmanın kapsamı ve kısıtları

Crunbase verileri incelendiğinde 2010 – 2015 arasında yatırım alan ve bulut hizmetleri, büyük veri ve makine öğrenmesi, e-ticaret, mobil uygulama ve sosyal medya sektörlerinde bulunan 29538 adet girişim şirketi bulunmaktadır. 29538 adet girişim şirketi bu çalışmanın anakütlesini oluşturmaktadır. Anakütle içerisinden anakütleyi temsil edecek 1675 adet girişim şirketi araştırmaya dahil edilmiştir. Araştırmaya dahil edilen şirketlerin günümüzde faaliyetlerini sürdürüp, sürdürmediği tespit edilmiştir.

Geçmiş çalışmalar incelendiğinde bir şirketin başarısız ya da başarılı olmasını etkileyen birçok faktörün olduğu görülmektedir. Bu çalışmada bir girişim şirketinin başarılı ya da başarısız olmasında aldıkları yatırımların etkilerinin varlığı ve sektörel farklılıklar incelenmiştir.

5.1.2.1. Girişim şirketleri

Evrensel olarak kabul edilen tek bir “girişim” tanımı bulunmamaktadır. Eisenmann ve arkadaşlarına (2011) göre, yeni başlayanların pazarda yeni

ürünler sunmak için oluşturduğu girişimlerdir. Ries (2011) girişim şirketlerini, büyük bir belirsizlikteki piyasa koşullarında yeni ürün veya hizmet yaratmak için tasarlanan teşebbüs olarak tanımlamaktadır.

Girişimcinin literatürdeki tanımına bakıldığında birçok tanımla karşılaşılsa da temel özellikleri itibari ile düş kuran, kurduğu düşlerin ardından giderek düşlerden proje üreten, projelerini gerçekleştirilebilir hale getiren, onları uygulayarak hayata taşıyan, zenginlik üreten ve inşa yaşamını kolaylaştıran bir toplumsal kaynak olarak açıklanabilir. Girişimcilik ise girişimci ekseninde ortaya çıkan bir parametredir. Bu parametre üretim fonksiyonunun en temel parçasıdır (Öz, 2019).

Günümüzde girişimci, fırsatları gören ve yakalayan, yenilikçi ve yaratıcı bir kişi olarak tanımlanır. Aynı şekilde günümüzde girişimciler, fırsatları pazarlayabilen fikirlere dönüştüren; zaman, çaba, para ya da beceri yolu ile değer yaratan bu fikirleri uygulamak için rekabetçi pazar risklerini üzerine alan ve tüm çabaları paraya çeviren kişilerdir (Semra Güney, 2019).

5.1.2.2. Girişim şirketlerinin yatırım aşamaları

Girişim şirketleri, dünya ekonomisinde büyüme, istihdam yatırım ve üretim konularında büyük öneme sahiptir. Girişim şirketlerinin dünya ekonomisindeki önemi büyük olmasına karşın girişim şirketlerinin kırılgan ve riskli bir yapısı mevcuttur. Birçok girişim şirketi kurulduktan itibaren ilk on yıl içerisinde ekonomik olarak başarısız olurlar.

Birçok girişim şirketi kurulma aşamasında yeterli sermayeye sahip değildir ve faaliyetlerini devam ettirebilmek için yeni kaynaklara ihtiyaç duyarlar. Girişim şirketleri ihtiyaç duydukları finansmanı karşılamak için şirketin belli bir hisse karşılığında yatırımcılardan destek alır.

Yatırım aşamaları genellikle üç aşamadan oluşmaktadır; Melek (Angel), Çekirdek (Seed) ve Risk (Venture).

Melek yatırım aşaması girişim şirketinin fikir ve kurulma aşamasında ihtiyaç duyduğu finansmanın karşılanması olarak adlandırılabilir. İlk aşamada prototip bir ürünün geliştirilmesi için kullanılacak bütçenin sağlandığı aşamadır.

Çekirdek aşaması fikir aşamasından çıkan ürünün daha fazla geliştirilme, üretim, satışa sunulma ve pazarlama dönemlerinde ihtiyaç duyulan yatırımın karşılandığı aşamadır. Yatırımcılar için en riskli aşamadır. Girişim şirketi henüz gelir elde edememiş ve geleceği de öngörülebilir değildir.

Risk aşaması girişim şirketlerinin büyüme aşamasına girdiği dönemde büyüme potansiyeline ulaşmak için ihtiyaç duyduğu finansmanların karşılandığı aşamadır. Risk aşamasında girişim şirketinin kurucu ortaklarının sahip oldukları vizyon ve büyüme hedefleri bu aşamadaki alacakları yatırım miktarını etkileyen en önemli faktörlerdendir.

5.1.2.3. Girişim şirketlerinin başarısız olması

Akademisyenler yaptıkları çalışmalar sayesinde uzun süredir genç firmaların yerleşik firmadan daha yüksek başarısızlık oranlarına sahip olduklarını belirtmişlerdir.

Yapılan çalışmalarda, yüksek başarısız olma eğiliminin, girişim şirketlerinde iş rollerinin etkili bir şekilde dağıtılmaması, iletişim sorunlarının bulunması ve dış alıcılar ve tedarikçiler ile yeterli ikili ilişkilerde bulunamaması olduğu öne sürülmüştür.

Bununla birlikte girişim şirketleri yeterli girişim kaynağa ve istikrarlı ilişkilere ulaşmakta zorluk çekmektedir. Bu da şirketler arasında önemli ölçüde farklılıklara yol açmaktadır.

Blank (2007) bir girişimin temel amacının tekrarlanabilir ve ölçeklenebilir bir iş modeli bulmak olması gerektiğini anlatmaktadır. Bir iş modeli, bir firmanın

ekonomik ve finansal açıdan sürdürülebilir olması için önemli bir paya sahiptir aynı zamanda müşterilerine nasıl değer katacağını tanımlar (Teece, 2010).

Picken'in (2017) iddia ettiği gibi, girişim şirketleri büyüme ve olgunlaşma sırasında çok sayıda engelin üstesinden gelmek zorundadır. Bir girişim şirketinin başarısı, iş modelinin dinamik ve etkili bir şekilde uyarlanma yeteneğine bağlıdır. Bundan dolayı da yatırım alan birçok girişim şirketi, uzun vadede dinamik yapısını kaybeder. Bu da birçok girişim şirketinin sonucunun başarısızlık olmasına neden olur.

Yeni bir girişim şirketinin oluşturulması zor, karmaşık ve riskli bir süreç olarak kabul edilmektedir. Bir girişimcinin karşılaşabileceği en büyük risk, piyasaya kimsenin istemediği veya ihtiyaç duymadığı bir ürün sunmaktır (Eisemann vd., 2013). Bu belirsizlik, girişimcilere kaynak sağlama konusunda yatırımcıları potansiyel çalışanları, tedarikçileri ve alıcıları tereddüt ettirmektedir. Bu koşullar, yeni kurulan girişim şirketlerinin yüksek başarısızlık oranlarına katkıda bulunmaktadır (Chang, 2004).

Bahsedilen faktörlerin yanı sıra, girişim şirketlerinin başarısızlığının önemli bir nedeninin, pazarı keşfetmek ve anlamak, müşterileri tanımlamak gibi süreçlerin hatalığı yapılması olduğu savunulmaktadır (Trimi ve Berbegal-Mirabent, 2012).

5.1.2.4 Araştırmaya dahil edilen sektörlerin belirlenmesi

Fikir aşamasındaki girişim şirketleri, profesyonel ve uzmanlaşmış yatırım fonlarının tercih ettiği ve özel yatırım alanı olarak dikkat çekmektedir. Yatırımcılar hızlı büyüme alanı olan ve nitelikli sektörlere yatırım yapmayı tercih etmektedir (Heard ve Sibert, 2000). Bu çalışmada da 2010 yılından 2015 yılına kadar en fazla yatırım alan ve hızlı büyüme potansiyeli olan 5 sektör seçilmiştir. Bu sektörler Bulut Hizmetleri, Büyük Veri ve Makine Öğrenmesi, E-Ticaret, Mobil Uygulama ve Sosyal Medya olarak belirlenmiştir.

Bulut hizmetleri, kullanıcıların internet olan her yerden erişim sağlayabildiği sistem olarak açıklanabilir. Son on yıl içerisinde hayatımıza giren akıllı cihazlar ile birlikte bulut hizmetleri de önem kazanmaya başladı. Kullanıcılar eskiden bilgisayarlarında veya telefonlarının hafızalarında sakladıkları belgeleri günümüzde bulut hizmetlerinde depolamaya ve her yerden ulaşmaya başladılar. Bu nedenle bulut hizmetleri hızlı büyüme potansiyeli günümüzde devam etmekte olan bir sektördür.

Son on yıl içerisinde sen çok kullanılan kavramlardan bir tanesi büyük veri diğeri de makine öğrenmesidir. İnternetin ve teknolojinin gelişmesi ile internette bırakılan ayak izleri gün geçtikçe arttı ve çok miktarda ham veriye dönüştü. Bundan dolayı da bu ham veriden büyük veri ve makine öğrenmesi kavramları oluştu. Aynı zamanda makine öğrenmesi kavramı cep telefonların kameralarında, otomobillerde ve birçok alanda kullanılmaya başlandı. Bu nedenle büyük veri ve makine öğrenmesi sektörü büyük bir potansiyel ve yatırım kazandı.

E-ticaret gün geçtikçe büyüyen ve hala potansiyeli bulunan bir sektördür. Bunun en büyük örneği internet üzerinden kitap satışı yapmakla sektöre giren daha sonrasında birçok ürünü internet üzerinden satmaya başlayan Amazon şirketi verilebilir. Günümüzde evden çıkmadan, mağazaya gitmeden cep telefonu ile neredeyse her şeyi sipariş vermek mümkün hale geldi. Bu nedenle son on yıl içerisinde gelişen sektörlerden biri ve en çok yatırım alan sektör olarak e-ticaret de araştırmaya dahil edilmiştir.

Akıllı telefonları akıllı halen getiren en önemli şeylerden biri de içerisinde yer alan mobil uygulamalardır. Hayatı kolaylaştıran birçok mobil uygulama geliştirildi ve milyonlarca kişi tarafından kullanılmaya başlandı. Akıllı telefonların içinde bulunan mobil uygulama mağazalarına her gün yüzlerce yeni uygulama dahil ediliyor ve bazıları siliniyor. Son on yıl içerisinde akıllı cihazların sayının artmasından dolayı mobil uygulama sektörü de çalışmaya dahil edildi.

İnsanların akıllı cihazları ve bilgisayarları ile en çok vakit geçirdiği yerlerden biri de sosyal medyadır. İnsanlığın en önemli ihtiyaçlarından biri de sosyalleşmedir

ancak teknolojinin gelişmesi ise sosyalleşme kavramı da gerçeklikten çıkıp sanal sosyal medya platformlarına taşınmıştır. Eskiden okunulan bir kitap bir arkadaşa tavsiye edilirken şimdi takipçilere tavsiye edilmeye ve çok daha fazla kişiye ulaşmaya başladı. Son on yılda birçok sosyal medya platformu oluşturuldu ancak bunlardan sadece çok az bir miktarı başarı yakalayabildi. Bu nedenle sosyal medya sektöründe yer alan girişim şirketleri de çalışmaya dahil edildi.

5.2. Araştırmanın Yöntemi

2010 – 2015 yılları arasında yatırım alan ve bulut hizmetleri, büyük veri ve makine öğrenmesi, e-ticaret, mobil uygulama ve sosyal medya sektörlerinde bulunan 29538 şirket çalışmanın anakütlesini oluşturmaktadır. Tüm şirketler içerisinden geneli yansıtabilecek her sektörden 335 şirket olmak üzere 1675 şirket rastgele seçilerek araştırmaya dahil edilmiştir. Araştırmada elde edilen veriler R programı kullanılarak analiz edilmiştir.

