



# Sınıflama Algoritmalarının Yağışın Varlığını Kestirme Konusundaki Performanslarının Karşılaştırması

Hakan Koçak<sup>1\*</sup> 

<sup>1</sup> Meteoroloji Genel Müdürlüğü, Ankara

hkocak@gmail.com

## Öz

Yağış tahmini başta hava tahmincileri, tarım ve ziraatla uğraşanlar olmak üzere tüm herkesi ilgilendiren önemli bir konudur. Son yıllarda büyük ivme yakalayan yapay zeka ve makine öğrenmesi uygulamaları diğer birçok alanda olduğu gibi yağış tahmininde de tatbik edilmekle beraber yüksek doğruluklu yağış kestirimi yapmak hala zorlu bir görev olarak karşımızda durmaktadır. Son yıllarda etiklerini daha fazla hissettiğimiz iklim değişikliği nedeniyle oluşan yağış rejimindeki değişiklikler bu zorlu görevi daha da zorlu hale getirmektedir.

Bu çalışmada veri seti üzerinde 5 farklı kategoriden 10 adet sınıflayıcı algoritma uygulanarak elde edilen performanslar karşılaştırılmıştır. Araştırmada ayrıca ana veri setinden bazı parametreler çıkarılarak farklı senaryolar oluşturulmuş, her bir senaryo için sınıflama algoritmaları uygulanarak performanslarındaki değişimler gözlemlenmiştir. Araştırma sonucunda tüm senaryolar göz önüne alındığında Fonksiyonlar kategorisi dört senaryodan üçünde en başarılı kategori olmuş ve en iyi performansa sahip sınıflayıcının da bu kategoriden MLP (Çok Katmanlı Yapay Sinir Ağı) sınıflayıcısı olduğu ortaya çıkmıştır. Araştırmada ayrıca oluşturulan senaryolar için en yüksek ortalama doğruluk oranlarının %83,4 ile %84,8 arasında değiştiği görülmüştür. Bu durum, veri setinden bazı parametrelerin çıkarılmasının sonuca büyük oranda etki etmediğini göstermektedir. Elde edilen bu sonuçlar, makine öğrenmesi yöntemlerinin yağışın varlığının kestirimi konusunda iyi derecede performans sağladığını ve bu amaçla kullanılabileceğini göstermiştir.

**Anahtar kelimeler:** Yağış Kestirimi, Sınıflandırma Algoritmaları, Weka, Yapay Zeka, Makine Öğrenmesi, Performans Karşılaştırması

## Performance Comparison of Classification Algorithms in Rainfall Prediction

### Abstract

Predicting rainfall is important issue that concerns everyone especially weather forecasters, farmers and those who work in agriculture sector. Although artificial intelligence and machine learning applications, which have gained great momentum in recent years, are applied in precipitation forecasting, as in many other areas, it is still a challenging task to make high-accuracy rainfall prediction. Changes in the precipitation regime due to climate change, the effects of which we have felt more and more in recent years, make this difficult task even more challenging.

In this study, the performances obtained by applying 10 classifier algorithms from 5 different categories on the data set were compared. In addition to that, different scenarios were created by removing some of the parameters from the original data set and the performance differences of the classification algorithms for each of the scenarios were noted. The results have shown that the Functions category was the most successful category in 3 of the 4 scenarios and MLP (Multi Layer Perceptron) algorithm which belongs to that category was the most successful classifier with the rate of 84.4%. Also, highest accuracy rates were between 83.4% and 84.8% considering all four scenarios. This shows that removing some of the parameters from the original parameter set does not have a significant impact on the classification accuracy. The study results have shown that machine learning techniques achieved good performance in predicting rainfall and could be used for that purpose.

**Keywords:** Rainfall Prediction, Classification Algorithms, Weka, Artificial Intelligence, Machine learning, Performance Comparison

\* Sorumlu yazar.  
E-posta adresi: hkocak@gmail.com

Alındı : 06 Ağustos 2021  
Revizyon : 01 Ekim 2021  
Kabul : 27 Ekim 2021

## 1. Giriş (Introduction)

Su, tüm canlıların hayatında vazgeçilmez bir öneme sahiptir. Su aynı zamanda tarım ve zirai faaliyetler için de şüphesiz çok önemli bir yere sahiptir. Bu nedenle su yeryüzünde en önemli doğal kaynaklardan biridir. Son yıllarda etkilerini daha fazla gördüğümüz küresel iklim değişikliği nedeniyle yağışlarda düzensizlik meydana gelmiştir. Bunun neticesinde ise yağışlarda genel olarak azalma, kuraklık ve bazen de kısa süreli aşırı yağış nedeni ile seller ve su taşkınları meydana gelmektedir. Tüm bunlar, değişen yağış rejiminin örüntüsünü daha yakından analiz etme ve beklenmeyen aşırı yağış nedeniyle meydana gelebilecek doğal afetleri önleme noktasında yağış tahmini konusunu önemli bir hale getirmiştir. Bu nedenle günümüzün bilgi teknolojilerindeki gelişmelere paralel olarak yağış tahmini konusu özellikle makine öğrenmesi ve yapay zeka disiplinlerinin ve alanlarda çalışan araştırmacıların da ilgisini çeken bir araştırma konusu haline gelmiştir.

Geçmiş yüzyıllardan günümüze doğal ve teknik bilimlerin değişim ve gelişimi deneysel çalışmalar ve doğal olayların modellenmesi ile yakından ilişkilidir. Bu değişim ve gelişim neticesinde ise doğaya dair bilginin temel kaynağının sadece ölçümlerin olabileceği ve aynı zamanda doğal olayların özelliklerinin de ölçülen gözlemler arasındaki ilişkiler ile açıklanabildiği temellerine dayanan fiziksel keşif kavramının temelleri ortaya çıkmıştır (Soucek, 1992). Doğa olaylarının fiziksel tasvirinin en önemli özelliği; eğer iki değişken arasındaki ilişki biliniyorsa bilinmeyen bir özelliğin veya değişkenin ilişkili olduğu diğer özellik veya değişen aracılığı ile nicel olarak kestirilebilmesidir. Değişkenler arasındaki ilişkilerin matematiksel soyut modeller olarak veya fizik kanunları şeklinde ifade edilebiliyor olması çözümlenmeli bir araştırma için en uygun olanıdır, fakat uygulama için mekanik veya elektriksel gibi diğer bazı modeller bu amaç için daha uygun olabilir (Grabec, 1990). Bilgi işleme ve makine öğrenmesi konusundaki son yıllardaki gelişmeler, bilimsel gelişim ve evrimin geleceğinin, doğal olayların modellerini kendileri geliştirip değiştiren zeki sistemlerin geliştirilmesi yönünde olacağı görüşünü ve beklentisini artırmıştır (Dibike ve Solomatine, 2001).

Yağış tahmini en zorlu görevlerden biridir. Bu konuda önerilen birçok algoritma olmasına rağmen yağışı yüksek doğrulukla tahmin etmek hala oldukça zordur. Ziraatın önemli önemli olduğu ülkelerde her yıl üretilen tahıl miktarı ve su kıtlığı her zaman büyük bir ilgi ile takip edilmektedir. Yağış miktarındaki mevsimsel ufak dalgalanmalar bile ziraat sektörü üzerinde büyük ve yıkıcı etkilere sebep olabilmektedir. Yağışın doğru tahmininin ayrıca doğal afetlerin neden olduğu can kayıplarını ve maddi zararları önleme konusunda potansiyel faydaları vardır. Taşkın ve kuraklık gibi bazı özel durumlarda yüksek doğruluklu yağış tahmini zirai yönetim ve doğal felaketleri önleme açısından oldukça önemlidir (Shoba, 2014).

Yağış tahmini aynı zamanda sel baskınlarının önlenmesi ve ayrıca su kaynaklarının yönetimi konusunda yardımcı olur. Yağışın zamanlaması ve miktarı zirai hasat konusunda belirleyici unsur olabilmektedir. Bu nedenle yağışa dair önceden bilgi sahibi olunması hem ziraatla uğraşanlara hem de karar verici konumundakilere zirai hasarın azaltılması konusunda yardımcı olabilir.

