Dicle University Journal of Engineering (DUJE)
web: http://dergipark.gov.tr/dumf

Araştırma Makalesi / Research Article

Uydu verilerinden karar ağaçları kullanarak orman yangını tahmini

Prediction of wildfire using decision trees from satellite data

Nurettin Beşli[1], Mehmet Emin Tenekeci[2*]

[1] Harran Üniversitesi, Elektrik-Elektronik Mühendisliği Bölümü, Şanlıurfa,nbesli@harran.edu.tr
[2] Harran Üniversitesi, Bilgisayar Mühendisliği Bölümü, Şanlıurfa,etenekeci@harran.edu.tr

MAKALE BİLGİLERİ

Makale geçmişi:

Geliş: 20 Aralık 2019
Düzeltme: 3 Mart 2020
Kabul: 5 Mart 2020

Anahtar kelimeler:

Orman yangınları tahmini, Uydu verileri, Karar ağaçları, Veri madenciliği

ÖZET

Tüm canlılık için önemli olan ormanların, yangınlar ile yok edilmesi doğaya verdiği zararın yanı sıra can ve mal güvenliğini de ciddi şekilde tehdit eder. Orman yangınları doğal yollarla veya bilinçli insan davranışları ile gerçekleşir. Orman yangınlarının önceden tahmini veya erken keşfi hızlı müdahale ve önlem almayı sağlayacaktır. Literatürde orman yangınlarını tahmin etmek için meteorolojik verileri ve uzaktan algılama verileri kullanılmaktadır. Bununla birlikte meteorolojik veriler ile mevcut orman yangının davranışı da belirlenebilmektedir. Bu çalışmada uydulardan alınan veriler ile orman yangınlarının tahmini yapılmıştır. Uydudan alınan verilerden hesaplanan Normalize Edilmiş Fark Bitki Örtüsü İndeksi (NVDI), Arazi Yüzeyi Sıcaklığı (LST) ve Termal anomali (TA) verileri kullanılarak orman yangınları tahmin edilmiştir. Bahsedilen verilerden tahmin yapmak için karar ağaçları kullanılmıştır. Karar ağaçlarının eğitiminde kullanılmak için veri setindeki verilerin %70' i kullanılmıştır. Geri kalan %30 veri ile oluşturulan modelin testi gerçekleştirilmiştir. Eğitim ve test işlemi farklı veriler ile 10 defa tekrarlı yapılarak uygulanan yöntemin ortalama performansı belirlenmiştir. Gerçekleştirilen denemelerde ortalama %98,62 duyarlılık oranı ile gerçekleşen yangınlar doğru tahmin edilmiştir. Tüm denemelerde yapılan tahminler için ortalama %93,11 doğruluk ile gerçek durum belirlenmiştir.

Doi: 10.24012/dumf.661925

ARTICLE INFO

Article history:

Received: 20 December 2019
Revised: 3 March 2020
Accepted: 5 March 2020

Keywords:

Wildfire prediction, Satelite data, Desicion tree, Data mining

ABSTRACT

Destruction of forests, which are important for all vitality, with fires, seriously threatens the safety of life and property. Wildfires occur by natural ways and conscious human behavior. The prediction and early detection of wildfires will allow for rapid intervention and prevention. In the literature, meteorological data and remote sensing data were used to predict forest fires. In addition, by using meteorological data, the behavior of existing wildfires could also be determined. In this study, wildfires were predicted from data received from satellites. Wildfires were predicted by using Normalized Differential Vegetation Index (NVDI), Land Surface Temperature (LST) and Thermal Anomaly (TA) data calculated from downloaded satellite data. Decision trees were used to make predictions using mentioned data. 70% of the data in the data set were used for the training of decision trees. The trained model was tested with the remaining 30% data. Training and testing process were repeated 10 times with different data and the average performance of the proposed model was determined. Occurred fires were correctly predicted on the average of 98.62% sensitivity in the experiments. For the predictions in all trials, the actual situation was determined with an average of 93.11% accuracy.

* Sorumlu yazar / Correspondence
Mehmet Emin Tenekeci
✉ etenekeci@harran.edu.tr

Please cite this article in press as N. Besli, M. E: Tenekeci, "Uydu verilerinden karar ağaçları kullanarak orman yangını tahmini", DUJE, vol. 11, no.3, pp. 899-906, September 2020.

Giriş

Ormanlar, dünyada karadaki doğal hayatın devamı ve çeşitliliği açısından en önemli habitattır. Hayvan ve bitki çeşitliliğinin en yoğun yerleşim alanı olan ve canlılığın devamı açısından en önemli madde olan oksijenin üretilmesini sağlayan ormanların korunması oldukça önemlidir. Ancak tüm dünyada olduğu gibi Türkiye'de de ormanların varlığı ciddi tehdit altındadır. Ormanlara en fazla zarar veren ise orman yangınlarıdır. [1] çalışmasında orman yangınlarının başlıca sebepleri genel olarak iki sınıfa ayrılmıştır. Bunlardan ilki yıldırım düşmesi, elektrik hatları, yanardağ patlaması ve aşırı derecede yükselen hava sıcaklıkları gibi doğal olaylardır. İkinci sebep olarak ise bilinçli ve bilinçsiz insan kaynaklı davranışlar olduğu ifade edilmiştir.