Girişim şirketlerinin başarısızlık tahmininin, şirketlerin bulundukları sektörler ve aldıkları yatırım miktarları ile gerçekleştirildiğinden dolayı yapay sinir ağları ve ikili lojistik regresyon analizi ile modellenmesi uygun görülmüştür.

Yapay sinir ağlarının kullanılmasındaki amaçlardan biri de şirket başarısızlıklarının tahmininde kullanılan klasik istatistiksel yöntemlere bir alternatif olabileceğinin gösterilmek istenmesidir.

Yapay sinir ağları, girdi ve çıktı değişkenleri arasındaki herhangi bir ön bilgiye ihtiyaç duymadan doğrusal ve doğrusal olmayan modellemeyi sağlayabilmektedir. Bu nedenle Yapay sinir ağları, tahmin aracı olarak diğer yöntemlere göre daha genel ve esnektir (Zhang vd., 1998).

Lojistik regresyon analizinin kullanım amacı diğer istatistiksel modelleme teknikleri ile aynıdır. En az değişken ile en iyi modeli oluşturmak şeklinde tanımlanabilir. Lojistik regresyonu diğer istatistiksel modellerden ayıran en

büyük özellik ise varsayımlarının bulunmamasıdır. Bundan dolayı da çalışmada ikili lojistik regresyon analizi tercih edilmiştir.

Şirket başarısızlıklarının tahmin edilmesi için yeni yeni kullanılan yapay sinir ağlarının tahmin gücünü karşılaştırmak amacı ile de ikili lojistik regresyon analizi kullanılmıştır. Sonuçların karşılaştırılabilmesi için yapay sinir ağları ve ikili lojistik regresyon analizi için benzer modeller oluşturulmuştur.

5.2.1. Araştırma verileri

Bu araştırma girişim şirketleri üzerine yapılmıştır. İncelenen çalışmalarda yeni kurulan bir şirketin 5-10 yıl arasında başarısız olma ihtimalinin çok yüksek olduğu tespit edilmiştir. Buradan yola çıkarak girişim şirketlerinin bulundukları sektörler ve aldıkları yatırım miktarlarına göre ileride başarısız olup olmayacağının tahmin edilebileceği düşünülmüştür. Bu nedenle Crunchbase web sitesinden 2010 – 2015 yılına ait girişim şirketlerini yatırım miktarları ve şirketlerin bulundukları sektörler alınmıştır.

2010 – 2015 yılları arasında en çok yatırım almış olan sektörler incelenmiş ve en çok yatırım alan ilk beş sektörün bulut hizmetleri, büyük veri ve makine öğrenmesi, e-ticaret, mobil uygulama ve sosyal medya olduğu tespit edilmiştir. Çalışmanın alınan verilerde bulunan tüm sektörler için gerçekleştirilmesinin uzun süreceği tespit edilmiştir. Bundan dolayı en çok yatırım alan ilk beş sektörün çalışmada kullanılması uygun bulunmuştur.

2010 – 2015 yılları arasında belirtilen sektörlerde bulunan toplam 29538 şirket yatırım aşamalarının tamamında yatırım almıştır. Cruncbase web sitesinde bulunan şirketlerin başarı durumu yeterince doğru ve güncel olmadığı için tek tek her şirket için kontrol edilmelidir. 29538 adet şirket için bu işlemin gerçekleştirilmesi uygun olmadığından dolayı anakütle içerisinden anakütleyi temsil edecek bir örneklem belirlenmiştir.

Çizelge 5.1. Tahmini örneklem büyüklüğü (Çıngı, 2009)

N (Örneklem Sayısı)	Sapma Miktarı				
	0,01	0,02	0,03	0,04	0,05
500					218
1000				375	278
3000		1334	787	500	341
5000		1622	880	536	357
10000	4899	1936	964	566	370
50000	8057	2291	1045	593	381
100000	8763	2345	1056	597	383
500000	9423	2390	1065	600	384

Örneklem sayısı Çizelge 5.1.'deki tahmini örneklem büyüklükleri değerlendirilerek 1675 olarak belirlenmiştir. Örnekleme yöntemi olarak tabakalı rastgele örnekleme yöntemi kullanılmıştır. Öncelikle anakütle sektör bazlı olarak tabakalara ayrılmıştır. Tabakalar incelendiğinde tabakalarda bulunan şirket sayıları birbirine yakın olduğu tespit edilmiştir. Bundan dolayı da her bir tabakadan eşit miktarda rastgele olarak 335'er şirket seçilmiştir. Böylelikle her sektörden 335 girişim şirketi olmak üzere toplam 1675 şirket olarak örneklem hazırlanmıştır.

Crunchbase web sitesinden alınan SQL dosyalarından gerekli sorgular yapılarak analizde kullanılacak veriler excel ortamına aktarılmıştır. Excelde oluşturulan dosya başlık hariç 1675 satır ve 9 sütundan oluşan bir matris yapısındadır.

Yapay sinir ağlarında analizin gerçekleştirilebilmesi için kategorik olan bağımsız değişkenler için dönüşüm işlemi gerçekleştirilmesi gerekmektedir. Örneğin sektör diye bir bağımsız değişken oluşturup, 0 (Bulut hizmetleri), 1 (Büyük veri ve Makine Öğrenmesi), 2 (E-Ticaret), 3 (Mobil Uygulama), 4(Sosyal Medya) olarak kodlandığı durumda yapay sinir ağı analizi gerçekleştirilememektedir. Bundan dolayı kategorik değişken içerisinde bulanan her bir kategori bir bağımsız değişken olarak analize dahil edilmelidir. Örneğin bulut hizmetleri sektörünü ele alınırsa, eğer şirketin sektörü bulut hizmeti ise bağımsız değişken 1 olarak kodlanmalı, şirket başka bir sektörde ise 0 olarak kodlanmalıdır. Bundan dolayı da her bir sektör analize ayrı birer bağımsız değişken olarak dahil

edilmiştir. Analizde kullanılan değişkenler ve değişkenlere ait bilgiler Çizelge 5.2. ve Çizelge 5.3.'te verilmiştir.

Çizelge 5.2. Bağımlı değişkene ait bilgiler

Değişken Adı	Değişken Yapısı	Kategori Kodu	Kategori Adı
Durum	Kategorik	0	Başarısız
		1	Başarılı

Çizelge 5.3. Bağımsız değişkenlere ait bilgiler

Değişken Adı	Değişken Yapısı	Kategori Kodu	Kategori Adı
Bulut Hizmetleri	Kategorik	0	Çalışmıyor
		1	Çalışıyor
Büyük Veri ve Makine Öğrenmesi	Kategorik	0	Çalışmıyor
		1	Çalışıyor
E-Ticaret	Kategorik	0	Çalışmıyor
		1	Çalışıyor
Mobil Uygulama	Kategorik	0	Çalışmıyor
		1	Çalışıyor
Sosyal Medya	Kategorik	0	Çalışmıyor
		1	Çalışıyor
Melek Yatırım Miktarı	Nümerik		
Çekirdek Yatırım Miktarı	Nümerik		
Risk Yatırım Miktarı	Nümerik		

5.3. Bulguların Değerlendirilmesi

Çalışmada bulguların değerlendirmesi dört aşamada gerçekleştirilecektir. Öncelikli olarak çalışmadaki verilere ait tanımlayıcı istatistik verileri incelenecek ve açıklanacaktır. Seçilen örneklemde bulunan şirketlerin ortalama aldıkları yatırım miktarı ve başarısız şirket sayıları incelenecektir. Daha sonrasında yapay sinir ağının ve ikili lojistik regresyon analizinin uygulanması gerçekleştirilmiş ve yöntemlere ait modeller ve modellerden oluşturulan sınıflandırma tabloları oluşturulmuştur. Bulguların değerlendirilmesi aşamasının en son adımında da her iki yöntemle oluşturulan modeller ve modellere ait sınıflandırma tabloları karşılaştırılmış ve elde edilen sonuçlar açıklanmıştır.

5.3.1. Tanımlayıcı istatistikler

Bu bölümde analizde kullanılan verilere ait tanımlayıcı istatistik tabloları oluşturulmuş ve tablolar hakkında kısa bilgiler verilmiştir. Öncelikle tanımlayıcı istatistikler sektör farkı gözetmeksizin incelenmiş daha sonrasında ise sektörler ayrı ayrı incelenerek tablolara ayrılmıştır.

Çizelge 5.4. Şirketlerin yatırım aşamasında aldıkları yatırımlara ilişkin tanımlayıcı istatistikler

Yatırım Aşamaları	Gözlem Sayısı	Toplam	Ortalama	Standart Sapma
Melek	1675	$1.745.501.397	$1.042.090	$609.628
Çekirdek	1675	$2.437.203.960	$1.455.047	$856.610
Risk	1675	$48.758.675.309	$29.109.657	$6.857.935

Çizelge 5.4. incelendiğinde analize dahil edilen 1675 adet girişim şirketinin melek yatırım aşamasında toplam yaklaşık olarak 1 milyar 745 milyon dolar yatırım aldığı ve ortalama şirket başına düşen yatırım miktarının da yaklaşık olarak 1 milyon dolar olduğu bulunmuştur. Çekirdek aşamasına bakıldığında ise şirketlerin aldıkları yatırım miktarları toplamı yaklaşık olarak 2 milyar 437 milyon dolar ve ortalaması da 1 milyon 455 bin dolar olarak bulunmuştur. Risk aşamasında alınan yatırımlara bakıldığında ise toplam yatırım miktarının 48 milyar 760 milyon dolar ve ortalaması da 29 milyon 100 bin dolar olarak tespit edilmiştir. Alınan yatırımların sektörel olarak farklılıklarının bulunup bulunmadığının tespiti için aynı istatistikler her sektör için ayrı ayrı hesaplanmış ve aşağıdaki tablolarda verilmiştir. Bulut hizmetleri sektöründe bulunan şirketlerin aldıkları yatırımlara ilişkin temel istatistikler Çizelge 5.5.'te verilmiştir.

Çizelge 5.5. Bulut hizmetlerinde bulunan şirketlerin yatırım aşamasında aldıkları yatırımlara ilişkin tanımlayıcı istatistikler

Yatırım Aşamaları	Gözlem Sayısı	Toplam	Ortalama	Standart Sapma
Melek	335	$365.410.226	$1.090.776	$574.895
Çekirdek	335	$576.538.632	$1.721.010	$1.057.351
Risk	335	$10.267.022.710	$30.647.829	$5.585.590

Çizelge 5.5. incelendiğinde bulut hizmetleri sektöründe bulunan şirketlerin melek yatırım aşamasında aldıkları toplam yatırım yaklaşık olarak 365 milyon 410 bin olarak, ortalama alınan yatırım miktarı ise yaklaşık olarak 1 milyon 90 bin olarak bulunmuştur. Çekirdek aşamasına bakıldığında ise toplam yatırım miktarı yaklaşık olarak 576 milyon 538 bin dolar olarak, ortalama ise yaklaşık olarak 1 milyon 721 bin olarak bulunmuştur. Risk aşamasında alınan yatırımlara bakıldığında ise toplam yaklaşık olarak 10 milyar 267 milyon olarak bulunmuş şirketlerin ortalama risk yatırımları ise yaklaşık olarak 30 milyon 650 bin dolar olarak bulunmuştur. Büyük veri ve makine öğrenmesi sektöründe bulunan şirketlerin yatırımlarına ait tanımlayıcı istatistik verileri Çizelge 4.6.'da verilmiştir.

