

# Orman Yangını Boyutunun Makine Öğrenmesi ile Tahmini

Mert KAYA  
TOBB University of Economics and  
Technology  
Department of Computer  
Engineering  
Ankara,06560,Turkey  
mertkaya@etu.edu.tr

## Abstract—

Orman yangınları doğal yaşamı ve ekolojik dengeyi tehdit eden en önemli etkenlerden biridir. Orman yangınlarının önceden tahmini yangınlara erken müdahale ve sonuçta oluşacak yıkımın azaltılması için çok önemlidir. Bu çalışmada UCI Makine Öğrenme Kütüphanesindeki Orman Yangını Dataset’ini kullanarak regresyon ve sınıflandırma üzerine çalışılmıştır. Sonuç olarak en iyi 5-fold CV accuracy sonuç SVM ile datasetteki Ay,Haftanın Günü,Sıcaklık,Bağıl Nem ve Rüzgar Hızı kullanıldığında ulaşılmıştır ve %61.32 dir.

**Keywords :** Orman Yangını Tahmini, Orman Yangını sınıflandırma, Multilayer Perceptron, Support Vector Machine

Fires Danger Rating System-CFFSDRS ve National Fire Danger Rating System-NFDRS verilebilir. Bu çalışmada oluşturulmaya çalışılan makine öğrenmesi modeli benzeri predictor’lar üzerine de çalışmalar olmuştur.

Bu çalışmalardan bazılarını örnek olarak Cortez ve Morais’in regresyon,karar ağacı, yapay sinir ağları ve support vektör makineleri kullanarak yaptığı çalışma %46 başarı vermiştir[4].

## III. ABOUT THE DATA

Orman yangınlarına etki edebilecek bir çok farklı etken gösterilebilir. Kullandığımız datada bulunan featurelar :

## I. INTRODUCTION

Orman yangınları dünya genelince kontrolsüz bir biçimde yayılabilmekte ve doğal yaşamı yoketmektedir. Her yıl ortalama 64,100 yangında toplam 6.8 milyon hektar orman yanmaktadır.[1] Bu yangınlar sadece doğal hayata etkileriyle değil ekonomiye de etkileri ile insan yaşamını her yönden tehdit etmektedir. Orman yangınları her yıl ortalama 91 milyar dolar maliyet oluşturmaktadır.[2]

Orman yangınları doğal yaşamın bir gerçeği olarak aşırı yoğun ormanlarda kendiliğinden oluşmakta ve ormanın kendiliğinden seyrekleşmesi sonucu doğal yaşamın devamlılığını sağlamaktadır. Bu çalışmada amaçlanan kontrolden çıkabilecek doğal sebepli yangınların ve insan kaynaklı yangınların erken tespit edilerek söndürülmesi için gerekli olan bilgileri sağlayacak bir model üretilmesidir.

Yangınlarda erken tespitin önemi yüksek olması ile beraber bunu sağlayabilecek cihazların maliyeti ve metotların kullanım sıklığının düşük olması günümüz problemlerindendir.

## II. LITERATURE REVIEW

Orman yangını tahmin sistemleri gerçek zamanlı veriye bağımlı bir şekilde çalışmaktadır. Bu veriyi sağlamak için ve karar destek için oluşturulmuş bir çok uygulama bulunmaktadır. Örnek olarak European Forest Fire Information Systems (EFFIS) [3], Canadian Forest

1. X : Ormanı 9x9 luk bir matrix olarak düşündüğümüzde (x,y) gösterimindeki x değeri
2. Y: Ormanı 9x9 luk bir matrix olarak düşündüğümüzde (x,y) gösterimindeki y değeri
3. Month: 12 aydan hangisinde verinin kaydedildiği bilgisi.
4. Day: Haftanın hangi günü olduğu bilgisi.
5. FFMCI: FWI[5] değerlerinden biridir. FWI değerleri kanada devleti tarafından orman yangınlarının tehlikesini sıralamak için oluşturulmuştur. FFMCI değeri orman çöp yakıtlarının orman gölgesi altındaki nemini açıklar. 0 – 101 arasında değişir.
6. DC: FWI değeridir. Toprak altındaki kuruluğu anlatır. 0 – 1000 arası değer alır.
7. ISI: FWI değeridir. Orman yangını yayılma potansiyelini açıklamak için kullanılır.
8. Temp: Celcius sıcaklık değeridir.
9. RH: Yüzdelik bağıl nemdir.
10. Wind: Rüzgar hızıdır .(km/h)
11. Rain: Yağmur miktarıdır.(mm/m<sup>2</sup>)
12. Area: Hektar bakımından yanmış olan orman miktarıdır.

