

RİSK FAKTÖRLERİ ve RİSK DEĞERLENDİRME MODELLERİNİN FARKLI VERİ SETLERİ ÜZERİNDE GERÇEKLENMESİ

Ahmet Unudulmaz¹, Oya Kalıpsız², ve M. Özgür Cingiz³

^{1,2,3} Bilgisayar Mühendisliği Bölümü, Yıldız Teknik Üniversitesi, İstanbul

¹ahmetunudulmaz@gmail.com

^{2,3} {kalipsiz,mozgur}@yildiz.edu.tr

Özet. Başarısız giden birçok proje, hatalı yönetilen süreçler, ürünlerin ve projelerin zamanında teslim edilememesi, maliyetlerde oluşan aşırı artışlar yazılım mühendisliğinde risk yönetiminin önemini günden güne arttırmaktadır. Özellikle risklerin düzgün bir şekilde ele alınamaması, projeler arası risk türlerinin ve gruplarının farklılaşması, şirket/kurum içi yapılan risk yönetimi ve risk analizi çalışmalarının ölçülememesi de bu durumu karmaşıklaştırmaktadır. Bu çalışmada risk faktörleri, sınıflandırıcı yöntemleri ve model performans yapıları incelenmekte, ayrıca 2 farklı şirketin veri setlerinin karşılaştırılmasına da yer verilmektedir. Gerçekleştirilen uygulamada, farklı firmaların veri setleri üzerinde faktör kararı verilerek gereksiz risk faktörlerinin projelerden çıkarılması sağlanmış, 32 risk faktörü kendi içinde sıralanarak hangi alt alanın en çok risk taşıdığı ortaya konulmuştur. Veri setleri içindeki risk faktörlerinin etki değerleri, risk faktörlerinin birbirleriyle kıyaslanmaları ve risk faktörü eleme yöntemlerinden bahsedilmiş olup, bu faktörler yardımıyla veriler çok kritik, kritik, koşullu hata, orta seviyeli ve ihmal edilebilir olarak sınıflandırılmıştır. Sınıflandırma işleminde Destek Vektör Makineleri, Karar Ağaçları ve Naive Bayes sınıflandırıcılarından faydalanılmıştır. Kullanılan yöntemlerin başarı oranları ve başarı yüzdeleri de değerlendirilmiştir.

Anahtar Kelimeler. Yazılım Risk Faktörleri, Riks Modelleri, Risk Tahmini, Risk Değerlendirme Modelleri

1 Giriş

Yazılım mühendisliğinde risk yönetiminin önemi son yıllarda daha çok ortaya çıkmıştır. Başarısız giden birçok proje, hatalı yönetilen süreçler, ürünlerin ve projelerin zamanında teslim edilememesi, maliyetlerde oluşan aşırı artışlar risk yönetiminin önemini günden güne arttırmaktadır. Özellikle risklerin düzgün bir şekilde ele alınamaması, projeler arası risk türlerinin ve gruplarının farklılaşması, şirket/kurum içi yapılan risk yönetimi ve risk analizi çalışmalarının ölçülememesi de bu durumu karmaşıklaştırmaktadır. Bildiri ise yukarıda belirtilen problemler üzerine yapılmıştır. Risklerin her projede standartlaştırılması, risk faktörlerinin düzgün bir şekilde analiz edilmesi ve uygulanan risk önleme ve risk analiz çalışmalarının başarı oranlarının ölçülebilmesi hedeflenmiştir.