Çizelge 5.6. Büyük veri ve makine öğrenmesinde bulunan şirketlerin yatırım aşamasında aldıkları yatırımlara ilişkin tanımlayıcı istatistikler

Yatırım Aşamaları	**Gözlem Sayısı**	**Toplam**	**Ortalama**	**Standart Sapma**
Melek	335	$333.872.709	$996.634	$641.951
Çekirdek	335	$779.215.286	$2.326.015	$1.357.682
Risk	335	$10.151.988.330	$30.304.443	$6.495.287

Çizelge 5.6. incelendiğinde büyük veri ve makine öğrenmesi sektöründe bulunan şirketlerin melek yatırım aşamasında aldıkları toplam yatırım yaklaşık olarak 333 milyon 872 bin olarak, ortalama alınan yatırım miktarı ise yaklaşık olarak 997 bin olarak bulunmuştur. Çekirdek aşamasına bakıldığında ise toplam yatırım miktarı yaklaşık olarak 779 milyon 215 bin dolar olarak, ortalama ise yaklaşık olarak 2 milyon 326 bin olarak bulunmuştur. Risk aşamasında alınan yatırımlara bakıldığında ise toplam yaklaşık olarak 10 milyar 152 milyon olarak bulunmuş şirketlerin ortalama risk yatırımları ise yaklaşık olarak 30 milyon 304 bin dolar olarak bulunmuştur. E-ticaret sektöründe bulunan şirketlerin yatırımlarına ait temel istatistik verileri Çizelge 5.7.'de verilmiştir.

Çizelge 5.7. E-Ticaret sektöründe bulunan şirketlerin yatırım aşamasında aldıkları yatırımlara ilişkin tanımlayıcı istatistikler

Değişkenler	Gözlem Sayısı	Toplam	Ortalama	Standart Sapma
Melek	335	$414.059.551	$1.235.998	$532.975
Çekirdek	335	$436.840.550	$1.304.001	$707.103
Risk	335	$6.877.290.225	$20.529.225	$7.456.559

Çizelge 5.7. incelendiğinde e-ticaret sektöründe bulunan şirketlerin melek yatırım aşamasında aldıkları toplam yatırım yaklaşık olarak 414 milyon 59 bin olarak, ortalama alınan yatırım miktarı ise yaklaşık olarak 1 milyon 235 bin olarak bulunmuştur. Çekirdek aşamasına bakıldığında ise toplam yatırım miktarı yaklaşık olarak 436 milyon 840 bin dolar olarak, ortalama ise yaklaşık olarak 1 milyon 304 bin olarak bulunmuştur. Risk aşamasında alınan yatırımlara bakıldığında ise toplam yaklaşık olarak 6 milyar 887 milyon olarak bulunmuş şirketlerin ortalama risk yatırımları ise yaklaşık olarak 20 milyon 529 bin dolar olarak bulunmuştur. Mobil Uygulama sektöründe bulunan şirketlerin yatırımlarına ait temel istatistik verileri Çizelge 5.8.'de verilmiştir.

Çizelge 5.8. Mobil uygulama sektöründe bulunan şirketlerin yatırım aşamasında aldıkları yatırımlara ilişkin temel istatistikler

Yatırım Aşamaları	Gözlem Sayısı	Toplam	Ortalama	Standart Sapma
Melek	335	$367.277.609	$1.096.351	$585.889
Çekirdek	335	$360.452.107	$1.075.976	$647.582
Risk	335	$7.004.150.847	$20.907.913	$5.896.173

Çizelge 5.8. incelendiğinde mobil uygulama sektöründe bulunan şirketlerin melek yatırım aşamasında aldıkları toplam yatırım yaklaşık olarak 367 milyon 227 bin olarak, ortalama alınan yatırım miktarı ise yaklaşık olarak 1 milyon 96 bin olarak bulunmuştur. Çekirdek aşamasına bakıldığında ise toplam yatırım miktarı yaklaşık olarak 360 milyon 452 bin dolar olarak, ortalama ise yaklaşık olarak 1 milyon 75 bin olarak bulunmuştur. Risk aşamasında alınan yatırımlara bakıldığında ise toplam yaklaşık olarak 7 milyar 4 milyon olarak bulunmuş şirketlerin ortalama risk yatırımları ise yaklaşık olarak 20 milyon 907 bin dolar olarak bulunmuştur. Sosyal Medya sektöründe bulunan şirketlerin yatırımlarına ait tanımlayıcı istatistik verileri Çizelge 5.9.'da verilmiştir.

Çizelge 5.9. Sosyal medya sektöründe bulunan şirketlerin yatırım aşamasında aldıkları yatırımlara ilişkin tanımlayıcı istatistikler

Yatırım Aşamaları	Gözlem Sayısı	Toplam	Ortalama	Standart Sapma
Melek	335	$264.881.303	$790.690	$712.480
Çekirdek	335	$284.157.385	$848.231	$513.300
Risk	335	$14.458.223.198	$43.158.875	$8.856.057

Çizelge 5.9. incelendiğinde sosyal medya sektöründe bulunan şirketlerin melek yatırım aşamasında aldıkları toplam yatırım yaklaşık olarak 264 milyon 880 bin olarak, ortalama alınan yatırım miktarı ise yaklaşık olarak 790 bin olarak bulunmuştur. Çekirdek aşamasına bakıldığında ise toplam yatırım miktarı yaklaşık olarak 284 milyon 157 bin dolar olarak, ortalama ise yaklaşık olarak 850 bin olarak bulunmuştur. Risk aşamasında alınan yatırımlara bakıldığında ise toplam yaklaşık olarak 14 milyar 460 milyon olarak bulunmuş şirketlerin ortalama risk yatırımları ise yaklaşık olarak 43 milyon 160 bin dolar olarak bulunmuştur.

Sektörler kendi aralarında karşılaştırıldığında her sektörün en çok yatırım aldığı aşama risk aşamasıdır. Bu beklenilen bir sonuçtur. Ancak sektörden sektöre melek ve çekirdek aşamasında aldıkları yatırım miktarı değişmektedir. Bulut hizmetleri ile büyük veri ve makine öğrenmesi sektörlerinde en az yatırım miktarı melek daha sonrasında çekirdek aşamasında iken diğer sektörler için melek ve çekirdek yatırım aşamalarında büyük farkların olmadığı tespit edilmiştir.

Analizde kullanılan şirketlerin başarısızlık durumları hem sektörler bazlı hem de toplu olarak aşağıdaki tablolarda gösterilmiştir.

Çizelge 5.10. Örneklemde bulunan girişim şirketlerinin başarı durumları

Sektörler	Kategori	Şirket Sayısı	Yüzde
Tüm Sektörler	Başarılı	892	0,53
	Başarısız	783	0,47
	Toplam	1675	1

Çizelge 5.10.'a bakıldığında analize dahil edilen 1675 adet girişim şirketinden %47'si başarısız olarak günümüzde faaliyetlerini sürdürmemektedir. Daha önceki çalışmalar incelendiğinde girişim şirketlerinin başarısız olma olasılığının yaklaşık olarak %60 civarında olduğu görülmüştür. Ancak bu çalışmada kullanılan sektörler yeni gelişmekte veya gelişmekte olan sektörler oldukları için başarısız olma ihtimalleri daha düşüktür. Aşağıda belirtilen tablolarda da sektörler tek tek incelenmiş ve başarı durumları açıklanmıştır.

Çizelge 5.11. Bulut hizmetleri sektöründe bulunan girişim şirketlerinin başarı durumları

Sektörler	Kategori	Şirket Sayısı	Yüzde
Bulut Hizmetleri	Başarılı	174	0,52
	Başarısız	161	0,48
	Toplam	335	1

Çizelge 5.11.'de bulut hizmetleri sektöründe bulunan şirketlerin başarısızlık istatistikleri gösterilmiştir. Tablo incelediğinde örnekleme dahil edilen bulut hizmetlerinde bulunan girişim şirketlerinin %48'i başarısız olmuştur.

Çizelge 5.12. Büyük veri ve makine öğrenmesi sektöründe bulunan girişim şirketlerinin başarı durumları

Sektörler	Kategori	Şirket Sayısı	Yüzde
Büyük Veri ve Makine Öğrenmesi	Başarılı	195	0,58
	Başarısız	140	0,42
	Toplam	335	1

Çizelge 5.12. incelendiğinde büyük veri ve makine öğrenmesi sektöründe bulunan şirketlerin %42 oranında başarısızlık oldukları görülmektedir. Başarı oranın belli oranda yüksek olmasının nedeni büyük veri ve makine öğrenmesinin 2010 yılından sonra daha da popüler hale gelmesi olarak açıklanabilir.

Çizelge 5.13. E-Ticaret sektöründe bulunan girişim şirketlerinin başarı durumları

Sektörler	Kategori	Şirket Sayısı	Yüzde
E-Ticaret	Başarılı	206	0,62
	Başarısız	129	0,38
	Toplam	335	1

Çizelge 5.13.'de e-ticaret sektöründe bulunan şirketlerin başarısızlık sayıları gösterilmiştir. Örneklemde bulunan e-ticaret sektöründeki şirketlerin

başarısızlık oranı %38 oranındadır. Günümüz birçok alışverişimizi bilgisayardan ya da telefondan yaptığımızı düşünürsek e-ticaret sektörü için bu oranın bu kadar düşük olması gayet normaldir. 2010 yılından itibaren hayatımıza çok fazla e-ticaret sitesi girmiş ve fiziksel mağazası bulunan şirketler bile e-ticaret üzerinden satışlarını yapmaya başlamıştır.

Çizelge 5.14. Mobil uygulama sektöründe bulunan girişim şirketlerinin başarısızlıklarına ait istatistikler

Sektörler	Kategori	Şirket Sayısı	Yüzde
Mobil Uygulama	Başarılı	172	0,51
	Başarısız	163	0,49
	Toplam	335	1

Çizelge 5.14. incelendiğinde mobil uygulama sektöründe bulunan şirketlerin %49 oranında başarısızlık oldukları görülmektedir. Başarısızlık ve başarı arasındaki oranın birbirine çok yakın olması mobil uygulama sektörünün dengesiz olduğunu ve her iki şirketten birinin başarısız olduğunu göstermektedir. Kullanılan akıllı cihazların mobil uygulama mağazalarına baktığımızda da her gün karşımıza yeni uygulamaların çıktığı ve birçok kullanılmayan uygulamanın mağazalardan hızlıca kalktığı görülmektedir.

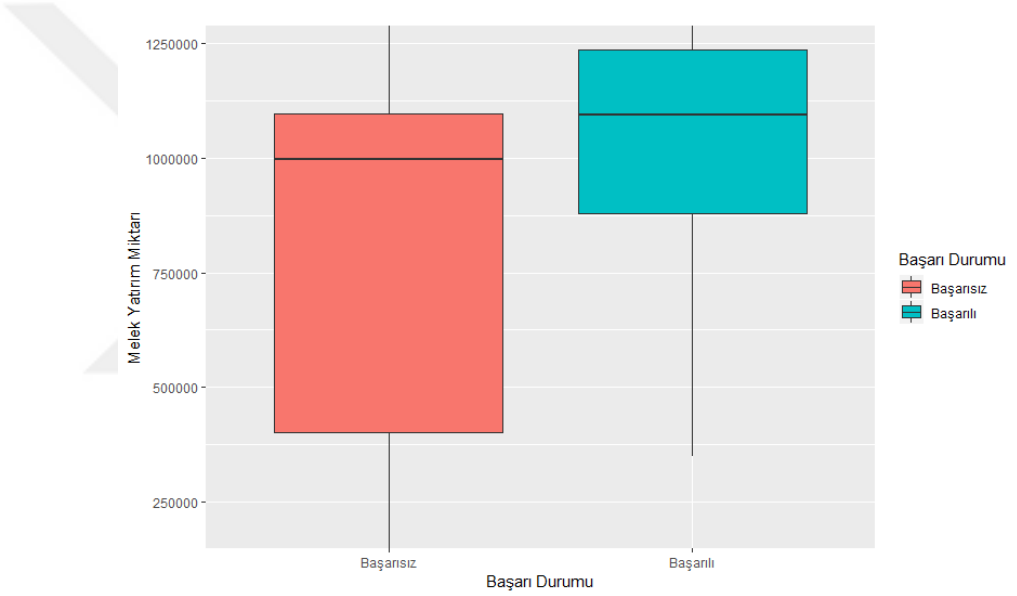
Çizelge 5.15. Sosyal medya sektöründe bulunan girişim şirketlerinin başarı durumları

Sektörler	Kategori	Şirket Sayısı	Yüzde
Sosyal Medya	Başarılı	145	0,43
	Başarısız	190	0,57
	Toplam	335	1

Çizelge 5.15. incelendiğinde örneklemde bulunan sosyal medya sektöründeki girişim şirketlerin başarısızlık oranı %57'dir. 2010 yılından itibaren hayatımıza birçok sosyal medya platformu girmiş ve birçoğu zamanla popülaritesini kaybederek yok olmuştur. Bir sosyal medya şirketinin başarılı olabilmesi aktif kullanıcı sayısına bağlıdır. Birçok sosyal medya şirketi ilk kurulum aşamasında yaptıkları reklamlar ile popüler hale gelmiş ve daha sonrasında bu popülerliklerini kaybetmişlerdir. Bundan dolayı da yeterli miktarda aktif kullanıcı sayısını koruyamayan sosyal medya girişim şirketleri başarısız olmuştur.