Bu çalışmada tüm canlılar için büyük öneme sahip yağışın varlığı farklı makine öğrenmesi metotları kullanılarak tahmin edilmeye çalışılmıştır. Ayrıca araştırmada kullanılan makine öğrenmesi metotlarının yağış tahminindeki başarı oranlarının çeşitli metrikler kullanılarak karşılaştırılması da araştırmanın bir diğer odak noktasını oluşturmaktadır.

## 2. Makine Öğrenmesi Teknikleri ile Yağış Tahmine Yönelik Önceki Çalışmalar (Previous Study)

Makine öğrenmesi teknikleri kullanarak yağış tahmini yapmaya yönelik olarak farklı yerli ve yabancı araştırmacılar tarafından yapılmış birçok araştırma bulunmakla birlikte bu araştırmaların farklı makine öğrenmesi metotlarının karşılaştırılmasına yönelik olmaktan ziyade az ve sabit sayıda girdi parametre seti üzerinde belirli bir veya birkaç makine öğrenmesi tekniğinin (Ör. yapay sinir ağları veya ağaç yapılı sınıflayıcılar) uygulandığı araştırmalar olduğu görülmektedir. Yapılan bu araştırmada ise çok sayıda (10) makine öğrenmesi metodu yağış tahmininin etki edeceği düşünülen yeterli sayıda meteorolojik parametre setine uygulanarak tahmin performansları karşılaştırılmış ve ayrıca parametre setinden bazı parametreler çıkarılarak oluşturulan farklı senaryolar için algoritmaların başarımlarındaki farklar gözlemlenmiştir. Aşağıda makine öğrenmesi metotları kullanılarak yağışın tahmin edilmesine yönelik bazı araştırmalar ve sonuçları hakkında özet olarak değinilmiştir.

Uzunali tarafından 2019 yılında yapılan bir araştırmada Kandilli Rasathanesi ve Deprem Araştırma Enstitüsü Müdürlüğü'nün, Kandilli Bölgesine ait Ocak 1918 ile Aralık 2018 yılları arasındaki 100 yıllık veri arşivi kullanılarak, önceki yıllara dayalı olarak, son yılların ortalama yağış değerlerinin tahmini yapılmıştır. Araştırmada yöntem olarak Yapay Sinir Ağı yöntemlerinden ANFIS (Adaptive Neuro-Fuzzy Interface System, Uyarlamalı Bulanık Ağ Çıkarım Sistemi) modeli kullanılmıştır. Araştırma sonucunda zaman serisi biçimindeki geçmiş yıllara ait yağış verilerini tek başına anlamlandırabilmek ne kadar zor olsa da veriyi yapay zekâ yöntemlerini uygulayarak kullandığımızda, bu verinin anlamlandırılmasında başarılı sonuçlar alınabileceği gözlemlenmiştir.

Xiao ve Chandrasekar tarafından 1997 yılında yapılan çalışmada ise YSA kullanılarak radar gözlemlerinden yağış tahmini yapılmıştır. Bu amaçla geliştirilen YSA modelinde girdi verisi olarak radar

gözlemleri; çıktı olarak ise yağış ölçer (rain gage) ölçümleri kullanılmıştır. Araştırmada yağış tahmini için YSA tekniğinin yanı sıra reflektivite-yağış miktarı (Z-R ilişkisi) ve reflektivite, diferansiyel reflektivite ve spesifik yayılım (propagation) gibi çoklu parametrelili ölçüm denklemleri teknikleri ile de aynı verilerle yağış tahmini yapılmıştır. Sonuçlar karşılaştırıldığında YSA kullanılarak elde edilen yağış tahminlerinin diğer tekniklerle elde edilen yağış tahminlerinden daha iyi olduğu görülmüştür.

Rani ve Govardhan, çok Katmanlı Algılayıcı Yapay Sinir Ağı (MLPNN) kullanarak yağış tahmini çalışması yapmışlar ve elde ettikleri sonuçları ARIMA tekniği kullanılarak bulunan sonuçlar ile karşılaştırmışlardır. Araştırmacılar sonuç olarak geri beslemeli algoritmasının en iyi sonucu verdiğini belirtmişlerdir.

Kannan, Prabhakaran ve Ramachandran 5 yıllık yağış verisi için Pearson katsayısı hesaplamışlar ve regresyon yaklaşımı ile yağış tahmini kestirimi yapmışlardır. Yağış tahmini kestirimi için çoklu doğrusal regresyon yönetimini kullanmışlardır. Araştırma sonucunda gerçek yağış değerleri hesaplanan yağış değerlerinden daha düşük çıkmıştır. Sonuçlara göre elde edilen modelin doğruluğu çok yüksek olmamakla birlikte kestirimi yapılan değerlere yakın değerler elde edilmiştir.

Soo-yeon, Sharad, Byunggu ve Dong (2012) yağış tahmini için CART ve C4.5 karar ağacı tekniklerini önermişlerdir. Yaptıkları çalışmada yağışı tahmin etmek için öncelikle yağış olasılığı belirlemişler; eğer yağış olasılığı var ise yağış tahmini yapmışlardır. Bu amaçla çalışmalarında rüzgar hızı, rüzgar yönü, ani rüzgar, nem, sıcaklık, buharlaşma, güneş ışınımı, rüzgar soğutma endeksi, çiy noktası, basınç irtifası, bulut tabanı, hava yoğunluğu, doymuş buhar basıncından oluşan 13 değişken kullanmışlardır. Önerilen model herhangi bir bölgedeki yağış olasılığını ve saatlik yağış miktarını zaman açısından etkili bir şekilde tahmin edebilen yararlı bir model olmuştur. Yağış olasılığını tahminde CART algoritmasının %99,2 ve C4.5 algoritmasının da %99,3 oranında başarılı olduğu görülmüştür. Saatlik yağış miktarı tahmininde ise CART ve C4.5 algoritmalarının doğruluk oranları sırasıyla %92,8 ve %93,4 olmuştur.

Sangari ve Balamurgan yağış tahmini konusunda K-En Yakın Komşu (KNN), Naive Bayes, Karar Ağaçları, Yapay Sinir Ağları ve Bulanık Mantık gibi farklı veri madenciliği tekniklerini karşılaştırmışlardır. Araştırmacıların kullandıkları farklı makine öğrenmesi tekniklerinin yağış tahmini konusundaki performanslarını da karşılaştırdıkları çalışma sonucunda yapay sinir ağlarının %85,77 oranı ile en iyi tahmin doğruluğuna sahip olduğu ortaya konmuştur.

Suhartono, Dwi, Bambang, Sutikno ve Heri Endozya'nın Pujon ve Wagus bölgeleri için yaptıkları çalışmada aylık yağış tahmini için ANFIS (Adaptive Neuro Fuzzy Inference System) ve ARIMA (Autoregressive Integrated Moving Average)

metotlarına dayanan bir grup (ensemble) metodu önermişlerdir. Çalışmada ARIMA, ANFIS ve bunların bileşkesinden oluşan grup (ensemble) metodlarının doğruluklarını karşılaştırmak amacıyla iki ampirik yağış veri seti kullanılmış ve sonuç olarak ANFIS ve ARIMA metodlarının tek başına grup (ensemble) metodundan daha başarılı sonuç verdiği görülmüştür. ANFIS metodu çalışma alanı olarak belirlenen iki alandan biri için daha başarılı sonuç verirken ARIMA metodu diğer alan için daha başarılı tahmin başarısı göstermiştir. 1975-2010 yılları arasına ait verilerin kullanıldığı çalışmada sonuç olarak ANFIS veya ARIMA metodlarının tek başına kullanılmasının daha başarılı sonuçlar verdiği ve karmaşık metotlar olan bileşke (ensemble) modellerin ise her zaman basit modellerden daha başarılı sonuç vermeyeceği ortaya konmuştur.

