



## Kitle Fonlaması Projelerinin Karar Ağacı ve Rastgele Orman Algoritmalarıyla Sınıflandırılması

Murat KILINÇ <sup>\*,a</sup>, Çiğdem TARHAN <sup>b</sup>, Can AYDIN <sup>b</sup>

<sup>a,\*</sup> Manisa Celal Bayar Üniversitesi, Bilgisayar Araştırma ve Uygulama Merkezi, MANİSA, 45140, TÜRKİYE

<sup>b</sup> Dokuz Eylül Üniversitesi, Yönetim Bilişim Sistemleri Bölümü, İZMİR, 35160, TÜRKİYE

### MAKALE BİLGİSİ

Alınma: 15.12.2020  
Kabul: 30.12.2020

**Anahtar Kelimeler:**  
Makine Öğrenmesi,  
Web Uygulamaları,  
Yönetim Bilişim  
Sistemleri

**\*Sorumlu Yazar**  
e-posta:  
[kilinc.murat@cbu.edu.tr](mailto:kilinc.murat@cbu.edu.tr)

### ÖZET

Kitle fonlaması platformları, internet üzerinden iş fikirlerini hayata geçirme ya da destek alabilme noktasında büyük olanaklar sağlayabilmektedir. Bu platformlarda destek beklenen projelerin başarısı, alınan finansal destek ile doğru orantılı bir şekilde artmaktadır. Fakat finansal destek alabilmek için projenin destekçilere iyi bir şekilde sunulması gerekir. Günümüzde bu platformlar iyi tasarlanmamış projelerle dolu olduğu için başarı oranı oldukça düşük göstermiştir. Bu sebeple, finansal destek alınabilmesi için projelerin başarı anlamında test edilmesi ve başarısız olarak sınıflandırılan projelerin eksiklerini gidererek destekçilere yeniden sunulması gerekmektedir. Bu kapsamda, ortaya koyduğumuz çalışmada birçok kategorideki Kickstarter projesi makine öğrenmesi yöntemleriyle sınıflandırılarak web arayüzünde son kullanıcıya sunulmuştur. Projelerin sınıflandırılması için, dağıtık veri setlerinde iyi sınıflandırma yapabilen Decision Tree ve Random Forest algoritmaları kullanılmıştır. Algoritmalar, sırasıyla %72.75 ve %81.55 oranında sınıflandırma yapabilmektedir. Ayrıca, yapılan sınıflandırmalar değerlendirme metrikleriyle de test edilerek ne kadar doğru sınıflandırma yapılabildiği ölçülmüştür. Bu sayede, kitle fonlaması platformlarına projelerini ekleyen veya ekleyecek olan girişimciler, finansal bir destek aramadan önce projelerini başarı anlamında test ederek eksiklerini görebileceklerdir.

## Classification of Crowdfunding Projects by Decision Tree and Random Forest Algorithms

### ARTICLE INFO

Received: 15.12.2020  
Accepted: 30.12.2020

**Keywords:**  
Machine Learning, Web  
Applications,  
Management  
Information Systems

**\*Corresponding Authors**  
e-mail:  
[kilinc.murat@cbu.edu.tr](mailto:kilinc.murat@cbu.edu.tr)

### ABSTRACT

Crowdfunding platforms can provide great opportunities to implement business ideas or get support over the internet. The success of projects that are expected to support these platforms increases in direct proportion to the financial support received. But in order to receive financial support, the project must be well presented to backers. Today, the success rate has declined considerably because these platforms are full of poorly designed projects. For this reason, in order to receive financial support, projects must be tested in terms of success and re-presented to supporters by eliminating the deficiencies of projects classified as unsuccessful. In this context, in our study, many categories of Kickstarter projects are classified by machine learning methods and presented to the end user in the web interface. For the classification of projects, decision Tree and Random Forest algorithms that can classify well in scattered data sets were used. Algorithms can classify by 73% and 81%, respectively. In addition, the classifications made were also tested with evaluation metrics and measured how accurate the classification can be made. In this way, entrepreneurs who add or will add their projects to crowdfunding platforms will be able to see their shortcomings by testing their projects for success before receiving financial support.

## 1. GİRİŞ (INTRODUCTION)

Veri kavramının öneminin giderek artmasıyla birlikte geçmişten günümüze süregelen birçok süreç, sektör ve kavram büyük bir değişim geçirmiştir. Özellikle verimliliğin esas olduğu zamanlarda bu durum kendini iyiden iyiye göstermektedir. Bu sebeple, değişimin temelini oluşturan verinin günümüzde çok iyi yorumlanması ve anlaşılır hale getirilmesi gerekir. Çünkü bir veri kümesinin nasıl depolandığının, düzenlendiğinin ve yönetildiğinin, söz konusu örgüt yetkinliği üzerinde büyük bir etkisi vardır [1]. Bu doğrultuda büyük veriyi işleyiş süreçleri bakımından incelendiğimizde, bu veri kümelerini insan eliyle düzenleyerek yorumlamak, görselleştirmek ya da anlaşılır hale getirerek raporlamak oldukça fazla zaman maliyeti anlamına gelmektedir. Çünkü veri sayısı sürekli artmış ve insan eliyle analizi zor bir hale gelmiştir. Bilgi teknolojileri tam da bu noktada büyük bir kolaylık sağlamaktadır. Bilgi teknolojilerinin gelişmesiyle birlikte 2000'li yılların başından itibaren birçok iş akışı daha hızlı ve verimli bir şekilde yapılabilmektedir. Bu gelişim, aynı zamanda bilgi teknolojilerinin değişerek, geçmişte ortaya koyulan kuralların değişmesine zemin hazırlamış bir şekilde, günümüze uyumlu hale gelmiştir [2]. Bu uyum içerisinde ortaya koyulan her teknolojinin fayda sağladığı farklı alanlar bulunmaktadır. Fayda sağlanan alanların temeli veri yapılarının işlenmesiyle ve yorumlanmasıyla oluşur. İş analitiği ve makine öğrenmesi yöntemleri de ortaya koyulan verinin işlenerek anlamlı hale getirilmesi sürecinde bilgi teknolojilerinin içerisinde kendisine yer bulmaktadır.