Türkiye de bulunduğu Akdeniz iklim kuşağından dolayı çokça orman yangınları ile karşı karşıya kalmaktadır [2]. Orman genel müdürlüğü verilerine göre 2015 yılında 22,3 milyon hektar orman alanı bulunmaktadır. 2017 yılında ise çıkan 2411 orman yangınında 119092 hektar ormanlık alan yanarak yok olmuştur [3].

Doğal sebeplerden çıkan yangınlarının önlenmesi mümkün olmamaktadır. İklim değişikliği sonucu yağmur oranında azalma ve yıllık ortalama sıcaklık değerinde artma ormanlarda meydana gelecek yangın ihtimalini arttırmaktadır. Bu nedenle yangına sebebiyet verecek doğal sebeplerin takibi ve yangın çıkma ihtimali olan alanların belirlenmesi, hızlıca müdahale imkânı sağlayacaktır. Bu şekilde etkilenecek orman alanı azaltılabilecektir. Böylece yetkili kuruluşları harekete geçirme ve gerekli kaynak planlaması ile müdahale ekiplerinin hızlı müdahalesi mümkün kılınacaktır [4].

Yangın meydana gelmeden tahmini kadar önemli bir diğer konu da yangının davranışının belirlenmesidir. Çünkü davranışını belirleme doğru müdahale ile yangının hızlı sonlandırılmasını sağlayacaktır. Ancak yangınların yayılma davranışını belirlemek oldukça zorlu bir işlemdir. Modellemeyi etkileyen birçok parametre bulunmaktadır. Özellikle hızlı değişim gösteren hava sıcaklığı, rüzgâr şiddeti, rüzgâr yönü, hava nemi gibi iklim parametreleri izlenmelidir. Bununla birlikte yangın olan bölgenin topoğrafik yapısı da oldukça önemlidir.

Orman yangınlarının tahmini için gelişen yeni yöntem ve verilerin kullanımı ile daha başarılı sonuçlar elde edilebilmektedir. Özellikle uydulardan alınan uzaktan algılama verilerinin kullanımı yaygınlaşmıştır. Yakın zamana kadar yüksek maliyetli olan bu verilerin elde edilmesi son yıllarda daha kolaylaşmıştır. Ayrıca alınan veriler, yüksek çözünürlüklü olup daha sık kaydedilmektedir. Alınan bu görüntüler yangının oluşumu ve yayılması için gerekli olan yangının dağılımı ve yoğunluğu, bölgeye en yakın su kaynağı, toprağın durumu, bitki örtüsünün durumu gibi ortam hakkında önemli bilgiler taşımaktadır. Meteorolojik verilerin birlikte kullanımı ile yangınlar daha doğru izlenerek müdahale daha etkin yapılabilmektedir. Bununla birlikte IoT teknolojisinin gelişimi ile sürekli ulaşılıp izlenemeyen alanlara yerleştirilen algılayıcılar ile takip edilmesi mümkün olacaktır [5].

Uydulardan alınan veriler, meteoroloji verileri, IoT cihazlarından toplanan veriler farklı özellikte ve boyutta olması modellemenin ve analizlerin bilgisayar ortamında yapılmasını zorunlu kılmıştır. Oluşan büyük veriden anlamlı ve değerli bilgi çıkarılması bilgi teknolojilerinin kullanımını zorunlu hale getirmiştir. Özellikle veri madenciliği ve yapay zekâ algoritmaları son yıllarda verileri işlemede oldukça yoğun kullanılmaktadır. Teknolojinin gelişmesi ile hayatın her alanında özellikle tıbbi çalışmalarda [6], hassas tarım uygulamalarında [7,8], klasik ve sosyal medya [9] alanında oluşan büyük veri yapay zekâ algoritmaları ile analiz edilerek anlamlı bilgiler çıkarılmaktadır.

Uzaktan algılama başlı başına büyük veri üreten bir veri toplama aracıdır. Özellikle farklı algılayıcıların kullanımı ve yüksek çözünürlüklü algılayıcıların sık veri kaydetmesi analiz edilmesi gereken veri miktarını arttırmaktadır. Orman yangınlarının uydu verileri ile tahmin edilmesi için kullanımında veri çeşitliliği de analizi zorlaştırıcı etkenlerdendir [10,11].