### 3. Araştırma Alanı ve Veriler (Study Area and Data)

#### 3.1. Otomatik meteorolojik gözlem istasyonu (Automatic meteorological observation station)

Araştırmada kullanılan yağış verileri Meteoroloji Genel Müdürlüğü (MGM) 17130 no'lu yağış gözlem istasyonuna ait veriler olup MGM veri tabanından temin edilmiştir. Bu yağış istasyonu, Kalaba Mahallesi Haliç Sokak Keçiören/Ankara'da bulunan Meteoroloji 9. Bölge Müdürlüğü yanında ölçüm yapmaktadır. 39°58'21.0" kuzey enleminde ve 32°51'50.0" doğu boylamında yer alan istasyonun bulunduğu rakım 883m ve rasat türü mm'dir.



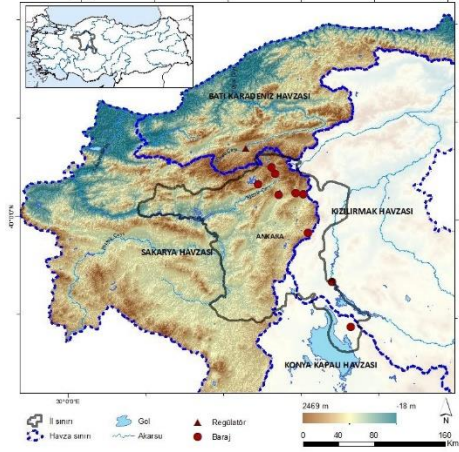
**Şekil 1.** Verilerin elde edildiği otomatik meteorolojik gözlem istasyonunun konumu (The location of meteorological observation station where the data were obtained from)

#### 3.2. Çalışma bölgesi bilgileri (Study area)

Bu çalışmaya konu olan Ankara ili, Türkiye nüfusunun yaklaşık %7 sine ev sahipliği yapmaktadır. Köppen (1968) iklim sınıflamasına göre BSk iklim tipine sahip Ankara, yıllık ortalama sıcaklığı 18,0°C'den düşük, yaz kuraklığının bulunduğu yarı kurak bir iklime sahiptir (Akman, 1990). Ankara ili, yenilenebilir su potansiyelleri birbirlerinden farklı olan su havzaları üzerinde yayılım göstermektedir. Artan sanayileşme,

kentleşme ve tarımsal aktivite ile birlikte küresel iklim değişikliği Ankara ili su kaynakları üzerindeki antropojen baskıyı arttırmaktadır (Kale, 2020).

Ankara ili idari sınırları, alansal olarak %70,6 oranında Sakarya Havzası, %20,8 oranında Kızılırmak Havzası, %8,4 oranında Konya Kapalı Havzası ve %0,2 oranında Batı Karadeniz Havzası sınırları içerisinde yer almaktadır (Şekil 2). Söz konusu havzalar sırası ile Türkiye yenilenebilir su potansiyelinin %3.44, %3.48, %2.43 ve %5.34'üne ev sahipliği yapmaktadır (Öziş, Barant, Durnabaşı ve Özdemir, 1997; Kale, 2020).



Şekil 2. Ankara ili su havzası haritası (Ankara province water basin map)

### 3.3. Veriler(Data)

Çalışmada kullanılan veriler, Meteoroloji 9. Bölge Müdürlüğü'ne ait 17130 no'lu Otomatik Meteorolojik Gözlem İstasyonuna (OMGİ) ait 2010-2020 yılları arası günlük olarak kaydedilen ortalama aktüel basınç, ortalama sıcaklık, ortalama nispi nem ve toplam yağış parametrelerine ait 3992 adet gözlem verisinden oluşmaktadır. Veri setinde yer alan meteorolojik gözlem parametreleri basınç, sıcaklık, rüzgar hızı, rüzgar yönü, iki gün öncesine ait nem, bir gün öncesine ait nem, mevcut güne ait nem, iki gün öncesine ait yağış bilgisi (var/yok), bir gün öncesine ait yağış bilgisi (var/yok), mevcut güne ait yağış bilgisi (var/yok) şeklinde seçilmiş ve Weka aracında işlenmek üzere bu araca özgü ARFF (Attribute-Relation File Format) dosya biçimi şeklinde düzenlenmiştir. ARFF dosya biçimine dönüştürme işleminde Python dilinde kısa bir program yazılarak Meteoroloji Genel Müdürlüğü arşivinden her bir parametre (sıcaklık, basınç, nem, yağış, vs.) için ayrı ayrı metin dosyası şeklinde elde edilen veriler bu program yardımıyla çeşitli işlemlerden geçirilmiş ve sadece gerekli bilgiler bu dosyalardan çekilerek tek bir ARFF dosyasında birleştirilmiştir. Bu parametrelerden basınç hPa cinsinden, sıcaklık °C cinsinden, rüzgar hızı m/sn cinsinden, rüzgar yönü olası tüm 16 yön isim kısaltmalarından (E, W, N, S, NW, NE, SW, SE, ENE, ESE, WNW, WSW, NNW, NNE, SSW, SSE) ve son olarak nispi nem % cinsinden MGM veri tabanında yer almaktadır ve bu birimler cinsinden veri tabanından

çekilmiştir. Araştırmada kullanılan veriler bu araştırma için MGM arşivinden elde edilmiş veriler olup daha önce başka herhangi bir araştırmada kullanılmamıştır.

Veriler, sınıflayıcılara girdi olarak kullanılmadan önce veri ön işleme işleminden geçirilerek eksik veriler ayıklanmış ve geriye 3992 adet girdiden oluşan çalışma veri seti kalmıştır. Veri ayıklama işleminden sonra nümerik tipte olan basınç, sıcaklık, nem, rüzgar hızı verilerine verilerini 0-1 arasında değerlere dönüştüren min-max normalizasyon işlemi uygulanmıştır. Bu amaçla Weka aracında "Preprocess" sekmesinde yer alan Filtrelerden Unsupervised -> attribute başlığı altında yer alan "Normalize" filtresi bahsi geçen bu nümerik veriler üzerinde uygulanarak veriler normalize edilmiştir.

## 4. Ampirik Çalışma (Empirical Study)

Araştırmada ilgili analizleri yapmak için açık kaynak kodlu bir makine öğrenmesi yazılımı olan ve 1997 yılında Yeni Zelanda'daki University of Waikato tarafından geliştirilen WEKA (v3.8.5) yazılımı kullanılmıştır. WEKA, makine öğrenmesi konusunda en yaygın kullanılan yazılım araçlarından biridir. Başlangıçta C dili kullanılarak geliştirilen WEKA daha sonra Java dili kullanılarak yeniden yazılmıştır. WEKA, makine öğrenmesi problemlerinin çözümüne yönelik pek çok farklı makine öğrenmesi algoritması içerir. WEKA ayrıca veri ön-işleme, regresyon, sınıflama, kümeleme, birliktelik kuralları, nitelik değerlendirme ve görüntüleme amacıyla kullanılabilir pek çok araca da sahiptir. Bununla beraber bu çalışmada Weka yazılımının sahip olduğu sınıflayıcı algoritmalarından sadece bazıları seçilerek kullanılmıştır.

Araştırmada kullanılan tüm makine öğrenmesi teknikleri performans açısından kolay ve doğru kıyaslanabilmeleri için sadece Weka aracı kullanılarak uygulanmışlardır. Weka aracılığı ile çalışma veri seti üzerinde araştırma için seçilen sınıflandırma algoritmaları uygulanma aşamasında öncelikle Weka "Classify" sekmesinde yer alan "Test options" kısmından veri setinin varsayılan oran olan %66 öğrenme (training) veri seti ve kalan %34'lük kısmının da test verisi olarak ayrılması ve sınıflandırmanın bu iki veri seti üzerinde uygulanması sağlanmıştır. Bu işlem, tüm 10 sınıflayıcı algoritma için algoritmalar uygulanmadan önce ayrı ayrı yapılmıştır.