Diğer taraftan, startup metodolojisine sahip girişimlerin popülaritesi geçtiğimiz yıllar içerisinde giderek artmaktadır [3]. Web 2.0 ile birlikte gelen teknolojilerin olgunluğu ve kaynak kullanımının başarısıyla birlikte girişimleri internet üzerinden finanse etme süreci olan kitle fonlaması büyük bir önem kazanmıştır [4]. Çoğunlukla mentorluk desteği verilen girişimler için günümüzde birçok platform geliştirilmiştir. Kickstarter gibi platformların popüler hale gelmesiyle birlikte proje geliştiricileri, her yıl yüz milyonlarca dolar destek toplayabilmektedir [5]. Yani bu mantıkla tasarlanmış platformlarda girişimciler fikirleri ve projeleri için destek arayışına girmektedirler. Platform üzerinden sağlanan finansal ve taktiksel desteklerden sonra ise, girişimcilerin fikirleri destekçiler, melek yatırımcılar ve sermaye fonları sayesinde hayata geçmektedir [6]. Beğenilmeyen ya da yeterli desteği bulamayan fikirler ise yeterli desteği bulamadığından dolayı başarısız olabilmektedir. Bu sebeple girişim fikri ekosisteminde yeterli desteği alabilmek, o fikri hayata geçirmek açısından oldukça önemlidir. Dolayısıyla, girişim fikrinin ya da projenin başarılı olup olmayacağı konusunda, projeyi destekçilere sunmadan önce bir öngörü sağlamak pozitif anlamda bir etki

sağlamaktadır. Bu öngörü, kitle fonlaması platformlarında destekçilerin ilgisini çekebilmek için kullanılabilir. Proje profilindeki bilgiler, içerik, metinler, kullanılan resimler gibi projenin dinamiklerini oluşturan etmenler, destekçilerin ilgisiyle doğrudan alakalıdır [7].

### 1.1 Literatür Taraması (Literature Review)

Kitle fonlama platformları, araştırmanın yöntem kısmını oluşturan iş analitiği ve makine öğrenmesi ile başarı tahminlemesi kapsamında incelendiğinde, birçok farklı çalışmaya rastlanılmıştır. 2013 yılında Chen tarafından bir Kickstarter projesinin önceden başarılı olup olmayacağını tahmin etmek amacıyla sistem geliştirme çalışmasında, destek vektörü makinesi yöntemi kullanılmıştır. Çalışma ile, projelerin başlangıçtaki özellikleri kullanılarak %67 doğruluk oranıyla başarı tahminlemesi yapılması sağlanmıştır [8]. 2019'da Kindler tarafından yürütülen araştırmada da kitlesel fonlama platformları olan Kickstarter, Indiegogo ve Sellaband gibi oluşumlar üzerindeki yayılım mekanizmaları araştırılmıştır. Çünkü projenin yayılımı, viralite ve başarıyla doğrudan alakalıdır [9]. Chung tarafından ortaya koyulan araştırmada ise Kickstarter veri setleri, destekçi-kampanya grafikleri ile birlikte Naive Bayes, Random Forest ve Adaboost sınıflandırma yöntemleri kullanılmıştır. En yüksek değeri veren Adaboost sınıflandırma yöntemi incelenen veri setine göre %76 oranında bir doğruluk oranıyla başarı tahminlemesi yapmıştır [10]. Rao ve ekibi tarafından 2014 yılında ortaya koyulan çalışmada ise kitle fonlamasına sahip projelerin başarı oranının %50'den az olduğu vurgulanmıştır. Ayrıca çalışma kapsamında karar ağaçları yöntemi kullanılarak verilen para taahhüdü ile kampanya başarısı arasındaki ilişki incelenmiştir. İncelemeye göre kampanya sürelerinin proje başarısında önemli bir etkiye sahip olduğu tespit edilmiştir. Ek olarak, kampanyalarda tohum para olarak kabul edilebilecek ilk %15'lik para girişini kullanılmasının, projelerin başarısının %84 oranında tahmin edilmesini sağladığı ortaya konulmuştur [11]. Jensen ve Özkil'in 2018 yılında ortaya koyduğu çalışmada ise kitle fonlaması platformlarında başarısızlığa sebep olabilecek etmenler incelenmiştir. İncelemeye göre kampanya başlatıcıların ürün özellikleriyle ilgili vaatlerde bulunma yeteneği ve bu kapsamda oluşturulan proje özellikleri, projenin başarısı için önemli rol oynamaktadır. Ayrıca çalışma, kitle fonlaması platformlarının bir araştırma içerisinde nasıl kullanılabileceğini de hem veri kütüphaneleriyle hem de ürün geliştirme vakaları ile göstermektedir [12]. Qianzhou ve ekibinin yaptığı araştırma ise temel olarak projelerin kategori ve hedef gibi temel noktalarına odaklanmaktadır. Bu kapsamda da Kickstarter üzerinden elde edilen büyük bir veri kümesi kullanılmıştır. Araştırma sonucuna göre proje açıklamalarında sağlanan bilgiler ile elde edilen

finansal destek arasında bir doğru orantı vardır. Ortaya koyulan model projenin finansal başarısını %73 doğruluk oranıyla tahmin edebilmektedir. Ayrıca araştırma, tahminleme doğruluk oranının daha da artması için Destek Vektör Makinesi (SVM) sınıflandırma yöntemini önermektedir [13]. Sheng Bi ve ekibinin 2016'da ortaya koyduğu bir başka çalışmada da ayrıntılı bir olasılık modeline dayanarak, girişimlerdeki çevrimiçi bilgilerin yatırımcıların kararlarını nasıl etkilediği araştırılmıştır. Çin'de faaliyet gösteren kitle fonlaması web sitelerinin verileriyle gerçekleştirilen çalışmada, daha yüksek beğeni sayısının, online geri bildirim, daha ayrıntılı proje açıklamasının ve projenin video ile tanıtılmasının fon yatırım kararlarında olumlu bir etki yaptığı ortaya konulmuştur. Çalışma içerisindeki veri analizi ile de, farklı proje kategorilerinin farklı perspektifler ile değerlendirilmesi gerektiği vurgulanmıştır [14]. Son olarak, Mortensen ve ekibinin yaptığı çalışmada da makine öğrenmesi yöntemleriyle başarı faktörünü tanımlama ve tahminleme çalışması yapılmıştır. Araştırmada, Fortune 500 listesindeki bir ambalaj şirketinin başarısını neyin yönlendirdiği incelenmiş ve bu doğrultuda bir model ortaya konulmuştur. İstatistiksel modelleme teknikleri kullanılmasıyla satış döngülerinin kısalmasına, satış maliyetlerinin azalmasına etki ederek hem gelir hem de kar hasılatını iyileştirdiği çalışma içerisinde görülmektedir. Makine öğrenmesi yöntemlerinden decision tree, gradient boost ve random forest algoritmalarının kullanılmasıyla en iyi model; accuracy %80, precision %86, recall %77 sonuçlarını vermiştir [15].