Yazılım risk yönetimi alanında yapılan bu çalışmada temel olarak risk faktörleri ve risk değerlendirme modellerinin birbirleriyle karşılaştırılmasından bahsedilecektir. İncelenen çalışmalar göz önünde bulundurulduğunda bu iki alanda şirketlerde yaşanan problemler ve zaman kayıpları dikkate alınmıştır. Risk faktörlerinin analiz edilmemesi veya hatalı analiz edilerek gerkesiz bir faktör için önlem alınması zaman kaybına ve maddi kayıplara yol açmaktadır. Bu nedenle faktörlerin doğru seçilmesi, kritik faktörlerin belirlenmesi ve etki değeri düşük faktörlerin elenmesi gerekmektedir. Bildiri çalışması kapsamında risk faktörlerinin detaylı analizi amaçlanmıştır. Diğer bir amaç ise yazılım mühendisliği yaklaşımlarını kullanarak indirgenmiş risk faktörleri ve veri setlerini doğru sınıflandırabilecek modeller oluşturmaktır. Kullanılan yöntemlerin başarı oranları ve başarı yüzdeleri de çalışma süresince incelenecektir.

Yapılacak çalışmada Turkcell Teknolojide gerçekleştirilen projelerdeki problem etki veri setleri ile yazılım sektöründe faaliyet gösteren bir firmanın hata veri setleri kullanılacaktır. Naive Bayes, destek vektör makineleri ve karar ağaçlarıyla oluşturulan model benzer yazılım projelerinde ortaya çıkacak aynı tip problemlerin projeye etkisine ait kestirim yapılmasını sağlamaktadır.

2 İlgili Çalışmalar

Yazılım mühendisliğinde risk yönetimi ile ilgili çalışmalara bakıldığında ise son yıllarda artış gözlemlenmektedir. Bu alanda yapılan çalışmalar geçmiş yıllara göre çeşitlilik göstermekte ve önem kazanmaktadır. Genel risk yönetimi, risk faktörleri, risk önlemek için modellerin kullanılması ile ilgili çalışmaların yayınlandığı, bu konular üzerinde daha detaylı durulduğu gözlemlenmiştir. Bu kısımda yazılım mühendisliğindeki risk yönetimi konusunda yapılan çalışmalar iki alt başlık altında toplanmıştır. İlk başlıkta risk yönetimi ve analizi ile ilgili dergi ve makale çalışmaları incelenmiş ve derlenmiştir. İkinci başlıkta ise risk faktörlerine yer verilmiş, faktörlerin birbirleriyle kıyaslanması, eleminize edilmesi ve sıralanması ile ilgili yapılan çalışmalardan bahsedilmiştir. Ayrıca yazılım mühendisliği yaklaşımları kullanılarak oluşturulan modeller ortaya konulmuş ve modellerinde birbirleriyle kıyaslanmasına yer verilmiştir.

2.1 Risk Yönetimi ve Analizi ile İlgili Yapılan Çalışmalar

Risk yönetimi ve analizi ile ilgili yapılan genel çalışmalar bu başlık altında toplanmıştır. Tahminler, belirsizlikler ve riskler ele alınmış ve bu belirsizliklerin/risklerin sebepleri incelenmiştir. Ayrıca risk yönetim metodolojilerinden ve risk adımlarından bahsedilmiş ve risklerin bağımlılıklarında bu kısımda değinilmiştir. Risk analizi alanında yapılan çalışmalarda bahsedilerek risk yönetimi ve analizi ile ilgili yapılan çalışmalar tamamlanmıştır.

Kitchenham ve Linkman bu çalışmasında tahminler, belirsizlik ve risk konularını ele almışlardır. Tahmin etme modellerini ise bu faktörleri göz önünde bulundurarak oluşturulması gerektiğine vurgu yapılmış ve belirli olmayan risk faktörleri için varsayımlarda bulunması gerektiğini belirtmiştir. Hatalı varsayımların etki değerleri göz

önünde bulundurulmakta değerlendirme modellerinin çıkarılabileceğinden bahsedilmiştir [1].

