A red text on a white background

Description automatically generated

**MÜHENDİSLİK VE DOĞA BİLİMLERİ FAKÜLTESİ**

**BYM446 - VERİ MADENCİLİĞİ DERSİ DÖNEM SONU PROJE ÖDEVİ**

**GEÇMİŞ VERİLERİ ÇEŞİTLİ ALGORİTMALARLA ANALİZ EDEREK YAĞIŞ TAHMİNLEME**

**HAZIRLAYANLAR**

MERT TOSUN- 210601027

MUZAFFER ENES YILDIRIM- 210601010

GÖKHAN ALTUNDAL – 210601017

**BİLGİSAYAR MÜHENDİSLİĞİ**

TESLİM TARİHİ

01.06.2025

ABSTRACT

Bu projede, meteorolojik verilere dayalı olarak yağış tahmini yapılabilmesi amacıyla bir makine öğrenmesi uygulaması geliştirilmiştir. Kullanılan veri seti, Avustralya Meteoroloji Bürosu tarafından sağlanan **WeatherAUS** veri setidir. Bu veri seti; sıcaklık, basınç, nem, rüzgar hızı ve yönü gibi birçok farklı meteorolojik ölçümü içermekte ve **"RainTomorrow"** adlı hedef değişken üzerinden ertesi gün yağış durumu ("Evet" veya "Hayır") tahminlenmektedir.

Proje kapsamında; eksik ve hatalı verilerin temizlenmesi, eksik değerlerin uygun yöntemlerle doldurulması, kategorik değişkenlerin sayısal forma dönüştürülmesi ve verilerin model eğitimine uygun hale getirilmesi gibi kapsamlı bir **veri ön işleme süreci** gerçekleştirilmiştir. Daha sonra, **Logistic Regression**, **Random Forest** ve **Decision Tree** gibi farklı makine öğrenmesi algoritmaları kullanılarak sınıflandırma modelleri oluşturulmuş ve bu modellerin performansları karşılaştırılmıştır.

Modellerin değerlendirilmesinde **doğruluk (accuracy)**, **hassasiyet (precision)**, **duyarlılık (recall)**, **F1 skoru** ve **ROC-AUC skoru** gibi metrikler kullanılmıştır. Elde edilen sonuçlara göre, **Random Forest Classifier** modeli, diğer modellere kıyasla en yüksek başarı oranını göstermiş; özellikle doğruluk, F1 skoru ve ROC-AUC skorlarında daha iyi performans sergilemiştir. Logistic Regression modeli dengeli ve hızlı bir sonuç verirken, Decision Tree modeli basitliğiyle öne çıkmış, ancak genelleme performansında daha zayıf kalmıştır.

Projede karşılaşılan başlıca zorluklar, eksik ve dengesiz veri problemleri, sınıflar arasındaki dengesizlik ve model optimizasyonu süreçleri olmuştur. Bu zorluklar, veri ön işleme teknikleri ve algoritmaların parametre ayarlamaları ile aşılmaya çalışılmıştır.

Çalışmanın sonunda, yağış tahmini üzerine kurulu bu modelin, hava durumu tahminleri ve tarımsal planlama gibi alanlarda fayda sağlayabileceği sonucuna varılmıştır. Gelecekteki çalışmalarda, modelin performansını artırmak amacıyla hiper parametre optimizasyonu, farklı algoritmaların denenmesi, zaman serisi analizleri ve daha geniş veri kaynaklarının kullanılması planlanmaktadır.

GİRİŞ

Günümüzde hava durumu tahminleri, tarım, ulaşım, enerji ve çevre yönetimi gibi birçok alanda kritik bir rol oynamaktadır. Özellikle yağış tahminleri, tarımsal üretim planlamasından şehir altyapı yönetimine kadar birçok sektörde önemli bir karar destek aracı olarak kullanılmaktadır. Doğru ve zamanında yapılan yağış tahminleri, tarımsal ürün kayıplarını önleyebilir, doğal afet risklerini azaltabilir ve insan hayatını doğrudan etkileyen çeşitli olumsuz durumların önüne geçebilir.

