



T.C.
BURSA TEKNİK ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

**MAKİNE ÖĞRENMESİ İLE GİRİŞİMLERİN (STARTUPLARIN)
BAŞARILARININ TAHMİN EDİLMESİ**

YÜKSEK LİSANS TEZİ

Melih BOZ

Akıllı Sistemler Anabilim Dalı

TEMMUZ 2023

T.C.
BURSA TEKNİK ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

MAKİNE ÖĞRENMESİ İLE GİRİŞİMLERİN (STARTUPLARIN)
BAŞARILARININ TAHMİN EDİLMESİ

YÜKSEK LİSANS TEZİ

Melih BOZ

Akıllı Sistemler Anabilim Dalı

Akıllı Sistemler Mühendisliği Tezli Yüksek Lisans Programı

Tez Danışmanı: Doç. Dr. Hasan ŞAHİN

TEMMUZ 2023

BTÜ, Lisansüstü Eğitim Enstitüsü'nün 20324814009 numaralı Yüksek Lisans Öğrencisi Melih BOZ, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı “MAKİNE ÖĞRENMESİ İLE GİRİŞİMLERİN BAŞARILARININ TAHMİN EDİLMESİ” başlıklı tezini aşağıda imzaları olan jüri önünde başarı ile sunmuştur.

Tez Danışmanı : **Doç. Dr. Hasan ŞAHİN**
Bursa Teknik Üniversitesi

Jüri Üyeleri : **Doç. Dr. Ömer Faruk Efe**
Bursa Teknik Üniversitesi

Doç. Dr. Halit Öztekin
Sakarya Uygulamaları Bilimler Üniversitesi

Teslim Tarihi **: 19 Temmuz 2023**
Savunma Tarihi **: 31 Temmuz 2023**



20.04.2016 tarihli Resmi Gazete’de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince; Bu Lisansüstü teze, Bursa Teknik Üniversitesi’nin aboneli olduğu intihal yazılım programı kullanılarak Lisansüstü Eğitim Enstitüsü’nün belirlemiş olduğu ölçütlere uygun rapor alınmıştır.

İNTİHAL BEYANI

Bu tezde görsel, işitsel ve yazılı biçimde sunulan tüm bilgi ve sonuçların akademik ve etik kurallara uyularak tarafımdan elde edildiğini, tez içinde yer alan ancak bu çalışmaya özgü olmayan tüm sonuç ve bilgileri tezde kaynak göstererek belgelediğimi, aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul ettiğimi beyan ederim.

Öğrencinin Adı Soyadı: Melih BOZ

İmzası:





Türkiye Girişimcilik Ekosisteminin Gelişmesine,

ÖNSÖZ

Bu tez çalışması, tamamlanması için beni cesaretlendiren, destekleyen ve ilham veren birçok kişiye ithaf edilmiştir.

Öncelikle tez danışmanım Doç. Dr. Hasan Şahin'e desteği ve yönlendirmesiyle tezi hazırlamamda yardımcı olduğu ve bana değerli önerilerde bulunup, beni her adımda desteklediği için teşekkürlerimi sunarım.

Bu yüksek lisans programına başlamamda ve devam ettirmemde büyük katkıları olan Coşkunöz Holding, C Ventures ve Yeşilova Holding yönetimine teşekkür etmek istiyorum. Girişimcilik ve yatırımcılık ekosistemini birlikte keşfettiğimiz değerli çalışma arkadaşım Efe Tokgöz'e katkıları için ayrıca teşekkür ederim.

Aileme ve sevgili eşim Semra'ya ve biricik kızım Elisa'ya sonsuz teşekkürlerimi sunmak istiyorum. Destekleri, anlayışları ve sabırları olmadan bu tezi tamamlamak imkansız olurdu. Her zaman yanımda oldukları ve beni motive ettikleri için minnettarım.

Temmuz 2023

Melih BOZ

İÇİNDEKİLER

Sayfa

ÖNSÖZ.....	vii
İÇİNDEKİLER	viii
KISALTMALAR	x
SEMBOLLER	xi
ÇİZELGE LİSTESİ.....	xii
ŞEKİL LİSTESİ.....	xiii
ÖZET.....	xiv
SUMMARY	xv
1. GİRİŞ	1
1.1 Risk Sermayesi (Venture Capital).....	2
1.2 Girişim Yatırım Süreçleri.....	3
1.3 Yatırım Süreçleri	5
1.3.1 Anlaşma kaynağı bulma.....	5
1.3.2 Yatırım	6
1.3.3 Destekleme.....	7
1.3.4 Çıkış (Exit).....	8
1.4 Girişimcilik Ekosistemi.....	8
1.4.1 Başarılı girişimler – unicornlar	8
1.4.2 Girişimlerin başarısız olma sebepleri.....	14
1.5 Tezin Amacı	15
2. LİTERATÜR ARAŞTIRMASI	16
2.1 Girişim Ekosistemi.....	16
2.2 Makine Öğrenmesi Literatür Araştırması.....	18
2.2.1 Lojistik Regresyon (LR)	18
2.2.1.1 Lojistik regresyon çalışma prensibi	18
2.2.1.2 Lojistik regresyon kullanım örnekleri.....	21
2.2.2 Destek vektör makineleri (support vector machine).....	21
2.2.2.1 Destek vektör makineleri çalışma prensibi	22
2.2.2.2 Destek vektör makineleri kullanım örnekleri.....	23
2.2.3 Rastgele orman (random forest).....	24
2.2.3.1 Rastgele orman çalışma prensibi.....	24
2.2.3.2 Rastgele orman kullanım örnekleri	26
2.2.4 Karar ağaçları (decision tree).....	26
2.2.5 Sinir ağları (neural network).....	26
2.2.6 Kümeleme (clustering).....	26
2.2.7 Takviyeli öğrenme (reinforce learning)	27
2.2.8 Gauss süreçleri (gaussian process).....	27
2.2.9 Ensemble öğrenme (ensemble learning).....	27
2.2.10 Derin öğrenme (deep learning)	27
2.2.11 Transfer öğrenme (transfer learning)	28

2.3 Düşük Kod ve Kodsuz Geliştirme Platformları (No-code /Low-code).....	30
2.3.1 Orange Data Mining.....	30
2.3.2 KNIME.....	32
3. ARAŞTIRMA MODELİ ve METOT.....	34
3.1 Verinin Kullanılacak Hale Getirilmesi.....	36
3.2 Ölçüm Metrikleri.....	37
4. ANALİZ ve BULGULAR.....	40
5. SONUÇ VE ÖNERİLER.....	43
KAYNAKLAR	45
ÖZGEÇMİŞ.....	Hata! Yer işareti tanımlanmamış.



KISALTMALAR

CVC	: Corporate Venture Capital (Kurumsal Risk Sermayesi)
VC	: Venture Capital (Risk Sermayesi)
PE	: Private Equity (Özel Sermaye)
ABD	: Amerika Birleşik Devletleri
VC	: Customer Relationship Management (Müşteri İlişkileri Yönetimi)
UNICORN	: Piyasa değeri 1 Milyar Doları Aşmış Girişim Şirketleri
DVM	: Destek Vektör Makineleri
LR	: Lojistik Regresyon
TNR	: True Negative Rate
TPR	: True Positive Rate

SEMBOLLER

USD : Amerikan Doları



ÇİZELGE LİSTESİ

Sayfa

Çizelge 1.1 : Hizmet edilen endüstriye göre unicornların toplam değerlemesi ve toplam sayıları.	9
Çizelge 1.2 : Ülkelerin sahip olduğu unicorn şirketlerin sayıları, değerleri ve GSYİH ile karşılaştırılması.	9
Çizelge 1.3 (devamı) : Ülkelerin sahip olduğu unicorn şirketlerin sayıları, değerleri ve GSYİH ile karşılaştırılması.	10
Çizelge 1.4 : Yıllara göre unicorn sayıları ve toplam değerlemeleri.	12
Çizelge 1.5 : Şekil 1.5 'e ait açıklama tablosu.	13
Çizelge 3.1 : Ham veri seti sütun isimleri.....	34
Çizelge 3.2 : Eğitimde kullanılacak veri setine ait sütunlar.	36
Çizelge 3.3 (devam) : Eğitimde kullanılacak veri setine ait sütunlar.	37

ŞEKİL LİSTESİ

Sayfa

Şekil 1.1 : Girişim sermayesi firmasının yapısı (Risk Sermayesi Diyagramı, 2023)..	2
Şekil 1.2 : Bir girişimin fikirden halka arza tüm süreci (A. Vital, How Startup Funding Works, 2013).....	4
Şekil 1.3 : Risk sermayesi yatırım hunisi (Fernandes, 2014)	5
Şekil 1.4 : Yıllara göre unicorn sayıları ve toplam değerlemelerinin görseli.....	12
Şekil 1.5 : Ülkelerin sektörlere göre sahip olduğu unicorn sayıları	13

MAKİNE ÖĞRENMESİ İLE GİRİŞİMLERİN (STARTUPLARIN) BAŞARILARININ TAHMİN EDİLMESİ

ÖZET

İyi girişimler genellikle basit bir fikirden ve belirli bir ihtiyaç ve mevcut bir pazar boşluğunu doldurmak için bir çözüm üreten birkaç kişiden oluşur. Diğer tarafta, üretilen çözümü gerçeğe dönüştürmek için sermaye ve bazen gerekli yardımı sağlayan melek yatırımcılar, risk sermayedarları ve kurumsal risk sermayeleri yani yatırımcılar vardır. Eğer bir yatırımcıysanız, her zaman bilmek isteyeceğiniz birkaç şey vardır: “İyi şirketler nerede?”, “Nasıl öğrenilir?” ve “Yatırım yapmaya değer mi?”. Yatırımcıların keşif yetenekleri genellikle yıllarca süren ağ (network) oluşturma ve markalaşma çabalarıyla oluşur. Doğru yatırımı yapmayı öğrenmek ise bir girişimin neden başarılı olurken diğerlerinin başarısız olduğunu sezgisel olarak çıkarmaya çalışmaktan ortaya çıkar.

Finansal hizmetler endüstrisinde girişim sermayesi yatırımları, yüksek riskli, yüksek getirili varlık sınıfı olarak kabul edilmektedir. Son yıllarda yatırım yapmak inanılmaz derecede rekabetçi hale gelmiş ve samanlıkta kişiyi zengin edecek iğneyi bulmak hiç bu kadar zor olmamıştır. Girişim yatırımlarında genellikle sektör kabulü olarak 6 adım bulunmaktadır. Bunlar; anlaşma kaynağı bulma (deal sourcing), anlaşma seçimi (deal selection), değerlendirme (valuation), anlaşma yapısı (deal structure), yatırım sonrası katma değer (post investment value added), çıkış (exits)’tir. Risk sermayesi analistleri mümkün olduğu kadar çok fazla girişim ile görüşüp yatırım tezlerine uygun olan girişimleri değerlendirmeleri gerekmektedir ve görüşülen çoğu girişim çoğu zaman yatırım tezine uygun olmamaktadır.

Herhangi bir girişim sermayesinin ya da fonun yaptığı temel varsayım, yatırımlarının çoğunun bir kayıp olacağı ve tüm geri dönüşün yapılan yatırımların en fazla %10'u tarafından yönlendirileceğidir. Risk-getiri profili bu nedenle çok istikrarsızdır, bu da yatırımcıları yalnızca “unicorn”a (değerlemesi 1 milyar doları geçen şirket) dönüşebilecek fırsatların peşine düşmeye ve iyi olmasına rağmen yapamayacakları yatırımları reddetmeye zorlar. Bununla birlikte, bir şirketin başarısını daha iyi tahmin etmenin bir yolu olsaydı, bir yatırımcı, unicorn aramak yerine, tümü 2 kat getiri sağlayan bir şirketler portföyüne kolayca yatırım yapabilir ve sonunda günümüzde çoğu fonun elde ettiğinden daha yüksek bir getiri elde edebilir.

Bu bağlamda bu tezin amacı; Türkiye’deki girişim ekosisteminde en büyük payı alan (Global Startup Ecosystem Report 2022, 2023) oyun, finansal teknolojiler ve yapay zekâ alanındaki girişimlerden yola çıkarak potansiyel yatırımların riskini daha iyi değerlendirmek için girişim sermayesi yatırımlarında karar verme süreçlerinde makine öğrenmesi algoritmalarının; oyun, finansal teknolojiler ve yapay zeka alanındaki girişimlerin gelecekteki başarılarının tahmin edilmesindeki etkiyi araştırmaktır.

Anahtar kelimeler: Makine Öğrenimi, Girişimcilik, Girişim Sermayesi, Tahmin

PREDICTING THE SUCCESS OF START-UPS WITH MACHINE LEARNING

SUMMARY

Good startups often originate from a simple idea and a few people identifying a solution to address a particular need and an existing market gap. On the other hand, there are angel investors, venture capitalists and corporate venture capitalists who provide the capital, and sometimes the necessary assistance, to make this dream a reality. If you are an investor, there are a few things you will always want to know: “Where are the good startups?”, “How to learn?” and “Is it worth the investment?”. Discovery capabilities are often built through years of networking and branding efforts. Making the right investment comes from intuitively trying to figure out why one venture succeeds and others fail.

Investing has become incredibly competitive in recent years and looking for a the needle in a haystack that will make you rich has never been more difficult. The results show that machine learning can support venture investors in their decision-making processes to find opportunities and better assess the risk of potential investments.

Good startups usually consist of a simple idea and a few people who come up with a solution to fill a specific need and an existing market gap. On the other hand, there are angel investors, venture capitalists and corporate venture capitalists who provide the capital and sometimes the necessary assistance to make the solution a reality. If you are an investor, there are a few things you will always want to know: “Where are the good startups?”, “How to learn?” and “Is it worth the investment?”. Investors often build discovery capabilities through years of networking and branding efforts. Learning to make the right investment comes from intuitively trying to figure out why one venture succeeds and others fail.

In the Financial Services Industry, venture capital investments are considered a high-risk, high-yielding asset class. Investing has become incredibly competitive in recent years, and “looking for a the needle in a haystack” that will make you rich has never been more difficult. In venture investments, there are usually 6 steps as sector acceptance. These are; deal sourcing, deal selection, valuation, deal structure, post investment value added, exit. Venture capital analysts need to interview as many startups as possible and evaluate the startups that fit their investment thesis, and most startups interviewed often do not fit the investment thesis.

The basic assumption any venture capital or fund makes is that most of their investment will be a loss and all return will be driven by no more than 10% of the investment made. The risk-return profile is therefore very unstable, forcing investors to seek opportunities that can only turn into “unicorns” (a company valued at more than \$1 billion) and reject investments that are good but cannot make. However, if there were a better way to predict a company's success, an investor, instead of looking

for a unicorn, could easily invest in a portfolio of companies all yielding 2x the return, and end up with a higher return than most funds today.

In this context, the aim of this study is to investigate the effect of machine learning algorithm in predicting the future success of venture capital investments in order to better evaluate the risk of potential investments based on the startups in the field of game, financial technologies and artificial intelligence, which have the largest share in the entrepreneurial ecosystem in Turkey.

