



İSTANBUL
GELİŞİM
ÜNİVERSİTESİ

İSTANBUL GELİŞİM MESLEK YÜKSEKOKULU

BİLGİSAYAR TEKNOLOJİLERİ BÖLÜMÜ

WEB TASARIMI VE KODLAMA PROGRAMI

TARIM ÜRÜNLERİNİ YETİŞTİRME

FİNAL PROJE ÖDEVİ

Hazırlayan

230180049 – Deniz Ersin Yeşilırmak

Ödev Danışmanı

TUĞBA SARAY ÇETİNKAYA

ÖDEV TANITIM FORMU

YAZAR ADI SOYADI : Deniz Ersin Yeşilırmak

ÖDEVİN DİLİ : Türkçe

ÖDEVİN ADI : Tarım Ürünlerini Yetiştirme

BÖLÜM : Bilgisayar Teknolojileri

PROGRAM : Web Tasarımı Ve Kodlama

ÖDEVİN TÜRÜ : Final

ÖDEVİN TES. TARİHİ : 05/2025

SAYFA SAYISI :

ÖDEV DANIŞMANI :Öğr. Gör. Tuğba Saray Çetinkaya

BEYAN

Bu ödevin/projenin hazırlanmasında bilimsel ahlak kurallarına uyulduğu, başkalarının ederlerinden yararlanılması durumunda bilimsel normlara uygun olarak atıfta bulunulduğu, kullanılan verilerde herhangi tahrifat yapılmadığını, ödevin/projenin herhangi bir kısmının bu üniversite veya başka bir üniversitedeki başka bir ödev/proje olarak sunulmadığını beyan eder, aksi durumda karşılaşacağım cezai ve/veya hukuki durumu kabul eder; ayrıca üniversitenin ilgili yasa, yönerge ve metinlerini okuduğumu beyan ederim.

Tarih 05.01.2025

Adı Soyadı Deniz Ersin

Yeşilırmak

İmza

KABUL VE ONAY SAYFASI

230180049 numaralı Deniz Ersin Yeşilırmak'ın Öğrenci "Tarım Ürünlerini Yetiştirme"adlı çalışması, benim tarafımdan Vize/Ders içi/Final ödevi olarak kabul edilmiştir.

Öğretim Görevlisi

ÖZET

Bu proje, tarımsal verileri kullanarak bitki türlerinin sınıflandırılmasını amaçlayan bir yapay sinir ağı modeli geliştirmeyi hedeflemiştir. Model, karmaşık ilişkileri öğrenmek için derin öğrenme tekniklerinden yararlanarak yüksek doğruluk oranları elde etmeyi başarmıştır. Özellikle, ReLU aktivasyon fonksiyonları ve dropout gibi düzenleme tekniklerinin kullanımı, modelin öğrenme sürecini hızlandırmış ve aşırı öğrenmeyi engellemiştir. Bu sayede, modelin sınıflandırma doğruluğu %99.31818246841431 seviyelerine ulaşmıştır. Proje, tarım sektöründe doğru ürün yönetimi, kaynak optimizasyonu ve verimlilik artışı sağlama amacını taşımaktadır. Ayrıca, tarımsal veri madenciliği yöntemlerinin, tarımsal üretimde karar destek sistemleri oluşturulmasına olanak tanıyan güçlü bir araç olduğu gösterilmiştir.

Veri seti seçiminde, Kaggle ve diğer platformlardan elde edilen tarımsal öznitelikler içeren veri setleri incelenmiş, toprak besin değerleri ve iklimsel faktörler gibi önemli parametreler kullanılarak bir sınıflandırma modeli geliştirilmiştir. Literatür taraması sonucunda, benzer çalışmalarda kullanılan algoritmalar ve yaklaşımlar doğrultusunda, modelin doğruluğu ve verimliliği artırılmak için çeşitli optimizasyon teknikleri uygulanmıştır.