Kwan ve Leung çalışmasında ise risk yönetim metodolojisini ve proje risk bağımlılıklarını ele almıştır. Konservatif metod, optimistik metod ve ağırlık metodları kullanılarak risk bağımlılığı grafi oluşturulmuştur. Risk grafları oluşturulduktan sonra ise risk yönetim paradigmaları risk bağımlılığı yönünden ele alınmıştır. Buna göre 3 farklı durum için kritiklik seviyeleri belirlenmiştir. Bunlardan ilki iç ciddiyet seviyesidir. Risk yönetim paradigmalarında direkt olarak kullanılmaktadır. İkinci ise diğer riskin oluşma durumuna için kullanılan ciddiyet seviyesidir. Üçüncü olarak ise tepki ciddiyet seviyesi kullanılmaktadır. Bu seviyelere göre farklı aksiyon planları alınmaktadır [2].

Ray ve Mohapatra ise risk analizi ile ilgili bir çalışma yayınlamışlardır. Yazılım geliştirme yaşam döngüsü içinde analiz ve dizayn adımları için durum bazlı risk önleme metodolojisi üzerine çalışmalar yapmışlardır. Risk tahmin metodu geliştirerek komponentler arası karmaşıklığı indirgeyip her bir bileşeni graf yöntemleriyle incelemişlerdir. Bunun yanında kritiklik analizinde yaparak sistem seviyelerindeki başarısızlık senaryolarını analizini yapmışlardır. Buna göre hata seviyelendirmeyi felaket, major, marjinal ve minör olarak 4 seviyede toplamışlardır [3].

2.2 Risk Faktörleri ve Risk Modellerinin Kullanılması ile İlgili Çalışmalar

Risk faktörlerinin değerlendirilmesi, sınıflandırılması ve eleminize edilmesi ile ilgili çalışmalarla risk modellerinin oluşturulması ve oluşturulan bu modellerin birbiriyle kıyaslanması ile ilgili çalışmalar bu başlık altında toplanmıştır. İncelenen çalışmalarda yapay sinir ağları, destek vektör makineleri, NaiveBayes, kNN yöntemi, bulanık mantık, çok katmanlı yapay sinir ağları kullanılmıştır. Bu sınıflandırma yöntemleri sayesinde risk verileri ve risk faktörleri sınıflandırılabilmiş, bilgisayar mühendisliği yaklaşımları kullanıldıktan sonra ise risk modelleri oluşturulmuştur. Yapılan çalışmalar aşağıda listelenmiştir.

Sun Yat Sen Üniversitesi'nden yayınlanan makalede yapay sinir ağları ve destek vektör makineleri yöntemleri kullanılarak yazılım projelerinin risk yönetiminin modellenmesi sağlanmıştır. 300 farklı firmaya gönderilen 194 anket sorusuyla birlikte kategoriler daha da netleştirilmiştir. Anket sonuçlarında 6 kategori bulunmuştur. Belirlenen 64 farklı risk yukarıda yer alan 6 ana risk grubu içerisinde tanımlanmıştır. Riskler standart yapay sinir ağlarıyla, genetik algoritma ile iyileştirilmiş yapay sinir ağlarıyla ve SVM yaklaşımlarıyla tahminleri yapılmıştır. NN elde edilen sonuçlar projelerin gerçek sonuçlarıyla kıyaslandığında %70 oranında bir doğruluk oranı elde edilmektedir. SVM kullanarak elde edilen başarı oranı ise %80'nin üzerinde yer almaktadır [4,5].

Windsor üniversitesinde yapılan çalışmada ise NaiveBayes sınıflandırıcısı ile K-en yakın komşuluk sınıflandırıcısının performansları incelenmiştir. Veri setlerine uygulanan bu iki yöntemin performansları belirtilmiştir. 671 veri 9 sınıfa ayrılmak istenmiş. NaiveBayes sınıflandırıcısı %12,43 hatayla verileri doğru sınıflandırabilmiştir. K en yakın komşuluk metodunda ise k değerine göre yüzdeler değişkenlik göstermiştir. K değerinin 5 seçildiği durumlarda hata oranı en düşük

çıkmiştir. %9,45 hata oranıyla sınıflandırma yapabilen K en yakın komşuluk yöntemi NaiveBayes yöntemine göre daha başarılı olmuştur [6].