Orman yangınlarının tahmini ve modellemesinin, makine öğrenme ve veri madenciliği algoritmaları ile yapıldığı birçok çalışma bulunmaktadır. [12-15] çalışmalarında makine öğrenme yöntemleri ile orman yangınının tahmini yapılmıştır. Padilla ve

arkadaşları insanların sebep olduğu orman yangınlarını tahmin etmek için mantıksal regresyon yöntemi kullanmıştır [12]. Cheng ve arkadaşları ise uzamsal olarak orman yangını tespit etmek için yinelenen yapay sinir ağı kullanmıştır [13]. Satir ve arkadaşları ise yapay sinir ağları kullanarak çalışma yapılan bölgeye ait orman yangın risk haritası çıkarmıştır [14]. Dieu ve arkadaşları SVM, Rasgele orman ve YSA yöntemleri ile orman yangınlarını tespit etmişlerdir [15] ve Arpaci ve arkadaşları ise en yüksek entropi ile yangın davranışını analiz etmiştir [16]. Castelli ve arkadaşları ise genetik programlama ile orman yangınlarını tahmin etmiştir [17]. Arrue ve arkadaşları ise kızılötesi kamera ile %90 doğru tespit oranında yangınları tespit etmiştir [18]. Cortez ve arkadaşları meteorolojik veriler kullanarak destek vektör makinaları ile orman yangınlarını %75 oranında tespit etmiştir [19]. Bu çalışmada yerel meteoroloji verileri gerçek zamanlı olarak toplanmıştır. Bu şekilde uydu verisine göre çok düşük maliyet ile elde edilen veriler ile tahmin işlemi yapılmaya çalışılmıştır. Kanada' da yapılan bu çalışmada beş meteoroloji verisi ile tahmin için Yangın Hava İndeksi (Fire Weather Index - FWI) önerilmiştir. Çalışma sonucu çok yüksek başarılı sonuçlar elde edilmemesine rağmen kullanılan verinin düşük maliyeti sebebi ile oldukça dikkat çekici olmuştur. Sayad ve arkadaşları uydu verileri kullanarak yapay sinir ağları ve destek vektör makinaları ile orman yangınlarını tahmin etmiştir.[20].

Yaptığımız bu çalışmada ise makine öğrenme algoritmalarından Karar Ağaçları yöntemi ile uydu verileri kullanılarak orman yangınlarının tahmini gerçekleştirilmiştir. Çalışmada elde edilecek sonuçlar literatürde kullanılan farklı veri madenciliği algoritmalarının sonuçları ile karşılaştırılarak etkinliği gösterilmiştir.

Çalışmamızda uydulardan alınan uzaktan algılama verilerinden hesaplanmış bazı bilgiler kullanılmıştır. Bunlar; Normalize Edilmiş Fark Bitki Örtüsü İndeksi (Normalized Difference Vegetation Index = NDVI), Arazi Yüzeyi Sıcaklığı (Land Surface Tempreture = LST) ve Termal anomali (Thermal Anomalies = TA) değeridir. Bu değerler ile karar ağaçları kullanılarak orman yangıları tahmin edilmiştir.

Yayının geri kalan kısmında; ikinci bölümde kullanılacak veri seti, üçüncü bölümde uygulanacak yöntemlerden bahsedilecektir. Dördüncü bölümde performans ölçütleri ifade edilirken, beşinci bölümde ise gerçekleştirilen denemeler ve elde edilen sonuçlar verilerek karşılaştırmalar yapılacaktır. Son bölümde çalışmanın katkıları ve daha sonra yapılabilecek çalışmalar hakkında bilgi verilecektir.

Veri Seti

Çalışmamızda, orman yangınlarının tahmini için kullanılacak veriler Kanada'nın ormanlık alanlarında 2013 – 2014 yılları arasında meydana gelmiş olan orman yangınları [21] ve uydulardan alınacak bilgilerden oluşmaktadır. Belirtilen tarih aralığındaki uydu verilerine göre eğitim ve tahmin işlemi gerçekleştirilmiştir. Kullanılan veri seti [20] çalışmasında kullanılarak Mendeley veri seti arşivinde erişime açılmıştır. Çalışmada bu veri seti alınarak kullanılmıştır.

Veriler Terra ve Aqua uydularında bulunan MODIS (Moderate Resolution Imaging Spectroradiometer) algılayıcıdan alınmıştır. MODIS kayıtları NASA arşivlerinde erişilebilir durumdadır [22]. Tahminde kullanılacak indeks ve veriler MODIS tarafından kaydedilmektedir. Çalışmamızda tahmin için kullanılacak olan veriler, NDVI MOD13Q1, LST MOD11A1 ve TA MOD14A1 kodu ile kaydedilen verilerden hesaplanmaktadır.

NDVI indeksi, bitki örtüsünün rengi ve canlılığı hakkında bilgi taşımaktadır. LST ise belirlenen bölgede toprak sıcaklığını belirlemek için kullanılmaktadır. Son olarak TA değeri ise sıcaklık anomalisi olarak ifade edilmek ve yangın göstergesi olarak işleme alınmaktadır. NDVI ve LST parametreleri yangın esnasında toprak ve bitki örtüsünde meydana gelecek farklılaşmayı belirlemek için seçilmiştir. TA parametresi ise direk yangın ile ilgili bilgi taşımakta ve ilk iki parametrede elde edilen sonucu kuvvetlendirmek için seçilmiştir.