Aşağıda Figure 1 da gösterildiği üzere elimizdeki verinin çok büyük bir kısmı 0 lardan oluşmaktadır.

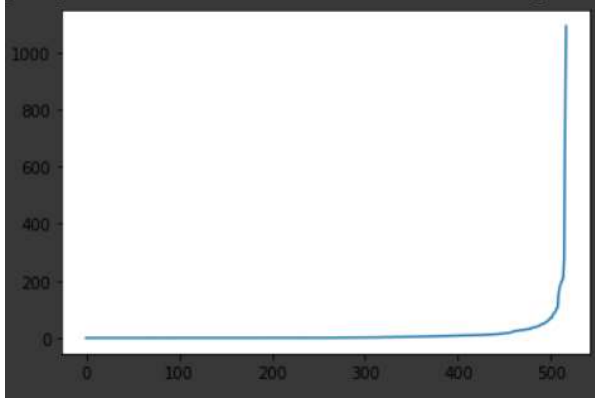


Figure 1

Target kısmı olarak kullanılacak “area” feature’ı ile alakalı daha fazla bilgi aşağıdaki Figure 2 kısmında görülebilir.

```
df['area'].describe()
```

count	517.000000
mean	12.847292
std	63.655818
min	0.000000
25%	0.000000
50%	0.520000
75%	6.570000
max	1090.840000

Figure 2

Yukarıda gösterilen Figure 2 den de anlaşılaacağı üzere 517 sample’ın ortalama yanan hektar miktarı 12.84 hektar olarak görülmüştür ve çok büyük bir kısmının 0 değeri olduğu Figure 1 den anlaşılabilmektedir.

Veri üzerine bu incelemeler yapıldıktan sonra Pearson Correlation grafiğini incelenmiştir. Bu grafik tüm bu çalışmanın kodları ile birlikte colab linkinden görülebilir.

[https://colab.research.google.com/drive/1UegC-LTkQ8j3SR\\_Bxg1G\\_pohe3XOrIOu?usp=sharing](https://colab.research.google.com/drive/1UegC-LTkQ8j3SR_Bxg1G_pohe3XOrIOu?usp=sharing)

Bu grafikten anlaşılması gerekenler :

Person correlation[6] :

- 0 a yakın değerler arada correlation olmadığını yada düşük bir correlation olduğunu açıklar.
- 1 e yakın değerler güçlü pozitif correlation gösterir.
- -1 e yakın değerler güçlü negative correlation gösterir.

#### IV. REGRESSION

Data üzerinde bir yanan hektar miktarının sayısal tahmini için linear regresyon üzerine uğraşmıştır. Regresyon yapılırken sıcaklık,nem,rüzgar,yağmur feature’ları kullanılmıştır alınan sonuçlar aşağıdaki gibidir :

Datanın %80 i train , %20 si test data olarak kullanıldığında,

- Least Squares Linear Regression :  
Mean absolute error : 25.8849
- Ridge Linear Regression (Learning rate = 0.01) :  
Mean Absolute Error: 25.8810
- Lasso Linear Regression (alpha = 0.1 , max iteration = 1000):  
Mean Absolute Error: 25.8830
- KNN Regression :  
Mean Absolute Error: 26.9093

Residual Plot :

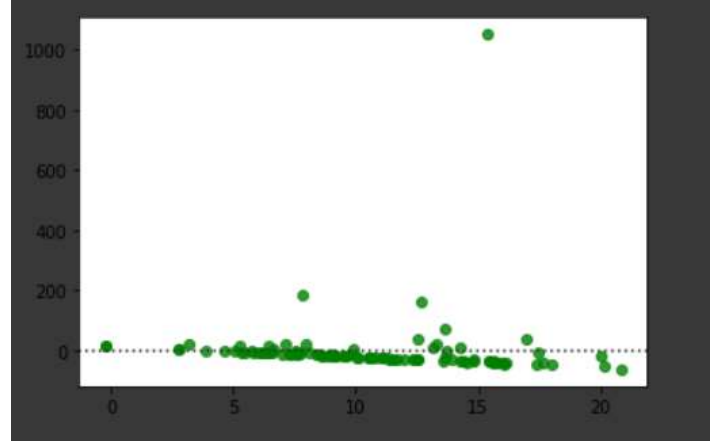


Figure 3

Residual Plot datanın lineer’liğini incelemekte önemli bilgiler vermektedir.Figure 4 te Residual Plot görülebilir.