Bulanık mantık alanında da risk değerlendirme alanında çalışmalar yapılmıştır. Hunan üniversitesinde yapılan bir çalışmada risklerde oluşan kayıplar üzerine yoğunlaşmıştır. Risk faktörlerini belirlemişler ve risklerin önem sırasına göre risk analiz ağı oluşturmuşlar(Semantik indirgemeler kullanılarak). Sonuç olarak yaptıkları çalışmada yazılım projelerindeki risk faktörlerin kritiklik seviyesini analiz etmişlerdir.Bulanık mantık alandaki bir diğer çalışmada ise yazılım projelerindeki tasarım risk analizi incelenmiştir. Yazılım projelerindeki risklerin tanımı ve kritiklik seviyelerinin farklılık göstermesinden dolayı böyle bir çalışma yayınlamışlardır. Proje riskleri ve ayrı bileşenler arasında bulanık mantık yöntemi kullanılmıştır. Bunun sonucunda da düşük, orta ve yüksek seviyeli riskleri tespit edebilmişlerdir [7,8].

3 Metodoloji

Yapılan çalışmada risk faktörleri belirlenmiş, risk verileri toplanmış ve model oluşturulmuştur. Yazılım sektöründe çalışan bir firmadan elde edilen, hataları seviyelendirmek için kullanılan alanlara göre faktörler belirlenecektir. Risk kestirimi sektörden sektöre farklılaşabileceği gibi, aynı işi yapan firmalar arasında bile riskler farklı olarak belirlenebilmektedir. Her yeni projeye göre önceki projeler için tanımlanan riskler farklılaşabilmektedir. Bu farklılık projelerde oluşan problemlerde dahi görülmektedir. Bu da risk yönetiminde standart oluşturmayı zorlaştırmaktadır. Bu nedenle bir projeye başlamadan önce ele alınması gereken ilk kısım risk faktörlerinin analizi olmalıdır.

Turkcell ICT'den elde edilen etki analizi için kullanılan veriler ile yazılım sektöründe faaliyet gösteren bir firmanın veri setlerine göre uygulama gerçekleştirilecektir. Yapılacak projede Turkcell ICT için temel olarak 6 risk faktörü ve problemin proje içindeki etkisini gösteren etki değeri belirlenmiştir. 384 problem etki değer veri seti üzerinde bilgisayar mühendisliği yöntemleri kullanılacaktır.

Yazılım sektöründe faaliyet gösteren firmanın verileri ise 8 projeden ve 32 risk faktöründen oluşmaktadır. Projeler yazılım alanında gerçekleşmiş olup veri setleri bu yazılımlarda görülen hatalardan oluşmaktadır. Temel olarak 32 risk faktörü etki değeri belirlenmiştir. Faktörlerden 19 tanesi tekil ve diğer projelerde ortak olmadığı için elenmiştir. Kalan 13 faktör ve alt alanları detaylı şekilde aşağıdaki tabloda belirtilmiştir. 6382 hata veri setleri üzerinde bilgisayar mühendisliği yöntemleri kullanılacaktır.

Her iki veri seti içinde etki değerlerine göre sınıflandırma algoritmaları kullanılarak başarı oranları karşılaştırılacaktır. İlk olarak Turkcell ICT risk faktörleri ve değerleri tablosu verilecek. Daha sonra ise yazılım firması risk faktörleri ve değerleri tablosu verilecektir.