Keywords: Machine Learning, Entrepreneurship, Venture Capital, Prediction



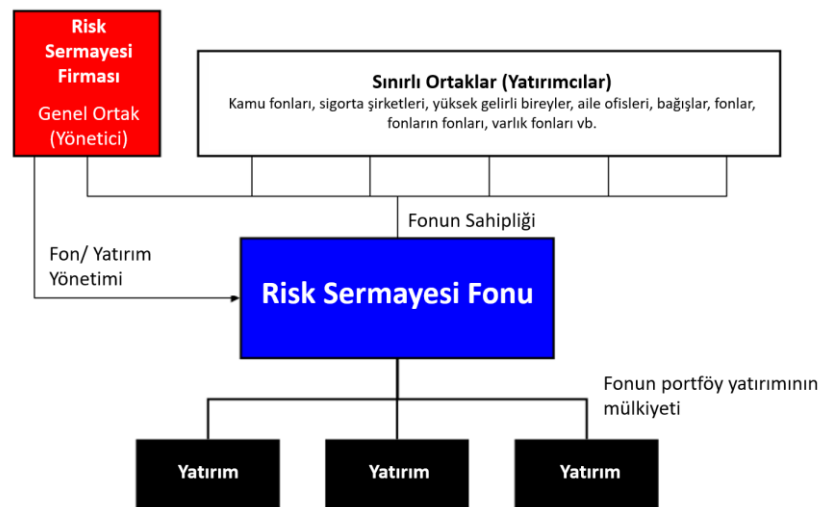
1. GİRİŞ

Yapay zeka (YZ), özellikle de Makine öğrenimi (MÖ), son yıllarda hayatımızı ve yaşamımızı her zamankinden daha fazla etkilemektedir. Yapay Zeka teknolojisinin birçok uygulama alanı ve sınırsız olanakları vardır. Bilgi işlem altyapısı iyileştirmeleri nedeniyle, belirli Yapay Zeka algoritmaları günümüzde insan uzmanlardan daha iyi performans göstermektedir. Kuralları çok net bir şekilde tanımlanmış, sürekli yapılan ve kompleks işlerde (otonom sürüş, tıbbi tanı koyma vb) yapay zekanın insanlar üzerindeki üstünlüğünü kanıtlanmıştır ve giderek daha iyi hale gelmektedir. Yapay zekanın farklı alanlarda artan kullanımı, bir çok iyileştirme (sıfır trafik kazası, doğru teşhis ve doğru tedavi) vaad etmektedir. Bu gelişmeleri finansal teknolojiler tarafında da görmek mümkündür. Örneğin bir girişime fon sağlamaya ilişkin karar verme süreçlerini otomatikleştirmek ve iyileştirmek için startup ekosisteminin dinamiklerini analiz etmek gibi mükemmel bir potansiyel sunuyor. İstatistiksel olarak, yaklaşık %20 girişim yatırımı başarılı oluyor, %5-10'u 10-20x'lik sonuçlar veriyor ve sadece %1'i 100x'ten daha yüksek sonuçlar veriyor (D. S. Hunter and T. Zaman, 2017). Dağıtılan sermayenin yaklaşık yüzde sekseni verimli bir şekilde yanlış uygulanır ve getiri sağlamaz. Küresel risk sermayesi pazarı, 2013 yılından 2021 yılına kadar artışını sürdürmüştür. 2022 yılında ise yaşanan küresel olaylar sebebi le 600 milyar dolarlık bir pazardan 400 milyar dolarlık bir pazara düştü. Bu değer sektör için halen umut verici. (Global Funding Slide In 2022 Sets Stage For Another Tough Year, 2023) bu da girişim yatırımlarında bir artış ve girişim şirketleri için görece finansman kolaylığı anlamına geliyor. Girişim sermayesi yatırımları, yüksek riskli, yüksek getirili varlık sınıfı olarak kabul edilir. Girişim yatırımına karar verme, önceki anlaşmalara, şirket verilerine, kurucu verilerine ve daha fazlasına uygulanan makine öğrenimi ile optimize edilebilir. Yatırım onay süreçleri şirketten şirkete değişse de girişim ile tanışmadan yatırımın tamamlanmasına kadar geçen süreç ortalama 3 ayı bulmaktadır. Burada en büyük dar boğaz doğru girişimin bulunmasıdır ve bunun için şirketlerdeki analistler sürekli yeni girişimlerle tanışmakta ve onları anlamaya çalışmaktadır. Ayrıca değerlendirilen verilere bakıldığında iki tip veri bulunmaktadır soyut veriler ve somut

veriler. Makineler, yeni bir girişimin başarısını gösteren büyük miktarda "somut" verileri (endüstri, çalışan sayısı, patent sayısı, alınan yatırım vb.) tutarlı bir şekilde işlemede özellikle faydalı olsa da, insanlar "soyut" verileri (inovatiflik seviyesi, girişimcinin karakteri ve tutarlılığı, ekibin uyumu vb.) yorumlamada üstündür (Dellermann, Lipusch, Ebel, & Popp, 2017). Bu sebeple ilk etapta geliştirilecek olan sistemlerin makine-insan işbirliğine ihtiyacı vardır. Bir yapay zeka algoritmasının bu somut verileri analiz etmesi oldukça olasıdır. Yalnızca en iyi adayları seçmekle kalmaz, aynı zamanda makine öğrenimi potansiyel olarak bir risk sermayesi şirketindeki diğer zorlukları çözmek için kullanılabilir.

1.1 Risk Sermayesi (Venture Capital)

Girişim Sermayesi, Sınırlı (Limited) Ortaklar olarak adlandırılan bireysel yatırımcılar tarafından, öz sermaye (işletmenin kısmi mülkiyeti) karşılığında başarılı işletmelere dönüşme potansiyeline sahip özel şirketlere yönelik sağlanan bir tür özel sermaye finansmanıdır. Bu yatırımcılar, risk sermayesi fonu çatısı altında yönetilen risk sermayedarları tarafından sağlanan sermayeyi kullanarak, yeni ve hızla büyüyen işletmeleri (startup şirketleri) finanse etmekte ve yeni ürün veya hizmetlerin geliştirilmesinde stratejik rehberlik ve yardım sağlamaktadır (Şekil 1.1). Risk sermayedarları, bu tür grupların yüksek düzeyde risk ve belirsizlik içerdiğinin farkında olduklarından dolayı, şirketi satma veya halka arz etme gibi uzun vadeli beklentilere dayalı olarak kısa vadeli getirilerden ziyade büyük miktarlarda özel sermaye yatırımı yapmaktadırlar (Risk Sermayesi, 2023).



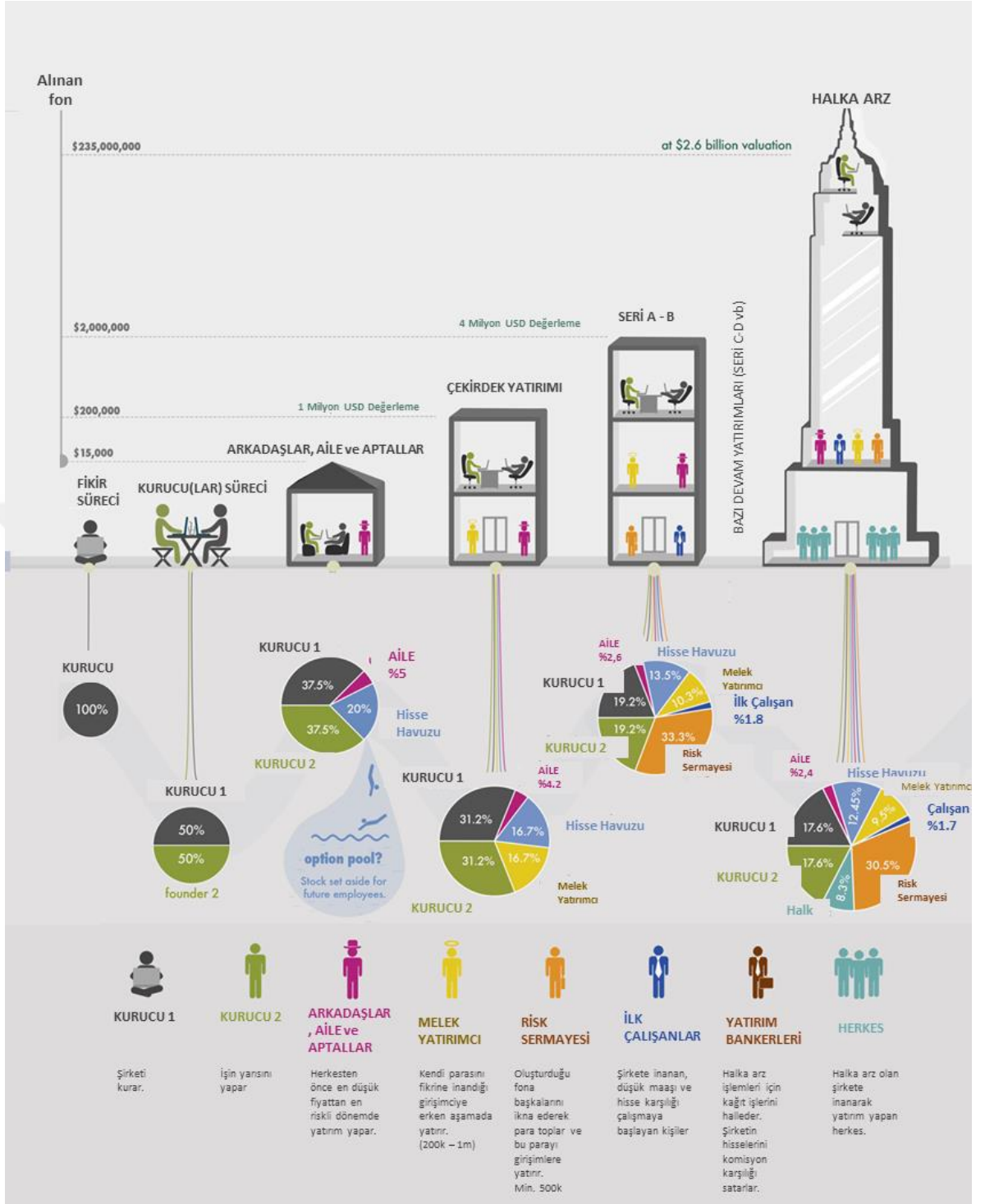
Şekil 1.1 : Girişim sermayesi firmasının yapısı (Risk Sermayesi Diyagramı, 2023).

1.2 Girişim Yatırım Süreçleri

Girişimcilik, temelde hayal kurma süreciyle başlayan ve değerlerin ortaya çıkmasıyla sonuçlanan bir süreçtir. İçinde düşünsellik ve eylem unsurlarını barındıran girişimcilik kavramı, ilk olarak 1730'lu yıllarda Richard Cantillon tarafından kâr elde etmek amacıyla işlerin düzenlenmesi ve iş riskinin üstlenilmesi olarak tanımlanmıştır. Girişimcilik kavramı, literatüre ilk giriş yıllarında bir işe başlama, bir iş kurma ve bu süreçte risk alarak kazanç elde etme şeklinde tanımlanmıştır. Daha sonraki dönemlerde ise tüketicilerin istek ve ihtiyaçlarının fark edilmesi, fikirlerin ortaya çıkarılması, fikirlerin ürün ve hizmetlere dönüştürülmesi ve pazara sunulması gibi faaliyetler de kavrama dahil edilerek genişletilmiştir (Zhao, 2005).

Bir girişim kurulduğu günden halka arz olduğu güne kadar bir çok aşamada fon toplayabilir. Risk sermayesinin girişim yatırım süreci gibi, girişimler de belirli dönüm noktalarında farklı kurumlardan hisse karşılığında fon toplayabilir. Girişim ekosisteminde kabul görmüş bir yaklaşım Şekil 1.2'de belirtilmiştir. Bu aşamalar:

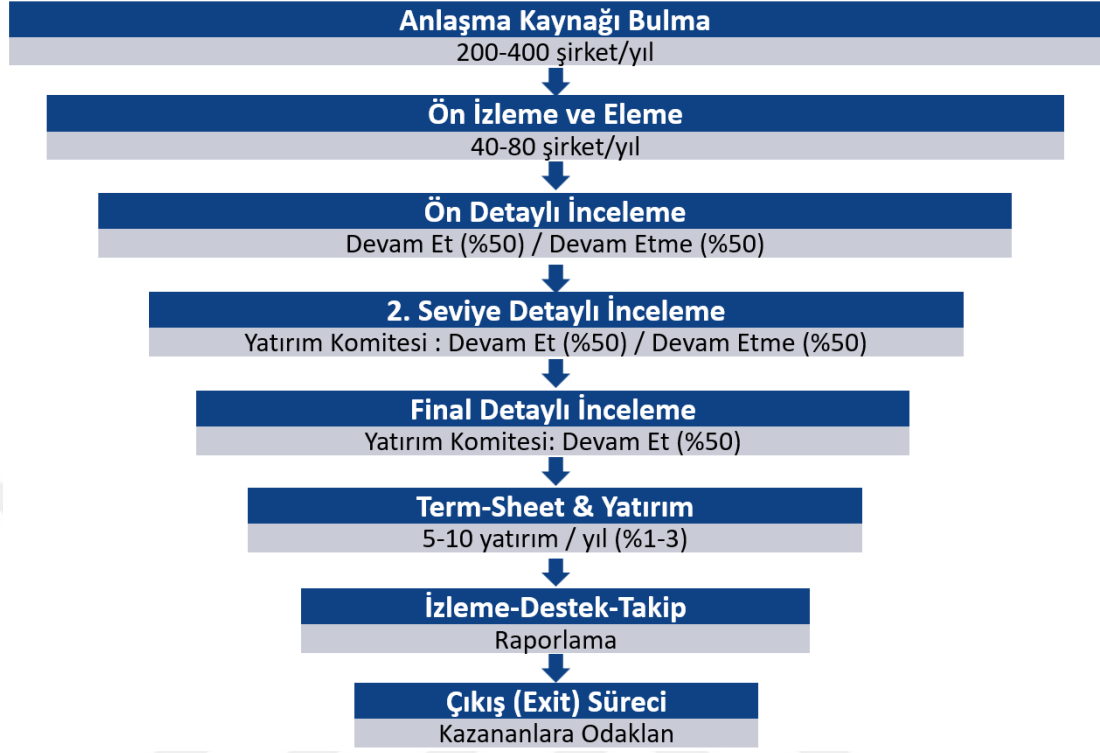
- Fikir Aşaması: Aile, Arkadaşlar ve Aptallar (3A)
- Kuluçka Yatırımlar: Melek Yatırımcılar, 3A
- Çekirdek Yatırımlar: Melek Yatırımcılar, Risk Sermayeleri
- Seri Yatırım Turları (A & B): Risk Sermayeleri, Kurumsal Risk Sermayeleri
- Seri Yatırım Turları (C & D): Yatırım Bankacılığı, Özel Sermayeler, Aile Ofisleri
- Halka Arz: Herkes



Şekil 1.2 : Bir girişimin fikirden halka arza tüm süreci (A. Vital, How Startup Funding Works, 2013).

Girişim yatırımlarında genellikle sektör kabulü olarak 6 adım bulunmaktadır. Bunlar; anlaşma kaynağı bulma (deal sourcing), anlaşma seçimi (deal selection), değerlendirme (valuation), anlaşma yapısı (deal structure), yatırım sonrası katma değer (post investment value added), çıkış (exits)tır. Bu aşamalarda görüşülen yatırım sayısını belirleyen etmenler fonun yapısı, yatırım tezi olmaktadır. Ortalama büyüklükte ve

erken aşama teknoloji girişimlerine yatırım yapan bir risk sermayesinin süreci Şekil 1.3'teki gibidir.



Şekil 1.3 : Risk sermayesi yatırım hunisi (Fernandes, 2014)

1.3 Yatırım Süreçleri

1.3.1 Anlaşma kaynağı bulma

Doğru girişimi bulmak yatırım sürecindeki en zahmetli ve en önemli adımdır. Ne kadar yüksek kaliteli girişimlere erişim sağlanabilirse sonraki adımlarda yapılacak olan faaliyetler de o kadar kolay olacaktır. Girişim sermayesi yatırımlarının 2013'ten beri yükselişte olduğu (Global Funding Slide In 2022 Sets Stage For Another Tough Year, 2023) düşünüldüğünde herkes en iyi girişimleri bulmaya çalışmaktadır. Bu aşamada yapılan bazı faaliyetler aşağıdaki gibi sıralanmaktadır:

- Oluşturulmuş İletişim ağları (network) sayesinde girişimlere dair bilgi edinmek
- Crunchbase, dealroom, startupwatch vb siteler üzerinden yeni girişimlere ulaşmak
- Fikir yarışmaları vasıtası ile iyi girişimleri bulmak

- Teknoloji Transfer Ofisleri ile grşmeler yapmak
- ok iyi bir marka algısı yaratıp girişimlerin gelmesini sağlamak (Sequoia Capital)

Bu aşamada grşlen girişimler mutlaka Mşteri İlişkileri Ynetimi (CRM) benzeri platformlar zerinde kaydedilmeli ve gelecekteki faaliyetleri takip edilmelidir.

1.3.2 Yatırım

Doğru girişimler bulunduktan sonra ilk olarak yatırım tezinin uygunluđuna bakılmaktadır:

- Girişimin faaliyet gsterdiđi endstri (finans, mobilite, sađlık vb)
- Girişimin bulunduđu aşama (kuluka, ekirdek, seri a vb)
- Yatırım byklđ; girişimin bulunduđu aşamaya gre verilmeye razı olunan tutar (ekirdek yatırımı 100bin USD - 250bin USD)
- Eş yatırımcı varlıđı
- rnn seviyesi (prototip, kullanılabilir en basit rn(MVP), bitmiř rn)
- Rekabet stnlkleri
- Ortaklık dađılımı
- Hedeflenen sermaye yzdesi (%5 - %10)
- Őirketin gelir retme durumu

Yukarıdaki aşamaya uygun girişimler bulunduktan sonra ařađıdaki konular hakkında daha spesifik sorular sorulmaktadır.

- Girişimin ve kurucuların gemiři
- Pazar
- rn
- Ekip
- retim
- Finansallar
- Yasal reglasyonlar

Bu aşamada da anlaşıldığı takdirde girişimin ürettiği ürün ile alakalı olarak alanında uzman kişilerin görüşleri alınır ve derin analiz (deep dive) denilen süreç başlar.

Derin analiz sürecini tamamlayan girişimler daha sonrasında durum tespiti (due-diligence) aşamasına geçerler burada bugüne kadar yaptığı tüm faaliyetler 3. Parti bir kuruluş tarafından kanuni açıdan incelenerek durum raporu hazırlanır.

Bir sonraki aşamada şirket değerlendirme süreci başlar. Şirket değerinin tespiti, oldukça zahmetli ve karmaşık bir süreçtir. Bu süreci daha da zorlaştıran etkenler arasında, şirketin mevcut durumu ve gelecekte yaratacağı değer birlikte ele alınması gerekliliği ile, şirket değerini etkileyen birçok değişkenin bulunması yer almaktadır. Literatürde mevcut olan ve uygulayıcılar tarafından aktif olarak kullanılan çeşitli değerlendirme yöntemleri, şirket değerinin belirlenmesi için kullanılabilir. Bununla birlikte, doğru değerlendirme yönteminin seçilmesi ve doğru verilerin kullanılması, şirket değerlemesinin yapılmasında son derece önemli bir rol oynamaktadır (Altundal, 2020). Girişimleri değerlemede kullanılabilecek bazı yöntemler aşağıdaki gibidir:

- Risk Sermayesi
- First Chicago
- Scorecard
- Berkus
- Risk Faktörleri Toplamı
- İndirgenmiş Nakit Akışı

Şirketin değeri ve verilecek olan pay üzerinde karşılıklı mutabakata varıldığı noktada ise sözleşme (term-sheet) hazırlanır ve karşılıklı olarak imzalanır. Bu noktadan sonra risk sermayesi şirketi artık girişimin anlaşıldığı oranda ortağı durumundadır.