Çalışma kodlarının paylaşımı ve kullanıcıların farklı değerlerin sonuçlarını görebilmesi için kullanıcı arayüzü oluşturulmuştur. Bu arayüzlerde colab linkinden ulaşıp , modellerin farklı değerlerdeki tahminleri denenebilir.

## V. CLASSIFICATION

Problem regresyon ile düzgün bir şekilde çözülememiştir bunun sebebi data'nın 0 a çok yakınsamakta olmasından kaynaklanıyor olabilir.

Problemi bir sınıflandırma problemine çevirmek için öncelikle kategorik değerlerin numerik değerlere dönüştürülmesi gerekmekte olduğu görülmüştür. Kategorik değerlerin numerik değerlere çevirilmesi için onehotencoding kullanılabilir. Kategorik-numerik dönüşüm yapılan kolonlar ay ve gün kolonları olduğundan bunlar arası ilişkiyi koruma amaçlı numerik dönüşüm elle yapılmıştır. :

- 1) Dayofweek feature'ına dönüşüm yapılırken hafta içi ve hafta sonu yangın ihtimalindeki değişimin incelenmesine karar verilmiş ve hafta içi:1 , haftasonu:2 olarak maplenmiştir.
- 2) Ay değerleri ise mevsimsel etkilerin incelenmesi için mevsimlere göre maplenmiş ve kuzey yarım küredeki ay-mevsim ilişkisi ile tasarlanmıştır. Kış ayları 1,Sonbahar ayları 2,Yaz ayları 3,İlkbahar ayları 4 olarak maplenmiştir.

- Data Scaling :

Elimizdeki verinin çok büyük bir kısmının 0 olmasının sonucu olarak target değerimiz olan “yanmış hektar miktarı” bölümünün scale edilmesi gerekmekte idi.

Bunun sağlanabilmesi için verinin devamlı artan ve sample sıraları değişmeyecek şekilde manipüle edilmiştir.

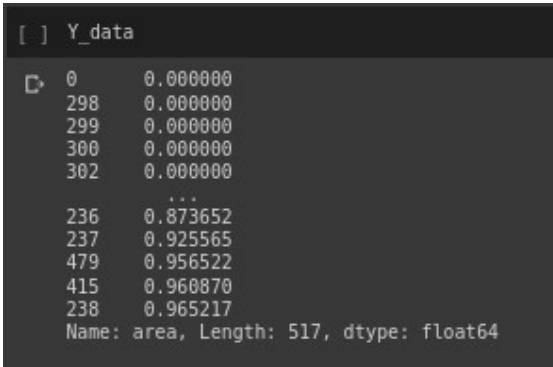


Figure 4

Yanan hektar'a göre sıralanıp scale edilen datanın görünümü Figure 5 tedir. 0-1 arasına sıkıştırılmıştır. Bu scaling'in sonucu olarak daha iyi dağılan data train edilmeye daha uygun hale gelmiştir.

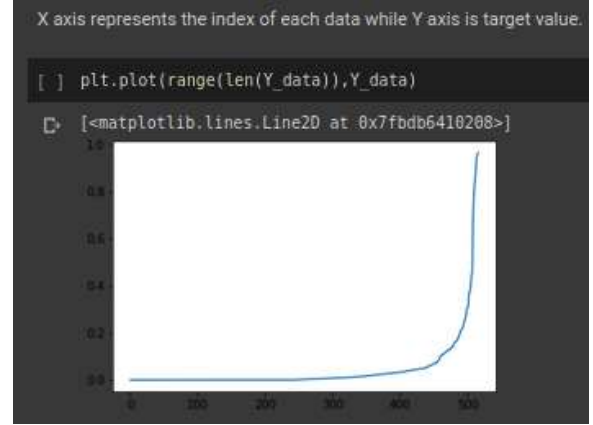


Figure 5

Figure 1 ve Figure 5 incelendiğinde verinin 0 a yatkınlığında azalma sağlandığı farkedilebilir.