Tablo 1. Turckell ICT Risk Faktörleri ve Değerleri

Faktörler	Değerler
FİNANSAL ETKİ	Gelir_Raporlamasına_ etkisi_var, Yok, BAPT_1/2000_under_charge_1/50000_over_charge
ETKİ İÇ MÜŞTERİ	Yok, 1-100, 101-1000, 1000_den_Fazla
ETKİ ABONE SAYISI	1-1000, 1001-10000, 10000_den_Fazla, Yok
REGULASYON ETKİSİ	Var, Yok
SOX ETKİSİ	Var, Yok
MARKA ETKİSİ	Var, Yok
ETKİ	HIGH, LOW, MEDIUM

Tablo 2. Yazılım Şirketi Risk Faktörleri ve Değerleri

Faktörler	Alt Alan	Değerler
Alt alanın öneri olup olmama durumu	2 alt alan	“Öneridir”, “Öneri değildir”
Proje içindeki durum	10 alt alan	“Yeni”, “Çözülen”, “Paydaşlara gönderilen”, “Tekrarlanamayan”, “Otomasyon” vs.
Görülme seviyesi	6 alt alan	“Her zaman”, “Bazen”, “Nadiren”, “Sık sık”, “Hiçbir zaman”, “Otomotize edildiğinde”
Kişi-grup-şirket bazında	Değişken	Kişi-grup-şirket sayısına göre değişkenlik gösteriyor
Kişi-grup-şirket bazında çözümleyen	Değişken	Kişi-grup-şirket sayısına göre değişkenlik gösteriyor
Kişi-grup-şirket bazında değiştiren	Değişken	Kişi-grup-şirket sayısına göre değişkenlik gösteriyor
Sınıf bilgisi	30 alt alan	“Fonksiyonel”, “Arayüz”, “Bağlantı”, “Kullanılabilirlik”, “Otomasyon”, “Faturalama”, “Güvenlik” vs.
Öncelik olma durumu	4 alt alan	“Öncelikli”, “Öncelikli değil”, “Koşullu öncelikli”, “Koşullu öncelikli değil”
İşletim sistemi	2 alt alan	“Windows”, “Linux”
Şirket bazlı mı?	3 alt alan	“Şirket bazlı”, “Paydaşlar”, “Outsource”
Hangi grupta gösterileceği	10 alt alan	Tür bilgisi ile paralellik gösteriyor
Yazılımsal-tasarımsal olma durumu	2 alt alan	“Yazılımsal”, “Donanımsal”
Kişi-grup-şirket bazında atanmış	Değişken	Kişi-grup-şirket sayısına göre değişken
Etki	6 alt alan	Çok kritik, kritik, orta, koşullu, düşük, ihmal edilebilir

Bilgi kazanımı bir özelliğin sınıf bilgisine ait belirsizliğinin, özellik bilgisi kullanılarak azaltılması olarak da tanımlanabilir. Belirsizlik ölçütü olarak rastsal

değişkenlerde kullanılan entropi, bilgi kazanımı hesabında da kullanılmaktadır. Kısaca, bilgi kazanımı özelliğın bir veya birden fazla sınıf için var olup olmamasına göre entropideki azalımı ölçmektedir.

Tablo 3. Bilgi Kazanımı Özellik Seçimine Göre Turkcell ICT Etki Değerleri Sıralaması

Risk Faktörleri	Etki Değerleri
Regulasyon Etkisi	Yüksek
Finansal Etki	Yüksek
Etki Abone Sayısı	Orta
Sox Etkisi	Orta
Etki İç Müşteri	Düşük
Marka Etkisi	Düşük

Tablo 4. Bilgi Kazanımı Özellik Seçimine Göre Yazılım Firması Etki Değerleri Sıralaması

Risk Faktörleri	Etki Değerleri
Kişi grup şirket bazında atanan	Yüksek
Kişi grup şirket bazında çözümleyen	Yüksek
Proje içindeki durum bilgisi	Yüksek
Kişi-grup-şirket bazında değiştiren	Yüksek
Kişi-grup-şirket bazında yaratan	Orta
Sınıf bilgisi	Orta
Grüldüğü işletim sistemi	Orta
Öncelik seviyesi	Düşük
Grup bilgisi	Düşük
Görülme sıklığı	Düşük
Öneri olma duru	Düşük
Yazılımsal-tasarımsal olma durumu	Düşük
Şirket bazlı olma durumu	Düşük

Bilgi kazanımı yüksek olan özellikler sınıflandırma için değerli özellikler olarak alınır ve belirli bir eşik değerinin altında kalan özellikler elenerek özellik seçimi gerçekleştirilmiş olur. Weka arayüzünde bilgi kazanımı kullanılarak bu faktör için yapılan çalışma sonrasında yukarıdaki tablolar elde edilmiştir.