1.3.3 Destekleme

Girişime yatırım yapan risk sermayesi şirketi artık şirketin ortağıdır. Başarıya da başarısızlığa da ortak olduğu için bu noktadan sonra girişimin faaliyetlerini etki alanı dahilinde desteklemesi girişimin başarısını artıracaktır. Yapılabilecek olan bazı faaliyetleri aşağıdaki gibi sıralayabiliriz.

- Risk sermayesi içerisinde bulunan portföy şirketleri ile işbirliği

- Kendi iletişim ağı içerisinde bulunan kurumlara tanıtılması
- Risk sermayesi şirketindeki uzmanların girişimden talep geldiğinde destekte bulunması.
- Gelecek yatırım turları için yatırımcı bulması

1.3.4 Çıkış (Exit)

Yatırım yapılan girişimdeki hisseler yatırım tezindeki hedeflenen getiri tutarına ulaştığında veya fon süresi dolduğunda bulunan hisselerin nakde çevrilmesi ve yatırımcılara yatırdığı paranın başlangıçta vaadedilen kazanç ile geri ödenmesi gerekmektedir. Burada risk sermayesi şirketi ne kadar fazla gelir elde ettiyse bir sonraki aşamada doğru orantılı olarak yatırımcılardan toplayabileceği meblağ artacaktır.

1.4 Girişimcilik Ekosistemi

Girişimler birçok alt sektörlere ayrılarak buldukları sektörlerde yıkıcı bir biçimde söz sahibi olmayı başarmışlardır

1.4.1 Başarılı girişimler – unicornlar

Son yıllarda girişim olarak kurulmuş ve piyasa değeri milyar hatta 10 milyar doların üzerine çıkmış binlerce girişim bulunmaktadır. Değeri 1 milyar doların üzerinde olan girişimlere ise unicorn (tek boynuzlu at) girişim denilmektedir. Şubat 2023 itibari ile Dünya üzerinde değerlemesi 1 milyar doları aşmış 1210 adet şirket bulunmaktadır. Bu sayı Temmuz 2022 tarihinde 1170 (kişisel araştırmam), Mayıs 2020 tarihinde ise 470 (Altundal V., 2020) olarak belirtiliyor. (The Complete List Of Unicorn Companies, 2023) Ülkemizde de girişimcilik ekosistemini geliştirmek ve unicornların çoğalması amacıyla Teknoloji ve Sanayi Bakanlığının altında “Turkish Unicorn” anlamına gelen “TURCORN 100” Programı kurulmuştur. (Turcorn 100, 2023) Türkiye’de Mart 2023 tarihi itibari ile Trendyol, Getir, Hepsiburada, Dream Games, Insider, Peak Games olmak üzere 6 adet TURCORNlar bulunmaktadır (Turcornlarımız, 2023).

Dünya üzerindeki değerlemesi 1 milyar doları aşmış şirketler CB Insights - Unicorn Companies List’ten elde edilen bilgiler ile analiz edildiğinde Çizelge 1.1’deki gibi bir tablo ortaya çıkmaktadır (The Complete List Of Unicorn Companies, 2023).

Çizelge 1.1 : Hizmet edilen endüstriye göre unicornların toplam değerlemesi ve toplam sayıları.

Endüstri Alanları	Toplam Değerleme (\$B)	Toplam Sayı (Adet)	Toplam Değerleme / Toplam Sayı
Finansal Teknolojiler	882	254	3,5
İnternet Yazılımları & Servisleri	661	229	2,9
E-ticaret	402	109	3,7
Yapay Zeka	342	90	3,8
Diğer	310	68	4,6
Sağlık	234	96	2,4
Tedarik Zinciri, lojistik & teslimat	214	67	3,2
Sibergüvenlik	148	58	2,6
Data Yönetimi & Analitiği	142	45	3,2
Otomotiv & Ulaştırma	119	37	3,2
Donanım	111	40	2,8
Eğitim Teknolojileri	98	32	3,1
Mobil & Telekomünikasyon	91	40	2,3
Tüketici & Perakende	86	31	2,8
Tatil	40	14	2,9
Genel Toplam	3.880	1.210	3,2

Yukarıdaki tablo değerlendirildiğinde en fazla finansal teknolojiler alanında unicornların olduğu görülmektedir. Bunu yazılım, e-ticaret, yapay zeka takip etmektedir. O endüstride bulunan toplam unicorn sayısını toplam değerlemeye böldüğümüzde ise diğer alanını yok sayarsak Yapay Zeka, E-ticaret, finansal teknolojiler olarak devam etmektedir. Tüm sektörlerdeki değerlemelerin toplamı ve toplam unicornların ortalamalarına bakıldığında ise dünyadaki ortalama girişim başına 3,2 milyar dolarlık bir şirket değerine ulaşılmaktadır.

Çizelge 1.2 : Ülkelerin sahip olduğu unicorn şirketlerin sayıları, değerleri ve GSYİH ile karşılaştırılması.

Sıralama	Ülkeler	Toplam Değerleme (\$B)	Girişim Sayısı	Değerleme / Toplam Sayı	GSYİH Milyar \$	GSYİH Sıralama
1	ABD	2.115	654	3,2	23.315	1
2	Çin	668	172	3,9	17.734	2
3	Birleşik Krallık	212	50	4,2	3.131	8
4	Hindistan	206	70	2,9	3.176	7
5	Almanya	83	29	2,9	4.260	6
6	Fransa	60	25	2,4	2.958	9

Çizelge 1.3 (devamı): Ülkelerin sahip olduğu unicorn şirketlerin sayıları, değerleri ve GSYİH ile karşılaştırılması.

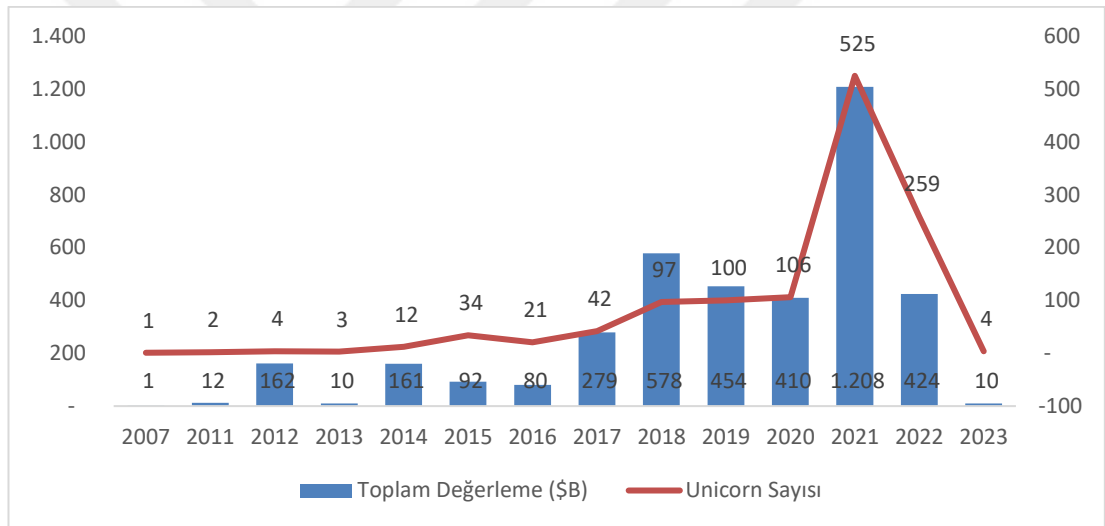
7	Avusturalya	55	8	6,9	1.553	15
8	Kanada	50	20	2,5	1.988	11
9	İsrail	49	23	2,1	489	29
10	Brezilya	39	16	2,4	1.609	14
11	Güney Kore	33	14	2,3	1.799	12
12	Endonezya	29	7	4,2	1.186	18
13	Singapur	23	14	1,6	397	39
14	İsveç	23	7	3,2	636	25
15	Hollanda	22	7	3,2	1.013	19
16	Hong Kong	20	7	2,9	368	42
17	Meksika	19	8	2,3	1.273	17
18	Türkiye	16	3	5,3	819	21
19	Finlandiya	12	4	3,1	297	47
20	İsviçre	12	6	2,1	801	22
21	Belçika	10	3	3,4	594	26
22	İrlanda	10	6	1,7	504	28
23	Şeyseller	10	1	10,0	1	183
24	Estonya	10	2	5,0	37	102
25	Japonya	9	6	1,5	4.941	5
26	Avusturya	8	2	3,8	480	32
27	Kolombiya	7	3	2,5	314	46
28	İspanya	7	5	1,4	1.427	16
29	Danimarka	7	2	3,4	398	38
30	Litvanya	6	2	3,1	66	83
31	Viyetnam	5	2	2,6	366	43
32	BAE	5	3	1,7	415	36
33	Norveç	5	4	1,1	482	31
34	Tayland	4	3	1,2	506	27
35	Yunanistan	3	2	1,7	215	54
36	Hırvatistan	3	2	1,5	69	79
37	Filipinler	3	2	1,5	394	40
38	Güney Afrika	3	2	1,3	419	34
39	Şili	3	2	1,3	317	45
40	Arjantin	2	1	2,5	487	30
41	İtalya	2	2	1,0	2.108	10
42	Nijerya	2	1	2,0	441	33
43	Lüksemburg	2	1	2,0	86	74
44	Malezya	2	1	1,7	373	41
45	Senegal	2	1	1,7	28	110
46	Bermuda	2	1	1,6	7	#YOK
47	Ekvator	2	1	1,5	106	68
48	Çek Cumhuriyeti	1	1	1,2	282	49
49	Mısır	1	1	1,0	404	37
50	Liechtenstein	1	1	1,0	#YOK	

CB Insights - Unicorn Companies List'e eklenen ilk girişim 2007 yılında olmuştur. 2007 yılından bu yana yukarıdaki analiz sonucunda görüldüğü üzere ABD ve Çin toplam sayısının yarısından fazlasını oluşturmaktadır. Girişimcilik ekosistemi, bir ülkenin ekonomisine büyük bir katkıları sağlamaktadır. Girişimlerin başarılı olması ve büyümesi, istihdam yaratır, yeni ürün ve hizmetlerin geliştirilmesine yol açar, vergi gelirlerini artırır ve diğer sektörleri canlandırır. Unicorn şirketleri, yüksek büyüme potansiyeline sahip genellikle teknoloji tabanlı şirketlerdir ve genellikle yatırımcıların büyük ilgisini çekerler. Bu şirketler, ülkelerin ekonomik kalkınmalarına katkıda bulunarak iş fırsatları yaratır, yeni teknolojiler geliştirir ve yenilikçi ürünler sunarlar. Aynı zamanda, unicorn şirketleri genellikle ülkelerin ekonomik potansiyelini yansıtırlar ve ülkelerin dijital dönüşümüne öncülük ederler. Özellikle son yıllarda, birçok ülke, teknolojik yeniliklere ve girişimciliğe odaklanarak ekonomik kalkınmalarını hızlandırmayı hedeflemektedir. Bu ülkeler, girişimcilik ekosistemlerini destekleyerek, girişimlere finansman sağlayarak ve yenilikçi fikirlerin geliştirilmesi için altyapı oluşturarak unicorn şirketlerinin ortaya çıkmasını teşvik etmektedirler. Bu süreçte, ülkelerin ekonomik kalkınmaları ve unicorn şirketlerinin ortaya çıkması arasında doğrudan bir korelasyon olmasa da güçlü ekonomilere sahip ülkelerin daha fazla unicorna sahip olma eğiliminde olduğunu görülmektedir. Bunun nedeni, elverişli bir iş ortamı, sermayeye erişim ve geniş bir tüketici pazarı gibi faktörler olabilir.

Örnek olarak ABD, dünya genelindeki en büyük girişimcilik ekosistemlerinden birine sahiptir. Girişimcilerin ve yenilikçi fikirlerin desteklendiği bir ortam yaratılmıştır ve ABD hükümeti, girişimcilik ekosistemi için önemli bir destek sağlar. ABD'deki girişimler, ekonomik büyümeye ve istihdama büyük katkı sağlamaktadır. 2019 yılında ABD'deki girişimler, toplam istihdamın yaklaşık %10'unu oluşturmakta ve ülke ekonomisindeki büyümeye önemli bir katkı sağlamaktadır.

Çizelge 1.4 : Yıllara göre unicorn sayıları ve toplam değerlemeleri.

Yıllar	Toplam Değerleme (\$B)	Unicorn Sayısı
2007	1	1
2011	12	2
2012	162	4
2013	10	3
2014	161	12
2015	92	34
2016	80	21
2017	279	42
2018	578	97
2019	454	100
2020	410	106
2021	1.208	525
2022	424	259
2023	10	4
Genel Toplam	3.880	1.210



Şekil 1.4 : Yıllara göre unicorn sayıları ve toplam değerlemelerinin görseli.

Unicorn sayıları yıllar bazında incelendiğinde ise sürekli bir artış trendinin olduğu gözlemlenmektedir. Zaman zaman sayılarında düşüşler yaşansa da gelecekte teknoloji tabanlı iş fikirlerine her zaman ihtiyaç olacağı bir gerçektir. Bu sebeple girişimlerin doğru bir şekilde seçilip uygun şekilde büyütülmesine ve unicorn hale gelmesine ihtiyaç vardır.

Ülkeler	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	Genel Toplam
ABD	53	10	14	45	35	27	9	134	14	74	162	13	37	26	1	654
Çin	19	17	8	1	2	29	12	7	18	10	15	13	6	12	3	172
Hindistan		2	1		1	17	7	17		2	12	2	1	7	1	70
Birleşik Krallık	4	1	2		1	2	1	27	1	2	2	1	5	1		50
Almanya		2	1		1	5		6	1	1	4		2	3	3	29
Fransa	2	1				6		6	2	2	2		3	1		25
İsrail	3			7	1		1	2	2	1	2	1	1	2		23
Kanada	2			3		1	1	3	1		6		1	1	1	20
Brezilya	1					5		6				1	1	2		16
Güney Kora			1			4		1			3	2	1	1	1	14
Singapur	2					3		5			2	1		1		14
Avustralya						1		2			5					8
Meksika						2		5						1		8
İsveç		1						1		1	1		2		1	7
Hollanda						1		4	1			1				7
Hong Kong			1			1		3						1	1	7
Endonezya			1			1		3						1	1	7
İrlanda								2		1	3					6
İsviçre				1	1			1		1	1			1		6
Japonya	1							2				1	2			6
İspanya		1									2		1		1	5
Norveç					2	1					1					4
Finlandiya										1	1	1		1		4
Türkiye											1	1		1		3
Tayland						1		1						1		3
BAE													2	1		3
Kolombiya								1					1	1		3
Belçika					1			1			1					3
Estonya	1	1														2
Hırvatistan		1										1				2
Güney Afrika			1									1				2
Avusturya							1	1								2
Filipinler								1					1			2
Danimarka								2								2
Yunanistan								1			1					2
Şili	1		1													2
Litvanya				1		1										2
Vietnam								1			1					2
İtalya								2								2
Nijerya								1								1
Şeysteller								1								1
Ekvator								1								1
Mısır								1								1
Bermuda	1															1
Arjantin								1								1
Lüksemburg													1			1
Senegal								1								1
Malezya						1										1
Liechtenstein											1					1
Çek Cumhuriyeti														1		1
Genel Toplam	90	37	31	58	45	109	32	254	40	96	229	40	68	67	14	1210

Şekil 1.5 : Ülkelerin sektörlere göre sahip olduğu unicorn sayıları

Çizelge 1.5 : Şekil 1.5 'e ait açıklama tablosu.

A	Yapay Zeka	F	E-ticaret	K	SaaS
B	Mobilite	G	Eğitim Teknolojileri	L	Mobil & Telekomünikasyon
C	Tüketici & Perakende	H	Finansal Teknolojiler	M	Diğer
D	Sibergüvenlik	I	Donanım	N	Tedarik Zinciri & Teslimat
E	Data Analitiği	J	Medikal Teknolojiler	O	Tatil

Şekil 1.5'te her biri farklı bir endüstriyi temsil eden toplamda 15 sütun vardır. Her sütun içerisinde ülkelerin birbiri ile kıyaslamaları bulunmaktadır. Tablo "Genel Toplam" sütununda büyükten küçüğe şeklinde sıralanmıştır. Yeşil renk o endüstrideki en yüksek sayıyı simgelerken kırmızı en düşük sayıyı göstermektedir. Örneğin; H ile adlandırılan Finansal Teknolojiler endüstrisinde ABD'de 134 adet unicorn bulunmaktadır. Takip eden ise 27 ile Birleşik Krallık, 17 ile Hindistan olduğu görülmektedir. Şekil 1.5 ile istenilen ülkedeki odaklanan sektör çıkarımları yapılabilmektedir.