İncelenen çalışmalar özetlenecek olursa kullanılan yöntemler; sistem geliştirme yaşam döngüsü (SDLC), veri madenciliği için endüstriler arası standart süreç (CRISP-DM), ve Örneklem-Keşfetme-Düzenleme-Modelleme-Değerlendirme (SEMMA) ekseninden oluşmaktadır. SEMMA, CRISP-DM yönteminin aksine problem tanımını içermediği için kitle fonlaması projelerini sınıflandırma noktasında yetersiz kalabilmektedir. Yine çalışmalar kapsamında, kitle fonlamasındaki başarı oranının bir düşüş eğiliminde olduğu göze çarpmaktadır. Bunun temel sebeplerinden birisi, kitle fonlamasına olan ilginin artması ve kitle fonlaması platformlarına yüklenen projelerin özensiz hazırlanması olarak nitelendirilebilir. Diğer taraftan, kitle fonlaması projeleriyle ilgili yapılan başarı sınıflandırma sonuçları yıllar geçtikçe bir artış göstermektedir. Çünkü özellikle 2018 ve sonrasında, birlikte kullanıldığında daha iyi sonuçlar verebilen (ensemble) makine öğrenmesi algoritmaları daha çok kullanılmaya başlanmıştır. Örneğin, Mortensen ve ekibinin yaptığı çalışmada; decision tree, gradient boost ve random forest algoritmaları birlikte kullanılarak %80'lik bir doğruluk oranı elde edilip, değerlendirme metrikleriyle değerlendirilmiştir [15].

İyi bir sınıflandırma yapılabilmesi için algoritmaları birlikte kullanmanın yanı sıra veri setine uygun algoritmayı seçip kullanmak da uygun bir çözümdür. Çünkü sınıflandırma kapsamında daha iyi sonuçlar alınabilmesi, veri setinin özellikleriyle de ilgili bir durumdur. Bu sebeple, veri setine uygun algoritma seçimleriyle daha iyi sınıflandırma yapılabileceği durumu göze çarpmaktadır.

Kickstarter projelerine başarı anlamında bir öngörü sağlayan çalışmamızda da, problemin tanımının yapılabildiği için CRISP-DM yöntemi kullanılarak bir ortam geliştirilmiş ve akabinde kitle fonlaması kullanıcılarının projelerinin başarısı hakkında bir öngörüye sahip olabilmesi amaçlanmıştır. Bu doğrultuda, kitle fonlaması kullanıcılarından projelerinin özellikleri alınarak makine öğrenmesi yöntemleriyle başarı tahminlemesi yapılmış ve bu sürecin web ortamında gösterilmesi sağlanmıştır. Böylece kitle fonlaması ekosistemine girmeden önce veya sonra projeler test edilebilir hale gelerek bir karar destek yapısı oluşturulmuştur. Çalışmamız ayrıca, dağınık veri seti için sınıflandırmaya uygun algoritmaların seçilerek uygulanması kapsamında literatüre bir katkı sağlamaktadır.

## 2. MATERYAL VE YÖNTEM (MATERIALS and METHODS)

### 2.1. Veri Seti (Data Set)

Kaggle tarafından sağlanan ve 300.000'den fazla Kickstarter girişimini çeşitli özellikleriyle beraber konu edinen büyük veri dosyaları, 2017 ve sonrasında oluşan verilerle filtrelenerek, yaklaşık 50 bin veri civarına indirgenmiştir [16]. Araştırmanın en güncel girişim proje verileriyle yapılabilmesi amacıyla bu filtreleme yapılmıştır. Belirtilen özelliklerde araştırmanın yapılabilmesi için, veri madenciliği proje döngüsü (CRISP-DM) kullanılmış ve döngü içerisindeki 6 aşama sırasıyla gerçekleştirilmiştir. Bu aşamalar; işin anlaşılması, verinin anlaşılması, verinin hazırlanması, modelleme, değerlendirme, konuşlandırma ve kontrol adımlarından oluşmaktadır [17][18]. Ayrıca, Veri Madenciliği proje döngüsü (CRISP-DM) oldukça esnek ve döngüsel bir model olarak kabul edilir. Veri bilimi projelerinde sıklıkla kullanılan model üzerindeki bu özellik sayesinde bir önceki adıma gitmek ve değişiklik yapmak mümkündür. Çünkü ihtiyaçlarda veya veri yapısında bir değişiklik olabilir. Bu sebeple, benzer durumlarla karşılaşıldığında ana çerçeveyi bozmadan modelin değerlendirilmesi gerekmektedir.

CRISP-DM uygulanırken, ilk aşamada muhtemel problemler değerlendirildikten sonra, literatür taraması yapılmış ve hangi yazılım kütüphanelerinin kullanılacağı belirlenmiştir (Tablo 1). İkinci aşamada ise verinin kalitesi, ulaşılabilirliği ve sürdürülebilirliği

tartışılmıştır. Üçüncü aşamada ise ön işlemeden geçen veri seti, modelleme aşamasında verilerin izleyeceği yol ve analiz yapıları bakımından bir model üzerinde uygulanmıştır.