Bu proje, **Avustralya Meteoroloji Bürosu tarafından sağlanan WeatherAUS veri seti** kullanılarak, ertesi gün yağış olup olmayacağını tahmin edebilen bir makine öğrenmesi modeli geliştirmeyi amaçlamaktadır. Projede kullanılan **RainTomorrow** değişkeni, sınıflandırma probleminin hedef değişkeni olarak belirlenmiştir. Projenin temel hedefi, bu değişkeni tahminleyebilen bir model oluşturmak ve farklı makine öğrenmesi algoritmalarının başarı oranlarını karşılaştırmaktır.

Çalışma kapsamında, farklı meteorolojik değişkenler (örneğin: sıcaklık, basınç, nem, rüzgar hızı ve yönü gibi) ile **Logistic Regression**, **Random Forest** ve **Decision Tree** gibi popüler sınıflandırma algoritmaları kullanılarak tahmin modelleri geliştirilmiştir. Modellerin doğruluk oranı, hassasiyet (precision), duyarlılık (recall), F1 skoru ve ROC-AUC gibi çeşitli metrikler üzerinden karşılaştırılması yapılmıştır.

Veri seti üzerinde gerçekleştirilen ön işlemler; eksik verilerin analiz edilmesi, verilerin temizlenmesi, eksik değerlerin uygun yöntemlerle doldurulması ve kategorik verilerin sayısal formatlara dönüştürülmesini içermektedir. Ayrıca, modelleme sürecinde eğitim ve test setleri oluşturularak veri seti bölünmüş, model değerlendirmeleri bu veri setleri üzerinden gerçekleştirilmiştir.

Bu proje ile amaçlanan, makine öğrenmesi tekniklerinin meteorolojik veriler üzerinde nasıl uygulanabileceğini ve yağış tahmini gibi gerçek dünya problemlerinin çözümünde ne derece etkili olabileceğini göstermektir. Aynı zamanda, bu proje farklı sınıflandırma algoritmalarının performansını değerlendirmek ve hangi algoritmanın yağış tahmini için daha uygun olduğunu belirlemek açısından önemli bir çalışma niteliği taşımaktadır.

İLGİLİ ÇALIŞMALAR

Yağış tahmini, meteoroloji alanında uzun süredir üzerinde çalışılan bir konudur. Geleneksel istatistiksel yöntemlerin sınırlamaları nedeniyle, son yıllarda makine öğrenmesi (ML) ve derin öğrenme (DL) teknikleri bu alanda önemli bir yer edinmiştir. Literatürde, çeşitli ML ve DL algoritmalarının yağış tahmini üzerindeki performansları incelenmiş ve karşılaştırılmıştır.

**Makine Öğrenmesi Yöntemleriyle Yağış Tahmini:**

Birçok çalışma, yağış tahmini için farklı makine öğrenmesi algoritmalarını kullanmıştır. Örneğin, Multiple Linear Regression (MLR), Decision Tree (DT), Random Forest (RF) ve Support Vector Machine (SVM) gibi algoritmalar, meteorolojik veriler kullanılarak yağış miktarını tahmin etmek için uygulanmıştır. Bu çalışmalar, özellikle sıcaklık, nem, basınç ve rüzgar hızı gibi atmosferik özelliklerin yağış tahmininde önemli faktörler olduğunu vurgulamaktadır.

**Derin Öğrenme Yöntemleriyle Yağış Tahmini:**

Derin öğrenme, karmaşık ve doğrusal olmayan ilişkileri modelleme yeteneği sayesinde yağış tahmininde etkili bir araç haline gelmiştir. Özellikle, Uzun Kısa Süreli Bellek (LSTM) ve Gated Recurrent Unit (GRU) gibi tekrarlayan sinir ağları, zaman serisi verilerindeki bağımlılıkları yakalayarak daha doğru tahminler yapabilmektedir. Bir çalışmada, DLSTM modelinin diğer derin öğrenme modellerine kıyasla daha düşük hata oranlarıyla en iyi performansı gösterdiği belirtilmiştir.