1.4.2 Girişimlerin başarısız olma sebepleri

Girişimlerin yaklaşık olarak %90'ı ilk beş yıl içinde faaliyetlerini sonlandırarak başarısız olmaktadır. Başarısızlık nedenleri çeşitlilik gösterse de, başarılı girişimlerle karşılaştırıldığında, ürün-pazar uyumu, girişimcinin deneyimi, uzmanlık alanı ve yaşının yanı sıra minimum uygulanabilir ürünün ortaya çıkma süresi, gelir oluşturma kapasitesi ve finansmana erişim gibi temel faktörlerin etkili olduğu görülmektedir (G. Kalyanasundaram, 2018).

Küresel ölçekte faaliyet gösteren en önemli yatırım ve danışmanlık firmalarından biri olan CB Insights tarafından 101 girişim üzerinde yapılan detaylı araştırmalar sonucu girişimlerin başarısız olmalarının en önemli 12 nedeni oranları ile birlikte tespit edilmiştir (The Top 12 Reasons Startups Fail, 2023). Yayınlanan bu rapora göre girişimlerin başarısız olmalarının en önemli 12 nedeni önem sırasına göre aşağıda sıralanmıştır.

- 1- Nakit tükenmesi/yeni sermaye artırılamaması
- 2- Pazarın üretilen ürüne ihtiyacının olmaması
- 3- Rekabete dayanamama
- 4- İş modelinin eksik/hatalı olması
- 5- Regülasyonlar ve yasal zorluklar
- 6- Fiyatlandırma / maliyet sorunları
- 7- Doğru ekibin kurulamaması
- 8- Ürünün yanlış zamanda piyasaya sunulması
- 9- Kötü/eksik ürün

- 10- Ekip/Yatırımcılar arasındaki uyumsuzluk
- 11- Pivotun (alan deęiřtirme) kötü sonuçlanması
- 12- Kuruculardaki tükenmişlik/tutku eksikliği

1.5 Tezin Amacı

Erken aşamadaki (early stage) girişimlerin başarısını tahmin etmek zorlu bir iştir ve yanlış yapılan seçimler daha sonrasında fırsat maliyetlerine sebep olmaktadır. Bir ekonomik seçim ile vazgeçilen en yüksek maliyetli alternatif, fırsat maliyetini oluşturur (Alkın, Yıldırım, Özer, & Mustafa, 2005). Yukarıda açıklanan girişim yatırım süreçleri zorlu ve uzun bir yolculuktur ve geçilmesi gereken bir çok aşama bulunmaktadır. Son dönemde bu tür bağlamlarda tahminlerde bulunmak için, makine ve insan zekasının tamamlayıcı yeteneklerini birleştirilmesinin daha iyi sonuçlar vereceği düşünülmektedir. Makineler, yeni bir girişimin başarısını gösteren büyük miktarda "somut" sinyalleri (endüstri, çalışan sayısı, patent sayısı vb.) tutarlı bir şekilde işlemede özellikle faydalı olsa da, insanlar "soyut" sinyalleri (inovatiflik seviyesi, girişimcinin karakteri, ekibin uyumu vb.) yorumlamada üstündür. Bireylerin sınırlı rasyonalitesinin kısıtlamalarının üstesinden gelmek için kolektif zekadan (makine + insan) yararlanmasının verimi artıracakı düşünölmektedir. Sonuç olarak bu tez kapsamında hem akademiye hem de uygulamaya sağlayabileceęi çeřitli katkılar ařaęıdaki gibi sıralanmıřtır:

- Bir girişimin nasıl başarılı olabileceęine iliřkin bir tahmin modeli,
- Yatırımcılara sınırlı bilgi ve yüksek riskli durumlarda "somut" sinyallere destek sağlayacak bir sistem tasarımı,
- Yapılan literatür araştırması sonucunda sınırlı sayıda Türkçe kaynaęa ulařılmış çalışmaların çoęunluęunu yabancı kaynaklarda olduęu görölmüřtür. Bu konuda Türkçe kaynaklara ihtiyaç olduęu düşünölmektedir. Bu çalışma Türkiye'de geliřmekte olan girişim ve yatırım ekosistemini destekleyecek ilgili akademik çalışmalara katkı sağlaması beklenmektedir.

2. LİTERATÜR ARAŞTIRMASI

2.1 Girişim Ekosistemi

Makine Öğrenmesi”, “Girişim”, “Yatırım”, “Yapay Zeka”, “Risk Sermayesi”, “Kurumsal Risk Sermayesi” anahtar kelimeleri ile yapılan literatür araştırmasının sonuçları aşağıdaki gibidir.

Popüler bir model, lojistik regresyona dayalı olarak sunulan başarı/başarısızlık modelidir (Lussier, 2001). Bu model Amerika Birleşik Devletleri, Şili veya Hırvatistan gibi farklı pazarlar için genişletildi ve doğrulandı (Halabí, 2014; Halabi, 2010). ABD pazarı durumunda, bu değişkenlerden sadece 4'ü istatistiksel olarak anlamlıydı (planlama, profesyonel tavsiye, eğitim ve personel).

Bir Risk Sermayesi (venture capital) yatırımcısının itibarını dikkate almanın ve ölçmenin farklı yolları da olabilir. Saygın Risk Sermayesi firmalarının şirketlerini başarılı çıkışlara götürme olasılığının daha yüksek olduğunu göstermektedir. Bu kanıt, bireysel Risk Sermayesi yatırımlarının performansını analiz ederken de geçerlidir (Nahata, 2008).

Başka bir çalışmada, 200 Fin firmasındaki büyümenin kurucuların güdüsü, arka plan özellikleri, yönetim tarzları vb. ile ilişkisini incelemek için lojistik regresyon analizini kullanmıştır (Littunen, 2010) .

Farklı bir yaklaşıma sahip bir çalışma, kuzeydoğu İngiltere'de yeni kurulan imalat firmasının hayatta kalmasını analiz etmektedir (P. Holmes, 2010). Çalışma, firmanın hayatta kalma süresi ile firma (tesis büyüklüğü) veya makroekonomik yönlerle ilgili sinyaller arasındaki ilişkiyi incelemek için log-lojistik tehlike modellerini kullanmıştır.

Diğer yaklaşım, doğrusal bir regresyon (Hoenen, 2012) ile sağlanan toplam Risk Sermayesi finansmanı miktarını modeller. Bu durumda, veri seti 1974'ten 2011'e kadar oluşturulan biyoteknoloji alanında çalışan ABD şirketleri hakkında bilgi içerir. Bu şirketleri tanımlayan değişkenler, alınan Risk Sermayesi yatırımlarının sayısı,

patentlerin sayısı, bu patentlerin alıntıları ve diğer coğrafi bilgilerle ilgilidir. İncelenen yaklaşımların çoğu regresyon analizine dayanmaktadır.

Bu akımdaki en eski çalışmalardan biri, şirketlerin satın alınmasını öngören kural tabanlı bir uzman sistem sunar (Ragothaman, 2003). Bu sistem, değerlendirme 200 şirketten oluşan bir veri seti ile sınırlı olmasına rağmen, %70'lik bir başarı oranına ulaşmıştır. Başarılı şirketleri seçmek için bir uzman sistem kullanılabilirse bile, makine öğrenimine dayalı veri odaklı yaklaşımlar daha popüler hale gelmektedir. Japonya'daki yaklaşık 600 birleşme ve devralma (merge & acquisition) vakasını tahmin etmek için topluluk sınıflandırıcılarının kullanılmasını önermektedir (Wei, 2009). Bu çalışmada, tahmin ediciler, patent analizinden ve hem yatırımcıların hem de aday hedef şirketlerin profillerinden elde edilen teknolojik değişkenlerdir. Yazarlar, bir edinimi tahmin ederken %88'lik bir küresel doğruluk ve %40'ın üzerinde bir kesinlik bildirmektedir.

Diğer bir çalışmada ise bir anketten elde edilen verileri kullanarak 142 Bulgar girişiminin başarısını tahmin etmek için çeşitli makine öğrenimi yöntemleri karşılaştırılmıştır (Haralampiev, 2014). Yazarlar, karar ağaçlarının en doğru yöntem olduğunu ve bunları, örneğin rekabet avantajının varlığı, kurucuların benzer bir pozisyonda deneyimi vb. gibi başlangıç başarı faktörlerini ortaya çıkarmak için kullandıklarını göstermektedir. Daha yüksek büyüme beklentileri ve hayatta kalma şansı olan iş modellerinin tahmini için daha karmaşık bir yaklaşım olarak denetimli ve denetimsiz öğrenmenin birlikte kullanılmasını araştırmışlardır. (Böhm) Çalışma, ABD ve Almanya'daki girişimleri dikkate almaktadır ve bir girişimin hayatta kalmasını tahmin etmeye çalışırken, ancak yine 181 şirketten oluşan küçük bir veri kümesini kullanarak %83,6'lık bir doğruluk elde etmişlerdir.

Diğer yazarlar, erken aşamadaki girişimlerin sonuçlarını tahmin etmeye çalışırken yatırım uzmanlarının ve makine öğrenme tekniklerinin performansını karşılaştırmışlardır (David McKenzie, 2019). Bir iş planı yarışmasına katılan 2.506 Nijeryalı girişim üzerinde yapılan çalışmada; makine öğrenimi yöntemlerinin, yatırım uzmanlarına kıyasla daha başarılı tahmin yaptıkları sonucuna varmışlardır. Makine öğrenimi için %63 ve insan uzmanlar için %58'lik bir başarı oranı bildirilmiştir, bu da onları yatırım uzmanlarının hangi girişimlerin başarılı olacağını belirlemede zorluk yaşadıkları sonucuna varmalarına yol açmaktadır.

2.2 Makine Öğrenmesi Literatür Araştırması

Makine öğrenimi, yapay zeka alanının bir alt dalıdır ve bilgisayar sistemlerinin açık programlama olmadan öğrenme, tahmin veya kararlar yapabilmesini sağlayan algoritmaların ve istatistiksel modellerin geliştirilmesine odaklanır. Son yıllarda, çeşitli endüstrileri devrimleştirebilecek potansiyeli nedeniyle büyük ilgi ve popülerlik kazanmıştır. Makine öğrenimi tekniklerinin uygulanması, sağlık, finans, görüntü tanıma, doğal dil işleme ve öneri sistemleri gibi farklı alanlarda umut verici sonuçlar göstermiştir. Büyük veri kümelerini işleyebilme ve karmaşık desenleri tanımlayabilme yeteneği sayesinde, makine öğrenimi, değerli içgörülerin elde edilmesi ve doğru tahminlerin yapılması için güçlü bir araç haline gelmiştir. Makine öğrenimi algoritmalarının kullanımı, farklı alanlarda önemli başarılar göstererek, dijital çağda yenilik ve ilerlemenin anahtarı olarak potansiyelini göstermektedir (Trevor Hastie, 2009).

2.2.1 Lojistik Regresyon (LR)

Lojistik regresyon, ikili sonuçları tahmin etmek için güçlü ve yaygın olarak kullanılan bir istatistiksel tekniktir. Hedefin, bir dizi öngörücü değişkene dayalı olarak başarı veya başarısızlık olasılığını belirlemek olduğu başarı tahmininde özellikle önemlidir. Lojistik regresyon modeli, girdi özelliklerini bir lojistik fonksiyona eşleyerek meydana gelen bir olayın olasılığını tahmin eder. Lojistik regresyondaki ufuk açıcı çalışmalardan biri, ikili sonuçlar için doğrusal regresyonun bir uzantısı olarak lojistik model kavramını tanıttığı makaledir (Cox, 1958). O zamandan beri lojistik regresyon, çeşitli disiplinlerde ikili sınıflandırma görevlerinin temel taşı haline gelmiştir.

2.2.1.1 Lojistik regresyon çalışma prensibi

Lojistik regresyon çalışma prensibi şu aşamalardan oluşmaktadır.

a. Veri Hazırlama:

Lojistik regresyon uygulamak için, sonucun tahmin edilmesine yardımcı olabilecek bir ikili sonuç değişkeni (ör. başarı veya başarısızlık) ve bir veya daha fazla öngörücü değişken (özellikler) içeren bir veri kümesine ihtiyacınız vardır. Veri seti, bireysel örnekleri temsil eden satırlar ve sonucu ve öngörücü değişkenleri temsil eden sütunlarla organize edilmelidir.

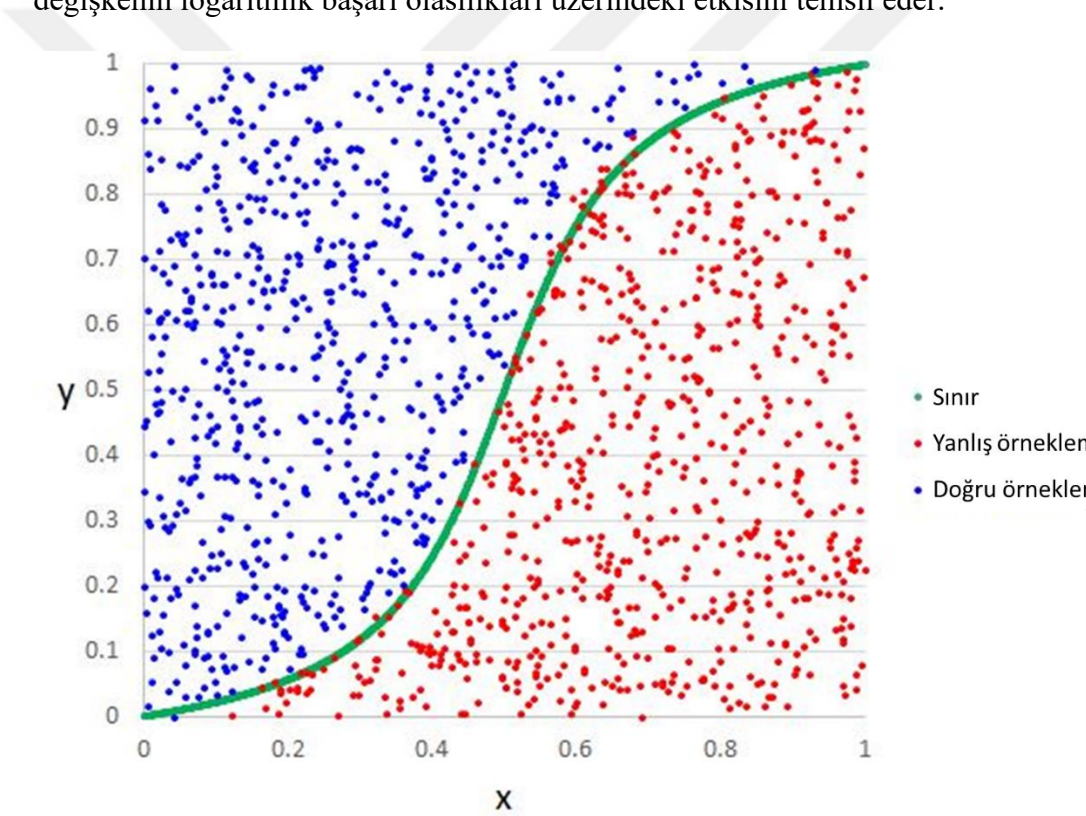
b. Sigmoid Fonksiyonu (Lojistik Fonksiyon):

Lojistik regresyon modeli, tahmin edilen log-oranları 0 ile 1 arasında bir olasılık değerine eşlemek için sigmoid işlevini (lojistik işlev olarak da bilinir) kullanır. Sigmoid işlevi aşağıdaki gibi tanımlanır:

$$P(Y = 1 | X) = 1 / (1 + e^{-(\beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_p * X_p)})$$

Burada ki $P(Y=1|X)$ X_1, X_2, \dots, X_p tahmin değişkenleri verildiğinde sonucun (başarının) 1 olma olasılığıdır.

$\beta_0, \beta_1, \dots, \beta_p$ lojistik regresyon modelinin katsayılarıdır (ağırlıklar), her öngörücü değişkenin logaritmik başarı olasılıkları üzerindeki etkisini temsil eder.



Şekil 2.1: Örnek Bir Sigmoid Fonksiyonu

c. Logit:

Sigmoid işlevini uygulamadan önce, lojistik regresyon, yordayıcı değişkenler verildiğinde sonucun log-olasılıklarını 1 olarak tahmin eder.

$$\log \frac{P(Y = 1 | X)}{1 - P(Y = 1 | X)} = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_p * X_p$$

Log-oran, başarısızlık sonucuna göre meydana gelen başarı sonucunun olasılıklarının logaritmasını temsil eder. Log-olasılıklar pozitif veya negatif olabilir ve karşılık gelen katsayılarla ağırlıklandırılmış öngörücü değişkenlerin doğrusal bir kombinasyonudur.

d. Model Eğitimi:

Lojistik regresyon modeli, verilere en iyi uyan katsayıların ($\beta_0, \beta_1, \dots, \beta_p$) optimal değerlerini bulmak için eğitilir. Bu, tipik olarak, tahmin edilen olasılıklar ile eğitim verilerindeki gerçek sonuçlar arasındaki farkı nicelleştiren bir maliyet fonksiyonunu en aza indiren gradyan iniş gibi bir optimizasyon algoritması aracılığıyla elde edilir.

e. Model Yorumu:

Model eğitildikten sonra, her tahmin değişkeninin başarının log-olasılıkları üzerindeki etkisini anlamak için katsayıları yorumlayabilirsiniz. Pozitif bir katsayı, karşılık gelen yordayıcı değişkendeki bir artışın başarının log-olasılıklarını (ve olasılığını) artıracakını gösterirken, negatif bir katsayı bunun tersini gösterir.

f. Tahminler Yapmak:

Model eğitilip katsayılar elde edildikten sonra, yeni veriler üzerinde tahminler yapmak için lojistik regresyon modeli kullanılabilir. Belirli bir yordayıcı değişken kümesi için, öğrenilen katsayıları kullanarak log-olasılıkları hesaplar ve ardından tahmin edilen başarı olasılığını elde etmek için sigmoid işlevi uygulanır.

g. Karar Sınırı:

İkili sınıflandırmada, karar sınırı, örnekleri iki sınıftan birine (başarılı veya başarısız) sınıflandırmak için kullanılan eşik olasılığıdır (genellikle 0,5). Öngörülen olasılık eşik üzerindeyse, örnek başarılı olarak sınıflandırılır ve eşik altındaysa başarısız olarak sınıflandırılır.