Araştırmada mevcut girdi parametreleri olan basınç, sıcaklık, rüzgar şiddeti, nem ve yağış parametrelerinin birbirleri ile olan ilişkilerini kaybetmeyecek şekilde bu parametrelere ilave olarak 1 ve 2 gün önceki nem ve yağış verileri de girdi parametreleri setine eklenmiştir. Girdi parametreleri arasındaki korelasyon ilişkisi Tablo 1.'de verilmiştir.

Araştırmada sınıflandırma algoritmalarına girdi olarak verilen Weka ARFF dosyasındaki parametre isimleri ve birimleri şu sırasıyla şekildedir:

1. Basınç (hPa)
2. Sıcaklık (°C)

3. Rüzgar şiddeti (m/sn)
4. Rüzgar Yönü (E, W, N, S, NW, NE, SW, SE, ENE, ESE, WNW, WSW, NNW, NNE, SSW, SSE)
5. Nem-2 (2 gün önceki nem değeri (%))
6. Nem-1 (1 gün önceki nem değeri (%))
7. Nem (Mevcut güne ait nem değeri (%))
8. Yağış-2 (2 gün öncesine yağış bilgisi (var/yok))
9. Yağış-1 (1 gün öncesine yağış bilgisi (var/yok))
10. Yağış (Kestirimi yapılacak çıktı parametresi (var/yok))

**Tablo 1.** Parametreler arası korelasyon ilişkisi (Correlations among parameters)

	B	S	R.Ş	N-2	N-1	N	Y-2	Y-1	Y
B	1	-0,33	-0,15	0,09	0,05	0,08	-0,26	-0,26	-0,13
S	-0,33	1	0,27	-0,73	-0,69	-0,66	-0,11	-0,06	-0,05
R.Ş	-0,15	0,27	1	-0,25	-0,24	-0,25	-0,04	-0,03	-0,04
N-2	0,09	-0,73	-0,25	1	0,85	0,73	0,37	0,25	0,12
N-1	0,05	-0,69	-0,24	0,85	1	0,85	0,25	0,37	0,25
N	0,08	-0,66	-0,25	0,73	0,85	1	0,19	0,25	0,37
Y-2	-0,26	-0,11	-0,04	0,37	0,25	0,19	1	0,16	0,09
Y-1	-0,26	-0,06	-0,03	0,25	0,37	0,25	0,16	1	0,16
Y	-0,13	-0,05	-0,04	0,12	0,25	0,37	0,09	0,16	1

Tabloda 1.'de B basınç, S sıcaklık, R.Ş rüzgar şiddeti, N-2 iki gün önceki nem, N-1 bir gün önceki nem, Y-2 iki gün önceki yağış, Y-1 bir önceki gün yağış ve Y bulunulan günün yağış parametresini ifade etmektedir. Tablo 1. incelendiğinde nem parametrelerinin kendi aralarında en yüksek korelasyon değerlerine sahip olduğu görülmektedir. Tüm tablo içerisinde ise mevcut güne ait nem (N) ile bir önceki güne ait nem değerinin en yüksek korelasyon değerine (0,85) sahip olduğu görülmektedir. Yine aynı tablodan sıcaklık (S) ile nem parametreleri (N, N-1, N-2) arasında yüksek sayılabilecek oranda aynı yönlü bir ilişki olduğu görülmektedir. Bunun yanında basınç (B) parametresi ile yağış parametrelerinin (Y, Y-1, Y-2) diğer parametrelerle korelasyon ilişkisinin genel olarak düşük olduğu Tablo 1.'den görülebilir.

#### 4.1. Sınıflandırma algoritmaları (Classification algorithms)

Yağış varlığı kestirimi için Weka aracında yer alan 5 adet ana sınıflayıcı kategorilerinin her birinden en az bir tane olmak üzere toplam 10 adet sınıflayıcı seçilerek çalışma veri seti üzerinde uygulanmıştır. Bu sınıflayıcı algoritmalar ve buldukları kategoriler (parantez içinde) şunlardır:

- Bayes (BayesNet, NaiveBayes)
- Fonksiyonlar (MultiLayerPeceptron, SMO)
- Lazy (IBk)
- Kurallar (DecisionTable, JRip, OneR)
- Ağaç (J48, NBTree)

Aşağıda alt başlıklar halinde bu sınıflayıcılara dair özet bilgiler verilmiştir.

##### 4.1.1. BayesNet

BayesNet, Bayes ağlarını özniteliklerin nominal olması ya da eğer nümerik iseler de önceden ayrıştırıldıkları (prediscretized) ve ayrıca eksik veri olmaması (eğer eksik veri varsa da bunlar global olarak değiştirilir) ön kabulü ile öğrenen bir algoritmadır. Ağın koşullu olasılık tablosunu tahminde iki farklı kısım bulunmaktadır. Bu çalışmada BayesNet SimpleEstimator kullanarak; K2 arama algoritması ise ADTree algoritması kullanılmadan çalıştırılmıştır (John ve Langley, 1995). Araştırmada BayesNet Algoritması şu Weka parametreleri ile uygulanmıştır: estimator=SimpleEstimator-A 0.5; searchAlgorithm=K2 -P 1 -S BAYES; useADTree=False

##### 4.1.2. NaiveBayes

NaiveBayes sınıflandırıcısı, olasılıklı bilginin öğrenilmesi ve temsil edilmesi için temiz anlamsallığa (semantik) sahip basit bir yaklaşım sunar. Algoritmanın naive yani saf (naif) olarak adlandırılmasının sebebi; belirli bir sınıfa ait tahmin edici özniteliklerin (attribute) koşullu olarak (belirli bir sınıfa ait olmaları koşulu) bağımsız olması ve gizli ve örtük özniteliklerin tahmin işlemini etkilemeyeceği şeklindeki iki varsayımı dayanıyor olmasından kaynaklanmaktadır (John ve Langley, 1995). Araştırmada kullanılan Weka NaiveBayes algoritma parametreleri şunlardır: useKernelEstimator=False; useSupervisedDiscretization=False



#### 4.1.3. MultilayerPerceptron (MLP)

MultilayerPerceptron, geri yayılım (backpropagation) algoritması kullanarak verileri sınıflandıran bir sınıflandırıcıdır. Bu yapay sinir ağı elle, algoritma ile ya da her ikisi kullanılarak oluşturulabilir. Oluşturulan ağ eğitim sırasında gözlemlenebilir ya da üzerinde değişiklikler yapılabilir. Bu ağda, sınıfın sayısal türde olmadığı durum haricinde tüm düğümler sigmoid fonksiyona sahiptir. Sınıf özneliğinin sayısal (nümerik) olması durumunda ise çıktı düğümleri belirli bir eşiğe sahip olmayan lineer birimler olurlar (George-Nektarios, 2013). Araştırmada MLP algoritması için kullanılan Weka parametreleri şunlardır: hiddenLayers=a; learningRate=0.3; momentum=0.2; seed=0; trainingTime=500; validationSetSize=0; validationThreshold=20

#### 4.1.4. SMO

SMO, bir destek vektör sınıflandırıcısının eğitimi için polinom veya Gauss çekirdeklerini (kernels) kullanarak sıralı minimal optimizasyon algoritmasını uygular. SMO sınıflandırıcısının değerlendirilmesi şu parametrelerle yapılır: c = 1.0; epsilon = 1.0E; kernel = PolyKernel; numFolds = -1; randomSeed = 1 (Witten ve Frank, 2005). Araştırmada SMO algoritmasının uygulanmasında kullanılan Weka parametreleri şunlardır: c=1.0; calibrator=Logistic -R 1.0E-8 -M -1; epsilon=1.0E-12; kernel=PolyKernel -E 1.0; numFolds=-1; randomSeed=1;

#### 4.1.5. IBk

IBk bir K en yakın komşu sınıflandırıcıdır. IBk sınıflandırıcısı uygun K değerini çapraz geçermeye dayanarak tayin edebilir. Bu sınıflandırıcı aynı zamanda uzaklık ağırlıklandırma da yapabilir (George-Nektarios, 2013). IBk algoritması araştırmada şu Weka parametreleri ile uygulanmıştır: KNN=1; crossValidate=False; distanceWeighting=No distance weighting; meanSquared=False; nearestNeighbourSearchAlgorithm=EuclideanDistance -R first-last; windowSize=0