NDVI bitkilerin foto sentetik aktiviteleri hakkında bilgi taşımaktadır. Ormandaki bitki örtüsünün ne kadar yeşil olduğunu belirtmektedir. Bu şekilde bölgede canlı bitki miktarını belirlemek mümkün olur. Bitkiler güneşten aldıkları elektromanyetik enerjinin 0,63 μm – 0,69 μm dalga boyu aralığındaki kırmızı ışığı absorve ederek fotosentezde kullanır. Buna karşılık 0,7 μm ve üzeri dalga boylarını yansıtırlar. Bu şekilde gözle yeşil olarak görülürler. NDVI değerinin hesaplanması için yansımalardan kırmızı ve yakın kızılötesi

bantları kullanılarak Eşitlik 1'e göre hesaplama yapılır. Bu değerler Terra uydusunda bulunan MODIS algılayıcıdan MOD13Q1 kodlu referans verisi kullanılarak elde edilmiştir.

$$NDVI = (NIR - RED)/(NIR + RED) \qquad (1)$$

NDVI değeri -1 ile 1 arasında hesaplanır. -1 değeri bitki örtüsü olmayan alanlarda hesaplanırken 1 değeri çok yoğun bitki örtüsünü ifade eder. Hesaplama için 0,76 – 0,90 aralığındaki yakın kızıl ötesi bant (NIR) ile 0,63 – 0,69 arası kırmızı bant (RED) kullanılır. Bitki örtüsünün canlılık durumu yana bilirliğini etkiler. Bu nedenle NDVI değeri yangın çıkma ihtimali konusunda bilgi taşımaktadır.

LST, toprağın ışınım sıcaklığını temsil eder. MODIS algılayıcıdan MOD11A1 verisi kullanılarak elde edilir. Gerçek LST verisi direk olarak MODIS'ten alınamamaktadır. HDF-EOS formatında belirli katsayılar ile birlikte alınan SDS değerinden hesaplamalar ile belirlenir. LST, Eşitlik 2 de verildiği şekli ile dönüşüm yapılarak Celcius olarak hesaplanmaktadır. SDS, algılayıcıdan 7500 ile 65535 aralığında alınmaktadır. Hesaplama için kullanılan k ölçeklendirme katsayısı ise 0,02 alınmaktadır. Bu katsayı HDF-EOS formatı şeklinde gelen veriden alınmaktadır. Elde edilen değer Kelvin değeri olduğu için 273 çıkarılarak Celcius' a dönüştürülmüştür. Hesaplanan LST, bitki örtüsü ve çıplak toprak sıcaklıklarının birleşimidir.

$$LST = (SDS * k) - 273 \qquad (2)$$

TA değeri, yüzeyden yansıyan güneş ışığı ile yüzey sıcaklığı toplamının zamanla değişimi bilgisini içermektedir. Diğer özellikler gibi MODIS' den alınan MOD14A1 verisi ile elde edilmektedir. 1 ile 10 arasında bir değer almaktadır. Elde edilen değerin yüksek olması ısı farkının yüksek olduğunu belirtmektedir. Çalışmamızda kullanılan veri setinde toplam 804 örnek bulunmaktadır. Örneklerin 386 tanesinde yangın bilgisi olmakta, 418 tanesinde ise yangın bilgisi bulunmamaktadır. Veri setinde bulunan öznitelikleri ve özellikleri Tablo 1 de verilmiştir.

Tahmin için karar ağaçlarının kullanımı

Genel olarak sınıflandırma algoritmalarından olan Karar ağaçları son yıllarda veri madenciliği alanında da yoğun olarak kullanılmaya başlanmıştır. Yaygın kullanımının en önemli sebebi kapalı kutu algoritmalara göre daha anlaşılır kurallar dizisi ile sınıflandırma ve bilgi çıkarma işlemi gerçekleştirmesidir [23]. Sınıflandırmayı aşamalı karar verme dizisi haline getirmektedir

Tablo 1. Veri setinin öznitelikleri
Table 1. Dataset features

Öznitelik	Üst Sınırlar	Veri Setindeki Sınırlar	Standart Sapma
NDVI	-1 – 1	0,030 – 0,781	0,1015
LST	-123 – 1037,7	-10,41 – 37,89	8,0743
TA	1 – 10	3 – 9	0,6567
Durum	0 / 1	0 / 1	0,5000