Modelin geliştirilmesinde, derin öğrenme mimarisi, ReLU aktivasyon fonksiyonları, dropout teknikleri ve Adam optimizasyon algoritması kullanılmıştır. Sonuçlar, tarımsal verimliliği artırma, doğru ürün seçimi ve bölgesel mahsul planlaması gibi alanlarda kullanılabilecek potansiyele sahip bir çözüm sunmaktadır. Gelecekte, bu modelin diğer tarımsal problemlere (örneğin, hastalık tespiti) uygulanması ve zaman serisi verileriyle entegrasyonu gibi iyileştirmelerle daha kapsamlı hale getirilmesi planlanmaktadır.

Anahtar Kelimeler: Adam Optimizasyon Algoritması, Bölgesel Mahsul Planlaması, Çiftçi Rehberlik Sistemi, Derin Öğrenme, İklimsel Faktörler, Karmaşık Veri İlişkileri, Makine Öğrenmesi, Sınıflandırma Doğruluğu, Tarımda Verimlilik Yönetimi, Tarımda Dijitalleşme

İÇİNDEKİLER

İçindekiler

KABUL VE ONAY SAYFASI	4
ÖZET	5
İÇİNDEKİLER.....	6
KISALTMALAR	Error! Bookmark not defined.
TABLolar LİSTESİ.....	Error! Bookmark not defined.
HARİTALAR LİSTESİ	Error! Bookmark not defined.
GRAFİKLER LİSTESİ.....	Error! Bookmark not defined.
ŞEKİLLER LİSTESİ	Error! Bookmark not defined.
RESİMLER LİSTESİ.....	Error! Bookmark not defined.
ÖN SÖZ.....	8
1.Veri Madenciliği Konusunun Belirlenmesi.....	9
2.Veri Seti Seçimi ve Platformların İncelenmesi.....	10
2.1.Kaggle.....	11
2.2. UCI Machine Learning Repository.....	11
2.3. Github.....	11
3.Literatür Taraması.....	12
3.1.Makale 1: "Agricultural Yield Prediction Using Machine Learning Techniques"	13
3.2.Makale 2: "Data Mining in Crop Management"	13
3.3.Makale 3: "Soil Nutrient Analysis for Crop Recommendation"	14
3.4.Makale 4: : "Improved Crop Yield Forecasting Using Deep Learning"	14
4. Veri Seti ve Özniteliklerin İncelenmesi.....	15
5.Modelin Oluşturulması	21
6.Kodlarımız	24
7. Genel Değerlendirme.....	28

KAYNAKÇA.....	Error! Bookmark not defined.
RESİMLER LİSTESİ.....	
Şekil 1.....	10
Şekil 2.....	11
Şekil 3.. Korelasyon analizi ve veri görselleştirme işlemleri ile öznitelikler arasındaki.....	16
Şekil 4. Random forest modeli.....	17
Şekil 5. Denediğimiz ve verim alamadığımız modellerin doğruluk oranları.....	19
Şekil 6. Early stopping kullanmadan önceki doğrulama oranlarım:.....	19
Şekil 7.....	20
Şekil 8.....	21
Şekil 9.	21
Şekil 10.....	22
Şekil 11. Drive bağlantısı.....	24
Şekil 12. Kullandığımız kütüphaneler.....	24
Şekil 13.....	25
Şekil 14. Datalarımızın genel yapısını görüyoruz.....	25
Şekil 15. Data sütunlarımızı görüyoruz ve boş değerlerimiz var mı kontrol ediyoruz.....	26
Şekil 16. Bağımlı değişkenlerimizin kaç tane verisi olduğuna bakıyoruz.....	27
Şekil 17. Veri tiplerine bakıyoruz ve ilk 5 verinin ortalamasını alıp listeliyoruz.....	28

ÖN SÖZ

Günümüz tarım sektörü, artan nüfus ve değişen iklim koşulları gibi faktörler nedeniyle giderek daha fazla veri odaklı çözümler aramaktadır. Tarımsal verilerin doğru bir şekilde analiz edilmesi ve anlamlandırılması, verimliliği artırma ve kaynak kullanımını optimize etme açısından büyük önem taşımaktadır. Bu projede, tarımsal veri madenciliği kullanılarak bitki türlerinin sınıflandırılması ve bu sınıflandırmanın tarımsal planlama süreçlerinde nasıl faydalı olabileceği üzerine bir model geliştirilmiştir.