Özet olarak, lojistik regresyon, öngörücü değişkenlerin doğrusal bir kombinasyonunu kullanarak başarının log-olasılıklarını tahmin ederek ve ardından log-olasılıkları sigmoid işlevini kullanarak bir olasılığa eşleyerek çalışır. Model, başarı olasılığını tahmin etmeye ve ikili sınıflandırmalar yapmaya izin veren en uygun katsayıları bulmak için eğitilmiş olacaktır.

2.2.1.2 Lojistik regresyon kullanım örnekleri

Makine öğreniminde LR, eğitilmesi ve uygulanması en basit ve en hızlı algoritmalarından biridir ve genellikle birçok sınıflandırma probleminde başlangıç noktası olarak kullanılır. Düşük varyansa sahip olduğu için aşırı uyuma daha az eğilimlidir, bu da onu sınıfların net bir şekilde ayrılmasıyla ikili sınıflandırma problemleri için çok uygundur. Diğer avantajları, özellik uzayında sınıfların dağılımı hakkında varsayımlarda bulunmama yeteneği ve ikili yerine çoklu sınıf sınıflandırma problemlerine genişletilebilmesidir.

LR'nin en büyük dezavantajları, belirli bağlamlarda uygulanmasını sınırlayabilen sınırlı sonuç ve birbiriyle ilişkili pek çok özelliğe sahip veri kümelerinde eğitilirse aşırı uydurmaya ne kadar yatkın olabilecekleridir (Howbert, 2012). LR, örneğin 500.000 özellik için 100.000 gözlem gibi yüksek boyutlu verilerle ilgili sorunları destekler.

En popüler makine öğrenimi algoritmalarından biri olan Lojistik Regresyon'un, dolandırıcılık sınıflandırmasından pazarlama kampanyalarındaki potansiyel müşterilerin sınıflandırılmasına kadar hemen hemen her çalışma alanında uygulamaları bulmak mümkündür.

2.2.2 Destek vektör makineleri (support vector machine)

Destek Vektör Makineleri (DVM), başarı tahmini görevlerinde yaygın olarak kullanılan sağlam ve popüler bir makine öğrenimi algoritmasıdır. DVM, bir veri kümesindeki iki sınıfı en iyi ayıran ve aralarındaki marjı en üst düzeye çıkaran en uygun hiperdüzlemi bulmayı amaçlar. DVM'nin etkinliği, çekirdek işlevlerinin kullanımı yoluyla yüksek boyutlu verileri ve doğrusal olmayan ilişkileri işleme yeteneğinde yatmaktadır.

DVM üzerine temel çalışmalardan biri, DVM için teorik çerçevenin tanıtıldığı (Vapnik, 1995) tarafından yazılan makaledir. O zamandan beri DVM, çeşitli alanlarda geniş çapta çalışılan ve uygulanan bir algoritma haline gelmiştir.

Destek Vektör Makineleri, sınıflandırma ve regresyon görevleri için kullanılan güçlü algoritmalar. DVM'ler, verileri yüksek boyutlu bir özellik uzayına haritalar ve farklı sınıfları maksimum marjla ayıran optimal bir hiper düzlem bulmaktadır. DVM, diğer makine öğrenimi algoritmalarından farklıdır. Yüksek boyutlu verileri/veri kümesini

(çok sayıda özellik içeren) DVM ile işleyebilir, ancak çok yoğun bellek nedeniyle uzun eğitim süresi gerektirdiğinden gözlem sayısının çok büyük olmadığı durumlar için daha uygundur (Manning, Raghavan, & Schütze, 2009; Statnikov, 2011).

2.2.2.1 Destek vektör makineleri çalışma prensibi

Destek vektör makineleri çalışma prensibi şu şekildedir.

a. Veri Hazırlama:

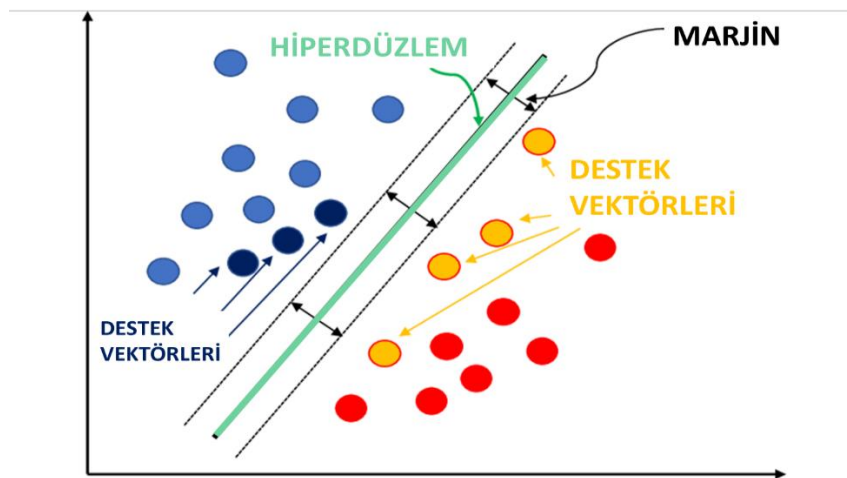
SVM'yi kullanmak için, farklı sınıf örnekleri içeren etiketli bir veri kümesine ihtiyacınız vardır. Her veri noktasının bir dizi özelliği (giriş değişkenleri) ve karşılık gelen bir sınıf etiketi (çıkış değişkeni) olmalıdır. DVM hem ikili sınıflandırma (iki sınıf) hem de çok sınıflı sınıflandırma (ikiden fazla sınıf) için uygundur.

b. Özellik Uzayı ve Hiperdüzlem:

SVM'de, her veri noktası bir özellik uzayında bir vektör olarak temsil edilir. Amaç, farklı sınıflara ait veri noktalarını en iyi şekilde ayırabilen bir hiper düzlem bulmaktır. Bir hiper düzlem, özellik uzayını farklı sınıflara karşılık gelen iki bölgeye bölen çok boyutlu bir düzlemdir.

c. Marj ve Destek Vektörleri:

DVM, iki sınıf arasındaki maksimum marja sahip hiperdüzlemi bulmayı amaçlar. Kenar boşluğu, hiper düzlem ile her sınıftan en yakın veri noktaları arasındaki mesafedir. Hiper düzleme en yakın olan veri noktalarına destek vektörleri denir. Bu destek vektörleri, optimal hiperdüzlemin belirlenmesinde çok önemli bir rol oynar.



Şekil 2.2: Destek Vektör Makinelerine ait görsel

d. Yumuşak Kenar Boşluğu ve C Parametresi:

Uygulamada, veriler her zaman tek bir hiperdüzlemle mükemmel şekilde ayrılamayabilir. DVM, bazı veri noktalarının yanlış sınıflandırılmasına veya marjın içinde kalmasına izin vererek bunu halleder. 'C' parametresi, marjı en üst düzeye çıkarmak ve sınıflandırma hatasını en aza indirmek arasındaki değiş tokuşu kontrol eder. Büyük bir 'C' değeri sert bir sınıra (daha az yanlış sınıflandırma ancak muhtemelen fazla uydurma) yol açarken, küçük bir 'C' değeri yumuşak bir sınıra (daha fazla yanlış sınıflandırma ancak daha iyi genelleme) yol açmaktadır.

e. Çekirdek (Kernel) Hilesi:

Verinin orijinal özellik uzayında doğrusal olarak ayrılamadığı durumlarda, SVM, veriyi ayrılabilir olabileceği daha yüksek boyutlu bir uzaya dönüştürmek için çekirdek hilesini kullanır. Yaygın çekirdek işlevleri, radyal temel işlevi (RBF), polinom ve sigmoid çekirdekleri içerir. Çekirdeğin seçimi, verilerin doğasına ve sınıflandırma problemine bağlıdır.

f. Optimizasyon ve İkili Problem:

SVM, hiperdüzlem ve destek vektörlerini bulmak için problemi bir optimizasyon görevi olarak formüle eder. Optimizasyon, birincil sorunu doğrudan çözmekten hesaplama açısından daha verimli olan ikili sorunu çözmeyi içerir. Bu, destek vektörleriyle ilişkili Lagrange çarpanlarının bulunmasını içerir.

g. Tahminler Yapmak:

Optimum hiperdüzlem ve destek vektörleri bulunduğunda, SVM yeni, görünmeyen veri noktalarında tahminler yapabilir. Yeni bir veri noktası için SVM, hiperdüzleme olan mesafesini hesaplar. Noktanın hiper düzlemin hangi tarafına düştüğüne bağlı olarak, ikili sınıflandırma için iki sınıftan birine veya çok sınıflı sınıflandırma için birden çok sınıftan birine sınıflandırılır.

2.2.2.2 Destek vektör makineleri kullanım örnekleri

Başarı tahmini bağlamında, DVM farklı alanlarda kullanılmıştır. Örneğin, finans sektöründe, (Zhang, 2005), çeşitli finansal göstergeler ve piyasa verilerine dayalı olarak hisse senedi fiyat hareketlerini tahmin etmek için DVM'yi kullanılmıştır.

Sağlık hizmetlerinde, tıbbi tedavilerin veya teşhislerin başarısını tahmin etmek için DVM uygulanmıştır. (Jianjiong Gao, 2013) klinik ve genetik özelliklere dayalı olarak kanser hastalarının hayatta kalma oranını tahmin etmek için DVM'yi kullanılmıştır.

Yüksek boyutlu problemlerin yaygın olduğu metin sınıflandırma problemlerinde (Statnikov, 2011) özellikle popülerdir. DVM'ler, metin kategorizasyonu (Joachims, 2005) ve görüntü tanıma (Cortes, 1995) gibi çeşitli uygulamalarda etkili oldukları kanıtlanmıştır.

2.2.3 Rastgele orman (random forest)

Rastgele Orman, başarı tahmini görevlerinde dikkate değer başarı gösteren bir topluluk öğrenme yöntemidir. Karmaşık verileri, yüksek boyutlu özellik alanlarını ve gürültülü verileri işleme yeteneği nedeniyle makine öğreniminde popüler bir seçimdir. Rastgele Orman, eğitim sırasında birden fazla karar ağacı oluşturarak ve daha doğru ve istikrarlı tahminler yapmak için tahminlerini birleştirerek çalışır.

Rastgele Orman üzerine ufuk açıcı çalışmalardan biri, Rastgele Orman algoritması kavramının tanıtıldığı (Breiman L. , 2001) makalesidir. O zamandan beri Rastgele Orman, çeşitli alanlarda geniş çapta çalışılan ve uygulanan bir algoritma haline gelmiştir.

Rastgele Orman, birden fazla karar ağacını birleştirerek tahminler yapmak için kullanılan bir ensemble (birleşik) öğrenme yöntemidir. Rastgele ormanda her ağaç, verinin rastgele bir alt kümesiyle eğitilir ve nihai tahmin oylama veya ortalamalama süreciyle elde edilir. Rastgele ormanlar, sağlamlıkları ve yüksek boyutlu verilerle başa çıkma yetenekleriyle bilinirler.

2.2.3.1 Rastgele orman çalışma prensibi

Rastgele orman çalışma prensibi şu şekildedir.

a. Veri hazırlama:

Rastgele Ormanı kullanmak için, farklı sınıf örnekleri (sınıflandırma için) veya sürekli hedef değerler (gerileme için) içeren etiketli bir veri kümesine ihtiyacınız vardır. Her veri noktasının bir dizi özelliği (giriş değişkenleri) ve karşılık gelen bir sınıf etiketi veya hedef değeri (çıkış değişkeni) olmalıdır.

b. Karar ağaçları oluşturma:

Rastgele Orman, eğitim sırasında birden fazla karar ağacı oluşturur. Her ağaç, verilerin rasgele bir alt kümesi ve rasgele bir özellik alt kümesi kullanılarak oluşturulur. Bu işlem, önyükleme toplama veya torbalama olarak bilinir.

c. Önyükleme örneği:

Her karar ağacı için, orijinal veri kümesinden değiştirilerek rastgele bir veri noktaları alt kümesi (örnekler) seçilir. Bu, bazı veri noktalarının örnekte tekrarlanabileceği, bazılarının ise dışarıda bırakılabileceği anlamına gelir. Ortalama olarak, her ağaçta orijinal veri setinin yaklaşık üçte ikisi kullanılır.

d. Öznitelik seçimi:

Bir karar ağacının her düğümünde, en iyi ayrımı yapmak için yalnızca rastgele bir özellik alt kümesi dikkate alınır. Bu, her ağaç farklı özelliklere odaklandığından ve fazla uydurmayı azalttığından, karar ağaçları arasında çeşitlilik sağlar.

e. Ağaç inşası:

Her karar ağacı, özyinelemeli bir süreç kullanılarak büyütülür. Ağaç, seçilen özelliklere göre her düğümdeki verileri bölerek genişletilir. Bölme, elde edilen her bir alt kümedeki verilerin homojenliğini en üst düzeye çıkarmak için seçilir. Bu süreç, maksimum ağaç derinliği veya yaprak düğüm başına minimum örnek sayısı gibi önceden tanımlanmış bir durdurma kriteri karşılanana kadar devam eder.

f. Tahminleri birleştirme:

Tüm karar ağaçlarını oluşturduktan sonra, Rastgele Orman son kararı vermek için tahminlerini birleştirir. Sınıflandırma görevleri için, her ağacın girdi veri noktasının sınıf etiketi için "oy kullandığı" oylamayı kullanır. En çok oyu alan sınıf son tahmindir. Regresyon görevleri için, her ağacın sayısal bir değer öngördüğü ve son tahminin tüm ağaç tahminlerinin ortalaması olduğu ortalamayı kullanır.

g. Performans arttırmak:

Rastgele Orman, aşırı uydurmayı azalttığı için bireysel karar ağaçlarına kıyasla daha iyi genelleme ve sağlamlık sağlar. Ortalama alma veya oylama işlemi, tek tek ağaç önyargılarını düzeltmeye yardımcı olur ve nihai tahminlerdeki varyansı azaltır.

2.2.3.2 Rastgele orman kullanım örnekleri

Biyoinformatik (Breiman, 2001) ve hisse senedi piyasası tahmini gibi çeşitli alanlarda yaygın olarak kullanılmışlardır.

Başarı tahmini bağlamında, Rastgele Orman farklı alanlarda kullanılmıştır. Örneğin, pazarlamada (Miron B. Kurs, 2010), demografik ve davranışsal verilere dayalı müşteri kaybını tahmin etmek için Rastgele Orman'ı kullanmıştır.

Finans alanında, (Pena, 2019) finansal göstergeleri ve geçmiş verileri kullanarak hisse senedi fiyat hareketlerini tahmin etmek için Rastgele Ormanı uygulamıştır.

Rastgele Orman, sağlık hizmetlerinde de kullanılmıştır. (Alba, 2017), klinik verilere ve demografik bilgilere dayalı olarak hasta yeniden kabul oranlarını tahmin etmek için Rastgele Orman'ı kullanmıştır.

2.2.4 Karar ağaçları (decision tree)

Karar ağaçları, kararları veya sınıflandırmaları modellemek için ağaç benzeri bir yapı kullanan popüler bir makine öğrenimi tekniğidir. Her iç düğüm bir özelliği veya değişkeni temsil ederken, her yaprak düğümü bir kararı veya sonucu temsil eder. Karar ağaçları yorumlanması ve görselleştirilmesi kolay olduğu için, karar verme süreçlerinde değerlidir. Tıbbi teşhis ve kredi riski değerlendirmesi gibi çeşitli alanlarda başarıyla uygulanmışlardır (Quinlan, 1986).

2.2.5 Sinir ağları (neural network)

Sinir ağları, insan beyninin yapısından esinlenerek, veriyi işlemek ve öğrenmek için birlikte çalışan bağlantılı düğümlerden (sinir hücreleri) oluşur. Karmaşık desenleri ele alma yetenekleri ve görüntü tanıma (Krizhevsky A. a., 2012) ve doğal dil işleme (Jacob Devlin, 2018) gibi çeşitli görevlerde son teknoloji performansı elde etmeleri nedeniyle son yıllarda büyük popülerlik kazanmışlardır.

2.2.6 Kümeleme (clustering)

Kümeleme, verileri benzer özelliklere sahip noktaları bir araya getirerek gruplandırmak için kullanılan bir denetimsiz öğrenme tekniğidir. Veri keşfi, desen tanıma ve müşteri segmentasyonu gibi alanlarda sıkça kullanılır. K-means (MacQueen, 1967) hiyerarşik kümeleme (Ward Jr, 1963) ve DBSCAN (Ester, 1996) gibi çeşitli kümeleme algoritmaları bulunmaktadır.