- Kmeans Clustering – Fuzzy Values :

Datanın sınıflandırma(classification) problemine çevirilmesi için target değerlerimizde numerik değerlerden sınıflara dönüştürülmesi gerekmektedir.

Bu dönüşüm için target (Yanmış hektar miktarı) değerleri kmeans clustering farklı değerler ile oluşturulmuş ve logistic regresyon ile en yüksek accuracy veren k=5 değeri kullanılmıştır.

Farklı Kmeans K değerleri için Logistic Regresyon : 5 Fold Cross Validation Score'ları :

Kmeans K = 5 , 5Fold CV = 0.6944

Kmeans K = 5 , 5Fold CV = 0.6944

Kmeans K = 5 , 5Fold CV = 0.6944

Kmeans k değeri 5 seçilerek, target değişkeni kolonu kullanılarak uygulandığında elde edilen cluster centroid değerleri fuzzy values için bize yol göstermiştir.

Fuzzy değerler dönüşümü bu cluster centroid'ler kullanıldığında bize Figure 6 daki gibi bir grafik vermektedir.

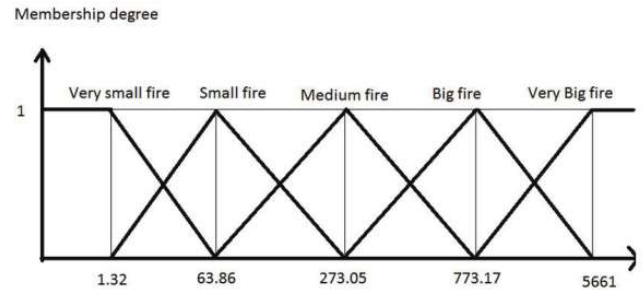


Figure 6

Numerik deęerler bu fuzzy deęerler dnüşümü uygulandıęında örnek olarak řu řekilde sınıflandırılabilir:

Elimizdeki target deęerine val dedięimizde :

$\alpha$  = en yakın alt sınır

$\xi$  = en yakın üst sınır

$X = (val - \alpha) / (\xi - \alpha)$

$Y = (\xi - val) / (\xi - \alpha)$

X ve Y deęerlerinden büyük olanı X ise en yakın alt sınır, Y ise en yakın üst sınır sınıfı deęer olarak atanır.

Bu dnüşüm uygulandıktan sonra Kmeans clusterlar kullanarak oluřturulmuř deęerlere gre ok kk orman yangını deęerlerinde azalma grlmüştür ve train edilmeye daha uygun bir veri elde edilmiştir.

Kmeans deęerlerinden fuzzy deęerlere geildięinde deęişim řu řekilde açıklanabilir:

ok kk orman yangını sayısı 368 den 317 ye dřmř.

Kk orman yangını sayısı 92 den 123 e ıkkmış.

Orta byklkte orman yangını sayısı 36 dan 46 ya ıkkmış.

Byk orman yangını sayısı 12 den 20 ye ıkkmış.

ok byk orman yangını sayısı 9 dan 11 e ıkkmıştır.

Bu dnüşmler sonucu elde edilen data 3 farklı řekilde paralanmış ve bu paralardan farklı modeller eęitilmiştir.Bu paralar:

1. Sadece Ay, Haftanın Gn, Sıcaklık, Nem ve Rzgar feature'ları
2. X,Y ve Ay, Haftanın Gn, Sıcaklık, Nem ve Rzgar feature'ları
3. FWI deęerleri ,X, Y ve Ay, Haftanın Gn, Sıcaklık, Nem ve Rzgar feature'ları

Bu data'lar ařaęıda listelenen yntemler ile eęitilmiştir sonuç kısmında daha detaylı açıklama grlebilir.

- KNN Classification
- Logistic Regression Classification
- Support Vector Machine Classification
- Multilayer Perceptron
- Gaussian Process Classification

Yukarıda bahsedilen her bir data parasının bu yntemlerle eęitilmiş hallerinin denenebileceęi kullanıcı arayz colab linkinde bulunabilir.

Logistic Regresyon ve Multilayer Perceptron farklı local minima lara fit edebilir.Accuracy'nin iyi bir řekilde anlaşılabilmesi iin bu modeller bir ok kez train edilmiş ve sonuları sonuç kısmında açıklanmıştır.