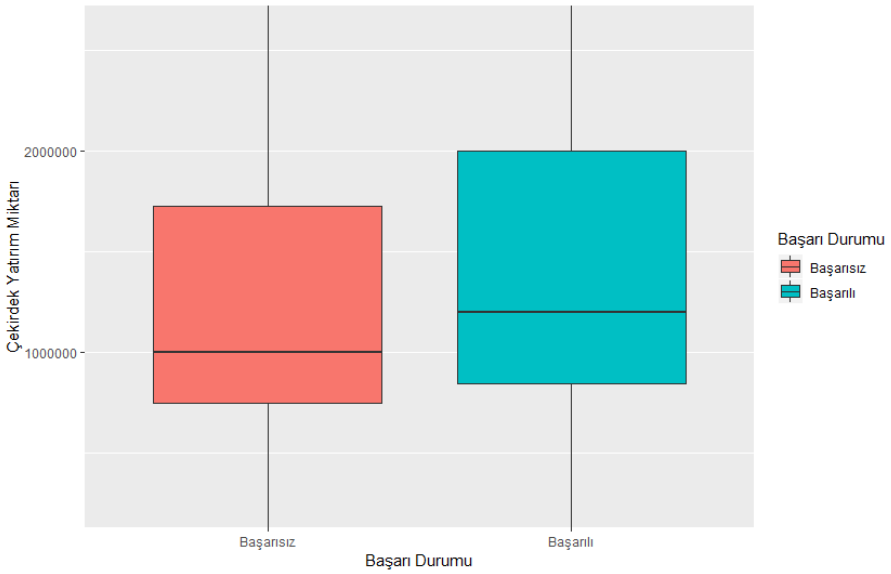
5.3.2. İkili lojistik regresyon analizinin uygulanması

Bu bölümde ikili lojistik regresyon analizi sonucunda elde edilen çıktıların tablolaştırılması ve yorumlanmasına ilişkin bilgiler verilmektedir. İkili lojistik regresyon analizini uygulamadan önce bağımlı değişken olan başarı durumu ile sürekli bağımsız değişkenler olan melek yatırım miktarı, çekirdek yatırım miktarı ve risk yatırımı miktarının dağılımlarının analiz edilmesi çalışma için faydalı olacaktır. Bundan dolayı Şekil 5.1., Şekil 5.2. ve Şekil 5.3'te sürekli bağımsız değişkenler ile bağımlı değişkene ait kutu grafikleri oluşturulmuştur.



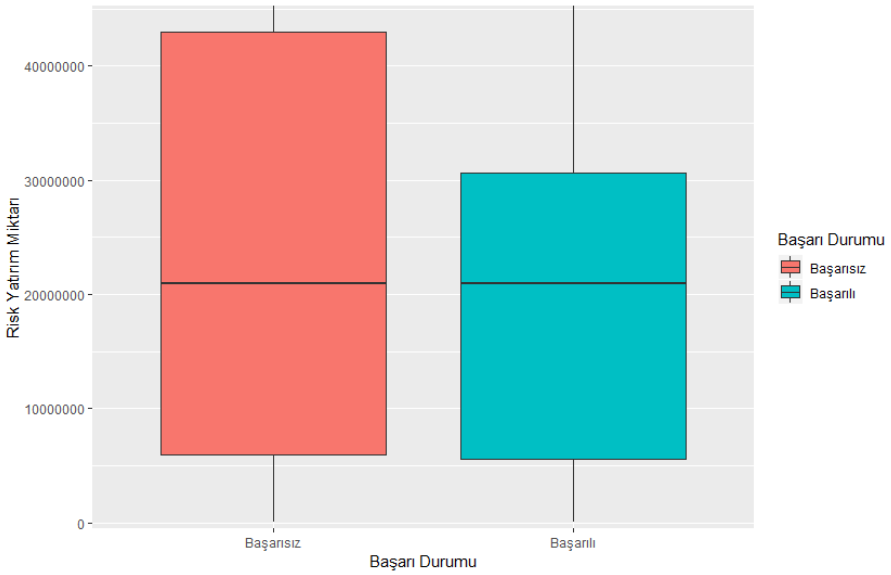
Şekil 5.1. Alınan melek yatırım miktarı ve başarı durumuna ait kutu grafiği

Şekil 5.1. incelendiğinde başarılı olan girişim şirketlerinin melek yatırımda aldıkları yatırımların daha normal dağıldığı görülmektedir. Başarısız olan girişim şirketlerine ait olan melek yatırım miktarları birinci kartil ile medyan arasında yoğunluklu olarak toplanmaktadır. Melek yatırım aşamasında yüksek yatırım alan şirketlerin başarılı olma durumunun daha iyi olduğu tespit edilmektedir.



Şekil 5.2. Alınan çekirdek yatırım ve başarı durumuna ait kutu grafiği

Şekil 5.2. incelendiğinde başarılı veya başarısız olan şirketlerin birbirine yakın miktarda çekirdek aşaması yatırımı aldığı görülmektedir. Başarılı ve başarısız olan girişim şirketlerinin çekirdek aşamasında aldıkları yatırım miktarları medyan ile üçüncü kartil arasında yoğunluk göstermektedir.



Şekil 5.3. Alınan risk yatırım ve başarı durumuna ait kutu grafiği

Şekil 5.3. incelendiğinde başarılı veya başarısız olan şirketlerin medyanlarının birbirine yakın olduğu görülmektedir. Başarısız olan girişim şirketlerinin risk aşamasında aldıkları yatırım miktarları medyan ile üçüncü kartil arasında

yoğunluk göstermektedir. Başarılı olan girişim şirketlerinin biraz daha normal dağılıma uygun olduğu görülmektedir.

Bağımlı değişken ile bağımsız değişkenler incelendikten sonra ikili lojistik regresyon analiz modeli oluşturulmuştur. Modele sektör ve alınan yatırım aşamalarının tamamı dahil edilmiştir. Oluşturulan modele ait istatistikler Çizelge 5.16.'da verilmiştir.

Çizelge 5.16. İkili lojistik regresyon modeline ait katsayılar

	Katsayı	**Standart Hata**	**Z Değeri**	**Sig.**
Sabit Terim	-0,0288	0,1322	-0,22	0,8277
Büyük Veri ve Makine Öğrenmesi	0,2623	0,1571	1,67	0,0950*
E-Ticaret	0,4143	0,1583	2,62	0,0089*
Mobil Uygulama	-0,0006	0,1566	00	0,9968
Sosyal Medya	-0,3699	0,1588	-2,33	0,0198*
Melek	3,36e-8	4,04e-8	0,83	0,4054
Çekirdek	3,96e-9	3,00e-8	-0,13	0,8950
Risk	2,56e-9	1,18e-9	2,16	0,0306*

* 0,90 güven oranı kabul edildiğinde

Çizelge 5.16.'da görüldüğü üzere oluşturulan modele ait sabit terim ve tüm değişkenlere ait katsayı, standart hata, Z değeri ve anlamlılık düzeyi yer almaktadır. Çizelge 5.16. incelendiğinde büyük veri ve makine öğrenmesi, e-ticaret, sosyal medya ve risk yatırımları değişkenleri %90 güven ile anlamlı değişkenler olarak modelde yer almaktadır. Diğer değişkenlerin katsayıları istatistiksel olarak anlamsız çıkmaktadır. Oluşturulan tablo incelendiğinde alınan melek yatırım değişkenine ait katsayı diğer yatırım değişkenlerine göre daha yüksek çıkmasına rağmen modelde anlamsız olarak bulunmuştur. Yatırım aşamaları içerisinde anlamlı olarak bulunan tek değişken risk yatırım miktarı değişkenidir.

Katsayıları anlamsız olan değişkenler çıkartılarak bir model daha oluşturulacak ve ikili lojistik regresyon modelinin doğru sınıflandırma oranı karşılaştırılacak ve daha iyi olan model tercih edilerek yapay sinir ağları ile karşılaştırılacaktır.

Katsayıları anlamsız olan değişkenler çıkartıldıktan sonra oluşturulan yeni modele ait katsayılar Çizelge 5.17.'de verilmiştir.

Çizelge 5.17. İkili lojistik regresyon modeline ait katsayılar

	Katsayı	**Standart Hata**	**Z Değeri**	**Sig.**
Sabit Terim	0,0015	0,0822	-0,02	0,9859
Büyük Veri ve Makine Öğrenmesi	0,2556	0,1354	1,89	0,0590*
E-Ticaret	0,4196	0,1368	3,07	0,0022*
Sosyal Medya	-0,3782	0,1361	-2,78	0,0055*
Risk	2,59e-9	1,18e-9	2,19	0,0282*

* 0,90 güven oranı kabul edildiğinde

Çizelge 5.17. incelendiğinde modele dahil edilen tüm değişkenler %90 güven ile anlamlıdır. Oluşturulan iki model iki farklı yöntem ile karşılaştırabilir. Bunlardan birincisi AIC (Akaike Bilgi Kriteri) dir. Doğrusal regresyon analizindeki düzeltilmiş R^2 değerine benzemektedir ancak AIC değeri düşük olan model tercih edilmelidir. Her iki modele ait AIC değerleri Çizelge 5.18.'de verilmiştir.

Çizelge 5.18. Modellere ait AIC değerleri

	AIC
Model 1	2296
Model 2	2291

Çizelge 5.18. incelendiğinde Model 1 (Katsayıları anlamsız değişkenlerin bulunduğu model) ve Model 2 (Katsayıları anlamsız değişkenlerin bulunmadığı model) için AIC değerleri birbirine çok yakındır. Buradan hangi modelin tercih edileceğine karar vermek doğru değildir. Bir diğer karşılaştırma yöntemi olan doğru sınıflandırma yöntemi uygulanmalıdır.

Çizelge 5.19. Model 1'e ait sınıflandırma tablosu

		Tahmin Edilen	
		Başarısız	**Başarılı**
Gözlenen	**Başarısız**	205	578
	Başarılı	149	743
Doğru Sınıflandırma Oranı = ((205+743)/1675)=0,57			

Çizelge 5.19.'a bakıldığında Model 1 gerçekte başarısız olan 783 girişim şirketinden sadece 205 tanesini doğru sınıflandırabilmiş, 578 tanesini başarılı olarak tahmin etmiştir. 892 tane başarılı olan girişim şirketinin 743 tanesini doğru tahmin etmiştir, bu iyi bir tahmin oranıdır. Ancak genel olarak doğru sınıflandırma oranına bakıldığında bu oran %57'dir.