Karar ağaçlarında, ağaç yapısı kök düğümü, iç düğümler (dallar) ve uç düğümlerden (yapraklar) oluşur. Ağaç yapısı, eğitim verilerine göre oluşturulmaktadır. Eğitim örneklerine ait öznitelikleri ağacın düğümlerinin yerleşimini ve eşik değer katsayılarını belirlemek için kullanılmaktadır. Oluşturulan karar ağacı öznitelik değerleri ve eğitim sürecinde belirlenen eşik değerlerine göre düğümlerde hareket ederek uç düğüme ulaşır. Bu şekilde örnek için sınıflandırma yapılmaktadır. Düğümlerden ayrılan dalların en yüksek farklılığa sahip hale getirilmesi veya geçiş yapılan düğümlerin benzerliğinin en düşük hale getirilmesi için en uygun eşik değerleri belirlenmelidir [23]. Her bir giriş örneği için uç düğüme ulaştıran tek bir yol bulunacak ve sonuçta karar bu yola göre verilecektir. Yapılacak uygulamaya göre tek veya çok bantlı karar verme yapılabilir [24]. Tek değişken ve ikiden fazla sınıf olursa aynı öznitelik farklı seviyelerde sorgulanmaktadır. Bu şekilde ikiden farklı alt küme veya sınıf oluşturulmaktadır.

Oluşturulacak ağaç yapısında düğümlerin hangi öznitelikler ile oluşturulacağı ve seçilecek eşik değerinin belirlenmesi için de alternatif yöntemler bulunmaktadır. Bunlardan en fazla kullanılanları bilgi kazancı, Gini Farklılık indeksi, Towing kuralı ve Ki-Kare olasılık istatistiğidir. Çalışmamızda karar ağacının oluşturulması ve eğitimi için Gini Farklılık İndeksi kullanılmıştır. Karar ağacı eşitlik 3 ve 4 te verildiği şekilde hesaplanan Gini Kazancına

göre dallar arası farklılığı en büyük yapacak şekilde oluşturulmuştur [20].

$$Gini\ Kazancı(T,A) = Gini(T) - P(T_l) * Gini(T_l) - P(T_r) * Gini(T_r) \quad (3)$$

$$Gini(T) = \sum p_i * (1 - p_i) = 1 - \sum_{j=1}^{J} {p_j}^2 \quad (4)$$

p_i ağaçta i. sınıfın bulunma olasılığını ifade ederken J değeri toplam sınıf sayısını belirtmektedir. $P(T_r)$ ve $P(T_l)$ ise düğümün sağ ve sol yanındaki dalların ağırlığını belirtmektedir.

Karar ağaçlarının uydu görüntülerinin sınıflandırılmasında kullanıldığı bazı çalışmalar literatürde bulunmaktadır [24 - 26]. Uzaktan algılamada kullanılacak veriler uydulardan alınan farklı bantlardaki görüntülerdir. Bu görüntülerden elde edilen indeks değerleri uydu görüntülerinin sınıflandırmasında öznitelikler olarak kullanılmıştır.

Çalışmamızda NDVI indeksi, LST değeri ve TA değeri parametre olarak kullanılmıştır. Eğitim örneklerinin değerlerine göre karar ağacı oluşturulacaktır. Oluşturulan karar ağacı eğitimde kullanılmayan örnekler ile test edilecektir.

Performans Ölçütleri

Yapılan çalışma sonucunda oluşturulacak modelin doğruluğunun belirlenmesi gerekmektedir. Bu amaç için veri setinin bir kısmı eğitim için kullanılmış ve geri kalan örnekler eğitime katılmadan modelin doğruluğunu test etmek için kullanılmıştır. Yapılan denemelerde çalışma performansı belirlemek için tahminlerin dört durumu belirlenmiştir. Bunlar;

- Doğru Pozitif (True Positive - TP): Doğru tahmin edilen yangın olmuş örnek
- Yanlış Pozitif (Fale Positive - FP): Yanlış olarak yangın olduğu tahmin edilen ancak yangın olmayan örnek
- Doğru Negatif (True Negative - TN): Doğru tahmin edilen yangın olmayan örnek
- Yanlış Negatif (False Negative - FN): Yanlış olarak yangın olmadı tahmin edilen ancak yangın olan örnek

Bu değerlerin değerlendirilmesi için Tablo 2' de gösterilen karşıtlık matrisi kullanılacaktır. Bu matris ile Eşitlik 5-9' da verilen performans ölçütleri hesaplanmıştır.

$$Duyarlılık = \frac{TP}{TP+FN} \quad (5)$$

$$Özgüllük = \frac{TN}{FP+TN} \quad (6)$$

$$Doğruluk = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

$$Hassasiyet = \frac{TP}{TP+FP} \quad (9)$$

Duyarlılık değeri oluşturulan modelin yangın olan alanları doğru tahmin performansını belirlemektedir. Özgüllük değeri ise yangın olmayan alanların doğru tahmin oranını vermektedir. Bunların birleşimi olarak da doğruluk değeri hesaplanmaktadır. Hassasiyet değeri ise yangın alanlarının tahmin edildiği durumdaki doğruluk değerini ifade etmektedir.