Tablo 1. Kullanılan Yazılım Kütüphaneleri  
(Software Libraries Used)

	Web	Makine Öğrenmesi	Veri Ön İşleme
<b>Python Kütüphanesi ve Yapıları</b>	Flask Framework	Sklearn Kütüphanesi	Pandas, Numpy
<b>Nesne Tabanlı Programlama Dilleri</b>	PHP, Javascript, Python (VS Code ile derlenmiştir.)	Python (Spyder VS Code ile derlenmiştir.)	Python (Jupyter ile derlenmiştir.)
<b>Diğer Araç ve Kütüphaneler</b>	HTML, CSS, Bootstrap, Chart.js	Google Colab	Weka

Değerlendirme aşamasında makine öğrenmesi yöntemlerinden alınan sınıflandırma algoritmalarının sonuçları, değerlendirme metrikleriyle test edilerek çalışma içerisine eklenilmiştir. Böylelikle konuşlandırma süreci için, geliştirilen web ortamında gösterilecek olan alanların hazırlanması sağlanmıştır. Bu noktada amaç, hangi sınıflandırma algoritmasının hangi sonucu verdiğini son kullanıcıya analiz ederek göstermektir. Dolayısıyla, konuşlandırma sürecinde de modelin web ortamında hayata geçmesi sağlanmıştır.

## 2.2. Veri Ön İşleme (Data Preprocessing)

11 sütundan oluşan veri setinde, name, category, main category, currency, deadline, goal, launched, pledged, state, backers, country gibi sütunlar bulunmaktadır (Tablo 2). Sağlıklı bir analiz için içeriği eksik olan veri kümeleri, veri setinin içerisinden temizlenerek, bir veri ön işleme yapılmıştır. Ayrıca string değerde olan sütunlar yine veri ön işlemede nümerik bir hale çevirilerek, veri ön işleme adımları tamamlanmıştır. Ön işleme sonucunda makine öğrenmesi yöntemleri için normalizasyonu yapılmış, nümerik veriler oluşturulmuştur. Veri ön işleme aşaması tamamlandıktan sonra Kickstarter veri seti analize ve makine öğrenmesi yöntemlerinin uygulanmasına hazır bir hale getirilmiştir. Ayrıca veri seti içerisindeki özelliklerin birbirlerini minimum düzeyde etkilemeleri için ölçeklendirme işlemleri gerçekleştirilmiştir. Böylelikle sınıflandırma aşamasında ortaya çıkabilecek aşırı öğrenme ya da eksik öğrenme gibi problemlerin önüne geçilebilmesi amaçlanmıştır ve web uygulamasındaki son kullanıcının karşısına daha anlamlı sınıflandırma sonuçları sunulmuştur.

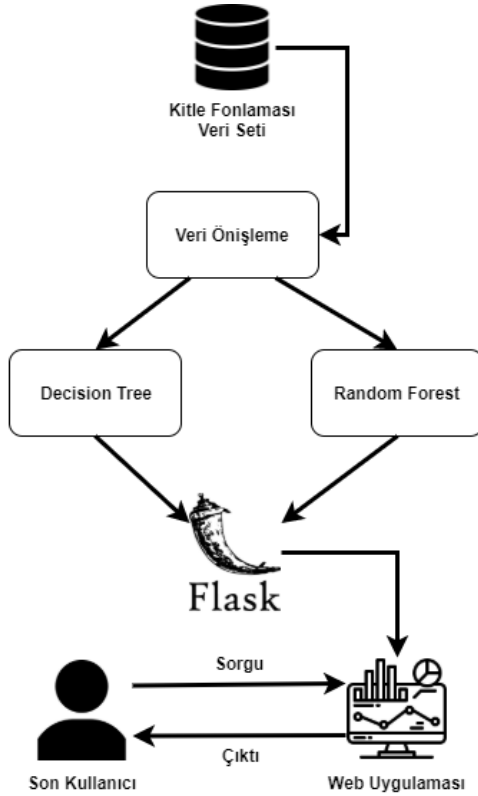
Tablo 2. Projelerin Özellikleri ve Açıklaması  
(Features and Description of the Projects)

Proje Özellik Adı	Açıklama
<i>name</i>	Proje isimleri
<i>category</i>	Projelerin alt kategorileri
<i>main category</i>	Alt kategorilerin daha genel olarak sunulduğu ana kategoriler
<i>currency</i>	Projeyi desteklemek için kullanılan para birimi
<i>deadline</i>	Projenin kitle fonlaması için son tarihi
<i>goal</i>	Bir proje oluşturucunun projesinin tamamlamak için ihtiyaç duyduğu para miktarı
<i>launched</i>	Projenin başlatıldığı tarih
<i>pledged</i>	Proje için taahhüt edilen miktar
<i>state</i>	Projenin içinde bulunduğu mevcut durum
<i>backers</i>	Projeye destek sağlayan kişi sayısı
<i>country</i>	Projenin ortaya koyulduğu ülke

## 2.3. Modelleme (Modelling)

Modelleme süreci, geliştirilmek istenen iş yapısına en uygun öğelerin dahil edilmesiyle oluşturulan bir karar destek sürecini ifade etmektedir. Bu kapsamda geliştirilen web uygulaması için makine öğrenmesi yöntemlerinden hangi algoritmaların kullanılacağı ve nasıl bir yol izleneceği bu aşamada belirlenmektedir (Şekil 1).

Modellemeye göre, girişimini sistem üzerinde test etmek isteyen kullanıcı, girdiği proje özellikleriyle birlikte sisteme dahil olmaktadır. Yaklaşık 50 bin veriden oluşan Kickstarter veri seti Python'ın Pandas ve Numpy kütüphaneleriyle bir ön işlemeden geçirildikten sonra web uygulamasının veritabanı oluşturulmuştur. Veritabanında nümerik olarak tutulan veriler, makine öğrenmesi algoritmalarıyla test edilerek yöntemlere göre doğruluk oranlarıyla beraber web arayüzünde gösterilmektedir. Bu sayede, makine öğrenmesi yöntemlerinin sınıflandırma yeteneğiyle oluşan web platformu kullanıcıya çıkarımlar sağlamaktadır.



Şekil 1. Uygulamanın Modellenmesi  
(Modeling the Application)

## 2.4. Makine Öğrenmesi (Machine Learning)

Çalışma kapsamında geliştirilen web uygulamasında gözetimli öğrenme yöntemlerinden karar ağacı ve rastgele orman sınıflandırıcısı kullanılmıştır.