**Hibrit Modeller ve Gelişmiş Yaklaşımlar:**

Bazı araştırmalar, makine öğrenmesi ve derin öğrenme tekniklerini birleştirerek hibrit modeller geliştirmiştir. Örneğin, ARIMA ve TBATS gibi zaman serisi modelleri ile ANN, SVR ve RF gibi makine öğrenmesi algoritmalarının kombinasyonu, yağış tahmininde daha yüksek doğruluk sağlamıştır. Ayrıca, encoder-decoder yapıları ve dikkat mekanizmaları kullanan derin öğrenme modelleri, çok adımlı yağış tahminlerinde başarılı sonuçlar elde etmiştir.

Örneğin, Kyner (2020) tarafından gerçekleştirilen **"Rain in Australia – Next-Day Prediction Model"** projesinde, **WeatherAUS** veri seti kullanılarak ertesi gün yağış tahmini yapılmıştır. Çalışmada Logistic Regression, Random Forest ve Decision Tree gibi çeşitli sınıflandırma algoritmaları kullanılmış ve bu modellerin doğruluk, F1 skoru ve ROC-AUC gibi performans metrikleri üzerinden karşılaştırılması yapılmıştır. Kyner’in çalışmasında, Random Forest modeli genel doğruluk açısından en yüksek başarıyı sağlamış, ancak Logistic Regression modeli yağışlı günleri tahmin etmede daha yüksek bir recall değerine ulaşmıştır ([Kyner, 2020](https://github.com/tjkyner/australian-rain-prediction)).

Ayrıca Oni ve arkadaşları (2024), Random Forest, Decision Tree ve Extra Tree algoritmalarını kullanarak bir karşılaştırmalı analiz gerçekleştirmiş ve Random Forest algoritmasının %66 doğruluk oranı ile en iyi performansı gösterdiğini belirtmiştir. Başka bir çalışmada, Rawat ve arkadaşları (2023), Logistic Regression, Random Forest, XGBoost, KNN ve SVM gibi algoritmalar üzerinde bir değerlendirme yapmış ve Random Forest ile XGBoost’un yüksek doğruluk oranları sağladığını ifade etmiştir.

Derin öğrenme yöntemleri de yağış tahmini çalışmalarında önemli bir yer tutmaktadır. Özellikle LSTM ve GRU gibi tekrarlayan sinir ağları, zaman serisi verilerindeki bağımlılıkları modelleyebilme yetenekleri sayesinde daha doğru tahminler yapılabilmesini sağlamaktadır. Bu doğrultuda Chittella ve arkadaşları (2023), geçmiş hava durumu verilerini kullanarak geliştirdikleri bir sinir ağı modeliyle yağış tahmini gerçekleştirmiş ve bu yaklaşımın tarım ve afet yönetimi gibi alanlarda değerli katkılar sağlayabileceğini ortaya koymuştur.

Bu literatür çalışmaları, **WeatherAUS** veri setinin ve benzeri meteorolojik veri setlerinin yağış tahmini çalışmalarında yaygın olarak kullanıldığını ve farklı makine öğrenmesi algoritmalarının bu problem üzerinde karşılaştırıldığını göstermektedir. Bu proje, Kyner’in çalışmasını temel alarak benzer bir veri seti üzerinde farklı modellerin değerlendirilmesini amaçlamış ve model performanslarını detaylı bir şekilde analiz etmiştir.

VERİ AÇIKLAMASI (DATA DESCRIPTION)

Bu projede kullanılan veri seti, **Avustralya Meteoroloji Bürosu** tarafından sağlanan ve **WeatherAUS** adıyla bilinen halka açık bir veri setidir. WeatherAUS veri seti, farklı şehir ve bölgelerdeki günlük meteorolojik gözlemleri içermektedir. Veri seti, Avustralya genelinde 49 farklı konumda ölçülen çeşitli hava durumu değişkenlerini ve ertesi gün yağış durumu bilgisini içermektedir.