Tablo 2. Karşıtlık matrisi
Table 2. Confusion matrix

		Tahmin Edilen Değerler	
		Yangın Olan Sınıf	Yangın Olmayan Sınıf
Gerçek Değerler	Yangın Olan Sınıf	TP	FN
	Yangın Olmayan Sınıf	FP	TN

Uygulama sonuçları ve tartışma

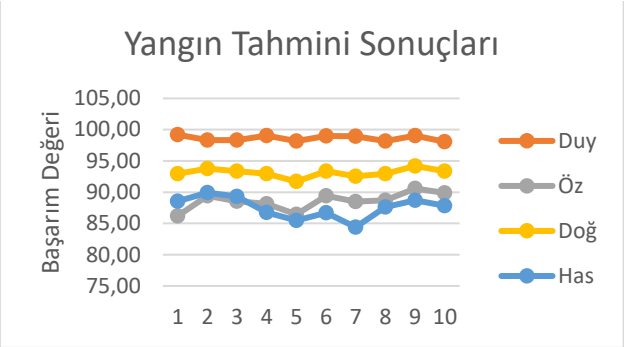
Bu çalışmada orman yangınlarının tahmini, uzaktan algılama verileri kullanılarak karar ağaçları ile gerçekleştirilmiştir. Orman yangının tahmini için NDVI indeksi, LST değeri ve TA değeri kullanılarak karar ağacı modellenmiştir.

Veri seti içerisinde bulunan 804 örneğin %70' i olan 563 tanesi rasgele seçilerek eğitime tabi tutulmuştur. Eğitilen karar ağacı geriye kalan 241 test örneği tahmin etmesi için kullanılmıştır. Yapılan uygulamada oluşturulan karar ağacının yapısını seçilecek eğitim örnekleri

belirleyecektir. Buna bağlı olarak test örnekleri tahmin edilecektir.

Denemelerde veri seti harmanlanarak 10 defa rastgele seçilen veriler ile eğitim ve test işlemi tekrar edilmiştir. Gerçekleştirilen 10 farklı deneme sonuçları ve sonuçlara göre hesaplanan performans kriterleri Tablo 3' de verilmiştir. Tablo 3'de ilk denemede 125 adet yangın ve 116 adet yangın olmayan örnek test için seçilmiştir. Bu örnekler rastgele seçilmiştir. Yapılan denemede 124 yangın olayı belirlenmiş, ancak 1 yangın olayı doğru tahmin edilememiştir. Benzer olarak 116 yangın olmayan örnekten 100 tanesi doğru tahmin edilirken 16 tanesi hatalı olarak yangın olduğu tahmin edilmiştir.

Elde edilen sonuçların farklılığından da anlaşıldığı eğitim ve test için seçilen verilerin farklılaşması sonuçları etkilemektedir. Duyarlılık değeri yangın olan örneklerin doğru bir şekilde tahminini ifade ettiğinden en değerli performans ölçütüdür. Elde edilen sonuçlarda duyarlılık %99 değerine kadar çıkmıştır. Ancak en yüksek değerin elde edildiği denemede test için seçilen örneklerde yangın olanların fazlalığı dikkat çekmektedir. Bu nedenle dikkate alacağımız tüm denemelerin ortalaması olacaktır. 10 deneme sonucunda ise ortalama %98,62 ile gayet tatmin edici değerler alınmıştır. Bu denemelerin sonuçları grafiksel olarak Şekil 1 gösterilmiştir.



Şekil 1. 10 denemede elde edilen performans kriterlerinin değişimi

Veri madenciliği ve makine öğrenme çalışmalarında aranan bilgilinin doğru tahmini kadar yanlış tahmin oranının da düşük olması istenir. Bu şekilde modelin performansı doğru olarak belirlenmiş olur. Bu iki bilgiyi birlikte veren ölçüt Doğruluk değeri olacaktır. Alınan sonuçlar incelendiğinde en başarılı sonuç %94,19 ile 9. denemede elde edilmiştir. Bununla birlikte tüm denemelerin ortalaması olarak %93.11 ile model doğru tahmin gerçekleştirilmiştir.

Çalışmamızda karar ağaçları kullanarak orman yangınlarının tespitinde elde edilen sonuçlar Tablo 4'de literatürdeki bazı çalışmalarda elde edilen sonuçlar ile birlikte verilmiştir. Diğer çalışmalarda sonuç olarak duyarlılık değeri verildiğinden karar ağaçları ile elde edilen duyarlılık değeri verilmiştir.

Tablo 3. Karar ağaçları ile orman yangını belirleme için 10 farklı deneme sonuçları
Table 3. 10 different trial results for determination of forest fire with decision trees

Deneme No	**TP**	**FN**	**FP**	**TN**	**Duyarlılık**	**Özgüllük**	**Doğruluk**	**Hassasiyet**
1	124	1	16	100	**99,20**	86,21	92,95	88,57
2	116	2	13	110	98,31	89,43	93,78	**89,92**
3	117	2	14	108	98,32	88,52	93,36	89,31
4	105	1	16	119	99,06	88,15	92,95	86,78
5	106	2	18	115	98,15	86,47	91,70	85,48
6	98	1	15	127	98,99	89,44	93,36	86,73
7	92	1	17	131	98,92	88,51	92,53	84,40
8	106	2	15	118	98,15	88,72	92,95	87,60
9	102	1	13	125	99,03	**90,58**	**94,19**	88,70
10	101	2	14	124	98,06	89,86	93,36	87,83
Ortalama					**98,62**	**88,59**	**93,11**	**87,53**
En Yüksek					99,20	90,58	94,19	89,92
En Düşük					98,06	86,21	91,70	84,40
Standart Sapma					0,46	1,39	0,68	1,72