Çizelge 5.20. Model 2'e ait sınıflandırma tablosu

		Tahmin Edilen	
		Başarısız	**Başarılı**
Gözlenen	**Başarısız**	190	593
	Başarılı	143	749
Doğru Sınıflandırma Oranı = ((190+749)/1675)=0,56			

Çizelge 5.20. incelendiğinde Model 2 gerçekte başarılı olan 892 girişim şirketinin 749 tanesini doğru tahmin etmiştir. Bu oran Model 1'den daha iyidir. Ancak başarısız olan şirketlere bakıldığında ise 783 girişim şirketinden sadece 190 tanesi Model 2 ile doğru tahmin edilebilmiştir ve toplam doğru sınıflandırma oranı da %56 olarak bulunmuştur.

Model 1 ve Model 2 için her iki karşılaştırma yöntemi de uygulanmış ve sonuçlar birbirine oldukça yakın çıkmıştır. Yapay sinir ağları ile karşılaştırmanın daha sağlıklı olması için daha yüksek sınıflandırma oranına sahip olan Model 1 çalışmada kullanılacaktır.

5.3.3. Yapay sinir ağlarının uygulanması

Bu bölümde yapay sinir ağları uygulamasının sonucunda elde edilen çıktıların tablolaştırılması ve yorumlanmasına ilişkin bilgiler verilmektedir. Yapay sinir ağları ile istenen performansın elde edilmesi için deneme yanılma yöntemi kullanılmıştır. En iyi performansı belirlemek için farklı katman ve nöron sayıları ile işlem tekrarlanmıştır. Elde edilen modellere ait MSE (hata kareler ortalaması) ve sınıflandırma başarı yüzdesi kullanılmıştır.

Yapay sinir ağları bölümünde anlatıldığı gibi, yapılan tahminler için oluşturulan veri kümesi; eğitim ve test verisi olmak üzere iki kısma ayrılmıştır. Literatür incelendiğinde, eğitim ve test verileri genellikle %70 eğitim, %30 test veya %80 eğitim, %20 test verisi olarak ayrılmaktadır. Yapılan çalışmada kullanılan veriler hem %70 eğitim, %30 test hem de %80 eğitim, %20 test verisi olarak ayrılmış ve veri setleri için yapay sinir ağları ayrı ayrı uygulanmıştır. Daha sonrasında her iki veri seti içinde elde edilen yapay sinir ağları modelleri ile hem eğitim verisi için hem de test verisi için doğru sınıflandırma tablosu oluşturulmuştur. Oluşturulan sınıflandırma tabloları birbiri ile karşılaştırılarak en iyi olan veri setine ait yapay sinir ağları modeli tercih edilmiş ve ikili lojistik regresyon analizi sonucu ile karşılaştırılmıştır.

Yapay sinir ağının eğitiminde iterasyon sayısı 10000 olarak belirlenmiştir. Ağın eğitimi iterasyon sayısı boyunca sürecek ve daha sonrasında bitecektir. Ancak 10000 iterasyondan önce eğitim verisi için MSE kümesinde artış olması durumunda eğitim o noktada durdurulur.

Yapay sinir ağı için uygun gizli katman sayısını ve nöron sayısını belirlemek için çeşitli modeller oluşturulmuş ve hata kareler ortalaması hesaplanmıştır. Oluşturulan modeller ve hata kareler ortalaması Çizelge 5.21.'de verişmiştir.

Çizelge 5.21. Modellere göre test setinin hata kareler ortalamaları

	Hata Kareler Ortalaması	
Modeller	**%70-%30**	**%80-%20**
3-3-3	0,4616	0,2721
5-5	0,6040	0,4514
3-3	**0,1948**	0,2561
3-2	0,2491	**0,2500**
2-2	0,61	0,2545

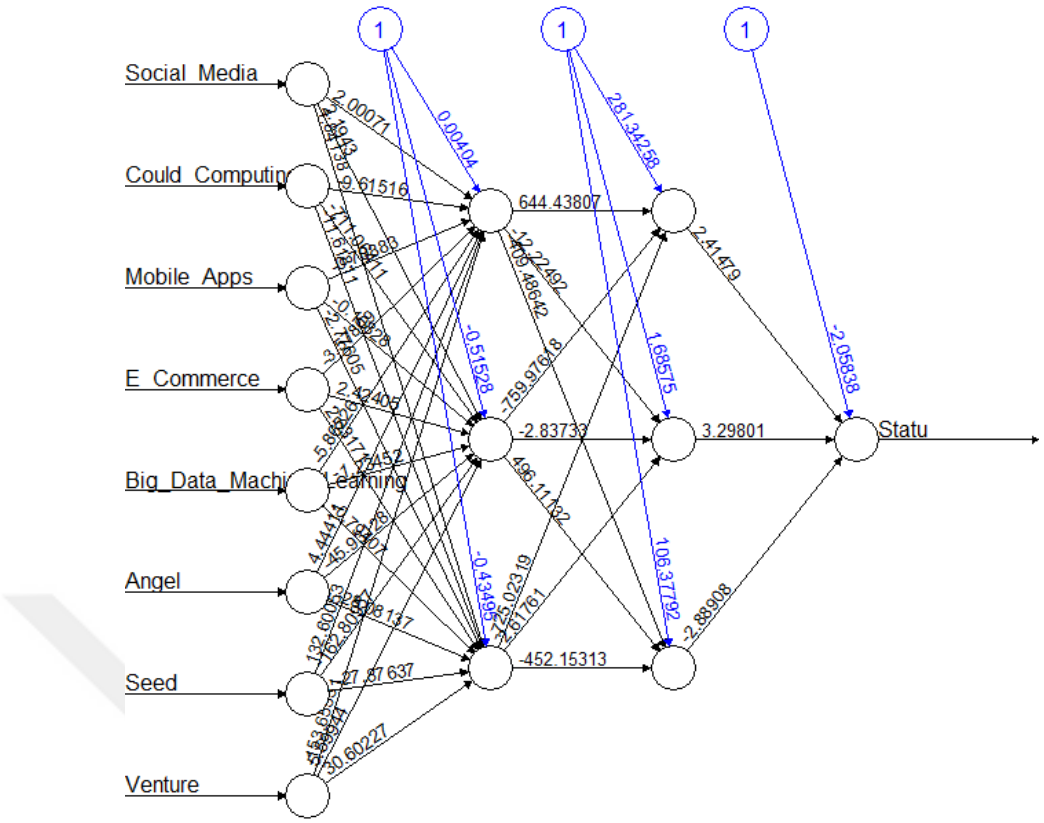
Çizelge 5.21. incelendiğinde %70 eğitim ve %30 test verisi için ve %80 eğitim ve %20 test verisi için en düşük hata kareler ortalamasına sahip 5 model en iyi modeller olarak seçilmiştir. %70 eğitim ve %30 test verisi için en iyi model iki gizli katmanlı ve üçer nöron sayısına sahip olan yapay sinir ağları modelidir. %80 eğitim ve %20 test verisi için en iyi model, iki gizli katmanlı ve ilk gizli

katmanında üç ikinci gizli katmanında iki nörona sahip olan yapay sinir ağları modelidir. Oluşturulan en iyi yapay sinir ağları modellerine ait özellikler Çizelge 5.22. ve Çizelge 5.23'te verilmiştir.

Çizelge 5.22. %70 - %30 için en iyi yapay sinir ağı yapısı

Ağ Modeli	Çok Katmanlı
Öğrenme Algoritması	İleri Beslemeli Geri Yayılımlı
Giriş Katmanındaki Nöron Sayısı	8
Gizli Katman Sayısı	2
1. Gizli Katmandaki Nöron Sayısı	3
2. Gizli Katmandaki Nöron Sayısı	3
Çıktı Katmanındaki Nöron Sayısı	1
Aktivasyon Fonksiyonu	Lojistik Fonksiyon

Çizelge 5.22. incelendiğinde %70-%30 en iyi yapay sinir ağı modeli çok katmanlı ve ileri beslemeli geri yayılımlı bir modeldir. Girdi katmanında sekiz nöron ilk gizli katmanında üç ikinci gizli katmanında üç nöron bulunan ve bir adet çıktı nöronu bulunan bir model yapısına sahiptir. Modelde aktivasyon fonksiyonu olarak lojistik fonksiyon kullanılmıştır. %70 - %30 için en iyi yapay sinir ağı yapısına ait grafik Şekil 5.4.'te verilmiştir.

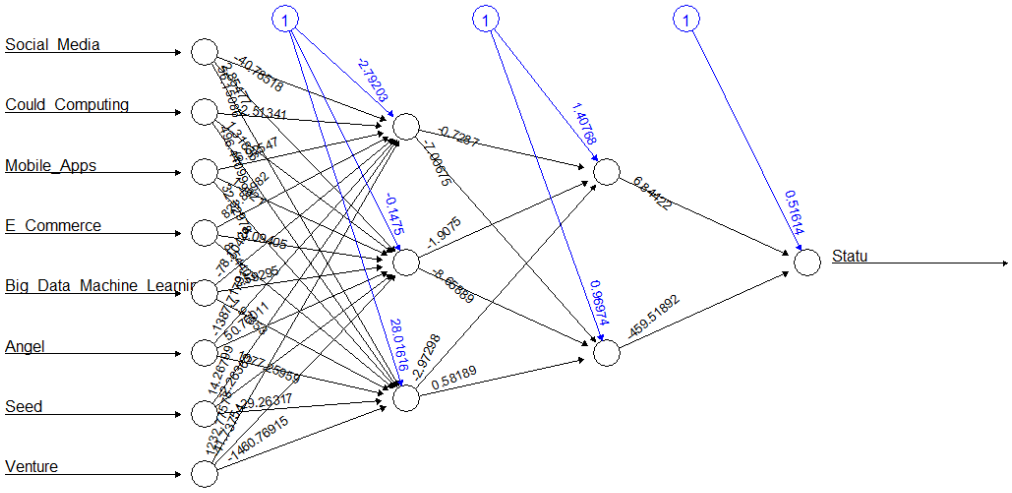


Şekil 5.4. %70 - %30 için en iyi yapay sinir ağı yapısı grafiği

Çizelge 5.23. %80 - %20 için en iyi yapay sinir ağı yapısı

Ağ Modeli	Çok Katmanlı
Öğrenme Algoritması	İleri Beslemeli Geri Yayılımlı
Giriş Katmanındaki Nöron Sayısı	8
Gizli Katman Sayısı	2
1. Gizli Katmandaki Nöron Sayısı	3
2. Gizli Katmandaki Nöron Sayısı	2
Çıktı Katmanındaki Nöron Sayısı	1
Aktivasyon Fonksiyonu	Lojistik Fonksiyon

Çizelge 5.23. incelendiğinde %80-%20 en iyi yapay sinir ağı modeli çok katmanlı ve ileri beslemeli geri yayılımlı bir modeldir. Girdi katmanında sekiz nöron ilk gizli katmanında üç ikinci gizli katmanında iki nöron bulunan ve bir adet çıktı nöronu bulunan bir model yapısına sahiptir. Modelde aktivasyon fonksiyonu olarak lojistik fonksiyon kullanılmıştır. %80 - %20 için en iyi yapay sinir ağı yapısına ait grafik Şekil 5.5.'te verilmiştir.



Şekil 5.5. %80 - %20 için en iyi yapay sinir ağı yapısı grafiği

Oluşturulan en iyi yapay sinir ağları modellerine ait eğitim ve test verileri için performans tabloları oluşturulmuştur. %70 - %30 modeline ait eğitim ve test verisi için oluşturulan performans tabloları Çizelge 5.21. ve Çizelge 5.22.'de verilmiştir. %80 - %20 modeline ait eğitim ve test verisi içi oluşturulan performans tabloları da Çizelge 5.23. ve Çizelge 5.24.'te verilmiştir. Modeller performans tablolarında bulunan doğru sınıflandırma oranlarına göre karşılaştırılmışlardır.