**Veri Seti Özellikleri:**

* **Kaynak:** Kaggle - Australian Weather Dataset
* **Boyut:** 142.193 satır (örnek/gözlem) ve 24 sütun (özellik).
* **Dönem:** 2008–2017 yılları arasındaki günlük gözlemler.
* **Hedef Değişken:**
  + **RainTomorrow** (Binary Sınıflandırma Problemi)
  + 0 = Yağmur yok, 1 = Yağmur var

### Ana Özellikler (Değişkenler):

| **Özellik Adı** | **Açıklama** |
| --- | --- |
| **Date** | Gözlemin yapıldığı tarih |
| **Location** | Ölçüm yapılan şehir/konum adı |
| **MinTemp** | Günün en düşük sıcaklığı (°C) |
| **MaxTemp** | Günün en yüksek sıcaklığı (°C) |
| **Rainfall** | O günkü toplam yağış miktarı (mm) |
| **Evaporation** | Buharlaşma miktarı (mm) |
| **Sunshine** | Güneşli saatlerin süresi (saat) |
| **WindGustDir** | En yüksek rüzgar esiş yönü |
| **WindGustSpeed** | En yüksek rüzgar hızı (km/h) |
| **WindDir9am** | 9:00’daki rüzgar yönü |
| **WindDir3pm** | 15:00’deki rüzgar yönü |
| **WindSpeed9am** | 9:00’daki rüzgar hızı (km/h) |
| **WindSpeed3pm** | 15:00’deki rüzgar hızı (km/h) |
| **Humidity9am** | 9:00’daki nem oranı (%) |
| **Humidity3pm** | 15:00’deki nem oranı (%) |
| **Pressure9am** | 9:00’daki atmosfer basıncı (hPa) |
| **Pressure3pm** | 15:00’deki atmosfer basıncı (hPa) |
| **Cloud9am** | 9:00’daki bulutluluk durumu (0–8 skalasında) |
| **Cloud3pm** | 15:00’deki bulutluluk durumu (0–8 skalasında) |
| **Temp9am** | 9:00’daki sıcaklık (°C) |
| **Temp3pm** | 15:00’deki sıcaklık (°C) |
| **RainToday** | O gün yağış durumu (Yes/No) |
| **RainTomorrow** | Ertesi gün yağış durumu (Yes/No) (Hedef değişken) |

**Eksik Veriler:**

Veri setinde özellikle bazı sütunlarda eksik değerler mevcuttur. Örneğin:

* **Sunshine**, **Evaporation**, **Cloud9am**, **Cloud3pm** gibi sütunlarda %30’un üzerinde eksik veri bulunmaktadır.
* Bu nedenle, projede eksik veri analizi yapılmış ve yüksek eksikliğe sahip sütunlar (örneğin Sunshine, Evaporation) veri setinden çıkarılmıştır.
* Geriye kalan sütunlardaki eksik veriler ise uygun doldurma teknikleriyle tamamlanmıştır (örneğin medyan veya mod değerleri ile).

**Veri Hazırlık Süreci:**

* Eksik verilerin analizi ve temizliği.
* Kategorik değişkenlerin sayısal değerlere dönüştürülmesi (Label Encoding).
* Tarih kolonundan **Yıl**, **Ay**, **Gün**, **Mevsim**, **GünAdı** gibi yeni özelliklerin türetilmesi.
* Hedef değişken **RainTomorrow**, 0 ve 1 olacak şekilde binary sınıflandırmaya uygun hale getirilmiştir.

YÖNTEM (METHOD)

Bu projede, **WeatherAUS** veri seti kullanılarak ertesi gün yağış tahmini (RainTomorrow) yapılabilmesi amacıyla farklı makine öğrenmesi modelleri uygulanmıştır. Model geliştirme süreci, aşağıdaki aşamalardan oluşmaktadır:

**1️.Veri Ön İşleme (Data Preprocessing)**

Veri setinde eksik veriler ve kategorik değişkenler bulunması nedeniyle öncelikle detaylı bir veri temizleme işlemi gerçekleştirilmiştir:

* **Eksik Veri Analizi ve Temizliği:**
  + Yüksek oranda eksik veri içeren sütunlar (**Sunshine**, **Evaporation**, **Cloud9am**, **Cloud3pm**) veri setinden çıkarılmıştır.
  + Sayısal sütunlardaki eksik değerler (örneğin **MaxTemp**) medyan değer ile, kategorik sütunlardaki eksik değerler (örneğin **WindGustDir**) ise mod değeri ile doldurulmuştur.
* **Kategorik Değişkenlerin Kodlanması:**
  + **WindGustDir**, **WindDir9am**, **WindDir3pm**, **RainToday** gibi kategorik değişkenler **Label Encoding** yöntemi ile sayısal değerlere dönüştürülmüştür.
* **Tarih Verisinin Özelliklerine Ayrılması:**
  + **Date** sütunundan **Yıl**, **Ay**, **Gün**, **Haftanın Günü (DayOfWeek)** ve **Mevsim (Season)** gibi yeni özellikler türetilmiştir.
* **Özellik Ölçeklendirme:**
  + **Logistic Regression modeli** için **StandardScaler** kullanılarak sayısal sütunlar ölçeklendirilmiştir.
  + **Random Forest ve Decision Tree** modelleri için ölçeklendirme yapılmamıştır (bu modeller ölçek duyarlı değildir).
* **Veri Setinin Bölünmesi:**
  + Veri seti **%80 eğitim** ve **%20 test** olarak ikiye ayrılmıştır.

**2️**. **Kullanılan Makine Öğrenmesi Modelleri (Algorithms)**

Proje kapsamında üç farklı sınıflandırma algoritması kullanılmış ve performansları karşılaştırılmıştır:

**Logistic Regression**

* Temel bir doğrusal sınıflandırma modelidir.
* Veri setindeki sınıf dengesizliği nedeniyle class\_weight='balanced' parametresi kullanılmıştır.
* Özellikle yağmurlu günlerin tahmin edilmesinde güçlü bir modeldir.

**Random Forest Classifier**

* Birden fazla karar ağacının (decision tree) birleştirilmesiyle oluşturulan bir topluluk modelidir.
* Özellikle karmaşık veri setlerinde güçlü bir performans sağlar.
* Bu modelde random\_state=42 parametresi kullanılmıştır.

**Decision Tree Classifier**

* Basit ve yorumlanabilir bir ağaç tabanlı modeldir.
* Hızlı çalışması ve kolay yorumlanabilir olması nedeniyle tercih edilmiştir.
* random\_state=42 parametresi ile çalıştırılmıştır.

**3. Model Değerlendirme Metrikleri (Evaluation Metrics)**

Modellerin başarısını ölçmek için aşağıdaki metrikler kullanılmıştır:

* **Accuracy (Doğruluk)**
* **Precision (Kesinlik)**
* **Recall (Duyarlılık)**
* **F1-Score**
* **ROC-AUC Skoru**
* **Confusion Matrix (Karışıklık Matrisi)**

Her modelin çıktısı bu metrikler üzerinden analiz edilmiş ve karşılaştırmalı sonuçlar elde edilmiştir.

DEĞERLENDİRME (EVALUATION)

Bu projede, farklı makine öğrenmesi algoritmalarının **yağış tahmini (RainTomorrow)** problemini ne kadar başarılı bir şekilde çözebileceği değerlendirilmiştir. Model performansları; **Accuracy (Doğruluk)**, **Precision (Kesinlik)**, **Recall (Duyarlılık)**, **F1-Score** ve **ROC-AUC Skoru** gibi metrikler üzerinden karşılaştırılmıştır.