#### 4.1.6. DecisionTable

DecisionTable, karar tablosu çoğunluk sınıflandırıcısını oluşturur. Bu sınıflandırıcı, best-first arama metodunu kullanarak özellik (feature) alt kümelerini değerlendirir ve bu amaçla çapraz geçermeye (cross-validation) kullanabilir. Bu sınıflandırıcı için arama safhasında kullanılacak BestFirst, RankSearch, GeneticSearch, vs. gibi birçok metod bulunmaktadır. Ayrıca IBk algoritması da sonuca yardımcı olmak amacıyla işleme dahil edilebilmektedir. (Witten ve Frank, 2005). DecisionTable algoritması araştırmada şu Weka parametreleriyle uygulanmıştır: crossVal=1; evaluationMeasure=accuracy (discrete class); RMSE; search=BestFirst -D 1 -N 5; useIBk=False

#### 4.1.7. JRip (RIPPER)

RIPPER en temel ve en popüler sınıflayıcı algoritmalarından biridir. Bu algoritmada sınıflar artan büyüklükte irdelenir ve sınıf için kademeli (incremental) azaltılmış hata budama yöntemi kullanılarak bir başlangıç kurallar seti oluşturulur (Witten ve Frank, 2005). JRip algoritması araştırmada şu Weka parametreleri ile uygulanmıştır: folds=3; minNo=2.0; optimizations=2; seed=1; usePruning=True

#### 4.1.8. OneR

OneR, Kural-tabanlı model kullanan bir diğer temel sınıflayıcı algoritmadır. Bu algoritma, her biri belirli bir özneliği (attribute) test eden kurallardan oluşan tek seviyeli (one-level) bir ağaç yapısı oluşturur. OneR algoritması basit, hesaplama olarak ucuz ve çoğu zaman veri içindeki yapıları karakterize etmek için oldukça iyi kurallar üretebilen bir algoritmadır (Witten ve Frank, 2005). Araştırmada OneR algoritması şu parametrelerle uygulanmıştır: bathSize=100; minBucketSize=6

#### 4.1.9. J48

Quinlan (Quinlan, 1993) tarafından geliştirilen C4.5 sınıflandırma algoritması günümüzde en popüler ağaç sınıflayıcılarından biri hatta belki de en popüleridir. J48 algoritması ise C4.5 algoritmasının optimize edilerek Weka'da uygulanmış halidir (Nguyen ve Choi, 2008). Araştırmada J48 algoritması şu Weka parametreleriyle uygulanmıştır: binarySplits=False; collapseTree=True; confidenceFactor=0.25; minNumObj=2; numFolds=3; reducedErrorPruning=False; seed=1; unpruned=False; useMDLcorrection=True

#### 4.1.10. NBTree

NBTree, Karar Ağaçları ve Naive Bayes algoritmaları kullanılarak elde edilen melez bir sınıflayıcı algoritmadır. Bu sınıflayıcı, yaprak düğümlere ulaşabilen veri örnekleri için yaprakların birer Naive Bayes sınıflayıcısı olduğu ağaç yapısı oluşturur. NBTree algoritmasının performans açısından Naive Bayes sınıflandırıcısından daha iyi olmasını beklemek oldukça mantıklı olmakla beraber bu performansın elde edilebilmesi için hızdan fedakarlık yapılmış olmaktadır (Kohavi, 1996). NBTree algoritması araştırmada şu parametrelerle uygulanmıştır: batchSize=100; doNotCheckCapabilities=False

#### 4.2. Performans Karşılaştırması

Her bir sınıflayıcının çalışma verileri ile Weka yazılımında uygulanmasından sonra elde edilen performans metriklerinden TP (True Positive – Doğru Pozitif), FP (False Positive – Yanlış Pozitif), TN (True Negative - Doğru Negatif) ve FN (False Negative – Yanlış Negatif) değerleri performans karşılaştırılması amacıyla kaydedilmiştir. Bu değerlerin kullanılarak

model sonuçlarına dair daha detaylı bilgiler veren Doğruluk (Accuracy), Hassasiyet (Recall veya Sensitivity) ve Kesinlik (Precision) gibi performans metrikleri aşağıdaki gibi hesaplanmaktadır (Gong, 2021):

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

$$\text{Kesinlik} = \frac{TP+TN}{TP+FP+TN+FN} \quad (2)$$

$$\text{Hassasiyet} = \frac{TP}{TP+FN} \quad (3)$$

Bu formüllerde geçen parametreler aşağıdaki gibi tanımlanmaktadır (Bilgin, 2021):

- TP (DP): Gerçekte pozitif sınıf içinde, tahminde ise pozitif sınıf içinde yer alan değer sayısı.
- FN (YN): Gerçekte pozitif sınıf içinde, tahminde ise negatif sınıf içinde yer alan değer sayısı.
- FP (YP): Gerçekte negatif sınıf içinde, tahminde ise pozitif sınıf içinde yer alan değer sayısı.
- TN (DN): Gerçekte negatif sınıf içinde, tahminde ise negatif sınıf içinde yer alan değer sayısı.

Doğruluk (Accuracy), doğru olarak sınıflandırılan örneklerin yüzdesidir. Hassasiyet (Recall), pozitif olarak tahmin etmemiz gereken işlemlerin ne kadarını pozitif olarak tahmin ettiğimizi gösteren bir metriktir. “Gerçek pozitiflerin ne kadarı doğru bir şekilde tanımlandı?”. Kesinlik (Precision), pozitif olarak tahmin edilen değerlerin gerçekten kaç adedinin pozitif olduğunu göstermektedir. (Ay, 2020).

Bunların yanı sıra sınıflandırma problemlerinde modelin performansını gösteren bazı metrikler de vardır; bu metriklerden bazıları ve en yaygın olarak kullanılanları F-Skoru (F-Measure) ve ROC (Receiver Operator Characteristic) performans metrikleridir.

F-Skoru, modelin doğruluğunun bir ölçüsüdür ve yukarıda formülleri verilen Kesinlik ve Doğruluk değerlerinin harmonik ortalaması alınarak hesaplanır. F-Skoru 1 ile 0 arasında bir değer alır; 1 değerine sahip olması modelin mükemmel olduğunu, 0 olması ise tamamen başarısız olduğunu gösterir. İyi bir F1-Skoru, yanlış pozitif ve yanlış negatif oranın az olduğunu ve örneklerin doğru biçimde sınıflandırıldığını gösterir. F-Skoru farklı düzeyde Doğruluk ve Kesinlik değerlerine sahip modelleri karşılaştırabilir hale getirebilmek için geliştirilmiş bir metriktir ve aşağıdaki şekilde hesaplanır (Nicholson, 2020):

$$\text{F-Skoru} = \frac{2 \times \text{Doğruluk} \times \text{Kesinlik}}{\text{Doğruluk} + \text{Kesinlik}} \quad (4)$$

ROC değeri ise bir sınıflama modelin performansını değerlendirme konusunda kullanılan en önemli

metriklerden biri olup farklı eşik değerlerinde modelin performansını gösterir. Bir olasılık eğrisi olan ROC sınıflama modelinin veri seti örneklerini farklı sınıflara ayırmada derecede başarılı olduğu hakkında bilgi verir. Bu eğrinin altında kalan alan AUC (Area Under Curve – Eğri Altında Kalan Alan) olarak adlandırılır ve bu alan ne kadar büyükse model veri seti örneklerini doğru sınıflara ayırmada o kadar başarılıdır (Narkhede, 2018).

ROC değerinin hesaplanmasında Gerçek Doğru Oranı (True Positive Rate - TPR) da denilen ve yukarıda açıklanan Hassasiyet metrik değeri ile Özgüllük denilen Yanlış Pozitif Oranı (False Positive Rate - FPR) metriklerinden yararlanır. Özgüllük ya da diğer adı ile Yanlış Pozitif Oranı, Yanlış Pozitiflerin Yanlış Pozitif ve Doğru Negatiflerin toplamına oranını ifade eder. Bu iki metriğin dağılımına çeşitli eşit değerleri uygulanarak bu değerler için modelin sınıflandırmayı ne derece iyi yaptığını gösteren ROC eğrisi elde edilir (Narkhede, 2018).