2.2.7 Takviyeli öğrenme (reinforce learning)

Takviyeli öğrenme, bir ajanın bir ortamda ardışık kararlar alarak kümülatif bir ödülü maksimize etmeyi öğrenmeye odaklanır. Bu yöntem, ajanın çevre ile etkileşim içinde öğrenmesini içerir ve ajanın eylemlerine bağlı olarak ödüller veya cezalar şeklinde geri bildirim alır. Takviyeli öğrenme, robotik (Kober, 2013) ve oyun oynama (Mnih, 2015) gibi alanlarda başarıyla uygulanmış ve karmaşık davranışları öğrenme ve optimal karar verme stratejilerine ulaşma yeteneği göstermiştir.

2.2.8 Gauss süreçleri (gaussian process)

Gauss süreçleri, esnek bir çerçeve sunan olasılıksal modellerdir ve regresyon ve olasılıksal tahminler için kullanılırlar. Fonksiyon dağılımını modellerler ve tahminlerdeki belirsizlikleri yakalayabilirler. Gauss süreçleri, mekânsal modelleme (Christopher K. I. Williams, 2005) ve zaman serisi analizi (Wilson, 2013) gibi çeşitli alanlarda uygulanmış ve doğru tahminler ve etkili belirsizlik hesaplama imkanı sağlamıştır.

2.2.9 Ensemble öğrenme (ensemble learning)

Ensemble öğrenme, birden fazla modeli veya sınıflandırıcıyı birleştirerek tahmin doğruluğunu ve genelleme yeteneğini artırmayı hedefler. Bireysel modeller arasındaki çeşitlilikten yararlanarak daha sağlam ve doğru tahminler yapmayı amaçlar. AdaBoost (Freund, 1997) ve Gradient Boosting (Friedman, 2001) gibi ensemble yöntemler, makine görüşü (Viola, 2001) ve sahtekârlık tespiti (Li, 2008) gibi alanlarda büyük başarılar elde etmiştir.

2.2.10 Derin öğrenme (deep learning)

Derin öğrenme, çok katmanlı derin sinir ağlarını eğiterek verinin hiyerarşik temsillerini öğrenmeyi hedefler. Doğal dil işleme (Vaswani, 2017) ve konuşma tanıma (Hinton, 2012) gibi çeşitli karmaşık görevlerde dikkate değer başarılar göstermiştir. Evrişimli sinir ağları (CNN) ve tekrarlayan sinir ağları (RNN) gibi derin öğrenme mimarileri, bilgisayar görüşü ve dizi modellemesi gibi alanları kökten değiştirmiştir. Derin öğrenme, bilgisayarlı görü ve ses tanıma, doğal dil işleme, tıbbi görüntü analizi ve oyun stratejileri gibi çeşitli alanlarda başarıyla kullanılmaktadır. Ayrıca, otomotiv, eğlence, finans ve sağlık gibi sektörlerde önemli rol oynar.

2.2.11 Transfer öğrenme (transfer learning)

Transfer öğrenme, bir görev veya alan üzerindeki bilgiyi başka bir ilgili ancak farklı görev veya alanda öğrenme ve performansı iyileştirme amacı güder. Öğrenilen temsilleri veya modelleri transfer ederek, sınırlı etiketli veriyle daha verimli ve etkili bir öğrenme sağlar. Transfer öğrenme, resim sınıflandırma (Krizhevsky A. , 2012) ve doğal dil işleme (Howard, 2018) gibi alanlarda uygulanmış ve büyük ölçekli veri kümelerinde önceden eğitilmiş modellerin performansı artırmış ve geliştirmeyi hızlandırmıştır.

Araştırmacılar ve uygulayıcılar, bu alt dallarda yeni teknikler keşfetmeye ve geliştirmeye devam ederek karmaşık sorunları çözmek ve makine öğreniminin potansiyelini açığa çıkarmak için çalışmaktadır.

Çizelge 2.1: Makine öğrenmesi algoritmalarının özet hali

Algoritma	Tanım	Uygulamalar	Avantajlar	Dezavantajları
Doğrusal Regresyon	Girdiler ile sürekli bir sayısal çıktı değişkeni arasındaki doğrusal ilişkiyi modelleyen basit bir algoritma	Hisse Senedi Fiyat Tahmini Konut fiyatlarını tahmin etmek Müşteri yaşam boyu değerini tahmin etme	Açıklanabilir yöntem Çıkış katsayısı ile yorumlanabilir sonuçlar Diğer makine öğrenimi modellerinden daha hızlı eğitilir	Girdiler ve çıktılar arasında doğrusallık olduğunu varsayar Aykırı değerlere duyarlı Küçük, yüksek boyutlu verilerle yetersiz kalabilir
Lojistik regresyon	Girdiler ile kategorik bir çıktı (1 veya 0) arasındaki doğrusal ilişkiyi modelleyen basit bir algoritma	Kredi risk skorunu tahmin etme Müşteri kaybı tahmini	Yorumlanabilir ve açıklanabilir Düzenleştirme kullanılırken fazla takmaya daha az eğilimli Çok sınıflı tahminler için uygulanabilir	Girdiler ve çıktılar arasında doğrusallık olduğunu varsayar Küçük, yüksek boyutlu verilerle uyum sağlayabilir
Ridge Regresyon	Regresyon ailesinin bir parçası — katsayılarını sifıra yaklaştırarak düşük öngörü sonuçları olan özellikleri cezalandırır. Sınıflandırma veya regresyon için kullanılabilir	Otomobiller için kestirimci bakım Satış geliri tahmini	Overfittinge daha az eğilimli Verilerin çoklu bağlantıdan muzdarip olduğu durumlar için en uygundur Açıklanabilir ve yorumlanabilir	Tüm öngörücüler son modelde tutulur. Özellik seçimi yapmıyor

Çizelge 2.1(Devamı) : Makine öğrenmesi algoritmalarının özet hali

Algoritma	Tanım	Uygulamalar	Avantajlar	Dezavantajlar
Karar ağacı	Karar Ağacı modelleri, tahminler üretmek için özellikler üzerinde karar kuralları oluşturur. Sınıflandırmaya veya regresyon için kullanılabilir.	Müşteri kaybı tahmini Kredi puanı modelleme hastalık tahmini	Açıklanabilir ve yorumlanabilir Eksik değerleri işleyebilir	Overfitting eğilimli Aykırı değerlere duyarlı
Rastgele Ormanlar	Birden fazla karar ağacının çıktısını birleştiren bir topluluk öğrenme yöntemi	Kredi puanı modelleme Konut fiyatlarını tahmin etmek	Fazla takmayı azaltır Diğer modellere kıyasla daha yüksek doğruluk	Eğitim karmaşıklığı yüksek olabilir Pek yorumlanamaz
Gradyan Boosting Regresyonu	Gradient Boosting Regresyon, tahmine dayalı zayıf öğrenenlerden oluşan bir topluluktan tahmine dayalı modeller oluşturmak için artırımı kullanır	Araba emisyonlarını tahmin etme Araç çağırma ücret tutarını tahmin etme	Diğer regresyon modellerine kıyasla daha iyi doğruluk Çoklu bağlantılılığı kaldırabilir Doğrusal olmayan ilişkileri yönetebilir	Aykırı değerlere karşı hassastır ve bu nedenle fazla uydurmaya neden olabilir Hesaplama açısından pahalı ve yüksek karmaşıklığa sahip
XGBoost	Verimli ve esnek Gradient Boosting algoritması. Hem sınıflandırma hem de regresyon görevleri için kullanılabilir	kayıp tahmini Sigortada hasar işlemler	Doğru sonuçlar sağlar Doğrusal olmayan ilişkileri yakalar	Hiperparametre ayarı karmaşık olabilir Seyrek veri kümelerinde iyi performans göstermez
K-Means	K-Means, en yaygın kullanılan kümeleme yaklaşımıdır; öklid mesafelerine dayalı olarak K kümesini belirler.	Müşteri segmentasyonu Öneri sistemleri	Büyük veri kümelerine ölçeklenir Uygulaması ve yorumlaması basit Sıkı kümelerdeki sonuçlar	Baştan itibaren beklenen sayıda küme gerektirir Değişken küme boyutları ve yoğunluklarıyla ilgili sorunları var

2.3 Düşük Kod ve Kodsuz Geliştirme Platformları (No-code /Low-code)

Düşük Kodlu Geliştirme Platformları (Low-code development platforms): Düşük kodlu platformlar, geliştiricilerin minimum el kodlaması ile uygulamalar oluşturmasını sağlayan görsel geliştirme araçlarıdır. Genellikle sürükle ve bırak arabirimleri ve önceden oluşturulmuş bileşenler sağlayarak, geliştiricilerin kapsamlı kod yazmak yerine uygulama mantığına ve işlevselliğine odaklanmasına olanak tanır.

Kodsuz Geliştirme Platformları (No-code development platforms): Kodsuz platformlar, düşük kod kavramını bir adım öteye taşır. Basit görsel arabirimler kullanarak uygulamalar oluşturmak için çok az kodlama deneyimi olan veya hiç olmayan kullanıcılara güç verirler. Kodsuz platformlar genellikle herhangi bir kod yazmadan uygulama bileşenlerinin birleştirilmesini ve ayarların yapılandırılmasını içerir.

2.3.1 Orange Data Mining

Orange Data Mining, açık kaynaklı veri analitiği ve makine öğrenimi aracıdır. 1996'da Slovenya, Ljubljana Üniversitesi, Bilgisayar ve Bilişim Bilimleri Fakültesi'ndeki Biyoinformatik Laboratuvarı tarafından geliştirilmiştir. Orange, kullanıcı dostu olacak şekilde tasarlanmıştır ve hem yeni başlayanlar hem de deneyimli veri analistleri için Makine Öğrenmesini erişilebilir olmasını sağlar. Kullanıcıların veri işleme ve analiz işlemlerini temsil eden çeşitli widget'ları birbirine bağlayarak veri iş akışları oluşturmasını sağlayan görsel bir programlama arabirimi sağlar (Orange, 2023).

Orange Data Mining'in temel özellikleri şunları içerir (Orange Data Mining Docs, 2023):

Görsel Programlama: Orange, kullanıcıların widget'ları tuvale sürükleyip bırakabilecekleri ve veri iş akışları oluşturmak için bunları bağlayabilecekleri görsel bir programlama ortamı sunar. Bu yaklaşım, kullanıcıların kod yazmadan karmaşık veri analizi süreçleri tasarlamasına olanak tanır.

Widget Kitaplığı: Platform, çok çeşitli veri ön işleme, analiz, görselleştirme ve modelleme görevlerini kapsayan zengin bir widget kitaplığı içerir. Kullanıcılar verileri işlemek, kalıpları keşfetmek ve tahmine dayalı modeller oluşturmak için bu pencere öğelerini kullanabilir.

Veri Keşfi ve Görselleştirme: Orange, veri keşfi ve görselleştirme için çeşitli widget'lar sunarak kullanıcıların etkileşimli çizelgeler, grafikler ve araziler aracılığıyla veri kümelerinden içgörüler elde etmelerini sağlar.

Makine Öğrenimi: Platform, sınıflandırma, regresyon, kümeleme ve daha fazlası için çeşitli makine öğrenimi algoritmalarını ve modellerini destekler. Kullanıcılar tahmine dayalı modelleme görevleri için bu modelleri verilerine uygulayabilir.

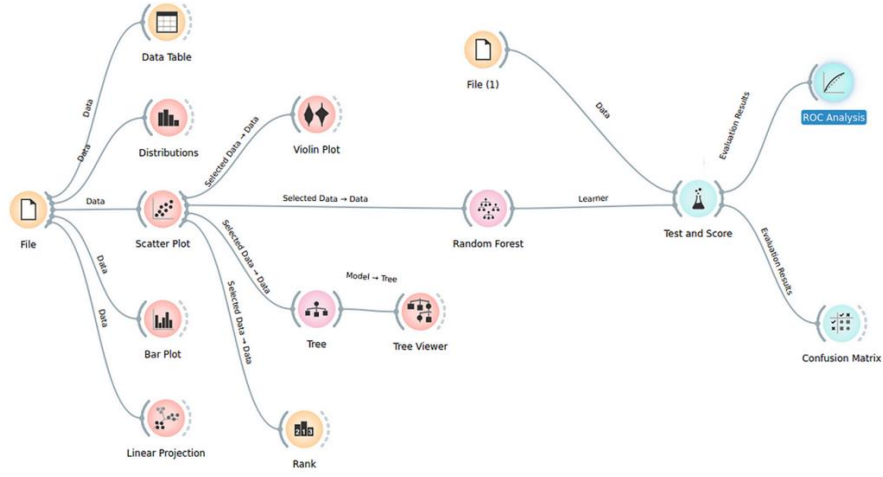
Etkileşimli İş Akışları: Orange, kullanıcıların veri iş akışlarıyla gerçek zamanlı olarak etkileşim kurmasına olanak tanır. Parametreleri ince ayar yapabilir, veri manipülasyonlarını ayarlayabilir ve sonuçları hemen görselleştirebilirler.

Veri Entegrasyonu: Orange, dosyalar, veritabanları ve web hizmetleri dahil olmak üzere çeşitli kaynaklardan veri entegrasyonunu destekler. Ayrıca, verileri analize hazırlamak için veri dönüştürme ve temizleme widget'ları sunar.

Eklentiler ve Uzantılar: Kullanıcılar, Orange topluluğundan eklentiler ve uzantılar yükleyerek Orange'ın işlevselliğini geliştirebilir. Bu eklentiler, özel görevler için ek pencere öğeleri ve işlevler sağlar.

Orange Veri Madenciliği, çeşitli veri analizi ve makine öğrenimi projeleri için akademi, araştırma ve endüstride yaygın olarak kullanılmaktadır. Kullanıcı dostu arayüzü ve güçlü veri analizi yetenekleri nedeniyle özellikle biyoinformatik, sosyal bilimler ve eğitim araştırmaları gibi alanlarda popülerdir.

Diğer açık kaynaklı araçlara benzer şekilde Orange, işbirliğini ve topluluk katkısını teşvik eder. Platform, kullanıcı topluluğu ve geliştiricilerin katkılarıyla sürekli güncellenmekte ve geliştirilmektedir. Açık kaynaklı bir yazılım olarak Orange ücretsiz olarak kullanılabilir ve kaynak kodu inceleme ve değişiklik için açıktır. Bu, kullanıcıların aracı kendi özel ihtiyaçlarına uyacak şekilde özelleştirmesine ve işlevselliğini gerektiği gibi genişletmesine olanak tanır.



Şekil 2.1 : Orange Data Mining programına ait bir arayüz.

2.3.2 KNIME

KNIME ("naim" olarak telaffuz edilir), açık kaynaklı bir veri analizi, raporlama ve entegrasyon platformudur. "Konstanz Information Miner" anlamına gelir ve başlangıçta Almanya'daki Konstanz Üniversitesi'nde geliştirilmiştir. KNIME, kullanıcıların, verilerin çeşitli veri işleme ve analiz işlemlerini temsil eden bir dizi düğüm aracılığıyla işlendiği veri iş akışlarını görsel olarak tasarlamasına olanak tanır. (KNIME, 2023)

KNIME'nin temel özellikleri şunları içerir:

Görsel Veri Analitiği: KNIME, kullanıcıların sürükle ve bırak yaklaşımı kullanarak veri iş akışlarını görsel olarak tasarlamasına olanak tanıyan bir grafik kullanıcı arabirimi (GUI) sağlar. Bu, programcı olmayanlar da dahil olmak üzere kullanıcıların karmaşık veri analizi süreçleri oluşturmasını kolaylaştırır.

Kapsamlı Düğüm (Node) Deposu: KNIME, veri işleme, veri madenciliği, makine öğrenimi, metin madenciliği ve görselleştirme düğümleri gibi çok çeşitli önceden oluşturulmuş düğümler sunar. Kullanıcılar bu düğümleri özelleştirebilir ve hatta belirli görevler için kendi özel düğümlerini oluşturabilir.

Veri Analitiği ve Makine Öğrenimi: KNIME, kullanıcıların kod yazmaya gerek kalmadan veri analizi, veri dönüştürme ve makine öğrenimi görevlerini gerçekleştirmesine olanak tanır. Tahmine dayalı modeller oluşturmak ve verilerden içgörüler elde etmek için çeşitli makine öğrenimi algoritmalarını ve araçlarını destekler.