**Modellerin Karşılaştırmalı Performansı**

**1️. Logistic Regression Sonuçları**

**metin, ekran görüntüsü, yazı tipi, yazılım içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.**

| **Metrik** | **Sonuç** | **Yorum** |
| --- | --- | --- |
| Accuracy | 0.78 | Genel doğruluk Random Forest ve Decision Tree’ye göre daha düşük. Veri dengesizliği etkili. |
| Precision (0) | 0.92 | Model "yağmur yok" sınıfında yüksek doğruluk sağlıyor. |
| Precision (1) | 0.50 | "Yağmur var" tahminlerinin yarısı doğru çıkıyor. |
| Recall (0) | 0.79 | "Yağmur yok" olanların %79’unu doğru tahmin ediyor. |
| Recall (1) | **0.76** | **En iyi recall değeri**. Yağmurlu günlerin %76’sını doğru tahmin ediyor. |
| F1-Score (1) | 0.60 | Dengeli bir başarı. Recall avantajı sayesinde fena değil. |
| ROC-AUC Skoru | 0.85 | İyi bir sınıf ayrıştırma yeteneği gösteriyor. |

Logistic Regression modeli, genel doğrulukta zayıf kalsa da (0.78), **yağmurlu günleri tahmin etmede en iyi performansı göstermiştir (Recall 0.76)**. Bu, sınıf dengesizliği olan bir veri setinde (RainTomorrow=0 çok daha fazla), modelin **daha fazla "yağmur var" tahmini yaparak hataya razı olması** anlamına gelir. Yani model, **hassasiyetten (precision)** feragat edip **duyarlılığı (recall)** artırmayı tercih etmiş bir davranış sergiler.

**2️. Random Forest Modeli**

**metin, ekran görüntüsü, yazılım, multimedya yazılımı içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.**

| **Metrik** | **Sonuç** | **Yorum** |
| --- | --- | --- |
| Accuracy | **0.86** | En yüksek doğruluk oranı. Genel tahmin başarısı güçlü. |
| Precision (0) | 0.87 | "Yağmur yok" tahminlerinde yüksek doğruluk. |
| Precision (1) | **0.76** | "Yağmur var" dediğinde doğru çıkma oranı daha iyi. |
| Recall (0) | **0.96** | "Yağmur yok" sınıfında neredeyse eksiksiz tahmin. |
| Recall (1) | 0.50 | Yağmurlu günlerin sadece yarısını yakalayabiliyor. |
| F1-Score (1) | **0.60** | Dengeli bir sonuç. |
| ROC-AUC Skoru | **0.88** | En yüksek ROC-AUC değeri. Sınıflar arasında iyi bir ayrım yeteneği var. |

Random Forest modeli, **genel doğruluk ve ROC-AUC skorunda en iyi performansı göstermiştir**. Bu, modelin "yağmur yok" olan günleri çok iyi tahmin ettiğini ve genel olarak daha istikrarlı bir sonuç verdiğini gösterir. Ancak, **yağmurlu günleri tahmin etmede (Recall 0.50)**, Logistic Regression’a göre daha zayıftır. Random Forest daha temkinli bir modeldir; yanlış pozitifleri azaltmaya çalışırken (yüksek precision), bazı yağmurlu günleri kaçırabilir (düşük recall).

**3️. Decision Tree Modeli**

metin, ekran görüntüsü, yazılım, multimedya yazılımı içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Decision Tree modeli, basitliği sayesinde hızlı ve kolay yorumlanabilir bir modeldir. Ancak, **ROC-AUC skorunun düşük olması (0.69)**, sınıfları ayırt etme gücünün zayıf olduğunu göstermektedir. Model, "yağmur yok" günlerini çok iyi tahmin ederken (Recall 0.96), "yağmur var" günlerinin yaklaşık yarısını (Recall 0.54) doğru tahmin edebilmiştir.

| **Model** | | **Avantajı** | **Dezavantajı** |
| --- | --- | --- | --- |
| Logistic Regression | | Yağmurlu günleri yakalamada iyi (Recall 0.76) | Genel doğruluk daha düşük |
| Random Forest | En yüksek doğruluk ve ROC-AUC | | Yağmurlu günleri tahmin etmede zayıf (Recall 0.50) |
| Decision Tree | | Hızlı ve kolay yorumlanabilir | ROC-AUC düşük, "yağmur var" tahminleri zayıf |

**Değerlendirme Grafikleri**

çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, metin, diyagram içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

metin, ekran görüntüsü, ekran, görüntüleme, sayı, numara içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