Araştırmada ele alınan 10 adet sınıflayıcı algoritmasına ait yağış tahmini performans metrikleri Tablo 2’de verilmiştir. Tablo 2 genel olarak incelendiğinde en iyi doğruluk oranını %84,24 ile Fonksiyonlar kategorisinde yer alan SMO algoritmasının verdiği görülmektedir. Bununla beraber bu performans değerinin diğer sonuçlardan açık ara önde olmadığı, Kurallar kategorisinde yer alan JRip (%83.0633) algoritması ile Ağaç sınıflayıcılar kategorisinde yer alan NBTree algoritmasının (%83.4315) da bu sonuca yakın performans gösterdiği söylenebilir. Ayrıca Kurallar kategorisinde yer alan OneR sınıflayıcısının tüm kategoriler içerisinde en düşük ortalama Doğruluk Oranına (%75,5) sahip olduğu görülmüştür. Bayes kategorisinde yer alan BayesNet ve NaiveBayes algoritmalarının performanslarının da OneR algoritmasından sonra en düşük performansa (76,6 ve 76,7) sahip sınıflayıcılar olduğu ortaya çıkmıştır.

Kategori bazında baktığımızda ise Bayes Sınıflayıcılar kategorisinde sonuçların birbirine çok yakın olduğu ve BayesNet ve NaiveBayes sınıflayıcıların birbirine çok yakın performans gösterdiği görülmektedir. Fonksiyon Sınıflayıcılar kategorisinde yer alan iki adet sınıflayıcı algoritmasından en başarılısının tüm sınıflayıcıların en başarılısı da olan SMO algoritması olduğu; buna mukabil diğer sınıflayıcı MLP algoritmasının da performansının (%81,95) iyi bir seviyede olduğu görülmektedir. Lazy kategorisinde yer alan tek sınıflayıcı olan IBk sınıflayıcısının performansının %78,05 ile fonksiyon sınıflayıcılardan biraz geride kaldığı görülmektedir. Kural tabanlı sınıflayıcıların ise genel olarak sınıflandırma performansının iyi olduğu, bu kategorinin en iyi performansına (%83.06) JRip algoritmasının sahip olduğu; DecisionTable algoritmasının buna yakın performans (%81.00) sergilediği, OneR algoritmasının performansının (%75,47) ise bunlardan biraz geride kaldığı söylenebilir. Son kategori olan Ağaç Sınıflayıcılar kategorisinde ise

J48 ve NBTree algoritmalarının performanslarının birbirine oldukça yakın olmakla beraber NBTree algoritmasının performansının J48 algoritmasının performansına nazaran biraz daha iyi olduğu görülmektedir.

Sınıflayıcı performanslarının Ortalama Doğruluk (OD) oranları haricinde tabloda yer alan Kesinlik, Hassasiyet, F-Skoru ve ROC Alanı değerlerine baktığımızda tabloda Doğruluk Oranı en yüksek olan

sınıflayıcı SMO algoritmasının aynı zamanda en yüksek ortalama Doğru Pozitif (DP), Kesinlik, Hassasiyet ve F-Skoru değerlerine sahip olan sınıflayıcı olduğunu görmek mümkündür. Buna mukabil Kurallar kategorisinde yer alan JRip algoritması en düşük ortalama Yanlış Pozitif (YP) oranına ve Fonksiyonlar kategorisinde yer alan MLP (MultiLayerPerceptron – Çok Katmanlı Yapay Sinir Ağı) algoritmasının ise en yüksek ROC Alanına sahip olduğu ortaya çıkmıştır.

**Tablo 2.** On adet sınıflayıcı almaya ait yağış tahmin performans metrikleri (Precipitation prediction performance metrics for the 10 classifiers)

Sınıflayıcı Kategorisi	Sınıflayıcı Algoritma	DP (%)	YP (%)	Doğruluk (%)	Kesinlik	Hassasiyet	F-Skoru	ROC Alanı
Bayes	BayesNet	76,7	26,0	76.730	0,789	0,767	0,774	0,827
	NaiveBayes	76,6	26,8	76.583	0,785	0,766	0,772	0,828
Fonksiyon	MLP	82,0	34,4	81.958	0,814	0,820	0,810	<b>0,858</b>
	SMO	<b>84,2</b>	27,7	<b>84.241</b>	<b>0,838</b>	<b>0,842</b>	<b>0,838</b>	0,783
Lazy	IBk	78,1	32,6	78.056	0,778	0,781	0,779	0,727
Kural	DecisionTable	81,0	27,8	81.001	0,809	0,810	0,810	0,854
	JRip	83,1	<b>25,7</b>	83.063	0,829	0,831	0,829	0,800
	OneR	75,5	35,2	75.478	0,755	0,755	0,755	0,701
Ağaç	J48	82,5	31,6	82.547	0,820	0,825	0,819	0,784
	NBTree	83,4	28,6	83.431	0,829	0,834	0,830	0,852

Araştırmada ayrıca farklı parametreler araştırma veri setinden çıkarılarak sınıflayıcıların performanslarındaki değişimler gözlemlenmiştir. Bu amaçla ilk önce veri setinden sadece rüzgar parametreleri (rüzgar yön ve şiddeti), sonrasında sadece önceki günlerin nem ve yağış parametreleri, en sonunda da hem rüzgar hem de önceki günlere ait nem ve yağış parametreleri çıkarılarak

sınıflayıcı performanslarındaki değişimlere bakılmıştır. Tablo 3, sadece rüzgar parametrelerinin (rüzgar yön ve şiddeti), Tablo 4, sadece önceki günlere ait nem ve yağış parametrelerinin ve son olarak Tablo 5 ise hem rüzgar hem de önceki günler nem ve yağış parametrelerinin çıkarıldığı durumda 10 sınıflayıcının performanslarını göstermektedir.

**Tablo 3.** Rüzgar parametreleri çıkarıldığında sınıflayıcıların yağış tahmin performans metrikleri (Precipitation prediction performance metrics after wind parameters removed)

Sınıflayıcı Kategorisi	Sınıflayıcı Algoritma	DP (%)	YP (%)	Doğruluk (%)	Kesinlik	Hassasiyet	F-Skoru	ROC Alanı
Bayes	BayesNet	76,2	<b>26,5</b>	76.215	0,785	0,762	0,769	0,824
	NaiveBayes	76,4	26,6	76.362	0,785	0,764	0,770	0,825
Fonksiyon	MLP	<b>84,8</b>	32,4	<b>84.757</b>	<b>0,849</b>	<b>0,848</b>	0,836	<b>0,897</b>
	SMO	84,5	28,1	84.462	0,840	0,845	<b>0,839</b>	0,782
Lazy	IBk	78,5	32,5	78.497	0,782	0,785	0,783	0,730
Kural	DecisionTable	81,0	27,8	81.001	0,809	0,810	0,810	0,854
	JRip	84,0	29,5	84.020	0,836	0,840	0,834	0,778
	OneR	75,5	35,2	75.478	0,755	0,755	0,755	0,701
Ağaç	J48	82,0	31,7	81.958	0,813	0,820	0,813	0,818
	NBTree	83,1	26,7	83.063	0,827	0,831	0,829	0,867



**Tablo 4.** Önceki günlerin nem ve yağış parametreleri çıkarıldığında sınıflayıcıların performans metrikleri (Precipitation prediction performance metrics after humidity and precipitation parameters belonging to previous days removed)