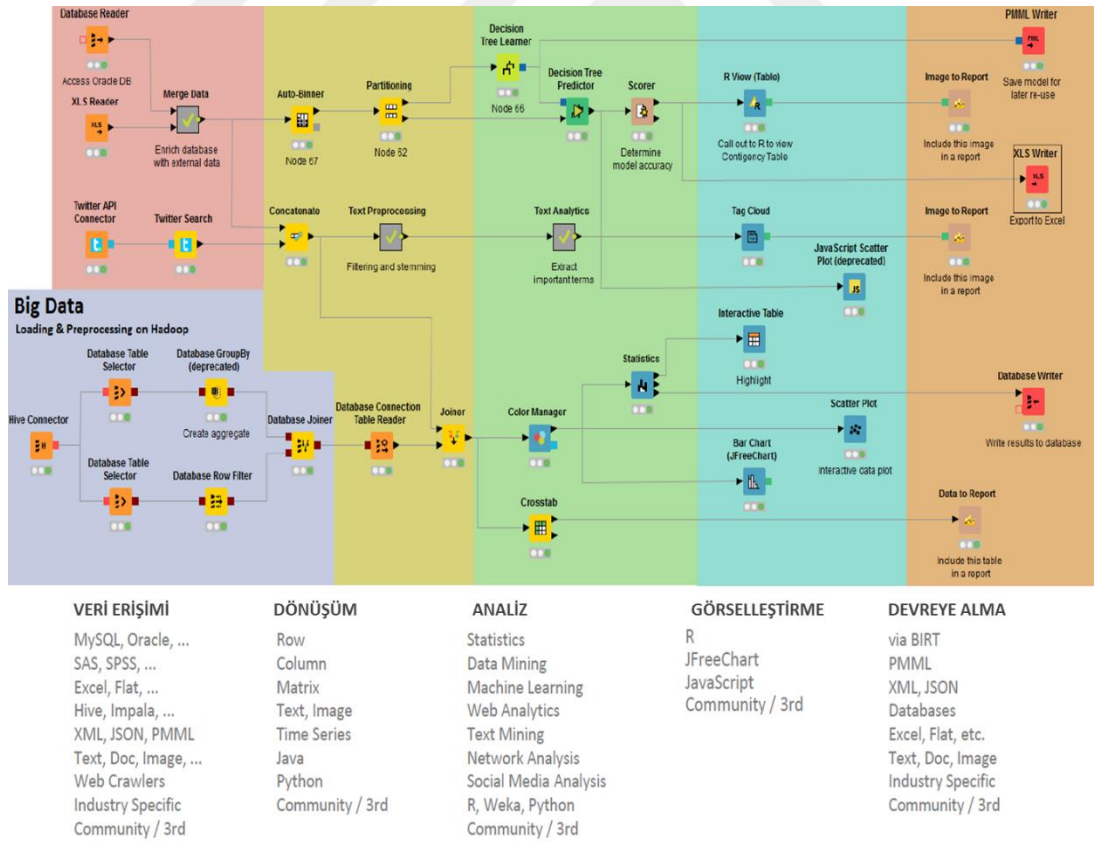
Raporlama ve Görselleştirme: KNIME, raporlar oluşturmak ve veri analizi sonuçlarını görselleştirmek için araçlar sağlar. Kullanıcılar, bulguları etkili bir şekilde iletmek için etkileşimli panolar ve görselleştirmeler oluşturabilir.

İş Akışı Paylaşımı ve İşbirliği: KNIME, kullanıcıların iş akışlarını başkalarıyla paylaşmasına izin vererek ekiplerin veri analizi projelerinde işbirliği yapmasını kolaylaştırır.

Genişletilebilirlik: KNIME oldukça genişletilebilir ve diğer veri analitiği araçları ve kitaplıklarıyla entegre edilebilir.

Ek olarak, kullanıcılar platformun işlevselliğini geliştirmek için özel uzantılar oluşturabilir. KNIME, veri bilimi, biyoinformatik, iş zekası ve araştırma dahil olmak üzere çeşitli alanlarda yaygın olarak kullanılmaktadır. Kullanıcı dostu arayüzü, kapsamlı düğüm deposu ve büyük veri kümelerini işleme yeteneği, onu veri analistleri, veri bilimcileri ve araştırmacılar için popüler bir seçim haline getiriyor.

KNIME'in topluluk sürümünün kullanımı ücretsizdir.



Şekil 2.2 : KNIME programı ile yapılabilecek çalışmalara dair özet

3. ARAŞTIRMA MODELİ ve METOT

Yapılan literatür araştırması ve pratik bilgidan esinlenerek bu çalışmada crunchbase sitesinin sağladığı verilerden faydalanılmıştır. Crunchbase, girişim ekosisteminde olan tüm firmaların kaydının tutulduğu bir internet sitesidir ve isteyen herkes bu siteye üye olup veritabanındaki girişimler hakkında çeşitli bilgiler elde edebilir. Kullanılan veri tabanı Türkiye'deki girişim ekosisteminde en büyük payı alan (The Global Startup Ecosystem Report 2023 : Syf 159) oyun, finansal teknolojiler ve yapay zekâ alanındaki girişimler filtrelenerek elde edilmiştir. Filtrelenmemiş hali ile 06.01.2023 tarihinde indirilen csv veritabanında 1.048.576 adet kayıt bulunmaktadır. Kapsamı daraltmak ve ihtiyacı doğru tespit edebilmek adına özellikle 3 lokomotif sektör seçilmiştir. 28.03.2023 tarihinde Türkiye ve yapay zeka, oyun ve finansal teknolojiler filtresindeki veriler excele aktarılmıştır. Excelde oluşturulan dosya başlık hariç 1000 satır ve 55 sütundan oluşan bir matris yapısındadır. Sütunları temsil eden başlıklar Çizelge 3.1'de gösterilmiştir.

Çizelge 3.1 : Ham veri seti sütun isimleri.

Kolon A
Organization Name
Organization Name URL
Investor Type
Full Description
Industries
Headquarters Location
Description
CB Rank (Company)
Founded Date
Founded Date Precision
Operating Status
Exit Date
Exit Date Precision
Closed Date
Closed Date Precision
Company Type
Investment Stage

Çizelge 3.1 (devam) : Ham veri seti sütun isimleri.

Accelerator Program Type
Accelerator Application Deadline
Accelerator Duration (in weeks)
School Type
School Program
Number of Enrollments
School Method
Number of Founders (Alumni)
Number of Founders
Founders
Number of Employees
Number of Funding Rounds
Funding Status
Last Funding Date
Last Funding Amount
Last Funding Amount Currency
Last Funding Amount Currency (in USD)
Total Funding Amount
Total Funding Amount Currency
Total Funding Amount Currency (in USD)
Number of Lead Investors
Number of Investors
Top 5 Investors
Stock Symbol
Stock Symbol URL
Valuation at IPO
Valuation at IPO Currency
Valuation at IPO Currency (in USD)
CB Rank (Organization)
CB Rank (School)
Similar Companies
IPqquery - Patents Granted

Diğer özniteliklerden farklı olarak CB Rank (Organization) ile ilgili olarak Crunchbase veri kümesindeki tüm varlıklar (şirket, kişiler, yatırımcılar vb.) için dinamik bir sıralamadır. Bir şirketin sıralaması, gerçek zamanlı olarak neyin en önemli olduğunu hızlı bir şekilde görebilmeniz için varlıkları puanlamak ve sıralamak için Crunchbase'in akıllı algoritmaları kullanılarak bulunur. Başka bir deyişle, bir varlığın önemini ölçmektedir.

Crunchbase sıralama algoritması, bir profilin sahip olduğu bağlantı sayısı, topluluk katılım düzeyi, finansman etkinlikleri, haber makaleleri ve satın almalar dahil olmak üzere birçok sinyali hesaba katmaktadır. Bir varlığın Crunchbase sıralaması değişkendir ve zaman içinde yükselip alçalabilir. Ürün lansmanları, finansman etkinlikleri, liderlik değişiklikleri ve haberler gibi zamana duyarlı olaylar bir şirketin Crunchbase sıralamasını etkiler (www.crunchbase.com, 2023).

Bu çalışmada veri temizleme, seçme ve dönüştürme için Microsoft Excel, denetimli makine öğrenmesi algoritmalarının çalıştırılması için no-code programlama yazılımı Orange, Knime ve denetimli makine öğrenmesi algoritmaları olan Lojistik Regresyon, Destek Vektör Makineleri, ve Rastgele Orman algoritmaları, yukarıdaki veri seti üzerinde kullanılmış ve algoritmaların sonuçları karşılaştırılmıştır.

3.1 Verinin Kullanılacak Hale Getirilmesi

Veri ön işleme, denetimli bir makine öğrenimi görevinin genel performansı üzerinde genellikle kritik bir etkiye sahip olabilir. Veri analizinin ilk aşamasında, CrunchBase veritabanının tam olarak anlaşılması ve ardından veri temizleme işlemi (eksik değerler, kopyalar, fazlalık veriler) yapılmıştır. 58 farklı sütundan oluşan veri seti ilk olarak 33 sütuna düşürülmüş daha sonrasında ise yeni türetilen sütunlar ve silinenler ile birlikte 25 sütuna düşürülmüştür. Filtrelenmeye hazır tam bir veritabanına sahip olmak, modelde kullanılacak verilerin kapsamını tanımlamak ve temel özelliklerin keşifsel bir analizini yapabilmek için önemlidir. Öğrenme görevinde kullanılacak son veri setini oluşturacak yeni özellikler tanımlanarak ve oluşturularak verilerin dönüştürülmesi sağlanmıştır.

Çizelge 3.2 : Eğitimde kullanılacak veri setine ait sütunlar.

Sütun İsimleri
Organization Name
Industries
City
CB Rank (Company)
CB Rank (Organization)
Founded Date
Exit Date
Exit Etmek için Geçen Yıl
Company Type
Number of Founders
Founders1

Çizelge 3.3 (devam) : Eğitimde kullanılacak veri setine ait sütunlar.

Sütun İsimleri

Founders2
Founders3
Founders4
Founders5
Founders6
Founders7
Number of Employees
Number of Funding Rounds
Funding Status
Last Funding Date
Last Funding Amount Currency (in USD)
Total Funding Amount Currency (in USD)
Number of Lead Investors
Number of Investors
Top 5 Investors 1
Top 5 Investors 2
Top 5 Investors 3
Top 5 Investors 4
Top 5 Investors 5
Organization Name URL
Full Description
Description
Similar Companies
IPqwery - Patents Granted

Gelinen son noktada veristesinin içerisinde bulunan bankalar, risk sermayeleri, sigorta şirketleri ve bunun gibi startup tanımına uymayan şirketler kaldırılmıştır. Bu haliyle Türkiye’deki Fintech, Yapay Zeka ve Oyun şirketlerinden oluşan veriseti 636 satır ve 25 sütundan oluşan bir hale getirilmiştir. Final verisetinde toplamda 333 farklı yatırımcı olduğu gözlemlenmiştir. 333 farklı yatırımcının 121’i bireysel (melek) yatırımcılar geri kalanlar ise Kurumsal (VC, CVC, PE) yatırımcılar olduğu gözlemlenmiştir. Ayrıca 688 adet kurucu bulunmaktadır.

3.2 Ölçüm Metrikleri

Çalışmalarda sonuçların doğruluğunun belirlenebilmesi için doğruluk (accuracy), kesinlik (precision), duyarlılık (recall), F1-skor ve karmaşıklık matrisi (confusion matrix), kullanılan ölçüm metrikleri arasında yer almaktadır.

	Tahminlenen (Predicted)	
	Gerçekleşen (Actual)	True Positives (TP) Şirket başarılı olarak sınıflandırıldı ve başarılı
False Positives (FP) Şirket Başarılı olarak sınıflandırıldı ve başarısız		True Negatives (TN) şirket başarısız olarak sınıflandırıldı ve başarısız oldu

Şekil 3.1: Karmaşıklık Matrisi

Karışıklık matrisi, modelin tahmin ettiği değerlerle gerçek değerleri karşılaştıran bir matris çıktısı verir. Bu matriste doğru pozitif (TP), doğru negatif (TN), yanlış pozitif (FP) ve yanlış negatif (FN) değerleri bulunmaktadır. TP, pozitif olup, modelin de pozitif tahmin ettiği; TN, negatif olup, modelin de negatif tahmin ettiği ; FP, aslında negatif değerde olup da modelin pozitif dediği; FN ise, aslında pozitif olup, modelin negatif dediği değerleri içerir. Bazı ölçüm metodları ve formülleri aşağıdaki gibidir:

Doğruluk metriği, TP ve TN değerlerinin toplanıp, toplam veri sayısına bölünmesiyle elde edilir.

$$\text{Doğruluk (Accuracy)} = \frac{TP + TN}{TP + TN + FP + FN}$$

Kesinlik metriği, TP değerinin tüm pozitif tahmin edilen değerlerin sayısına bölünmesiyle bulunur.

$$\text{Kesinlik (Precision)} = \frac{TP}{TP + FP}$$

Duyarlılık metriği ise TP değerinin gerçekte pozitif olan tüm değerlere bölünmesiyle hesaplanır.

$$\text{Duyarlılık (Sensevity yada Recall yada TPR)} = \frac{TP}{TP + FN}$$

Özgüllük metriği ise TN değerinin gerçekte negatif olan tüm değerlere bölünmesiyle hesaplanır.

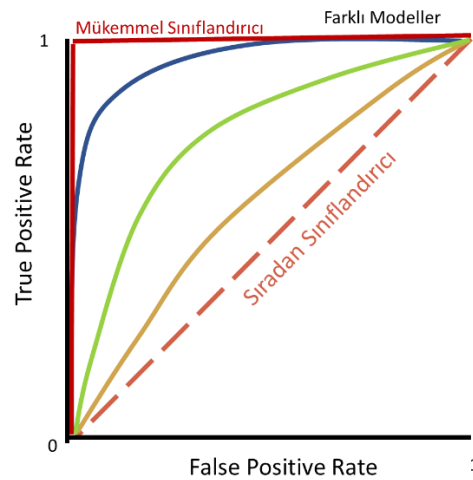
$$\text{Özgüllük (Specificity yada TNR)} = \frac{TN}{TN + FP}$$

F1-Skor metriği, kesinlik ve duyarlılık metriklerin harmonik ortalamasıyla elde edilen bir metriktir.

$$F1\ Skor = 2 * \frac{Kesinlik * Duyarlilik}{Kesinlik + Duyarlilik}$$

Çapraz doğrulama, belirlenen k değeri kadar veri setinin bölünmesi ve bir parçanın test olarak kullanılması, diğer parçaların eğitim için kullanılması şeklinde kurgulanan bir yöntemdir. Her parça bir kez test verisi olarak kullanılır ve her adımda doğruluk değerleri kaydedilir. Bulunan tüm doğruluk değerlerinin bir ortalaması alınır. Bu değer, modelin ortalama doğruluk değeridir.

Ayrıca diğer bir ölçüm metodu da Alıcı İşletim Karakteristiği (Receiver Operating Characteristic – ROC) eğrileri ve Eğrinin Altında Kalan Alan (Area Under the Curve – AUC). ROC eğrisi, ikili sınıflandırma sistemlerinde ayırım eşik değerinin farklılık gösterdiği durumlarda, hassasiyetin kesinliliğe olan oranıyla ortaya çıkmaktadır. ROC daha basit anlamda doğru pozitiflerin, yanlış pozitiflere olan kesri olarak da ifade edilebilir (Swets, 1996). Eğri, kesinlik ve hassasiyet arasındaki dengeyi değerlendirmek için kullanılmıştır. ROC eğrisi altında kalan alan ROC puanı olarak tanımlanabilir. ROC eğrisi değişen sınıflandırma eşik değerlerine göre doğru pozitiflerin sayısının, yanlış pozitiflerin bir fonksiyonu olarak çizilmesiyle oluşmaktadır. ROC puanı 1 (bir) olduğunda anlamı, pozitifler mükemmel bir şekilde negatiflerden ayrılmıştır, olmaktadır. ROC puanı 0 (sıfır) olduğunda ise herhangi bir pozitif bulunamadı anlamına gelir.

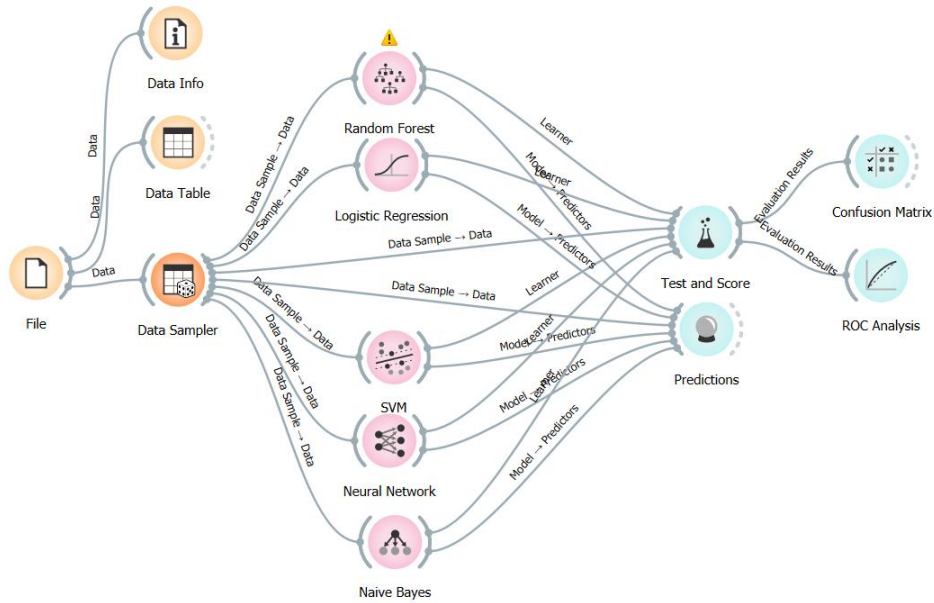


Şekil 3.2 : ROC Eğrilerinin çalışma prensibini anlatan bir görsel

4. ANALİZ ve BULGULAR

Crunchbase'den alınan Türkiye'deki Oyun, Finansal Teknolojiler ve Yapay Zeka alanındaki girişimlerin başarı tahminlemede hangi makine öğrenimi algoritmasının amacımıza daha iyi uyduğunu görmek için değerlendirmenin ilk aşamasında, birkaç algoritmanın doğruluğu hesaplandı. Bir makine öğrenimi algoritmasını değerlendirmek ve seçmek için tek başına doğruluk önerilmese de hızlı ve kolay yorumlanabilen bir metrik olarak kullanılabilir.