## VI. RESULTS

- Sadece Ay,Haftanın Gn,Sıcaklık,Nem ve Rzgar feature'ları kullanılarak eęitilen modellerin 5Fold CV Accuracy deęerleri :

KNN Classification :	0.50
SVM Classification (lineer kernel):	0.6131
SVM Classification (poly kernel):	0.6131
Multilayer Perceptron (Learning rate = adaptive , maximum iteration = 20000 , solver = stochastic gradient descent) (50 kere fit edilme sonrası ortalama accuracy ) :	0.61008
Gaussian Process Classification :	0.4816
Logistic Regresyon (stochastic average gradient) (50 kere fit edilme sonrası ortalama accuracy ) :	0.6256

Figure 7

- X,Y ve Ay,Haftanın Gn, Sıcaklık, Nem ve Rzgar feature'ları kullanılarak eęitilen modellerin 5Fold CV Accuracy deęerleri :

KNN Classification :	0.4894
SVM Classification (poly kernel):	0.6131
Multilayer Perceptron (Learning rate = adaptive , maximum iteration = 20000 , solver = stochastic gradient descent) (50 kere fit edilme sonrası ortalama accuracy ) :	0.61002
Logistic Regresyon (stochastic average gradient) (50 kere fit edilme sonrası ortalama accuracy ) :	0.6112

Figure 8

- X,Y ve Ay,Haftanın Günü, Sıcaklık,

Nem ve Rüzgar feature'ları kullanılarak eğitilen modellerin 5Fold CV Accuracy değerleri :

KNN Classification :	0.4523
SVM Classification (poly kernel):	0.6131
Multilayer Perceptron (Learning rate = adaptive , maximum iteration = 20000 , solver = stochastic gradient descent) (50 kere fit edilme sonrası ortalama accuracy ) :	0.5800
Logistic Regresyon (stochastic average gradient) (50 kere fit edilme sonrası ortalama accuracy ) :	0.6073

Figure 9

Figure 7, Figure 8 ve Figure 9 daki verilen incelendiğinde Logistic Regresyon ve SVM'in en yüksek accuracy score'ları vermiştir. Ulaşılan en yüksek accuracy %61.31 bulunmuştur.

#### ACKNOWLEDGMENT

Orman yangınları gibi çok az sıklıkla olan olayların verisi incelendiğinde scale etmenin modelleri eğitmek için çok anlaşılmıştır.

Orman yangını boyutu tahminindeki önemli feature'lardan biri ormanın boyutudur. Elimizdeki veri bu bilgiyi sağlamadığından özelliklerin farklı etkilerini incelemek ve bu farklı özelliklere göre tahmin etmek amaçlanmıştır.

Eğitilen modeller kullanıcı arayüzü oluşturularak paylaşılmış ve yangınların önceden tahmininde kullanılabilir hale getirilmiştir.

#### REFERENCES

- [1] Congress, US. "Wildfire Statistics." <https://fas.org/sgp/crs/misc/IF10244.Pdf>, [fas.org/sgp/crs/misc/IF10244.pdf](https://fas.org/sgp/crs/misc/IF10244.pdf).
- [2] Bartz, Kelsey. "Record Wildfires Push 2018 Disaster Costs to \$91 Billion." *Center for Climate and Energy Solutions*, 12 July 2019, [www.c2es.org/2019/02/record-wildfires-push-2018-disaster-costs-to-91-billion/](http://www.c2es.org/2019/02/record-wildfires-push-2018-disaster-costs-to-91-billion/).
- [3] Commission, European. "Copernicus." EFFIS, 2020, [effis.jrc.ec.europa.eu/applications/seasonal-forecast/](http://effis.jrc.ec.europa.eu/applications/seasonal-forecast/).
- [4] P. Cortez, A. Morais, A data mining approach to predict forest fires using meteorological data, In Proceedings of the 13th Portuguese Conference on Artificial Intelligence, 512-523, (2007).
- [5] National Wildfire Coordinating Group, Canadian. "Fire Weather Index (FWI) System." *NWCG*, 2019, [www.nwcg.gov/publications/pms437/cffdrs/fire-weather-index-system](http://www.nwcg.gov/publications/pms437/cffdrs/fire-weather-index-system).