Çizelge 5.24. %70 – %30 Eğitim seti performans sonuçları

		Tahmin Edilen	
		Başarısız	**Başarılı**
Gözlenen	**Başarısız**	355	192
	Başarılı	12	613
Doğru Sınıflandırma Oranı = ((335+613)/1172) = 0,8088 ≅ 0,81			

1172 adet şirket rastgele olarak seçilerek eğitim seti oluşturulmuştur. Geriye kalan 503 şirket verisi ile de test seti oluşturulmuştur. %70 - %30 için belirlenen eğitim verisi ile seçilen yapay sinir ağı modeli eğitilmiş ve eğitim tamamlandıktan sonra performans tablosu oluşturulmuştur. Çizelge 5.24. incelendiğinde eğitimi tamamlanan yapay sinir ağı modeli seçilen 547 başarısız şirketin 355 tanesini başarısız olarak tahmin ederken 192 tanesini başarılı olarak tahmin etmiştir. Seçilen 625 başarılı olan şirketin de 613 tanesini doğru sınıflandırırken 12 tanesini yanlış sınıflandırmıştır. Sadece başarılı şirketler ele alındığında 625

şirketten 613 tanesini doğru sınıflandırmıştır. Başarılı şirketler için doğru tahmin etme oranı yaklaşık olarak %98 olarak bulunmuştur. Sadece başarısız şirketler ele alındığında 547 adet şirketten 355 tanesini doğru sınıflandırabilmiştir ve doğru sınıflandırma oranı başarısız şirketler için yaklaşık olarak %65 olarak bulunmuştur. Bakıldığında model başarılı şirketleri oldukça yüksek oranda doğru sınıflandırırken, başarısız şirketleri sınıflandırmada yeterli seviyede başarılı olmuştur. Genel orana baktığımızda ise doğru sınıflandırma oranı eğitim verisi için yaklaşık olarak %81 oranındadır.

Çizelge 5.25. %70 – %30 Test seti performans sonuçları

		Tahmin Edilen	
		Başarısız	**Başarılı**
Gözlenen	**Başarısız**	245	81
	Başarılı	33	144
Doğru Sınıflandırma Oranı = ((245+144)/503) = 0,7733 ≅ 0,77			

Çizelge 5.25. incelendiğinde test veri setinde seçilen başarısız 326 şirketin 245 tanesi doğru olarak sınıflandırılmış 81 tanesini yapay sinir ağı başarılı olarak sınıflandırmıştır. Başarısız şirketler için bakıldığında doğru sınıflandırma oranı yaklaşık olarak %75 olarak bulunmuştur. Başarılı olarak test verisine seçilen 177 şirketten 144 tanesi yapay sinir ağı modeli ile doğru sınıflandırılmış ve 33 tanesi başarısız olarak sınıflandırılmıştır. Başarılı şirketler için doğru sınıflandırma oranına bakıldığında %81 olarak bulunmuştur. Test seti için başarılı ve başarısız şirketlerin tamamının doğru sınıflandırılma oranına bakıldığında ise 503 şirketten başarısız olanların 245 tanesi başarılı olanların 144 tanesi toplamda 389 tane şirket doğru sınıflandırılmıştır. Test seti için doğru sınıflandırılma oranına bakıldığında ise %77 olarak bulunmuştur. %80 - %20 yapay sinir ağları modeline ait eğitim ve test veri için oluşturulan performans sonuç tabloları Çizelge 5.26. ve Çizelge 5.27.'de verilmiştir.

Çizelge 5.26. %80 – %20 Eğitim seti performans sonuçları

		Tahmin Edilen	
		Başarısız	**Başarılı**
Gözlenen	**Başarısız**	388	237
	Başarılı	149	566
Doğru Sınıflandırma Oranı = ((388+566)/1340) = 0,7044 ≅ 0,70			

1340 adet şirket rastgele olarak seçilerek eğitim seti oluşturulmuştur. Geriye kalan 335 şirket verisi ile de test seti oluşturulmuştur. %80 - %20 için belirlenen eğitim verisi ile seçilen yapay sinir ağı modeli eğitilmiş ve eğitim tamamlandıktan sonra performans tablosu oluşturulmuştur. Çizelge 5.26. incelendiğinde eğitimi tamamlanan yapay sinir ağı modeli seçilen 625 başarısız şirketin 388 tanesini başarısız olarak tahmin ederken 237 tanesini başarılı olarak tahmin etmiştir. Seçilen 715 başarılı olan şirketin de 566 tanesini doğru sınıflandırırken 149 tanesini yanlış sınıflandırmıştır. Sadece başarılı şirketler ele alındığında 715 şirketten 566 tanesini doğru sınıflandırmıştır. Başarılı şirketler için doğru tahmin etme oranı yaklaşık olarak %79 olarak bulunmuştur. Sadece başarısız şirketler ele alındığında 625 adet şirketten 388 tanesini doğru sınıflandırabilmiştir ve doğru sınıflandırma oranı başarısız şirketler için yaklaşık olarak %62 olarak bulunmuştur. Bakıldığında model başarılı şirketleri oldukça yüksek oranda doğru sınıflandırırken başarısız şirketleri sınıflandırmada yeterli seviyede başarılı olmuştur. Genel orana baktığımızda ise doğru sınıflandırma oranı eğitim verisi için yaklaşık olarak %70 oranındadır.

Çizelge 5.27. %80 – %20 Test seti performans sonuçları

		Tahmin Edilen	
		Başarısız	**Başarılı**
Gözlenen	**Başarısız**	101	57
	Başarılı	50	127
Doğru Sınıflandırma Oranı = ((101+127)/335) = 0,6805 ≅ 0,68			

Çizelge 5.27. incelendiğinde test veri setinde seçilen başarısız 158 şirketin 101 tanesi doğru olarak sınıflandırılmış 57 tanesini yapay sinir ağı başarılı olarak sınıflandırmıştır. Başarısız şirketler için bakıldığında doğru sınıflandırma oranı yaklaşık olarak %64 olarak bulunmuştur. Başarılı olarak test verisine seçilen 177

şirketten 127 tanesi yapay sinir ağı modeli ile doğru sınıflandırılmış ve 50 tanesi başarısız olarak sınıflandırılmıştır. Başarılı şirketler için doğru sınıflandırma oranına bakıldığında %72 olarak bulunmuştur. Test seti için başarılı ve başarısız şirketlerin tamamının doğru sınıflandırılma oranına bakıldığında ise 335 şirketten başarısız olanların 101 tanesi başarılı olanların 127 tanesi toplamda 228 tane şirket doğru sınıflandırılmıştır. Test seti için doğru sınıflandırılma oranına bakıldığında ise %68 olarak bulunmuştur.

Oluşturulan yapay sinir ağlarının Hata Kareler Ortalamasına (MSE) bakıldığında %70 - %30 test ve eğitim verisi için daha düşük bulunmuştu. Doğru sınıflandırma oranlarına da baktığımızda %70 - %30 test ve eğitim verisi için oluşturulan yapay sinir ağı modeli daha doğru sınıflandırma işlemi gerçekleştirmektedir. %70 - %30 için doğru sınıflandırma oranı yaklaşık olarak %77, %80 - %20 için doğru sınıflandırma oranı yaklaşık olarak %68 olarak bulunmuştur. Lojistik regresyon analizi ile karşılaştırma işleminde daha iyi sonucu veren %70 - %30 eğitim ve test verisi ile oluşturulan yapay sinir ağı modeli kullanılmıştır.

5.3.4. Bulguların karşılaştırılması

Bu bölümde lojistik regresyon analizi bölümünde seçilen en iyi model ile yapay sini ağları bölümünde seçilen en iyi model karşılaştırılmış ve sonuçları yorumlanmıştır. Lojistik regresyon analizi kısmında en iyi model olarak seçilen modele ilişkin performans tablosu Çizelge 5.28.'de verilmiştir.

Çizelge 5.28. Lojistik regresyon analizi performans tablosu

		Tahmin Edilen	
		Başarısız	**Başarılı**
Gözlenen	**Başarısız**	205	578
	Başarılı	149	743
Doğru Sınıflandırma Oranı = ((205+743)/1675)=0,57			

Çizelge 5.28.'de ikili lojistik regresyon analizi ile en yüksek sınıflandırma oranının %57 olduğu tespit edilmiştir. Sadece başarılı olan şirketler için doğru sınıflandırma oranına bakıldığında 892 başarılı olan şirketten 743 tanesini doğru

sınıflandırmış, 149 şirketi başarısız olarak sınıflandırmıştır. Sadece başarılı olan şirketler için ikili lojistik regresyon analizinin doğru sınıflandırma oranı yaklaşık olarak %83 olarak tespit edilmiştir. Başarısız olan şirketler için baktığımız da ise 783 başarısız olan girişim şirketinden sadece 205 tanesi doğru sınıflandırabilmiştir ve doğru sınıflandırma oranı yaklaşık olarak %26 olarak tespit edilmiştir. İkili lojistik regresyon analizinde kategorileri ayrı ayrı değerlendirdiğimizde başarılı olan şirketler için %83 doğru tahmin ile iyi bir sonuç vermişken başarısız olan şirketler için aynı şeyi söylemek mümkün değildir. %26 oranı ile başarısız olan şirketlerin tahmininde ikili lojistik regresyon analizi iyi bir sonuç vermemiştir. Yapay sinir ağları kısmında en iyi model olarak seçilen modele ilişkin performans tablosu Çizelge 5.29.'da verilmiştir.

Çizelge 5.29. Yapay sinir ağları performans sonuçları

		Tahmin Edilen	
		Başarısız	**Başarılı**
Gözlenen	**Başarısız**	245	81
	Başarılı	33	144
Doğru Sınıflandırma Oranı = ((245+144)/503) = 0,7733 ≅ 0,77			

Çizelge 5.29.'da yapay sinir ağları ile en yüksek sınıflandırma oranının yaklaşık olarak %77 olduğu tespit edilmiştir. Sadece başarılı olan şirketler için doğru sınıflandırma oranına bakıldığında 177 başarılı olan şirketten 144 tanesini doğru sınıflandırmış, 33 şirketi başarısız olarak sınıflandırmıştır. Sadece başarılı olan şirketler için yapay sinir ağlarının doğru sınıflandırma oranı yaklaşık olarak %81 olarak tespit edilmiştir. Başarısız olan şirketler için baktığımız da ise 326 başarısız olan girişim şirketinden 245 tanesi doğru sınıflandırılmıştır ve doğru sınıflandırma oranı yaklaşık olarak %75 olarak tespit edilmiştir.

İkili lojistik regresyon analiz ve yapay sinir ağları performansları karşılaştırıldığında başarılı ve başarısız olan şirketler için ayrı ayrı bakıldığında ikili lojistik regresyon analizi başarılı şirketleri %83 oranında doğru sınıflandırırken, yapay sinir ağlarında bu oran %81 olarak tespit edilmiştir. Başarılı olan şirketler için her iyi analiz yönteminde birbirine yakın ve yüksek

oranda doğru sınıflandırma işlemi yaptığı söylenebilmektedir. Ancak başarısız olan şirketlere bakıldığında ise yapay sinir ağlarının doğru sınıflandırma oranı %75'e, ikili lojistik regresyon analizinin doğru sınıflandırma oranı ise %26'ya kadar düşmüştür. İki analiz yöntemi de başarılı olan şirketleri tahmin etmede daha başarılı iken başarısız olan şirketler için doğru sınıflandırma oranları düşmüştür.

6. SONUÇ VE ÖNERİLER

Girişim şirketleri ve yatırımcılar için girişim şirketlerinin başarılı olması ve çıkış yapması oldukça önemlidir. Girişim şirketlerinin başarılı olması ekonomik ve teknolojik açıdan oldukça önemli bir yer tutmaktadır. Bir girişim şirketinin başarılı olma olasılığı oldukça düşüktür. Son on yıl içerisinde kurulan girişim şirketi sayısı oldukça artmıştır. Bir girişim şirketinin başarılı olabilmesi ve girişim ekosisteminde hayat sürdürebilmesi için en önemli nokta yatırım almasıdır. Birçok girişim şirketi yatırım aşamalarından elleri boş dönerken bazı girişim şirketleri istedikleri yatırım miktarlarını yatırımcılardan alabilirler. Yatırım alan girişim şirketlerinden birçoğu aldıkları yatırımları doğru kullanmadıkları için başarısız olurlar.