YSA ve SVM yöntemleri [20] çalışmasında birlikte denenerek önerilmiştir. Çalışmamızda [20] çalışmasında kullanılan veri seti kullanılmış ve daha başarılı sonuç alınmıştır. Bu çalışma orman yangınlarını tahmin için uzaktan algılama verilerini kullanmıştır. Bu çalışmada kullanılan veri setinin aynısı kullanılmıştır. [27] çalışmasında önerilen ISTIFF ise geçmişteki orman yangınları bilgisinden dinamik geri beslemeleri sinir ağı ile zaman serisi modellemesi yapmıştır. Oluşturduğu model ile gelecekteki orman yangınlarını tahmin etmiştir. [28] çalışmasında kullanılan STIFF modeli [27] çalışması gibi önceden gerçekleşen orman yangınlarının konumlarına göre zaman serisi modellemiştir. Ek olarak veri setinin YSA modelini oluşturmuştur. Her iki modelden elde edilen sonuçları kullanarak regresyon ile tahminlerini gerçekleştirmiştir.

Elde edilen sonuçlar çalışmamızda uygulanan karar ağaçları yönteminin diğer çalışmalardan daha başarılı olduğunu göstermiştir. Deneme ve test veri setlerinin rasgele seçilmesinden dolayı yöntemin genel başarısını ortaya koymak için 10 deneme gerçekleştirilerek ortalama sonuç verilmiştir.

Ayrıca çalışmamızda kullanılan veri seti [20] ile aynı veri seti olduğu için karşılaştırma bu çalışma ile olması daha doğru olacaktır. Çalışmamızda kullanılan karar ağaçları ile elde edilen %98,62 doğruluk oranı [20] çalışmasından daha yüksek doğruluk ile yangınları tahmin etmiştir. [20] çalışmasında diğer performans parametreleri verilmemesinden dolayı karşılaştırma yapılamamıştır. Ancak duyarlılık değeri diğer parametrelerin hesaplanmasından etkili olduğundan çok yüksek değişiklik göstermeyeceği varsayılabilir.

Sonuç ve Öneriler

Yapılan çalışmada Kanada için alınmış uzaktan algılama verilerinden orman yangınlarının tahmini gerçekleştirilmiştir. Gayet başarılı ve cesaret verici sonuçlar elde edilmiştir. Benzer yapılan çalışmalara göre daha başarılı sonuçlara ulaşıldığı görülmüştür. Bu çalışmada uzaktan algılama bilgilerinin benzer uygulamalar için kullanılabileceği anlaşılmıştır. Ayrıca tahmin başarısını arttırmak için farklı uygulamalarda kullanılan meteorolojik veriler ve eski orman yangını kayıtları da kullanılabilir. Bu şekilde daha fazla parametre ile daha yüksek başarı oranına ulaşılabilir. Ayrıca uygulanılan yöntem farklı veri madenciliği yöntemleri ile birlikte hibrid bir şekilde kullanılarak ta daha başarılı tahminler gerçekleştirilebilir.

Tablo 4. Farklı çalışmalarda elde edilen sonuçlar
Table 4. Results obtained from different studies

	Yöntem	Duyarlılık (%)
[20]	YSA	98,32
[20]	SVM	97,48
[27]	ISTFF	95
[28]	STIFF	65
	Karar Ağaçları	98,62

Kaynaklar

[1] Dayananda P. W. A., "Stochastic models for forest fires", Ecological Modeling, Volume 3 (1977), 309-313.

[2] Altan, G., Türkeş, M., Tatlı, H., "Çanakkale ve Muğla 2009 yılı orman yangınlarının Keetch-Byram Kuraklık İndisi ile klimatolojik ve meteorolojik analizi." In: 5th Atmospheric Science Symposium Proceedings Book: 263-274. Istanbul Technical University, 27-29 April 2011, Istanbul. Turkey.

[3] Türkiye Cumhuriyeti Orman genel müdürlüğü, https://www.ogm.gov.tr/ekutuphane/Sayfalar/Istatistikler.aspx (Erişim 10/01/2019)

[4] C.S. Eastaugh, H. Hasenauer, "Deriving forest fire ignition risk with biogeochemical process modelling Environ." Model. Softw., 55 (2014), pp. 132-142

[5] Tedim F., Leone V., Amraoui M., Bouillon C., Coughlan M., Delogu G., Fernandes P., Ferreira C., McCaffrey S., McGee T., Parente J., Paton D., Pereira M., Ribeiro L., Viegas D., Xanthopoulos G., "Defining extreme wildfire events: difficulties, challenges, and impacts Fire" 1 (1) (2018), p. 9

[6] Gümüşçü A., Tenekeci M. E., "Estimation of active sperm count in spermiogram using motion detection methods", Journal of the Faculty of Engineering and Architecture of Gazi University (2018), https://doi.or./10.17341/gazimmfd.460524

[7] Karadağ K., Tenekeci M.E., Taşaltın R., Bilgili A., "Detection of pepper fusarium disease using

machine learning algorithms based on spectral reflectance", Sustainable Computing: Informatics and Systems, (2019), Article in Press.