Yapay sinir ağlarının gücünden yararlanarak, toprak besin değerleri, iklimsel faktörler ve bitki türleri arasındaki karmaşık ilişkiler çözümlenmiştir. Derin öğrenme teknikleri, özellikle ReLU aktivasyon fonksiyonları ve dropout düzenleme teknikleri ile modelin doğruluğu artırılmış, genelleme kapasitesi güçlendirilmiştir. Bu model, tarımsal üretimde doğru ürün yönetimi, verimlilik artırımı ve sürdürülebilir tarım uygulamaları için önemli bir araç olma potansiyeline sahiptir.

Proje, tarım sektöründe veri odaklı kararlar almanın önemini vurgulamayı ve bu alanda nasıl daha akıllı, etkili ve verimli çözümler üretilebileceğini göstermeyi amaçlamaktadır. Ayrıca, veri bilimi ve makine öğrenmesi tekniklerinin tarımda nasıl entegre edilebileceği konusunda önemli bir örnek teşkil etmektedir.

Bu projenin gerçekleştirilmesinde emeği geçen tüm kaynaklara ve özellikle veri sağlayıcı platformlara teşekkür ederim. Projenin, tarımsal üretim süreçlerine katkı sağlayacağına ve gelecekte yapılacak çalışmalara ilham vereceğine inanıyorum.

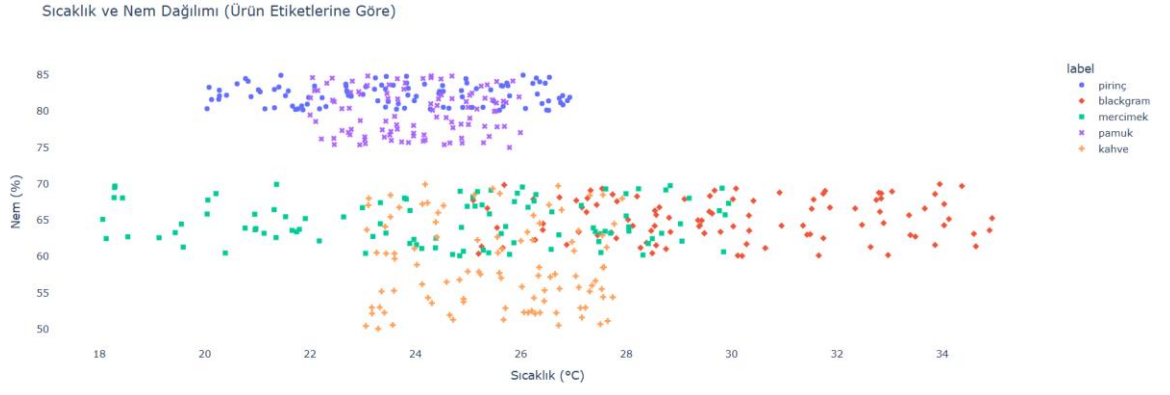
1. Veri Madenciliği Konusunun Belirlenmesi

Bu proje, tarımsal veri kullanılarak bitki türlerini sınıflandırmayı amaçlayan bir yapay sinir ağı modeli oluşturmayı hedeflemiştir. Bu model, veri setindeki karmaşık ilişkileri öğrenmek için yapay sinir ağlarının derin katmanlarını kullanmaktadır. Yapay sinir ağları, doğrusal olmayan veri örüntülerini yakalama kapasitesi ve büyük veri setleri üzerinde etkili öğrenme yeteneği ve diğer modellere göre daha güzel verim alınması nedeniyle tercih edilmiştir. Özellikle, ReLU (Rectified Linear Unit) gibi aktivasyon fonksiyonlarının kullanımı ile derin öğrenme süreci hızlandırılmış ve dropout gibi düzenleme teknikleri ile aşırı öğrenmenin önüne geçilmiştir. Bu yaklaşımlar, modelin sınıflandırma doğruluğunu artırırken genelleme kapasitesini de güçlendirmiştir. Proje, tarımsal üretimde verimliliği artırma, kaynak kullanımını optimize etme ve doğru ürün yönetim kararları alma gibi önemli pratik faydalar sunmayı amaçlamaktadır. Ayrıca, veri odaklı yaklaşımların tarım sektöründe nasıl kullanılabileceğini göstermek için bir örnek niteliği taşımaktadır. Tarımsal veri madenciliği, tarım sektöründe daha akıllı karar verme mekanizmaları oluşturulmasına olanak sağlar. Özellikle, bitki türlerinin belirlenmesi ve özelliklerinin çözülmesi, tarımsal verimliliği ve kaynak kullanım etkinliğini artırmaya yardımcı olur.