3.1 Destek Vektör Makineleri

Risk analizi için en çok kullanılan sınıflandırma yöntemlerinden biri destek vektör makinesidir. Destek vektör makineleri metin sınıflandırma, yüz tanıma, el yazısı tanıma ve buna benzer makine öğrenmesi algoritmalarının kullanıldığı alanlarda başarılı sonuçlar vermesinden dolayı çalışmada tercih edilen sınıflandırma yöntemlerinden biridir. Destek Vektör Makineleri karar düzlemlerinin ideal şekilde belirlenmesi prensibine dayanır. İdeal karar sınırı ayırdığı sınıflara ait verilere mümkün olduğu kadar karar düzleminde uzak olmalıdır.

Weka arayüzünde sınıflandırıcı olarak destek vektör makinelerinin kullanımı sonucunda Turkcell ICT için 384 problemin etki analiz verisinin 371 adedini doğru olarak belirleyebilmiştir. Modelin doğruluk oranı ise %96,61'dir.

Yazılım sektöründeki firmanın veri setlerine bakıldığında 8135 hata veri setinin 6382 adedini doğru olarak sınıflandırmıştır. Modelin doğruluk oranı %78,45'dir. Test sonuçlarının doğrulama yöntemi 10 kere çapraz doğrulama olarak uygulanmıştır. Karışıklık matrisine bakıldığında ise kritik, koşullu hata, hata değil, orta derece hatalar genel olarak doğru olarak sınıflandırılmasına rağmen çok kritik ve ihmal edilebilir hatalardaki sınıflandırma başarı yüzdelerinde düşüşler gözlemlenmiştir.

3.2 Karar Ağaçları

Karar ağaçları örnek verileri bir ağaç yapısına benzemektedir. Ağacın kökünden ağacın alt dallarına kadar sıralayarak verilerin sınıflandırılmasını sağlar. Ağaçtaki düğümler özellik, dallar özellik değer bilgisini ve yaprak düğümler ise sınıf etiketini belirtmektedir. Karar ağaçlarının yaygın kullanımının nedeni ağaç yapılarının kurallarla ve sade bir şekilde tanımlanabilmesidir. Bu şekilde öğrenilen kurallar kolay bir şekilde aktarılmış olur. Karar ağaçları kullandığı algoritma farklılıklarına göre en çok kullanılanlar ID3, ASSISTANT ve J48'dir.

Karar ağacı sınıflandırıcısıyla veriler aç gözlü yaklaşım (greedy) olan belirli kriterleri en iyi özelliğe dayalı bölünürler. Düğümler bölünme şekline göre çoklu veya ikili bölünmektedir. En iyi düğümler bölünmesi için homojen sınıf dağılımına sahip düğümler tercih edilir. Homojenliği belirlemek için düğümün saf olmama değeri (impurity) hesaplanır. Saf olmama değerini ölçmek için gini indeksi, entropi gibi değerler kullanılır. Tüm veriler karar ağacında doğru şekilde sınıflandırılana kadar karar ağacı dallandırılır.