Karar ağacı algoritması içerisinde öncelikle hiyerarşik bir ağaç yapısı oluşturulmaktadır [19]. Ağacın yaprak seviyesinden itibaren oluşturulan yapı ile sınıfların entropi değerlerine göre ağaç üzerine dizilimi gerçekleştirilir. Daha sonrasında ise kontrol edilmek istenen değerler kök yapısından başlayarak dallara doğru ilerlenerek sınıflandırma sonucuna ulaşılır. Bu sayede test için ayrılan veriler ile modelin doğruluğu kontrol edilir. Ayrıca yeni gelecek olan gerçek değerler de bu sayede sınıflandırılır. Ayrıca karar ağaçları, temel sınıflandırma algoritmalarından biri olup, birden fazla karar ağacı biraraya gelerek Random Forest algoritmasını oluşturur [20]. Algoritmanın en büyük problemlerinden birisi olan overfitting (aşırı öğrenme) olduğu durumlarda ise, Random Forest yöntemi içerisinde karar ağacı sayısını belirleyerek ortaya çıkan sonuçlara da bakılması gerekir.

Random Forest Classification (Rastgele Orman Sınıflandırıcısı), en çok kullanılan başarılı sınıflandırma yöntemlerinden birisi olarak göze

çarpmaktadır. Algoritma, her sınıflandırıcının girdi vektöründen bağımsız bir şekilde karar ağaçlarından oluşur. Akış içerisindeki her bir ağaç giriş vektörünü sınıflandırmak için birim oyu verir [21]. Ayrıca geleneksel yöntemlerden biri olarak kabul edilen Decision Tree'nin en büyük problemlerinden birisi overfitting (aşırı öğrenme) olarak bilinmektedir. Rastgele Orman Algoritması'nda bu problem, öznelik setinden rassal olarak birçok alt setin seçilmesiyle aşılmaktadır. Seçilen alt setlerin ürettiği sonuçlar arasında en fazla birim oyunu alan değer, Rastgele Orman Algoritmasının sonucunu oluşturur. Bu sınıflandırma yöntemi, dengesiz bir dağılım gösteren veri setlerinde ve kategorik değişkenlerde daha iyi sonuç vermektedir. Uygulama geliştirilirken kullanılan veri seti kategoriksel olarak birçok alana sahip olup dengesiz bir dağılım gösterdiğinden dolayı bu yöntem tercih edilmiştir.

## 2.5. Değerlendirme (Evaluation)

Veri madenciliği için endüstriler arası standart sürecin 5. kısmını oluşturan değerlendirme aşamasında, araştırmada kullanılan yöntemler, çıktıları ve performansları ele alınmıştır. Girişimcilerin projelerini başarı anlamında test edecekleri web platformu, makine öğrenmesi yöntemleriyle geliştirilmiştir. Veri setinin uyumluluğu sebebiyle makine öğrenmesi sınıflandırma yöntemlerinden olan Decision Tree ve Random Forest Algoritmaları tercih edilmiştir. Uygulama içerisindeki verilerin %70'i eğitim için, %30 ise test için ayrılmıştır. Gerçek değerlerle sonuç elde etme aşamasında ise girişimcilerden alınan proje özellikleriyle projenin başarı durumu tahminlenmiş ve sonuçları Flask Framework ile kullanıcı arayüzüne tekrar geri döndürülmüştür.

Tablo 3. Karmaşıklık Matrisi  
(Confusion Matrix)

		Gerçek Değerler	
		Pozitif (1)	Negatif (0)
Tahmin Edilen Değerler	Pozitif (1)	<b>TP</b>	<b>FP</b>
	Negatif (0)	<b>FN</b>	<b>TN</b>

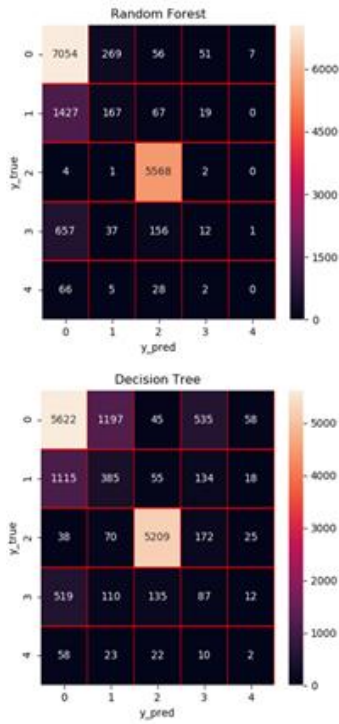
**TP** : True Positive (Doğru Pozitif) → Algoritma evet sonucunu verdi, asıl sonuç da evet

**FP** : False Positive (Yanlış Pozitif) → Algoritma evet sonucunu verdi, ama asıl sonuç hayır

**FN** : False Negative (Yanlış Negatif) → Algoritma hayır sonucunu verdi, asıl sonuç da evet

**TN** : True Negative (Doğru Negatif) → Algoritma hayır sonucunu verdi, ama asıl sonuç hayır

Bu noktada, kullanılan bu sınıflandırma yöntemlerinin ne kadar doğrulukla sınıflandırma yapabildiğinin ölçülmesi gerekir. Sınıflandırma yöntemlerinin test edilebilmesi amacıyla karmaşıklık matrisi kullanılmıştır (Tablo 3). Karmaşıklık matrisinin özellikleri, makine öğrenmesi sınıflandırma yöntemlerine entegrasyon açısından uyumlu bir durumdadır ve daha fazla anlamsal açıklama sağlar [22]. Ayrıca karmaşıklık matrisi gerçek değerlerin bulunduğu bir test verisi üzerinde sınıflandırma yönteminin performansını tanımlamak amacıyla kullanılır. Bu doğrultuda 5 sınıf içeren veri setinin Decision Tree ve Random Forest algoritmalarıyla analizi yapılmıştır. Sonrasında ise sıcaklık haritasıyla analizler görselleştirilerek karmaşıklık matrisleri ortaya çıkmıştır (Şekil 2).