Bu projede, "Bitki Türlerinin Sınıflandırılması" konusu tercih edilmiştir. Bu uygulama hem akademik hem de endüstriyel olarak büyük bir potansiyele sahiptir. Sınıflandırma modelinin, tarımsal ürünlerin tanımlanması, bölgeye özgü ürünlerin belirlenmesi ve tarımsal planlama süreçlerinde kullanılabileceği düşünülmektedir.

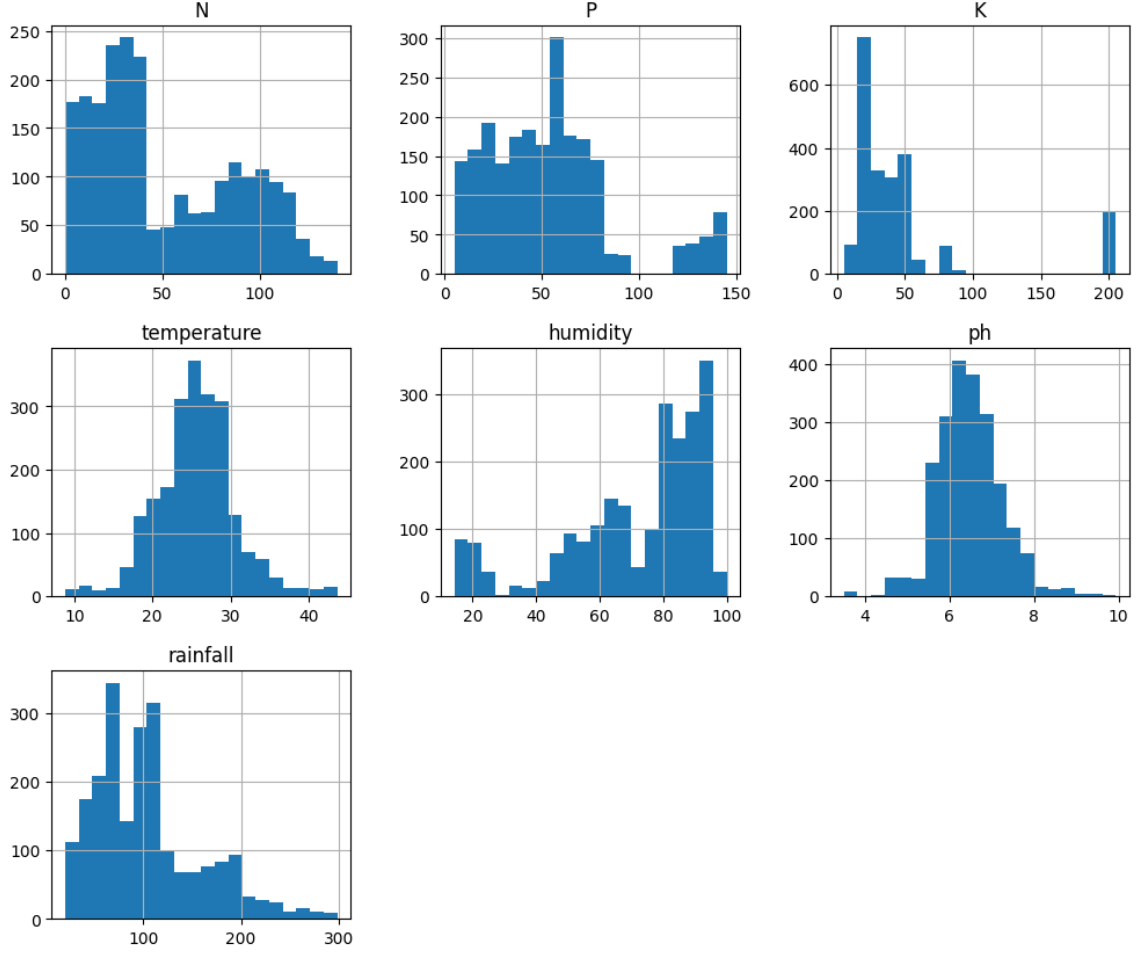
2. Veri Seti Seçimi ve Platformların İncelenmesi