Sınıflayıcı Kategorisi	Sınıflayıcı Algoritma	DP (%)	YP (%)	Doğruluk (%)	Kesinlik	Hassasiyet	F-Skoru	ROC Alanı
Bayes	BayesNet	79,2	32,2	79.234	0,788	0,792	0,790	0,833
	NaiveBayes	79,8	30,2	79.823	0,796	0,798	0,797	0,843
Fonksiyon	MLP	82,5	33,7	82.547	0,820	0,825	0,816	<b>0,860</b>
	SMO	83,3	30,5	83.284	0,828	0,833	0,826	0,764
Lazy	IBk	77,1	33,2	77.098	0,770	0,771	0,771	0,720
	DecisionTable	81,4	28,4	81.443	0,812	0,814	0,813	0,846
Kural	JRip	<b>83,9</b>	<b>27,8</b>	<b>83.873</b>	<b>0,834</b>	<b>0,839</b>	0,834	0,782
	OneR	75,5	35,2	75.478	0,755	0,755	0,755	0,700
Ağaç	J48	80,7	32,9	80.706	0,800	0,807	0,801	0,790
	NBTree	80,8	33,5	80.780	0,800	0,808	0,801	0,832

**Tablo 5.** Rüzgar ve önceki günlerin nem ve yağış parametreleri çıkarıldığında sınıflayıcıların yağış tahmin performans metrikleri (Precipitation prediction performance metrics after humidity and precipitation parameters belonging to previous days along with wind parameter removed)

Sınıflayıcı Kategorisi	Sınıflayıcı Algoritma	DP (%)	YP (%)	Doğruluk (%)	Kesinlik	Hassasiyet	F-Skoru	ROC Alanı
Bayes	BayesNet	80,8	29,7	80.780	0,804	0,808	0,806	0,843
	NaiveBayes	81,1	30,8	81.075	0,805	0,811	0,807	0,854
Fonksiyon	MLP	<b>83,4</b>	33,2	<b>83.431</b>	<b>0,832</b>	<b>0,834</b>	0,824	<b>0,873</b>
	SMO	<b>83,4</b>	30,6	<b>83.431</b>	0,829	<b>0,834</b>	0,827	0,764
Lazy	IBk	76,3	33,4	76.288	0,765	0,763	0,764	0,715
	DecisionTable	81,4	28,4	81.443	0,812	0,814	0,813	0,846
Kural	JRip	81,8	35,4	81.811	0,812	0,818	0,807	0,736
	OneR	75,5	35,2	75.478	0,755	0,755	0,755	0,701
Ağaç	J48	81,7	30,6	81.664	0,811	0,817	0,812	0,822
	NBTree	83,1	<b>26,3</b>	83.063	0,828	0,831	<b>0,829</b>	0,854

Tablo 3'ten de görüleceği üzere veri setinden rüzgar parametreleri çıkarıldığında da ortalama Doğruluk Oranı en yüksek olan algoritma yine Fonksiyonlar kategorisine ait bir algoritma olmuş ancak bu sefer SMO yerine MLP algoritması en yüksek doğruluk performansı göstermiştir. Bununla beraber SMO algoritmasının performansının (%84,5) MLP algoritması performansına (%84,8) çok yakın olduğu görülmüştür. Önceki seferde olduğu gibi veri setinden rüzgar parametreleri çıkarıldığında da Bayes kategorisine ait sınıflayıcılar en düşük performansa (%76,2 ve %76,4) sahip olmuştur. İlginç bir sonuç olarak, rüzgar parametrelerinin veri setinden çıkarılmasının aslında sınıflama performansını artırdığı görülmüş; tam veri setinde en yüksek doğruluk oranı %84,2 iken rüzgar verileri çıkarıldığında doğruluk oranı %84,8'e çıkmıştır. Bununla birlikte doğruluk oranındaki bu artış önemli derecede bir artış değildir.

Rüzgar parametrelerini tekrar dahil edip veri setinden önceki günlere ait nem ve yağış parametrelerini çıkardığımızda ise Tablo 4'te görülen durum meydana gelmektedir. Tablo 4'ten de görüleceği üzere yeni durumda en yüksek performans sağlayan algoritma Kurallar kategorisinden JRip algoritması olmuştur. Ancak bu sefer yeni performansta (%83,9) önekilere

göre az da olsa bir düşüş olmuştur. Bu yeni senaryoda en düşük sınıflandırma performansı gösteren algoritma ise %75,5 oranı ile yine Kurallar kategorisinden OneR algoritması olmuştur.

Son durumda yani yağış veri setinden hem rüzgar hem de önceki günlerin nem ve yağış parametrelerinin çıkarıldığı durumda bakıldığında (Tablo 5); bu senaryoda en iyi performans gösteren algoritmanın kategorisinin değiştiği ve bir önceki durumdakinden farklı olarak Fonksiyonlar kategorisine ait MLP ve SMO algoritmalarının eşit (%83,4) ve en iyi sınıflama performansı sağladığı görülmektedir. Yeni performans değerinin diğerlerine çok yakın ve onlardan çok az daha düşük değerde olduğunu söylemek mümkündür. Bu son durumda en düşük performansı ise %75,5 ile Kurallar kategorisinden OneR sınıflayıcısının gösterdiği ortaya çıkmıştır.

## 5. Sonuç ve Tartışma (Results and Discussion)

Bu çalışmada 2010-2020 yılları arası için Ankara ili Keçiören ilçesinde yer alan Meteoroloji Genel Müdürlüğü'ne ait 17130 no'lu yağış istasyonundan elde edilen günlük basınç, sıcaklık, nem, rüzgar ve yağış verileri üzerinde çeşitli makine öğrenmesi algoritmaları uygulanarak yağışın varlığının kestirimi konusunda bu

sınıflayıcıların performanslarının karşılaştırılması amaçlanmıştır. Bu amaçla aynı veri setinde çeşitli değişiklikler uygulanarak dört farklı senaryoda 10 adet sınıflayıcı algoritma eldeki veriler üzerinde uygulanmış ve her senaryo sonucu elde edilen performans metrikleri performans karşılaştırması yapmak üzere kaydedilmiştir.

Çalışmada birinci senaryo olarak tüm verilere (basınç, sıcaklık, rüzgar, nem ve yağış) ilave olarak önceki iki günün nem ve yağış bilgileri de veri setine ayrı parametreler olarak dahil edilmiştir. İkinci senaryoda veri setinden sadece rüzgar parametreleri (rüzgar yön ve şiddeti) çıkarılarak bu yeni veri seti üzerinde 10 adet sınıflayıcı uygulanarak performansları kaydedilmiştir. Üçüncü senaryoda rüzgar parametreleri veri setine tekrar dahil edilmiş, ancak bu sefer önceki günlere ait nem ve yağış parametreleri veri setinden çıkarılmıştır. Son ve dördüncü senaryoda ise hem rüzgar

parametreleri hem de önceki günlere ait nem ve yağış parametreleri veri setinden çıkarılmış ve bu haliyle sınıflayıcı performanslarındaki değişimler not edilmiştir. Elde edilen sonuçlara göre ilk senaryoda en iyi performans gösteren sınıflayıcı %84,2 doğruluk oranı ile Fonksiyonlar kategorisinden SMO algoritması olmuştur. İkinci senaryoda ise en iyi performans gösteren algoritma yine Fonksiyonlar kategorisinden %84,8 ortalama doğruluk oranı ile MLP (MultiLayerPerceptron – Çok Katmanlı Yapay Sinir Ağı) algoritması olmuştur. Üçüncü senaryoda en iyi doğruluk oranına sahip sınıflayıcı algoritma Kurallar kategorisinden JRip algoritması olmuştur. Son yani dördüncü senaryoda ise Fonksiyonlar kategorisinden MLP ve SMO algoritmaları birbirine eşit ve %83,4 ortalama doğruluk oranı ile en iyi performansa sahip algoritmalar olmuşlardır. Bu dört senaryoya ait sonuçlar Tablo 5’te özetlenmiştir.