Şekil 2. Sınıflandırma Problemlerinin Karmaşıklık Matrisleri

(Confusion Matrix for Classification Problem)

Karmaşıklık matrislerinin ortaya koyulmasından sonra sınıflandırma metriklerinin değerlendirilmesi gerekmektedir. Bu doğrultuda, ilk olarak ortaya koyulan modelden çıkan sonuçların ne kadarının doğru tahmin edildiğinin belirtilmesi gerekir. Bunun için de sınıflandırma yöntemleri için en çok doğruluk oran ölçümü kullanılmaktadır [23]. Doğruluk oranı, doğru bir şekilde sınıflandırılan verinin tüm verilere oranlamasıyla elde edilir (1).

$$Accuracy (Doğruluk Oranı) = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

Doğruluk oranından sonra, karmaşıklık matrisi kullanılarak yapılan hesaplamalardan diğer ikisi Recall (Geri Çağırma) ve Precision (Hassasiyet) ölçümleridir. Bu doğrultuda, doğru sınıflandırılan pozitif örneklem sayısının, doğru sınıflandırılmış toplam pozitif tahmin edilen örnek ile yanlış sınıflandırılmış negatif örnek sayısının toplamına oranına duyarlılık ölçümü denilir (2). Ölçüm, pozitif olarak tahmin edilmesi gereken değerlerin, ne kadarının pozitif olarak tahmin edildiğini gösteren bir metrik olarak değerlendirilir. Özellikle yanlış negatif tahminlemenin maliyetinin yüksek olduğu durumlarda kullanılmaktadır.

$$Recall (Duyarlılık) = \frac{TP}{TP + FN} \quad (2)$$

Karmaşıklık matrisinden elde edilen sonuçlar baz alındığında hassasiyet ölçümü, doğru bir şekilde sınıflandırılan pozitif örneklem sayısının, toplam pozitif örneklere oranı olarak hesaplanmaktadır (3). Diğer bir deyişle, pozitif olarak tahmin edilen değerlerin gerçekten kaç adedinin pozitif olduğunu gösteren bir ölçüm parametresidir.

$$Precision (Kesinlik) = \frac{TP}{TP + FP} \quad (3)$$

F1 Skoru, uç durumların göz ardı edilmediği bir sınıflandırma yöntemi değerlendirme metriğidir. Metriğin ölçümü precision ve recall değerlerinin harmonik ortalamasıyla elde edilir (4). Sınıflandırma yöntem seçimlerinde doğruluk oranı yerine F1 skor değerine daha önce bakılmasının temel nedeni eşit dağılmamış olan veri kümelerinde yanlış bir model seçimi yapmamaktır. Kickstarter veri seti de eşit dağılmamış örneklemelerden oluştuğu için kullanılan 2 sınıflandırma yönteminin de F1 skoru belirlenmiştir.

$$F1 Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (4)$$

Karmaşıklık matrisinden ayrı tutularak, Cohen'in Kappa katsayısı ( $\kappa$ ) ise iki farklı değerleyici arasındaki uyumayı ölçen bir istatistik metriktir [24]. Ortaya çıkan ölçüm değeri -1 ile +1 arasında bulunmaktadır.  $\kappa$  değeri +1 değerine ne kadar yaklaşırsa iki farklı değerleyici arasındaki uyumluluk o derecede iyidir.

$$Kappa Score (\kappa) = \frac{P_o - P_c}{(1 - P_o)} \quad (5)$$

Benzer şekilde  $\kappa$  değeri -1 değerine ne kadar yakınsa iki değerleyici arasındaki uyumsuzluk yüksek seviyededir ve güvenilirlik açısından bir anlam taşımamaktadır.  $\kappa = 0$  durumunda ise iki değerleyici arasında uyumun şansa bağlı olabileceği belirtilir.

Tablo 4. Kappa Değerinin Değerlendirilmesi  
(Evaluation of the Kappa Score)

$\kappa$ (Kappa Score)	Evaluation
$\kappa < 0$	Hiç Uyuşma Olmaması
$\kappa > 0.0$ ve $\kappa \leq 0.20$	Önemsiz Uyuşma Olması
$\kappa \geq 0.21$ ve $\kappa \leq 0.40$	Orta Derecede Uyuşma Olması
$\kappa \geq 0.41$ ve $\kappa \leq 0.60$	Çoğunlukla Uyumlu Olması
$\kappa \geq 0.61$ ve $\kappa \leq 0.80$	Önemli Derecede Uyuşma Olması
$\kappa \geq 0.81$ ve $\kappa \leq 1$	Mükemmel Uyuşma olması

$\kappa$  Kappa değeri,  $P_0 = \text{Kabul Edilen Oran}$ ,  $P_c = \text{Beklenen Oran}$  olmak üzere yukarıdaki denklem ile bulunur (5).  $\kappa \geq 0.4$  olması uygun bir değer olarak görülmektedir (Tablo 4). Model değerlendirme metriklerinin Kickstarter girişimleri için kullanılan sınıflandırma yöntemlerini test etmede kullanılmasıyla Python programlama dili aracılığıyla sınıflandırma raporları elde edilmiştir.

Tablo 5. Random Forest Algoritması Sınıflandırma Raporu

(Random Forest Algorithm Classification Report)

Proje Durumu	Precision	Recall	F1 Score	Support
Başarısız	0.76	0.95	0.85	7539
İptal Edilen	0.32	0.08	0.13	1758
Başarılı	0.95	1.00	0.97	5423
Devam Eden	0.16	0.01	0.03	839
Askıya Alınan	0.11	0.01	0.02	97
Avg / Total	0.75	0.82	0.77	15656

Rapor içerisinde 5 farklı proje durum sınıfının elde ettiği precision, recall, F1 Score değerleri listelenmektedir. Buna göre Random Forest sınıflandırma yönteminin elde ettiği değerler doğruluk oranları %81 iken precision ortalaması %75, recall %82 ve F1 Score %77 olarak sıralanmaktadır (Tablo 5).

Tablo 6. Decision Tree Algoritması Sınıflandırma Raporu

(Decision Tree Algorithm Classification Report)

Proje Durumu	Precision	Recall	F1 Score	Support
Başarısız	0.77	0.76	0.76	7465
İptal Edilen	0.23	0.22	0.22	1772
Başarılı	0.95	0.94	0.95	5483
Devam Eden	0.08	0.09	0.08	850
Askıya Alınan	0.02	0.03	0.03	86
Avg / Total	0.73	0.72	0.73	15656

Karar ağaçlarında ise %73 oranında bir doğruluk oranı elde edilirken, precision ölçüm ortalaması %73, recall ortalaması %72, F1 değer ortalaması %77 olarak bulunmuştur (Tablo 6). Değerlendirilen sınıflandırma yöntemleri, ölçümleri ile birlikte karşılaştırılmıştır (Tablo 7). Precision, Recall ve F1 skorlarındaki micro parametresi, toplam gerçek pozitifleri, yanlış negatifleri ve yanlış pozitifleri sayarak metrikleri global olarak hesaplamaktadır. Macro parametresi, her bir etiket için metrikleri hesaplayarak ağırlıksız ortalamayı bulan bir parametredir. Weighted parametresi ise, her bir etiket için metrikleri hesaplar ve ortalamalarını her bir etiket için gerçek örnek sayısını alarak destekler. Bu parametre etiket dengesizliğini hesaba katmak için macro parametresini değiştirir. Dolayısıyla precision ve recall arasında olmayan bir F1 skoru ortaya çıkarabilir.