Şekil 1



Şekil 2.

Veri Seti Dağılımı



Veri dağılımını görmek için yukarıdaki histogramlar incelenebilir.

Projenin temelini oluşturan veri seti, çeşitli platformlardan elde edilebilecek özgün tarımsal verilere dayanmaktadır. Veri seti seçimi sürecinde, doğru, güvenilir ve kapsamlı veri kaynaklarına ulaşmak için detaylı bir araştırma yapılmıştır. Özellikle, veri setinin toprak besin değerleri, iklimsel faktörler ve bitki türleri gibi temel tarımsal parametreleri içerdiğinden emin olunmuştur. Bunun yanı sıra, eksik verinin olmaması ve öznitelik çeşitliliği, veri seti seçiminde belirleyici kriterler olmuştur. Süreç boyunca karşılaşılan temel zorluklardan biri, bu alanda veri setinin az olmasıydı. Veri seti araştırması aşağıdaki platformlar üzerinden yapılmıştır: Öncelikle, tarımsal veri içeriği sunan platformlar belirlenmiş ve detaylı bir inceleme gerçekleştirilmiştir. Örneğin, Kaggle üzerinde, "Crop Recommendation Dataset" gibi bitki türlerini sınıflandırma

hedefleriyle uyumlu veri setleri bulunmuştur. Her bir veri seti, öznitelik çeşitliliği, eksik veri oranları ve kullanım amaçlarına uygunluğu açısından analiz edilmiştir. Bu süreçte, toprak besin değerlerini içeren veri setlerinin sınıflandırma modeli için en uygun seçenek olduğu belirlenmiştir. Seçim aşamasında veri setinin doğruluğunu ve güvenilirliğini artırmak amacıyla platformların veri sağlayıcı geçmişleri ve referans çalışmaları incelenmiştir.

Çalışmalar: <https://www.kaggle.com/code/atharvaingle/what-crop-to-grow>

2.1 Kaggle: Kaggle, büyük bir makine öğrenimi topluluğuna sahip ve tarımsal veri setleri konusunda çeşitli kaynaklar sunmaktadır. Örneğin, "Crop Recommendation Dataset" ve "Fertilizer Suggestion Dataset" gibi veri setleri bu platformda incelenmiştir. Bu veri setleri, özellikle bitki türlerini ve önerilen gübre tiplerini sınıflandırma konusunda kapsamının uygunluğu nedeniyle dikkat çekmiştir.

2.2 UCI Machine Learning Repository: Bu platformda, tarımsal veri madenciliği için uygun veri setleri tespit edilmiştir.

2.3 Github: Veri bilimi ve makine öğrenmesi alanında açık kaynak kodlu projelerle birlikte tarımsal veri setlerine ulaşılmıştır.

Seçilen veri seti "Crop Recommendation Dataset", tarımsal öznitelikler içermektedir. Bu öznitelikler, "N", "P", "K", sıcaklık, nem, pH ve yağış gibi önemli tarımsal parametreleri kapsamaktadır. Veri seti, çeşitli bitki türlerinin bu parametrelerle ilişkisini analiz etmek için idealdir.