çizgi, metin, ekran görüntüsü, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

KARŞILAŞILAN ZORLUKLAR VE GELECEK ÇALIŞMALAR

Proje sürecinde en önemli zorluklardan biri, **veri setindeki eksik değerler** ve **sınıf dengesizliği** olmuştur. Özellikle bazı sütunlarda (%30'dan fazla) eksik veri bulunması, modelin başarısını olumsuz etkileyebilecek bir durum oluşturmuş ve bu nedenle eksik sütunlar çıkarılmış veya uygun yöntemlerle doldurulmuştur. Ayrıca, veri setinde "yağmur yok" sınıfı, "yağmur var" sınıfına göre çok daha fazla sayıda olduğu için modellerin dengesiz veriyle başa çıkabilmesi için ek önlemler(balanced) alınmıştır.

Gelecekte, model performansını daha da artırmak amacıyla **hiperparametre optimizasyonu**, **XGBoost, LightGBM gibi daha gelişmiş algoritmaların kullanımı** ve **zaman serisi tahminleri üzerine odaklanan LSTM gibi derin öğrenme yöntemlerinin** uygulanması planlanmaktadır. Ayrıca, modelin pratikte kullanılabilirliğini artırmak için bir web arayüzü geliştirilebilir ve tahminlerin görselleştirilmesi sağlanabilir.

EKİP ÜYELERİNİN PROJEYE KATKILARI;

Muzaffer Enes Yıldırım: Literatür taraması, sonuçların karşılaştırılması , Random Forest ve Decision Tree modellerinin eğitilmesi ve test edilmesi süreçleri, sunum karşılaştırılması, veri ön işleme aşamalarına katkı

Gökhan Altundal : Random Forest ve Decision Tree modellerinin eğitilmesi, modellerin performans analizleri, Confusion Matrix ve ROC-AUC çıktılarının yorumlanması

Mert Tosun : Veri setinin analizi, eksik verilerin temizlenmesi ve veri ön işleme süreci, Logistic Regression modelinin kurulumu, eğitimi ve performans değerlendirmesi, raporun yazımı.

**KAYNAKÇA**

1. **Kyner, T.J. (2020). *Rain in Australia – Next-Day Prediction Model*. GitHubRepository.Retrievedfrom** [**https://github.com/tjkyner/australian-rain-prediction**](https://github.com/tjkyner/australian-rain-prediction)
2. **Oni, O., & Oyediran, M. (2024). *Comparative Analysis of Weather Prediction Using Classification Algorithms: Random Forest, Decision Tree,andExtraTree*.ResearchGate.Retrievedfrom** [**https://www.researchgate.net/publication/380691926\_Comparative\_Analysis\_of\_Weather\_Prediction**](https://www.researchgate.net/publication/380691926_Comparative_Analysis_of_Weather_Prediction)
3. **Rawat, A., & Singh, V. (2023). *Weather Prediction Using Machine Learning Models: Logistic Regression, Random Forest, and XGBoost*. SSRN. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=4489040**
4. **Chittella, A., & Kiran, A. (2023). *Rainfall Prediction using Deep Learning and Machine Learning Techniques*. ResearchGate. Retrieved from** [**https://www.researchgate.net/publication/385527886\_Rainfall\_Prediction\_using\_Deep\_Learning\_and\_Machine\_Learning\_Techniques**](https://www.researchgate.net/publication/385527886_Rainfall_Prediction_using_Deep_Learning_and_Machine_Learning_Techniques)
5. **Australian Bureau of Meteorology. (2020). *WeatherAUS Dataset*. Kaggle.Retrievedfrom https://www.kaggle.com/datasets/jsphyg/weather-dataset-rattle-package**
6. **Scikit-learn Developers. (2023). *Scikit-learn: Machine Learning in Python*. Retrieved from https://scikit-learn.org/stable/**
7. **Python Software Foundation. (2023). *Python Documentation*. Retrieved from** [**https://docs.python.org/3/**](https://docs.python.org/3/)