Girişim şirketlerinde yaşanan başarısızlıklar ülke ekonomisi üzerinde de olumsuz etkiler yaratmaktadır. Girişim şirketinin başarısızlığı melek yatırımcıları, risk sermayecilerini ve yatırım bankalarını da olumsuz yönde etkilemektedir. Bu nedenle bir girişim şirketinin aldıkları yatırımlar ile ileride başarılı olup olamayacağının öngörülmesi, girişim sektörü ve yatırım sektörü için oldukça önemlidir.

Bu çalışmada sınıflandırma problemlerinde kullanılan çok değişkenli istatistik yöntemlerinden lojistik regresyon analizi ve makine öğrenmesi tekniklerinden yapay sinir ağları yöntemlerinin sınıflandırma performanslarının karşılaştırılması amaçlanmıştır.

Bu çalışmada; melek, çekirdek ve risk aşamalarında yatırım alan ve bulut hizmetleri, büyük veri ve makine öğrenmesi, e-ticaret, mobil uygulama ve sosyal medya sektörlerinden birinde bulunan girişim şirketlerinin önceden başarılı olup olamayacağı ikili lojistik regresyon analizi ve ileri beslemeli geri yayılımlı yapay sinir ağları ile tespit edilmiş, girişim şirketlerinin başarılı olması ile aldıkları yatırım miktarları arasındaki ilişki ve girişim şirketinin başarılı olması ile bulunduğu sektör arasındaki ilişki incelenmiştir. Yapay sinir ağları ve ikili lojistik regresyon analizi yöntemleri kullanılarak girişim şirketlerinin başarılı olup

olmayacağının öngörülmesine yönelik erken uyarı modelleri tahmin edilmeye çalışılmıştır.

Girişim şirketinin bulunduğu sektör ve melek, çekirdek ve risk aşamalarında aldıkları yatırım miktarları bağımsız değişkenler olarak belirlendikten sonra girişim şirketlerinin başarı tahmini ikili lojistik regresyon analizi ile gerçekleştirilmiştir. İkili lojistik regresyon analizi ile girişim şirketlerin başarılı olup olmadığı %57 oranında doğru sınıflandırılmıştır. Başarılı olan 892 girişim şirketinden 149 tanesi ikili lojistik regresyon analizi tarafından başarısız olarak sınıflandırılmıştır. Başarısız olan 783 girişim şirketinden 578 tanesini yanlış sınıflandırarak başarılı olarak tahmin etmiştir. İkili lojistik regresyon analizi başarılı olan girişim şirketleri için %83 gibi yüksek bir oranda doğru sınıflandırma oranına sahipken, başarısız olan şirketlerin için bu oran %26 gibi oldukça düşük bir oran olarak bulunmuştur.

Yapay sinir ağları yöntemi uygulanmasında ağ ileri beslemeli geri yayılımlı sinir ağı algoritmasına göre %70 oranındaki eğitim verisi ile eğitilmiştir. Buna göre ağın eğitimi %81 başarı oranı ile tamamlanmıştır. %30 oranındaki test verisi eğitine ağ modelini test etmek için analiz edildiğinde, başarılı olan 177 şirketin 33 tanesi, başarısız olan 326 şirketin 81 tanesini yanlış sınıflandırmış ve doğru sınıflandırma oranı %77 olarak bulunmuştur. Yapay sinir ağları yönteminde başarılı olan girişim şirketlerinin doğru sınıflandırma oranı %81, başarısız olan girişim şirketlerinin doğru sınıflandırma oranı ise %75 olarak bulunmuştur.

İki yöntemin sonuçları karşılaştırıldığında; ikili lojistik regresyon analizi başarılı olan girişim şirketlerini tahmin etmede yapay sinir ağlarından çok küçük bir oranda daha iyidir. Ancak başarısız olan girişim şirketlerini tahmin etmede oldukça düşük bir başarıya sahip olduğu için genel başarı oranı oldukça düşük çıkmaktadır. Yapay sinir ağlarında ise başarılı olan şirketlerin doğru tahmin oranı ile başarısız olan şirketlerin doğru tahmin oranı birbirine oldukça yakındır. Bundan dolayı ikisinin arasında ve yüksek oranda kabul edilebilir bir sonuç elde edilebilmiştir.

Lojistik regresyon analizinin sahip olduğu %57'lik doğru sınıflandırma oranı neredeyse her 4 girişim şirketten 2 tanesini doğru 2 tanesini yanlış sınıflandırdığı anlamına gelmektedir. Girişim şirketlerinin melek aşamasında aldıkları ortalama 1 milyon dolar, çekirdek aşamasında aldıkları ortalama 1,5 milyon dolar ve risk aşamasında aldıkları ortalama 30 milyon dolar düşünüldüğünde girişim şirketinin başarılı olup olmayacağını %57'lik bir başarı ile tahmin edilmesi kabul edilebilir bir durum değildir. Aynı zamanda bir yatırımcı için başarılı şirketlerin doğru sınıflandırılmasından ziyade başarısız olması beklenen şirketlerin doğru sınıflandırılması çok daha önemlidir. Çünkü başarısız olacağı düşünülen bir şirketin başarılı olarak tahmin edilmesi, girişim şirketinin istediği miktardaki yatırımı alması ve daha sonrasında başarısız olması ile sonuçlanabilmektedir. Bu durum da yatırımcılar için oldukça riskli bir durumdur. Lojistik regresyon analizinde başarısız olması beklenen şirketlerin %74'ü başarılı olacak diye tahmin edilmektedir. Bu durum yatırımcılar için kabul edilemez bir durumdur.

Yapay sinir ağlarına bakıldığında ise %77'lik bir doğru sınıflandırma oranı mevcuttur. Bu oranda her 4 girişim şirketinden 3 tanesini doğru 1 tanesini yanlış sınıflandırdığı anlamına gelmektedir. Başarısız olması beklenen şirketlerin tahminde de %75'lik doğru tahmin de yatırımcılar için kabul edilebilir bir durumdur.

Yapay sinir ağlarının doğru sınıflandırma başarısının ikili lojistik regresyon analizine oranla daha yüksek olduğu, bu nedenle ikili lojistik regresyon analizi yerine yapay sinir ağlarının tercih edilebileceği söylenebilir.

Yapay sinir ağları lojistik regresyon analizinin aksine belli bir algoritma üzerine analiz yapmak yerine analizde kullanılan örneklemden kendine bir algoritma oluşturmaktadır. Bundan dolayı çalışmada kullanılan verilere daha iyi uyum sağladığı ve daha iyi bir sonuç verdiği söylenebilir.

Lojistik regresyon analizi ile uygun bir model oldukça kısa bir sürede oluşturulmaktadır. Yapay sinir ağlarında veriye uygun bir modelin oluşturulması oldukça fazla zaman almaktadır.

Yapay sinir ağları ile sınıflandırmanın lojistik regresyon analizi kullanılarak yapılan sınıflandırmadan daha iyi sonuçlar verme eğiliminde olduğu ayrıca yine aşırı eğitim, mimarinin hatalı oluşturulması vb. problemleri olmayan yapay sinir ağları modellerinin daha iyi öngörü performansı sağlayabildiği görülmektedir (Yurtoğlu, 2005).

Tahmin edilen modellerden lojistik regresyon analizi modelinin girişim şirketlerinin başarısızlığının öngörülmesinde kullanılmasının uygun olmadığı ancak yapay sinir ağları tarafından oluşturulan modelin girişim şirketlerinin başarısızlıklarının öngörülmesinde kullanılabilir olduğu söylenebilir.

KAYNAKLAR

Akgül, A., 2003. Tıbbi Araştırmalarda İstatistiksel Analiz Teknikleri: SPSS Uygulamaları. Yeni Mustafa Kitabevi,Ankara.

Alpar, R., 2013. Uygulamalı Çok Değişkenli İstatistiksel Yöntemler. Detay Yayıncılık, Ankara.

Altinirmak, S., Karamaşa, Ç., 2016. Comparison Of Machine Learning Techniques For Analyzing Banks' Financial Distress. Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 19(36), 291-304.

Altunöz, U., 2015. Lojistik Regresyon ve Diskriminant Yöntemleriyle Banka Başarısızlıklarının Tahmini: Türk Bankaları Deneyimi. Kahramanmaraş Sütçü İmam Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 5(1), 45-54.

Arıkan Kargı, V. S., 2015. Yapay sinir ağ modelleri ve bir tekstil firmasında uygulama. Ekin Kitapevi Yayınları.

Atasoy, D., 2001. Lojistik Regresyon Analizinin İncelenmesi ve Bir Uygulaması. Cumhuriyet Üniversitesi, Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, Sivas..

Ayan, T. Y., Değirmenci, N., 2018. Firma Finansal Başarısızlık Öngörüsü İçin Bir Lojistik Regresyon Modeli. Uluslararası İktisadi ve İdari İncelemeler Dergisi, 77-88.

Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., Vanthienen, J., 2003. Benchmarking state-of-the-art classification algorithms for credit scoring. Journal of the Operational Research Society, 54(6), 627-635.

Baron, R. A. 1998., Cognitive mechanisms in entrepreneurship. Journal of Business Venturing, 13(4), 275-294.

Berenson, M. L., Levine, D. M., 1996. Basic business statistics: Concepts and applications (6th ed). Prentice Hall.

Blank, S. G., 2007. The four steps to the epiphany: Successful strategies for products that win (3rd ed). S. G. Blank.

Budak, H., Erpolat, S., 2012. Kredi Riski Tahmininde Yapay Sinir Ağları ve Lojistik Regresyon Analizi Karşılaştırılması. AJIT-e: Bilişim Teknolojileri Online Dergisi, 3(9), 23-30.

Chang, S. J., 2004. Venture capital financing, strategic alliances, and the initial public offerings of Internet startups. Journal of Business Venturing, 19(5), 721-741.

Cheng, B., Titterington, D. M., 1994. Neural Networks: A Review from a Statistical Perspective. Statistical Science, 9(1), 2-30.

Cramer, J. S., 2003. The Origins of Logistic Regression. SSRN Electronic Journal.

Çelik, M. K., 2010. Bankaların Finansal Başarısızlıklarının Geleneksel ve Yeni Yöntemlerle Öngörüsü. Yönetim ve Ekonomi: Celal Bayar Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 17(2), 129-143.

Çıngı, H., 2009. Örnekleme Kuramı (3. bs). Bizim Büro Basımevi Yayın Dağ. Tic. Ltd. Şti.

Dasgupta, C. G., Dispensa, G. S., Ghose, S., 1994. Comparing the Predictive Performance of a Neural Network Model with Some Traditional Market Response Models. International Journal of Forecasting, 10(2), 235-244.

Dreiseitl, S., Ohno-Machado, L., 2002. Logistic Regression and Artificial Neural Network Classification Models: A Methodology Review. Journal of Biomedical Informatics, 35(5-6), 352-359.

Eberhart, R. C., Shi, Y., 2007. Computational Intelligence: Concepts to Implementations. Elsevier/Morgan Kaufmann Publishers.

Eisemann, T., Ries, E., Dillard, S., School, H. B., 2013. Hypothesis-driven Entrepreneurship: The Lean Startup. President and Fellows of Harvard College.

Elmas, Ç., 2011. Yapay Zekâ Uygulamaları: Yapay Sinir Ağı, Bulanık Mantık, Genetik Algoritma. Seçkin Yayıncılık.

Funding Rounds. (t.y.). Crunchbase. Geliş tarihi 15 Kasım 2019, gönderen https://www.crunchbase.com/search/funding_rounds

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep learning. The MIT Press.