**Tablo 6.** Dört senaryoya ait özet en iyi performans metrikleri (Summary of best performance metrics belonging to the four scenario)

	En iyi Sınıflayıcı	Kategori	DP (%)	YP (%)	DO (%)	Kes.	Hass.	F-Skoru	ROC Alanı
1. Senaryo	SMO	Fonksiyon	84,2	27,7	84.24	0,838	0,842	0,838	0,783
2. Senaryo	MLP	Fonksiyon	84,8	32,4	84.75	0,849	0,848	0,836	0,897
3. Senaryo	JRip	Kural	83,9	27,8	83.87	0,834	0,839	0,834	0,782
4. Senaryo	MLP	Fonksiyon	83,4	33,2	83.43	0,832	0,834	0,824	0,873
	SMO	Fonksiyon	83,4	33,2	83.43	0,832	0,834	0,824	0,873

Tablo 6’dan de görüleceği üzere en yüksek ortalama doğruluk oranları %83,4 ila %84,8 arasında değişmektedir -ki bu da senaryolar arasındaki performans farkının çok fazla olmadığını göstermektedir. Dolayısıyla basınç, sıcaklık, rüzgar yönü, rüzgar şiddeti, nem ve yağış parametrelerinden oluşan ana veri setinden rüzgar parametrelerinin yahut önceki güne ait nem ve yağış parametrelerinin çıkarılması ya da bunların tekrar dahil edilmesi sınıflayıcı performanslarında çok büyük oranda bir değişime neden olmamaktadır. Bununla birlikte %80’ler civarında elde edilen bu performansların yağış tahmininde, diğer bir ifade ile yağışın varlığını kestirmede iyi derecede performanslar olduğu söylenebilir. Yine Tablo 6’dan görüleceği üzere yağışın varlığını tespit konusunda en başarılı kategori Fonksiyonlar kategorisi ve en başarılı algoritmalar ise bu kategorideki SMO (Sequential Minimal Optimization) ve MLP (Çok Katmanlı Yapay Sinir Ağı) algoritmaları olmuştur. Dört senaryo içerisinde bu durumun dışında kalan tek senaryo üçüncü yani rüzgar parametrelerinin dahil olduğu ama önceki günlere ait nem ve yağış parametrelerinin çıkarıldığı senaryo olmuştur. Bu senaryoda ise en başarılı kategori Kurallar kategorisi ve en başarılı sınıflayıcı algoritma ise bu kategorideki JRip algoritması olmuştur.

Elde edilen bu sonuçlar makine öğrenmesi algoritmalarının Meteoroloji biliminin bir konusu olan

yağışın varlığının kestiriminde başarılı şekilde uygulanabileceğini göstermektedir. Elde edilen sonuçların hava tahmincilerinin, Meteoroloji bilimiyle ilgili kişiler ve uygulamacılar ve genel olarak makine öğrenmesi konusunda çalışan araştırmacılar için yararlı olacağı düşünülmektedir. Konu ile ilgili yapılacak gelecekteki çalışmalarda araştırmacılara yağışa etki eden farklı parametreleri de dahil ederek sınıflayıcı performanslarını karşılaştırmaları ve ayrıca yağışın varlığının kestirimini yanında yağış miktarının kestirimi konusunda araştırma yapmaları önerilmektedir.

#### Kaynaklar (References)

- Akman, Y. (1990). İklim ve Biyoiklim (Biyoiklim Metotları ve Türkiye İklimleri). Ankara: Palme Yayınları.
- Ay, Ş. (2020). Model performansını değerlendirmek Metrikler. Erişim adresi: <https://medium.com/deeplearning-turkiye/model-performans%C4%B1n%C4%B1-de%C4%9Ferlendirmek-metrikler-cb6568705b1>
- Bilgin, G. (2021). Makine öğrenmesi algoritmaları kullanarak erken dönemde diyabet hastalığı riskinin araştırılması. Zeki sistemler teori ve uygulamaları dergisi 4(1), 55-64. DOI: 10.38016/jista.877292

- Dibike, Y.B., Solomatine, D.P. (2001). River flow forecasting using artificial neural networks. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, Volume 26, Issue 1, Pages 1-7.
- Gong, M. (February 2021). A novel performance measure for machine learning classification, *International Journal of Managing Information Technology (IJMIT)*, Vol.13, No.1
- Grabec, I. (1990). Empirical modelling of natural phenomena by a self-organizing system. *Proc. Neural Network Conf.* 90, Vol. 2, 529-532.
- John, G. H., Langley, P (1995). Estimating Continuous Distributions in Bayesian Classifiers. *Proc. of the 11th Conf. on Uncertainty in Artificial Intelligence* içinde.
- Kale, M. M. (2020). İklim Değişikliği Çerçevesinde Ankara İli Ana Su Havzaları Gelecek Projeksiyonu: Sakarya ve Batı Karadeniz Havzaları, *Coğrafi Bilimler Dergisi/ Turkish Journal of Geographical Sciences*, 18(2), 191-215, doi: 10.33688/aucbd.732831.
- Kannan, M., Prabhakaran, S. and Ramachandran, P. (2010). Rainfall Forecasting Using Data Mining Technique. *International Journal of Engineering and Technology*, Vol.2 (6), 397-401
- Kohavi, R. (1996). Scaling up the accuracy of naïve-bayes classifier: A decision-tree hybrid. *Proc. of the 2nd International Conference on Knowledge Discovery and Data Mining* içinde (202–207). Menlo Park: AAAI Press.
- Narkhede (2018). Understanding AUC-ROC curve. Erişim adresi <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>
- Nguyen, H. A., Choi, D. (2008). Application of Data Mining to Network Intrusion Detection: Classifier Selection Model. *11th Asia-Pacific Network Operations and Management Symposium (APNOMS 2008)* içinde (399-408 ss.).
- Nicholson (2020). Evaluation metrics for machine learning – Accuracy, precision, recall, and F1 defined. Erişim adresi <https://wiki.pathmind.com/accuracy-precision-recall-f1>
- Öziş, Ü., Barant, T., Durnabaşı, İ., Özdemir, Y. (1997). Türkiye'nin su kaynakları potansiyeli. *Meteoroloji Mühendisliği*, 2, 40-45.
- Quinlan, J. (1993). *C4.5: Programs for Machine Learning*. San Mateo: Morgan Kaufmann
- Rani, B. K. and Govardhan, A. (2013). Rainfall prediction using Data Mining techniques-A Survey. *Computer Science and Information Technology*, pp. 23-30.
- Sangari, R. S. and Balamurugan, M. (2014). A Survey on rainfall prediction using Data Mining. *International Journal of Computer Science and Mobile Applications*, vol. 2, no. 2, pp. 84-88.
- Shoba G (2014). Rainfall prediction using Data Mining techniques: A Survey. *International Journal of Engineering and Computer Science*, vol. 3, no. 5, pp. 6206-6211.
- Soo-Yeon, J., Sharad, S., Byunggu, Y. and Dong, H. J. (2012). Designing a Rule-Based Hourly Rainfall Prediction Model. *IEEE IRI*, August 8-19.
- Soucek, B. (1992). Prediction of Chaotic Dynamical Phenomena by a Neural Network. *Dynamic, Genetic and Chaotic Programming içinde* (471-500). U.S.A: John Wiley & Sons Inc.
- Suhartono, S., R. Faulina, D. A. Lusia, B. W. Otok, Sutikno and H. Kuswanto (2012). Ensemble method based on ANFIS-ARIMA for rainfall prediction. *2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE)*, pp. 1-4, doi: 10.1109/ICSSBE.2012.6396564.
- Theofilis, G. (2013). Weka classifier summary. Erişim adresi [https://www.academia.edu/5167325/Weka\\_Classifiers\\_Summary](https://www.academia.edu/5167325/Weka_Classifiers_Summary)
- Uzunali, A. (2019). Yapay sinir ağlarına dayalı yağış tahmin ve analizi. (Yayımlanmamış yüksek lisans tezi). İstanbul Kültür Üniversitesi Lisansüstü Eğitim Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı, İstanbul
- Witten, I.H., Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd edn. San Francisco: Morgan Kaufmann.
- Xiao, R. And V. Chandrasekar. (1995). Multiparameter radar rainfall estimation using neural network techniques. *Preprints, 27th Conf. on Radar Meteorology*, Vail, CO, Amer. Meteor. Soc., 199–204.