[8] Gümüşçü A., Tenekeci M.E., Bilgili A.V. "Estimation of wheat planting date using machine learning algorithms based on available climate data", (2019) Sustainable Computing: Informatics and Systems. Article in Press.

[9] Lejdel B., "Conceptual Framework for Analyzing Knowledge in Social Big Data.", Proceeding Big Data and Smart Digital Environment. ICBDSDE (2018). Studies in Big Data, vol 53, pp 347-358.

[10] Chi M., Plaza A., Benediktsson J.A., Sun Z., Shen J., Zhu Y., "Big data for remote sensing: challenges and opportunities", Proc. IEEE, 104 (11) (2016), pp. 2207-2219

[11] Ramapriyan H., Brennan J., Walter J., Behnke J., "Managing big Data: NASA tackles complex Data challenges", Earth Imaging J. (2013) [Online].

[12] Padilla M., Vega-García C. (2011) On the comparative importance of fire danger rating indices and their integration with spatial and temporal variables for predicting daily human-caused fire occurrences in Spain. International Journal of Wildland Fire 20, 46-58.

[13] Cheng T., Wang J., "Integrated spatio-temporal data mining for forest fire prediction", Trans. GIS, 12 (5) (2008), pp. 591-611

[14] Satir O., Berberoglu S., Donmez C., "Mapping regional forest fire probability using artificial neural network model in a Mediterranean forest ecosystem Geomatics", Natural Hazards and Risk (2015), pp. 1-14

[15] Dieu Tien Bui, Quang-Thanh Bui, Quoc-Phi Nguyen, Biswajeet Pradhan, Haleh Nampak, Phan Trong Trinh,"A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area",Agricultural and Forest Meteorology, Volume 233, (2017), Pages 32-44,ISSN 0168-1923.

[16] Arpaci A., Malowerschnig B., Sass O., Vacik H., "Using multi variate data mining techniques for estimating fire susceptibility of Tyrolean", forests Appl. Geogr., 53 (2014), pp. 258-270

[17] Castelli M., Vanneschi L., Popovic A., "Predicting burned areas of forest fires: an artificial intelligence approach", Fire Ecol., 11 (2015), pp. 106-118

[18] Arrue B., Ollero A., Matinez de Dios J., "An intelligent system for false alarm reduction in infrared forest-fire detection", IEEE Intell. Syst., 15 (3) (2000), pp. 64-73

[19] Cortez P., Morais A., "A data mining approach to predict wildfires using meteorological data", Proc. 13th Port. Conf. Artif. Intell (2007), pp. 512-523

[20] Sayad, Y. O., Mousannif, H., Moatassime, H. A., "Predictive modeling of wildfires: A new dataset and machine learning approach", Fire Safety Journal, Volume 104, (2019), Pages 130-146

[21] The Canadian Wildland Fire Information System (CWFIS) http://cwfis.cfs.nrcan.gc.ca/.

[22] MODIS data products, Courtesy of the NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, https://lpdaac.usgs.gov.

[23] Quinlan J.R., "C4.5: Programs for Machine Learning", Morgan Kaufmann, San Mateo, CA, (1993),302 s

[24] Zhang J., Gruenwald L., Gertz M., VDM-RS: A visual data mining system for exploring and classifying remotely sensed images,Computers & Geosciences, Volume 35, Issue 9, (2009),Pages 1827-1836, ISSN 0098-3004, https://doi.org/10.1016/j.cageo.2009.02.006.

[25] Hansen, M. C., Stehman, S. V., Potapov, P. V., Arunarwati, B., Stolle, F., & Pittman, K. "Quantifying changes in the rates of forest clearing in Indonesia from 1990 to 2005 using remotely sensed data sets." Environmental Research Letters 4.3 (2009): 034001.

[26] Pal M., Mather P.M., "An assessment of the effectiveness of decision tree methods for land cover classification", Remote Sensing of Environment, 86, (2003), 554-565

[27] Li Z., Dunham M.H., Xiao Y., Zaïane O.R., Simoff S.J., Djeraba C. (Eds.), "STIFF: a forecasting framework for SpatioTemporal data", Mining Multimedia and Complex Data. PAKDD 2002. Lecture Notes in Computer Science, vol. 2797, Springer, Berlin, Heidelberg (2003)

[28] Cheng T., Wang J., "Integrated spatiotemporal Data Mining for forest fire prediction", Trans. GIS, 12 (5) (2008), pp. 591-611