Veri setinin doğası gereği, ilk izlenim olarak çok fazla öznitelik olması ve verisetinin az olmasından kaynaklı olarak sınıflandırıcıların (karar ağaçları, rastgele orman vb) daha iyi çalışacağı öngörülerek 5 farklı algoritma üzerinde ön denemeler yapılmıştır.



Şekil 4.1: Orange üzerinde yapılan makine öğrenmesi çalışması

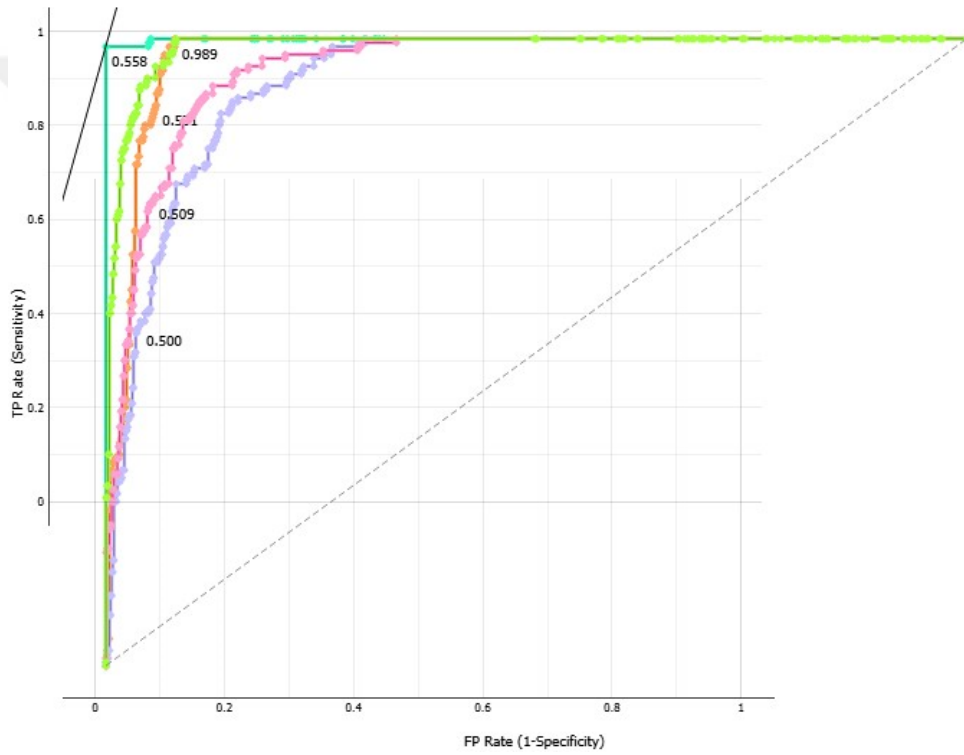
Tam veri setinin %80'lik kısmı eğitim için seçildi ve 10 kez çapraz kat doğrulama kullanılarak aşağıdaki sonuçlar elde edilmiştir.

Test edilen tüm makine öğrenimi algoritmaları, tümü çok yüksek bir genel doğruluk elde edilmiştir. Rastgele Ormanlar, Şekil 4.1'de gösterildiği gibi tüm ölçüm metriklerinde en yüksek değerleri elde etmiştir.

Çizelge 4.1 : Sınıflandırma Raporu

Algoritma	Doğruluk (Accuracy)	F1 Score	Kesinlik (Precision)	Duyarlılık (Recall - TPR)	Özgüllük (TNR)
Rastgele Orman	0,996	0,996	0,996	0,996	0,989
Lojistik Regresyon	0,933	0,935	0,939	0,933	0,902
Yapay Sinir Ağları	0,910	0,910	0,910	0,910	0,779
Naive Bayes	0,900	0,908	0,936	0,900	0,979
Destek Vektör Makineleri	0,873	0,857	0,862	0,873	0,511

Rastgele Orman en yüksek Recall oranına sahip olması ve Kesinlik ile Recall arasındaki iyi dengesi nedeniyle seçilen algoritmadır.



Şekil 4.2: Algoritma Sonuçlarının ROC Eğrileri

Test edilen tüm algoritmaları karşılaştırarak, Şekil 4.1'de gösterildiği gibi beş makine öğrenimi algoritması için AUC puanlarının çok iyi olduğu değerlendirilebilir.

Rastgele Ormanlar, diğer algoritmaların herhangi birinden daha sağlam bir model oluşturur. Lojistik Regresyon, verilerde öngörülebilir bir doğrusallık olduğu için iyi bir çözüm olduğunu kanıtıyor. DVM, gözlemleri çok boyutlu bir özellik uzayında sınıflandırma kabiliyeti nedeniyle, eğitim için daha yüksek bir süre pahasına da olsa mükemmel sonuçlar sağlar. Rastgele Ormanlar, belirli bir şirketin kategorisi için

diğerlerinden daha önemli olan özellikler gibi daha karmaşık bağımlılıkları keşfedebileceği için nihai olarak seçildi. Geçerli veri kümesi, herhangi bir kayıp değeri olmayan yalnızca ikili özelliklerden oluşmasına rağmen, Rastgele Orman aynı zamanda kategorik veriler, büyük aykırı değerler içeren veriler ve seyrek veriler aracılığıyla da çalışır.



5. SONUÇ VE ÖNERİLER

Yapay zeka teknolojileri, birçok farklı endüstride bir şirketin değer zincirindeki karar alma senaryolarında geniş çapta benimseniyor. Spesifik olarak, finansal hizmetlerde yapay zekanın benimsenmesi, müşteri odaklı, operasyon tabanlı, ticaret tabanlı ve düzenleyici tabanlı kullanım durumlarında karar vermeyi iyileştirmek için hızlı bir şekilde artıyor. Girişim firmaları (özellikle yeni çağ firmaları), karar vermek için veriye dayalı yaklaşımları aktif olarak kullanıyor ve ayrıca rekabet avantajı oluşturmak için yapay zeka teknikleriyle deneyler yapmaktadır. Genç risk sermayedarları, risk sermayesi karar verme sürecinde yapay zekanın kullanılması fikrine, bir miktar direnen ve daha az hevesli (AI/ML modelleri oluşturmak için yeterli veri bulunmadığını öne sürerek) eski ve deneyimli risk sermayedarlarına kıyasla daha açık oldukları görülmektedir.

Girişim şirketleri ve yatırımcılar için girişim şirketlerinin başarısı ve çıkışı büyük önem taşımaktadır. Girişim şirketlerinin başarısı hem ekonomik hem de teknolojik açıdan çok önemli bir rol oynamaktadır. Bir girişim şirketinin başarılı olma olasılığı oldukça düşüktür. Son on yılda, kurulan girişim şirketlerinin sayısında önemli bir artış olmuştur. Bir girişim şirketinin girişimcilik ekosisteminde başarılı olması ve varlığını sürdürmesi için kilit faktör, yatırımları güvence altına almaktır. Birçok girişim şirketi yatırım aşamalarından eli boş dönerken, bazıları da yatırımcılardan istediği yatırım tutarlarını alabiliyor. Ancak yatırım alan girişim şirketlerinin önemli bir kısmı bu yatırımları etkin bir şekilde kullanamamakta ve başarılı olamamaktadır.

Bu çalışmada; Türkiye’de faaliyet gösteren ve Yapay Zeka, Oyun ve Fintech sektörlerinden birinde bulunan girişim şirketlerinin önceden başarılı olup olamayacağı lojistik regresyon, destek vektör makineleri ve rastgele ormanlar algoritmaları ile tespit edilmeye çalışılmıştır. Girişim şirketlerinin başarılı olması ile aldıkları yatırım miktarları arasındaki ilişki ve girişim şirketinin başarılı olması ile bulunduğu sektör arasındaki ilişki incelenmiştir.

Bu çalışmanın temel amacı, başarılı şirketleri veya start-up'ları sınıflandırmak için bir model oluşturmaktır. %99,6 Gerçek Pozitif Oranı (TPR) ve %99,6 Kesinlik ve %99 AUC ile %98,9 Yanlış Pozitif Oranı ile bir şirketi başarılı veya başarısız olarak sınıflandırmak için bir ikili sınıflandırıcı oluşturularak, amaca ulaşılmaya çalışılmıştır. Kullanılan makine öğrenimi algoritması, hızlı ve kolay yorumlanabilen ve olumlu

sonular veren bir model saėlayan Rastgele Ormanlardır ama diėer makine ğrenmesi algoritmaları da iyi sonular verdiėi grlmektedir. Giriřimlerin bařarısını etkileyen unsurlar arasında ynetim becerileri, rekabet ortamı, pazar deėiřiklikleri, ekonomik kořullar ve řans faktrleri gibi birok etken bulunmaktadır.

Sonu olarak bu tez ile bir giriřimin nasıl bařarılı olabileceėine iliřkin bir tahmin modeli oluřturulmuřtur. Bu sayede yapay zeka tahminleri, diėer analiz yntemleri ve insan deėerlendirmeleri ile birlikte kullanıldıėında daha gvenilir sonular retebilir.

Ayrıca yatırımcılara sınırlı bilgi ve yksek riskli durumlarda “somut” sinyallere destek saėlayacak bir sistem tasarımı yapılmaya alıřılmıřtır.

Yapılan literatr arařtırması sonucunda sınırlı sayıda Trke kaynaėa ulařılmıř alıřmaların oėunluėunu yabancı kaynaklarda olduėu grlmřtr. Bu konuda Trke kaynaklara ihtiya olduėu dřnlmektedir. Bu alıřma ile Trkiye’de geliřmekte olan giriřim ve yatırım ekosistemini destekleyecek ilgili akademik alıřmalara katkı saėlaması beklenmektedir.

Sonu olarak bu tez alıřması hem akademiye hem de uygulamaya katkı saėlayacaėı dřnlmektedir.

Gelecekte yapılacak alıřmalarda tez kapsamında yrtlen yaklařıma ek olarak, her sektrn kendine ait dinamiklerine dair verilerin toplanması ve bunların da hesaba katılması gerektiėidir. nk giriřim ekosistemi sektrler bazında farklı deėerlendirme dinamiklerine sahiptir. Tm bunlara ek olarak doėal dil iřleme ile řirketlerin ve kurucularının hakkında ıkan haberler deėerlendirmeye dahil edilip bařarı tahmini artırılabilir.

KAYNAKLAR

- A. Vital, M. V.** (2013, Eylül). *How Startup Funding Works*. Adioma: <https://blog.adioma.com/how-funding-works-splitting-equity-infographic/> adresinden alındı
- Alkın, E., Yıldırım, K., Özer, & Mustafa.** (2005). Fırsat Maliyeti Nedir? *İktisada Giriş* (s. 36). içinde Eskişehir.
- Altundal, V.** (2020, Ekim). ERKEN AŞAMA STARTUPLARIN FİNANSMANI VE DEĞERLEMESİ.
- Böhm, M. W.** (tarih yok). The Business Model DNA: Towards an Approach for Predicting Business Model Success. *Wirtschaftsinformatik und Angewandte Informatik*, s. 1006-1020.
- Breiman, L.** (2001). Random Forests. *Machine Learning*, s. 5-32.
- Christopher K. I. Williams, C. E.** (2005). *Gaussian Processes for Machine Learning*. MIT Press.
- Cortes, C. V.** (1995, Eylül). Support-vector networks. *Mach Learn*, s. 273–297.
- D. S. Hunter and T. Zaman.** (2017, Haziran). Picking Winners: A Framework For Venture Capital. *ArXiv/70604229*.
- David McKenzie, D. S.** (2019). Predicting entrepreneurial success is hard: Evidence from a business plan competition in Nigeria. *Journal of Development Economics*.
- Dellermann, D., Lipusch, N., Ebel, P., & Popp, K. M.** (2017). Finding the Unicorn: Predicting Early Stage Startup Success through a Hybrid Intelligence Method. *International Conference on Information Systems*. Seul.
- Ester, M. K.** (1996). A density-based algorithm for discovering clusters in large spatial databases with noise.
- Fernandes, J. S.** (2014). *Venture Capital Process*. Slideshare: <https://www.slideshare.net/Sky7777/the-venture-capital-process> adresinden alındı
- Freund, Y.** (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, s. 119-139.
- Friedman, J. H.** (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, s. 1189-1232.
- G. Kalyanasundaram.** (2018). Why Do Startups Fail? A Case Study Based Empirical Analysis in Bangalore. *Asian Journal of Innovation and Policy*, s. 79-102.

- Global Funding Slide In 2022 Sets Stage For Another Tough Year.** (2023, Mayıs). Crunchbase: <https://news.crunchbase.com/venture/global-vc-funding-slide-q4-2022/> adresinden alındı
- Global Startup Ecosystem Report 2022.** (2023, Mayıs). Startup Genome: <https://startupgenome.com/report/gser2022> adresinden alındı
- Halabí, C.** (2014). A model for predicting small firm performance: Increasing the probability of entrepreneurial success in Chile. *Journal of Small Business and Enterprise Development*.
- Halabi, R. N.** (2010). A Three-Country Comparison of the Business Success versus Failure Prediction Model. *Journal of Small Business Management*, s. 360-377.
- Haralampiev, K. Y.** (2014). Models and Tools for Technology Start-Up Companies Success Analysis. *Economic Alternatives*, s. 15-24.
- Hinton, G. E.** (2012). Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Processing Magazine*, s. 82-97.
- Hoenen, S.** (2012). Do Patents Increase Venture Capital Investments between Rounds of Financing?
- Howard, J.** (2018). Universal language model fine-tuning for text classification. *In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, s. 328-339.
- Jacob Devlin, M.-W. C.** (2018, Ocak). BERT: Pre-training of deep bidirectional transformers for language understanding.
- Joachims, T.** (2005, Haziran). Text categorization with support vector machines: Learning with many relevant features. *Machine Learning: European Conference on Machine Learning (ECML-98)*, s. 137–142.
- Kober, J. B.** (2013). Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research*, s. 1238-1274.
- Krizhevsky, A.** (2012). Learning multiple layers of features from tiny images. *Technical report, University of Toronto*.
- Krizhevsky, A.** (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*.
- Li, Y. L.** (2008). Fault detection and diagnosis for semiconductor manufacturing using support vector machines. *IEEE Transactions on Semiconductor Manufacturing*, s. 178-187.
- Littunen, H.** (2010). The rapid growth of young firms during various stages of entrepreneurship. *Journal of Small Business and Enterprise Development*, s. 8-31.
- Lussier, R.** (2001). A Cross-National Prediction Model for Business Success. *Journal of Small Business Management*, s. 228 - 239.
- MacQueen, J.** (1967). Some methods for classification and analysis of multivariate observations. *In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, s. 281-297.

- Mnih, V. K.** (2015). Human-level control through deep reinforcement learning. *Nature*, s. 529-533.
- Nahata, R.** (2008). Venture capital reputation and investment performance. *Journal of Financial Economics*, s. 127-151.
- P. Holmes, A. H.** (2010). An analysis of new firm survival using a hazard function. *Applied Economics*, s. 185-195.
- Quinlan, J.** (1986, Mart). Induction of decision trees. Machine learning. *Mach Learn*, s. 81–106.
- Ragothaman, S. N.** (2003). Predicting Corporate Acquisitions: An Application of Uncertain Reasoning Using Rule Induction. *Information Systems Frontiers*, s. 401–412.
- Risk Sermayesi.** (2023, Mayıs 17). Wikipedia: https://en.wikipedia.org/wiki/Venture_capital adresinden alındı
- Risk Sermayesi Diyagramı.** (2023, Mayıs 17). Wikipedia: https://upload.wikimedia.org/wikipedia/commons/7/79/Venture_Capital_Fund_Diagram.png adresinden alındı
- The Complete List Of Unicorn Companies.** (2023, Şubat). CB Insights: <https://www.cbinsights.com/research-unicorn-companies> adresinden alındı
- The Top 12 Reasons Startups Fail.** (2023, Mayıs). CB Insight: <https://www.cbinsights.com/research/report/startup-failure-reasons-top/> adresinden alındı
- Trevor Hastie, R. T.** (2009). *The Elements of Statistical Learning : Data Mining, Inference, and Prediction*. Springer.
- Turcorn 100.** (2023, şubat). Turcorn: <https://turcorn.gov.tr/> adresinden alındı
- Turcornlarımız.**(2023,Şubat).Turcorn:<https://turcorn.gov.tr/ekosistemimiz/turcornlarımız/> adresinden alındı
- Vaswani, A. S.** (2017). Attention is all you need.
- Viola, P.** (2001). Rapid object detection using a boosted cascade of simple features. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Ward Jr, J. H.** (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American statistical association*, s. 236-244.
- Wei, C.-P. &.S.-S.** (2009). Patent Analysis for Supporting Merger and Acquisition (M&A) Prediction: A Data Mining Approach. *Lecture Notes in Business Information Processing*, s. 187-200.
- Wilson, A. G.** (2013). Gaussian process kernels for pattern discovery and extrapolation. *The Journal of Machine Learning Research*, s. 669-693.
- Zhao, F.** (2005). Exploring the Synergy Between Entrepreneurship and. *International Journal of Entrepreneurial Behavior &*, 25-41.