Güney, S., 2019. Girişimcilik: Temel Kavramlar Ve Bazı Güncel Konular. Ankara; eBook Collection (EBSCOhost).

Heard, R., Sibert, J., 2000. Growing New Businesses with Seed and Venture Capital: State Experiences and Options.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer Feedforward Networks are Universal Approximators. Neural Networks, 2(5), 359-366.

Kaelbling, L. P., Littman, M. L., Moore, A. W., 1996. Reinforcement Learning: A Survey. Journal of Artificial Intelligence Research, 4, 237-285.

Kaygin, C. Y., Tazegül, A., Yazarkan, H., 2016. Estimation Capability of Financial Failures and Successes of Enterprises Using Data Mining and Logistic Regression Analysis. Ege Akademik Bakış Dergisi, 16(1), 147-159.

Kuzu, S., 2017. Yapay Sinir Ağı Modeli Yöntemiyle İşletme Başarısının Tahmin Edilmesi: 2008 Küresel Finansal Kriz Örneği. Siirt Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 5(10), 676-690.

Liou, F., 2008. Fraudulent Financial Reporting Detection and Business Failure Prediction Models: A comparison. Managerial Auditing Journal, 23(7), 650-662.

Nevruz, C. A., 2019. Lojistik Regresyon ile Kobi'lerin Kredibilite Ölçümü. T.C. Marmara Üniversitesi, Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, İstanbul.

Osborne, J. W., 2017. Regression linear Modeling: Best Practices and Modern Methods. Thousand Oaks, CA: SAGE Publications.

Öz, S., 2019. Girişimcilik: Girişimcinin Seyir Defteri. İstanbul; eBook Collection (EBSCOhost).

Öztemel, E., 2012. Yapay Sinir Ağları. Papatya Yayıncılık.

Paliwal, M., Kumar, U. A., 2009. Neural networks and statistical techniques: A review of applications. Expert Systems with Applications, 36(1), 2-17.

Perols, J., 2011. Financial Statement Fraud Detection: An Analysis of Statistical and Machine Learning Algorithms. AUDITING: A Journal of Practice & Theory, 30(2), 19-50.

Picken, J. C., 2017. From startup to scalable enterprise: Laying the foundation. Business Horizons, 60(5), 587-595.

Ries, E., 2011. The Lean Startup: How Today's Entrepreneurs Use Continuous İnnovation to Create Radically Successful Businesses (1st ed). Crown Business.

Ripley, B. D., 1994. Neural Networks: A Review from Statistical Perspective: Comment. Statistical Science, 9(1), 45-48.

Russel R. Russo., 2019. Neural Networks for Beginners: An Easy Textbook for Machine Learning Fundamentals to Guide You Implementing Neural Networks with Python and Deep Learning (Artificial Intelligence).

Simpson, P. K., 1990. Artificial Neural Systems: Foundations, Paradigms, Applications, and Implementations (1st ed). Pergamon Press.

Swingler, K., 2001. Applying Neural Networks: A Practical Guide (3. printing). Kaufman.

Tabachnick, B. G., Fidell, L. S. 2015. Çok Değişkenli İstatistiklerin Kullanımı (M. Baloğlu, Çev.). Nobel.

Tam, K. Y., Kiang, M. Y., 1992. Managerial Applications of Neural Networks: The Case of Bank Failure Predictions. Management Science, 38(7), 926-947.

Tatlıdil, H., 2002. Uygulamalı Çok Değişkenli İstatistiksel Analiz. Akademi Matbaası.

Teece, D. J., 2010. Business Models, Business Strategy and Innovation. Long Range Planning, 43(2-3), 172-194.

Trimi, S., Berbegal-Mirabent, J., 2012. Business Model Innovation in Entrepreneurship. International Entrepreneurship and Management Journal, 8(4), 449-465.

Tu, J. V., 1996. Advantages and Disadvantages of Using Artificial Neural Networks Versus Logistic Regression for Predicting Medical Outcomes. Journal of Clinical Epidemiology, 49(11), 1225-1231.

White, H., 1989. Learning in Artificial Neural Networks: A Statistical Perspective. Neural Computation, 1(4), 425-464.

Yurtoğlu, H., 2005. Yapay Sinir Ağları Metodolojisi İle Öngörü Modellemesi: Bazı Makroeknomik Değişkenler İçin Türkiye Örneği. DPT.

Zhang, G., Eddy Patuwo, B., Y. Hu, M., 1998. Forecasting with Artificial Neural Networks: International Journal of Forecasting, 14(1), 35-62.

EKLER

EK A. Lojistik Regresyon Analizi Kodları

```
#Verilerin R Studio'ya Aktarılması

> library(readxl)
> Data=read_excel("Data.xlsx")
>
Data$Big_Data_Machine_Learning=as.factor(Data$Big_Data_Machine_Learn
ing)
> Data$Cloud_Computing=as.factor(Data$Cloud_Computing)
> Data$E_Commerce=as.factor(Data$E_Commerce)
> Data$Mobile_Apps=as.factor(Data$Mobile_Apps)
> Data$Social_Media=as.factor(Data$Social_Media)

#Şekiller
> library(ggplot2)

#Şekil 4.1
> Box_Angel <- ggplot(Data,aes(x=Statu, y=Angel, fill=Statu),group =
1) +
  geom_boxplot(outlier.shape=NA)+coord_cartesian(ylim =
quantile(Data$Angel, c(0.1, 0.9)))+ labs(title="",x="Başarı Durumu",
y = "Melek Yatırım Miktarı")
> Box_Angel + scale_fill_discrete(name = "Başarı Durumu", labels =
c("Başarısız", "Başarılı")) + scale_x_discrete(labels=c("Başarısız",
"Başarılı"))

#Şekil 4.2
> Box_Seed <- ggplot(Data,aes(x=Statu, y=Seed, fill=Statu),group =
1) +
  geom_boxplot(outlier.shape=NA)+coord_cartesian(ylim =
quantile(Data$Seed, c(0.1, 0.9)))+ labs(title="",x="Başarı Durumu",
y = "Çekirdek Yatırım Miktarı")
> Box_Seed + scale_fill_discrete(name = "Başarı Durumu", labels =
c("Başarısız", "Başarılı")) + scale_x_discrete(labels=c("Başarısız",
"Başarılı"))

#Şekil 4.3
> Box_Venture <- ggplot(Data,aes(x=Statu, y=Angel, fill=Statu),group
= 1) +
  geom_boxplot(outlier.shape=NA)+coord_cartesian(ylim =
quantile(Data$Venture, c(0.1, 0.9)))+ labs(title="",x="Başarı
Durumu", y = "Risk Yatırım Miktarı")
> Box_Venture + scale_fill_discrete(name = "Başarı Durumu", labels =
c("Başarısız", "Başarılı")) + scale_x_discrete(labels=c("Başarısız",
"Başarılı"))



#Lojistik Regresyon Analizi

#Model 1
> Logit_Model_1 = glm(Statu~ Big_Data_Machine_Learning +
Cloud_Computing + E_Commerce + Mobile_Apps + Social_Media + Angel +
Seed + Venture ,family = "binomial", data = Data )
> summary(Logit_Model_1)
> Data$Predict_1 <- ifelse(Logit_Model_1$fitted.values >0.5,"1","0")
> Table_1 <- table(Data$Statu,Data$Predict_1)
```

```
> confusionMatrix(Table_1)

#Model 2

Logit_Model_2 = glm(Statu~ Big_Data_Machine_Learning + E_Commerce +
Social_Media + Venture ,family = "binomial", data = Data )
summary(Logit_Model_2)
Data$Predict_2 <- ifelse(Logit_Model_2$fitted.values >0.5,"1","0")
Table_2 <- table(Data$Statu,Data$Predict_2)
confusionMatrix(Table_2)

#Model 1 ve Model 2 AIC
> library(stats)
> AIC(Logit_Model_1,Logit_Model_2)
```

EK B. Yapay Sinir Ağları Kodları

```
#Verilerin R Studio'ya Aktarılması

> library(readxl)
> Data=read_excel("Data.xlsx")
> Data_Normal = Data
> for(i in 1:9) {Data[,i] =(Data[,i]-min(Data[,i]))/(max(Data[,i])-
min(Data[,i])) }

#70-30 Verileri
> Index_70 = sample(1:nrow(Data),1172)
> Train_Data_70 = Data[Index_70,]
> Test_Data_70 = Data[-Index_70,]

#80-30 Verileri
> Index_80 = sample(1:nrow(Data),1340)
> Train_Data_80 = Data[Index_80,]
> Test_Data_80 = Data[-Index_80,]

#Model Oluşturma
> library(neuralnet)

> Model_70_3_3 = neuralnet(Statu~ Social_Media + Could_Computing +
Mobile_Apps + E_Commerce + Big_Data_Machine_Learning + Angel + Seed
+ Venture, data = train_data, hidden = c(3,3), linear.output =
FALSE, stepmax = 1e6 )
> plot(Model_70_3_3)

> Model_80_3_2 = neuralnet(Statu~ Social_Media + Could_Computing +
Mobile_Apps + E_Commerce + Big_Data_Machine_Learning + Angel + Seed
+ Venture, data = train_data, hidden = c(3,2), linear.output =
FALSE, stepmax = 1e6 )
> plot(Model_80_3_2)

#Eğitim Verisi İçin Gerçek, Tahmini Verilerin ve MSE Değerinin
Oluşturması

> Output_70_3_3 = compute(Model_70_3_3, Train_Data_70[,-9])
> Prediction_70_3_3 = Output_70_3_3$net.result * (max(Data_Normal[-
Index_70,9])-min(Data_Normal[-Index_70,,9]))+min(Data_Normal[-
Index_70,9])
> Actual_70_3_3 = Data_Normal[-Index_70,9]
> Actual_70_3_3 = as.numeric(Actual_70_3_3$Statu)
> MSE_70_3_3 = sum((Prediction_70_3_3 -
Actual_70_3_3)^2)/nrow(Train_Data_70)
> table(Actual_70_3_3,round(Prediction_70_3_3))
> MSE_70_3_3

#Test Verisi İçin Gerçek ve Tahmini Verilerin Oluşturulması
> Output_70_3_3_Test = compute(Model_70_3_3, Test_Data_70[,-9])
> Prediction_70_3_3_Test = Output_70_3_3_Test$net.result *
(max(Data_Normal[-ind,9])-min(Data_Normal[-
ind,9]))+min(Data_Normal[-ind,9])
> Actual_70_3_3_Test = Data_Normal[Index_70,9]
> Actual_70_3_3_Test = as.numeric(Actual_70_3_3_Test$Statu)
> table(Actual_70_3_3_Test,round(Prediction_70_3_3_Test))
```

ÖZGEÇMİŞ



Adı Soyadı : Kubilay ERİŞLİK

Doğum Yeri ve Yılı : Osmangazi, 23/04/1994

Medeni Hali : Bekar

Yabancı Dili : İngilizce

E-posta : kubilayerislik@ticaret.edu.tr

Eğitim Durumu

Lise : Küçükçekmece Gazi Anadolu Lisesi, 2012

Lisans : İstanbul Ticaret Üniversitesi,
İnsan ve Toplum Bilimleri Fakültesi, İstatistik Bölümü

Mesleki Deneyim

Kuramsoft Yazılım Bilgisayar
Güvenlik Reklam Danışmanlık Ltd. Şti. 2018-2019

İstanbul Ticaret Üniversitesi,
İnsan ve Toplum Bilimleri Fakültesi 2019-(devam ediyor)

Yayınları

Erişlik, K., Deniz Başar, Ö., 2020. Girişim Şirketlerine Yapılan Yatırımların Sektörlerinin Yapay Sinir Ağları ve Lojistik Regresyon Analizi ile Tahmin Edilmesi, 20. Uluslararası Ekonometri, Yöneylem Araştırması ve İstatistik Sempozyumu, Ankara, 12-14 Şubat, 282-283.