Weka arayüzünde sınıflandırma fonksiyonlarından Turkcell ICT veri setinde karar ağacı yaklaşımı olarak J48 karar ağacı sınıflandırıcı kullanılmıştır. 384 problem etki analiz verisinin 350 adedini doğru olarak belirleyebilmiştir. 34 tanesini ise farklı etiketlemiştir. Modelin doğruluk oranı ise %91,14'dür. Yazılım firma veri setlerinde ise Weka arayüzünde sınıflandırma fonksiyonlarından Karar ağacı yaklaşımı olarak J48 karar ağacı sınıflandırıcı kullanılmıştır. 8135 hatanın 6231 adedini doğru olarak belirleyebilmiştir. 1904 tanesini ise farklı etiketlemiştir. Modelin doğruluk oranı ise %76,59'dur. Test sonuçlarının doğrulama yöntemi 10 kere çapraz doğrulama olarak uygulanmıştır. Karışıklık matrisine bakıldığında ise kritik, koşullu hata, hata değil,

orta derece hatalar genel olarak doğru olarak sınıflandırılmasına rağmen çok kritik ve ihmal edilebilir hatalardaki sınıflandırma başarı yüzdelerinde düşüşler gözlemlenmiştir.

3.3 Naive Bayes Yönetimi

Klasik Naive Bayes uzun yıllardır farklı sınıflandırma problemlerinde yaygın olarak kullanılan makine öğrenmesi yaklaşımıdır. Klasik Naive Bayes bağımsızlık varsayımına dayanmaktadır. Klasik Naive Bayes'e göre özellikler sınıftan bağımsızdır. Klasik Naive Bayes tüm verilere ait özelliklerden verinin sınıf bilgisi olarak varsayımda bulunduğu sınıfa ait olma olasılık değerini her sınıfı etiketi için hesaplar ve en yüksek olasılık değerini veren sınıfa ait varsayımı, doğru olarak kabul ederek sınıflandırma işlemini gerçekleştirir.

Naive Bayes Yöntemi kullanılarak Weka arayüzünde sınıflandırma fonksiyonlarından yararlanılarak Turkcell ICT veri setinde Naive Bayes risk modeli oluşturulmuştur. 384 problem etki analiz verisinin 352 adedini doğru olarak modelleyebilmiştir. 32 tanesini ise farklı etiketlemiştir. Modelin doğruluk oranı ise %91,66 dır.

Yazılım firmasının veri setlerinde ise NaiveBayes yöntemi kullanıp Weka arayüzünde sınıflandırma fonksiyonlarından yararlanılarak risk modeli oluşturulmuştur. 8135 problem etki analiz verisinin 5914 adedini doğru olarak modelleyebilmiştir. 2221 tanesini ise farklı etiketlemiştir. Modelin doğruluk oranı ise %72,69dur. Test sonuçlarının doğrulama yöntemi 10 kere çapraz doğrulama olarak uygulanmıştır. . Karışıklık matrisine bakıldığında ise kritik, koşullu hata, hata değil, orta derece hatalar genel olarak doğru olarak sınıflandırılmasına rağmen çok kritik ve ihmal edilebilir hatalardaki sınıflandırma başarı yüzdelerinde düşüşler gözlemlenmiştir.

4 Sonuçlar ve Öneriler

Yapılan çalışmada risk faktörleri belirlenerek Turkcell ICT için veri setinde yer alan problemlerin projelerdeki etki değerlerine ve önemine bakılmıştır. Yazılım sektöründe faaliyet gösteren firmanın veri setine bakıldığında ise veri setinde yer alan hata değerlerin incelenmiştir. Veri setlerindeki bu değerlere göre risk modelleri oluşturulmuş ve oluşturulan bu modeller yardımıyla gelecek problemlerin etki değerlerini kestirilmiştir. Veri setleri üzerinde Naive-Bayes, destek vektör makineleri ve karar ağaçları yöntemleri kullanılarak oluşturulan modeller incelenmiştir.

Turkcell Teknolojiden elde edilen veri kümesinde 6 risk faktörü ve problemin etkisinden oluşan 7 özellikten oluşan 384 veri çalışmada kullanılmıştır. Yazılım firmasındaki veri setlerinde ise 13 risk faktörü ve hata bilgilerinden oluşan 14 özellikten oluşan 8135 veri çalışmada kullanılmıştır.

Destek Vektör Makineleri(DVM), Karar Ağacı(j48) ve Naive-Bayes(NB) yaklaşımlarıyla projede yer alan veriler eğitilerek 10 kere çapraz doğrulama yöntemiyle test edilmiştir.