Ayrıca sınıflandırma yöntemlerinin kappa skorları gözlemciler arası ilişki bazında Random Forest için 0.681 (önemli derecede uyuma), Decision Tree için 0.576 (çoğunlukla uyumlu olma durumu) olarak bulunmuştur (Tablo 3, Tablo 7).

Random Forest ve Decision Tree sınıflandırma algoritmaları değerlendirildiğinde sırasıyla %81 ve %73 oranında sınıflandırma yapabilmektedirler. Flask Framework'ü ile web ortamında girişimcilerden alınan proje verileri başarı tahminlemesi için sorgulanabilir bir duruma gelmiştir. Uygulama üzerinde veri artışı sağlandıkça ortaya çıkan değerler de değişiklik gösterecektir. Bu değişiklik yine web arayüzünde yüzdesel olarak kullanıcı tarafında değerlendirilir.

Tablo 7. Sınıflandırma Yöntemlerinin Değerlendirme Metriklerine Göre Sonuçlarının Karşılaştırılması  
(Comparison of the Results of Classification Methods According to Evaluation Metrics)

Sınıflandırma Yöntemleri	Doğruluk Değeri	Kesinlik Değeri			Duyarlılık Değeri			F1 Skoru			Kappa Skoru
		Macro	Micro	Weighted	Macro	Micro	Weighted	Macro	Micro	Weighted	
Random Forest	0.81	0.462	0.815	0.741	0.411	0.815	0.815	0.400	0.815	0.764	0.681
Decision Tree	0.73	0.411	0.731	0.732	0.410	0.731	0.731	0.410	0.731	0.731	0.576

### 2.5. Konuşlandırma ve Kontrol (Deployment and Control)

Veri madenciliği proje döngüsünün konuşlandırma sürecinde uygulama kullanıma hazır bir şekilde hayata geçirilmektedir. Bu kapsamda, kullanılan makine öğrenmesi sınıflandırma yöntemleri Flask Framework'ü kullanılarak web arayüzünde gösterilmiştir (Şekil 3). Ortaya çıkan sonuçlar proje özellikleri değiştikçe değişmekte ve projenin başarı durumu için önceki veriler göz önüne alınarak

varsayımsal bir tahmin yapılmaktadır. Dolayısıyla, makine öğrenmesinin yöntemlerinin iş zekasına katkısının örnekleri görülmüştür [25]. Karar desteğinin etkileştirilmesi için web ortamında kullanıcı arayüzü oluşturma kapsamında karmaşık problemlere de çözüm sağlandığı görülmektedir [26]. Öyle ki dağınık özelliklere sahip olan Kickstarter veri seti eğitilerek sınıflandırma için hazır bir şekilde web ortamında kullanılmıştır.

Classification Methods	Accuracy Score	F1 Score (Macro)	F1 Score (Micro)	F1 Score (Weighted)	Precision Score (Macro)	Precision Score (Micro)	Precision Score (Weighted)	Recall Score (Macro)	Recall Score (Micro)	Recall Score (Weighted)	Kappa Score	Classification Results
Random Forest	% 81.55	% 39.89	% 81.55	% 76.41	% 43.76	% 81.55	% 74.39	% 41.14	% 81.55	% 81.55	% 68.36	Successful Project / Class: [2]
Decision Tree	% 72.75	% 41.05	% 72.75	% 72.95	% 41.06	% 72.75	% 73.16	% 41.05	% 72.75	% 72.75	% 57.12	Successful Project / Class: [2]

Şekil 3. Flask ile Sınıflandırma Yöntemlerinin Kullanıcı Arayüzünde Gösterilmesi  
(Showing Classification Methods with Flask on User Interface)

### 3. TARTIŞMA ve SONUÇ (DISCUSSION and CONCLUSION)

Makine öğrenmesi yöntemleriyle kitle fonlaması projelerinin sınıflandırılmasıyla ilgili zamanla daha iyi doğruluk değerleri üreten çalışmalar yapılmış ve gözlemlenmiştir. Yapılan çalışmalarda, birlikte kullanılan algoritmalar ile yapılan sınıflandırmalarda başarı oranının arttığı gözlemlenebilmektedir. Fakat veri setine uygun algoritmanın tespit edilerek sınıflandırma yapılabilmesi, algoritmaların birlikte kullanımı yerine daha iyi sonuçlar ortaya koyabilmektedir. Nitekim çalışmamızda da %81.55 ile literatürdeki çalışmalardan daha iyi bir sonuç elde edilmiş ve değerlendirme metrikleriyle bu sonucun sınıflandırma yetenekleri test edilmiştir. Yani, veri setine uygun algoritma seçimi yapıldığında birlikte algoritma kullanımına gerek kalmadığı sonucu ortaya çıkmaktadır. Bu yüzden makine öğrenmesi

yöntemleriyle yapılacak çalışmalarda veri setine uygun algoritmanın bulunması büyük önem teşkil eder. Ayrıca çalışmamızın çıktılarından bir diğeri de ortaya koyulan sınıflandırma sonuçlarının son kullanıcıya sunulmasıdır. Böylelikle dinamik bir yapı kurulmuştur. Kullanıcı sistem üzerinde sorgulama yaptıkça proje özellikleri sisteme dahil olabilecektir. Dolayısıyla öğrenme yeteneği artabilen bir yapı ortaya çıkmaktadır.