Yazılım sektöründeki firma verilerinin sonuçlarına bakıldığında bulunan sonuçlar da da görüldüğü üzere DVM ile sınıflandırılan verilerde doğru sınıflandırma oranı karar ağaçlarından ve NaiveBayes sınıflandırıcılarından daha iyi sonuçlar verdiği gözlemlenmiştir. DVM %78,45 oranında başarı gösterirken, KA %76,59, NB ise %72,69 oranında başarılı sınıflandırma yapmıştır.

Turkcell ICT verilerinin sonuçlarına bakıldığında bulunan sonuçlarda da görüldüğü üzere DVM ile sınıflandırılan verilerde doğru sınıflandırma oranı karar ağaçlarından ve NaiveBayes sınıflandırıcılarından daha iyi sonuçlar verdiği gözlemlenmiştir. DVM %91,66 oranında başarı gösterirken, KA %91,14, NB ise %91,66 oranında başarılı sınıflandırma yapmıştır.

DVM her iki veri setinde en iyi sınıflandırma yöntemi olarak ortaya çıkmıştır. İlgili çalışmalar kısmında incelenen makaleler ve projelerde elde edilen sonuçlara paralellik göstermektedir [9,10]. Turkcell ICT'deki başarı yüzdeleri daha yüksek olmasına rağmen daha az veri ve faktör kullanılmıştır. Yazılım sektöründeki firmanın veri sayıları, faktör sayıları ve etki değerleri ise daha fazla olarak yer almaktadır.

Kaynaklar

1. Kitchenham,B. ve Linkman,S., "Estimates,Uncertainty,andRisk",Software IEEE,14(3):69-74 (1997)
2. Kwan,T.W. ve Leung,H.K.N., "A Risk Management Methodologyfor Project Risk Dependencies",IEEE TRANSACTIONS ON SOFTWARE ENGINEERING,14(5) (2011)
3. Ray,M. ve Mohapatra,D.P., "Risk analysis: a guidingforce in theimprovement of testing",Software IET,7(1):29-46 (2013)
4. Hu,Y.,Huang,J.,Chen,J. ve Liu,M., "Software Project Risk Management Modelingwith- Neural Network andSupportVector Machine Approaches",Natural Computation, 2007.ICNC,Haikou (2007)
5. Hu,Y.,Zhang,X.,Sun,X. ve Liu,M., "An Intelligent Model for Software Project Risk Prediction",Information Management, Innovation Management andIndustrialEngineering, 2009 International Conference,Xi'an (2009)
6. Islam M.J.,Wu,Q.M.J. ve M.Ahmadi, "InvestigatingthePerformance of Naive- BayesClassifiersand K- NearestNeighborClassifiers",Convergence Information Technology, 2007. International Conference,Gyeongju (2007)
7. Tang,A. ve Wang,R., "Software project risk assessment model based on fuzzytheory",ComputerandCommunication Technologies in AgricultureEngineering (CCTAE), 2010 International Conference,Chengdu (2010)
8. Bragina,T. ve Tabunshcyk,G., "Fuzzy model forthe software projectsdesign risk analysis",CAD Systems in Microelectronics (CADSM), 11th International Conference TheExperience of DesigningandApplication,Polyana-Svalyava (2011)
9. Cingiz, M. Ö., Unudulmaz, A. ve Kalıpsız, O., "Farklı Kestirim Yaklaşımları Kullanılarak Yazılım Projelerindeki Problem Etkilerinin Risk Yönetiminin Gerçekleşmesi", Ulusal Yazılım Mühendisliği Sempozyumu (UYMS), Ankara, 2012
10. Cingiz, M. Ö., Unudulmaz, A. ve Kalıpsız, O., "Prediction of Project Problem Effects on Software Risk Factors", The 12th IEEE International Conference on Intelligent Software Methodologies tools and Techniques, Hungary, 2013