Sonuç olarak, gözetimli öğrenme kapsamındaki sınıflandırma yöntemlerinin bir araya gelmesiyle geliştirilen web ortamında, Kickstarter üzerinde projeleri için destek bekleyen girişimciler için bir karar destek yapısı oluşturulmuştur. Oluşturulan yapı girişimcilerin projelerinin özelliklerine odaklanmış ve çalışmanın temeli bu olmuştur. Proje özelliklerine dayanarak yapılan sınıflandırma işlemleri, Random Forest için %81.55, Decision Tree için ise %72.75



sonuçlarını vermiştir. Sonuçlar karmaşıklık matrisi ile test edilerek F1 skoru, Kesinlik ve Duyarlılık skorlarıyla değerlendirilmiştir. Ayrıca değişkenler arasındaki ilişkilerinde ortaya koyulabilmesi için de Kappa değeri kullanılmış, çok yüksek ve önemli derecede uyumlar ile anlamlı sonuçlar üretilmiştir. Ortaya koyulan platformun daha da işlevsel olabilmesi için gelecek çalışmalarda kullanıcı karar desteğinin artırılması gerekmektedir. Bu gereksinim yapay sinir ağlarının sağlayabileceği olanakların son kullanıcı tarafına aktarılmasıyla mümkün olabilir. Bu sebeple benzer çalışmaların titizlikle gerçekleştirilerek yeni yöntemlerin uygulanması literatüre büyük bir katkı sağlayacaktır.

#### 4. KAYNAKLAR (REFERENCES)

- [1] Laudon, K. C. (2007). *Management information systems: Managing the digital firm*. Pearson Education India.
- [2] Theis, T. N., & Wong, H. S. P. (2017). The end of moore's law: A new beginning for information technology. *Computing in Science & Engineering*, 19(2), 41.
- [3] Mansoori, Y., Karlsson, T., & Lundqvist, M. (2019). The influence of the lean startup methodology on entrepreneur-coach relationships in the context of a startup accelerator. *Technovation*, 84, 37-47.
- [4] Zvilichovsky, D., Inbar, Y., & Barzilay, O. (2015). Playing both sides of the market: Success and reciprocity on crowdfunding platforms. Available at SSRN 2304101.
- [5] Etter, V., Grossglauser, M., & Thiran, P. (2013, October). Launch hard or go home! Predicting the success of Kickstarter campaigns. In *Proceedings of the first ACM conference on Online social networks* (pp. 177-182).
- [6] Kuppuswamy, V., & Bayus, B. L. (2018). Crowdfunding creative ideas: The dynamics of project backers. In *The economics of crowdfunding* (pp. 151-182). Palgrave Macmillan, Cham.
- [7] Cheng, C., Tan, F., Hou, X., & Wei, Z. (2019, August). Success prediction on crowdfunding with multimodal deep learning. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence, Macao, China* (pp. 10-16).
- [8] Chen, K., Jones, B., Kim, I., & Schlamp, B. (2013). Kickpredict: Predicting Kickstarter Success. *Technical report, California Institute of Technology*.
- [9] Kindler, A., Golosovsky, M., & Solomon, S. (2019). Early Prediction of the Outcome of Kickstarter Campaigns: Is the Success due to Virality? *Palgrave Communications*, 5(1), 1-6.
- [10] Chung, J., & Lee, K. (2015, August). A long-term study of a crowdfunding platform: Predicting project success and fundraising amount. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media* (pp. 211-220).
- [11] Rao, H., Xu, A., Yang, X., & Fu, W. T. (2014, April). Emerging dynamics in crowdfunding campaigns. In *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction* (pp. 333-340). Springer, Cham.
- [12] Jensen, L. S., & Özkil, A. G. (2018). Identifying challenges in crowdfunded product development: a review of Kickstarter projects. *Design Science*, 4.
- [13] Du, Q., Fan, W., Qiao, Z., Wang, G., Zhang, X., & Zhou, M. (2015). Money talks: a predictive model on crowdfunding success using project description.
- [14] Bi, S., Liu, Z., & Usman, K. (2017). The influence of online information on investing decisions of reward-based crowdfunding. *Journal of Business Research*, 71, 10-18.
- [15] Mortensen, S., Christison, M., Li, B., Z hu, A., & Venkatesan, R. (2019, April). Predicting and Defining B2B Sales Success with Machine Learning. In *2019 Systems and Information Engineering Design Symposium (SIEDS)* (pp. 1-5). IEEE.
- [16] Mouillé, M. (2018). Kickstarter Projects Dataset, Kaggle. More than 300,000 kickstarter projects (Version 7). Access address: <https://www.kaggle.com/kemical/kickstarter-projects>
- [17] Huber, S., Wiemer, H., Schneider, D., & Ihlenfeldt, S. (2019). DMME: Data mining methodology for engineering applications—a holistic extension to the CRISP-DM model. *Procedia CIRP*, 79, 403-408.
- [18] Fahmy, A. F., Mohamed, H. K., & Yousef, A. H. (2017). A data mining experimentation framework to improve six sigma projects. In *2017 13th International Computer Engineering Conference (ICENCO)* (pp. 243-249). IEEE.
- [19] Berhane, T. M., Lane, C. R., Wu, Q., Autrey, B. C., Anenkhonov, O. A., Chepinoga, V. V., & Liu, H. (2018). Decision-tree, rule-based, and random forest classification of high-resolution multispectral imagery for wetland mapping and inventory. *Remote sensing*, 10(4), 580.

[20] Leiva, R. G., Anta, A. F., Mancuso, V., & Casari, P. (2019). A Novel Hyperparameter-Free Approach to Decision Tree Construction That Avoids Overfitting by Design. *IEEE Access*, 7, 99978-99987.

[21] Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1), 217-222.

[22] Xu, J., Zhang, Y., & Miao, D. (2020). Three-way confusion matrix for classification: a measure driven view. *Information Sciences*, 507, 772-794.

[23] Yang, T. L., Lin, C. H., Chen, W. L., Lin, H. Y., Su, C. S., & Liang, C. K. (2019). Hash Transformation and Machine Learning-based Decision-Making Classifier Improved the Accuracy Rate of Automated Parkinson's Disease Screening. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*.

[24] Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1), 37-46.

[25] Wang, J., Wu, X., & Zhang, C. (2005). Support vector machines based on K-means clustering for real-time business intelligence systems. *International Journal of Business Intelligence and Data Mining*, 1(1), 54-64.

[26] Cook, A., Wu, P., & Mengersen, K. (2015). Machine learning and visual analytics for consulting business decision support. In *2015 Big Data Visual Analytics (BDVA)* (pp. 1-2). IEEE.