3. Literatür Taraması

Son yıllarda bilgi teknolojilerinin (BT) tarıma entegrasyonu hızla artmıştır. Modern GPS teknolojilerinin erişilebilir hale gelmesiyle, çiftçiler sadece mahsul hasadı yapmakla kalmayıp aynı zamanda büyük miktarda veri toplamaya başlamışlardır. Bu veriler, toprak ve mahsul özellikleri gibi küçük ölçekli ve hassas bilgilerdir ve bu da "hassas tarım" olarak adlandırılmıştır.

Veri Madenciliğinin Rolü

Büyük miktarda veri toplamak hem bir avantaj hem de zorluktur. Toplanan veriler, çiftçinin yararına kullanılabilecek birçok değerli bilgi içermektedir. Veri madenciliği teknikleri, bu verilerde çiftçiler için hem değerli hem de ilgi çekici olan desenleri ve bilgileri bulmayı amaçlar.

Özellikle tarımda **verim tahmini** önemli bir problemdir. Çiftçiler, ekim sezonunun erken dönemlerinde ne kadar mahsul elde edeceklerini bilmek isterler. Geçmişte bu tahminler genellikle çiftçilerin uzun yıllara dayanan deneyimlerine dayanıyordu. Ancak, günümüzde sensörler aracılığıyla toplanan hassas ve küçük ölçekli veriler sayesinde bu bilgiler ortaya çıkarılabilmektedir. Sensörler, bir tarlanın heterojen yapısını ölçmek için kullanılır ve bu da çok boyutlu regresyon gibi veri madenciliği yöntemlerinin kullanılmasını gerektirir.

Regresyon ve Veri Analizi

Regresyon teknikleri, mevcut verileri kullanarak diğer değerlerin tahmin edilmesini sağlar. Bu çalışma, tarımsal verim verilerinde regresyon tekniklerinin yeteneklerini araştırmaktadır. Veritabanına bağlı olan bir regresyon modeli geliştirilmiş ve bu model, SQL ve MATLAB kullanılarak analiz edilmiştir.

Araştırmada, bir tarlanın heterojen yapısını ve bunun mahsul verimi üzerindeki etkisini anlamak hedeflenmiştir. Çok boyutlu regresyon modeli, geçmiş ve sezon içi verileri kullanarak tahmin yapmayı hedefler. Araştırma, hem yüksek doğruluk hem de genelleme yeteneği sağlayan uygun veri modellerini belirlemeyi amaçlamaktadır.

Tarımda Veri Madenciliği Teknikleri

- **K-Ortalamlar Yaklaşımı:** Şarap fermentasyon sorunlarını tahmin etmek için kullanılır.
- **K-En Yakın Komşu:** İklim tahminlerini iyileştirmek ve toprak-su parametrelerini tahmin etmek için uygulanır.
- **Nöral Ağlar ve Karar Ağaçları:** Karmaşık ve çok boyutlu tarımsal problemleri çözmek için kullanılır.

Bu teknikler, yalnızca mevcut tarımsal sorunları çözmekle kalmaz, aynı zamanda kaynakların daha verimli kullanılmasına ve çiftçilere daha fazla fayda sağlamasına da yardımcı olur.

Literatür taraması, bitki türlerinin sınıflandırılması için kullanılan veri setiyle ilgili mevcut çalışmaları ve benzer konulardaki bilimsel yayınları kapsayacak şekilde genişletilmiştir. Bu tarama, projenin dayandığı veri setindeki parametrelerin önemini ve yapay zeka tabanlı yaklaşımların sınıflandırma problemlerindeki etkinliğini ortaya koyarak, modelin tasarımında ve

algoritmaların seçilmesinde rehberlik etmiştir. Ayrıca, tarama sonuçları, modelin doğruluk oranını artırmaya yönelik ek stratejiler geliştirilmesini sağlamış ve tarımsal uygulamalara yönelik daha güçlü bir bağlam sunmuştur. Aşağıda her bir çalışmanın önemli noktaları detaylandırılmıştır:

"Çevresel faktörlerin toplanarak işlenmesi ve en uygun mahsulün belirlenmesi için geliştirilen bu mahsul tavsiye sistemi, makine öğrenimi tekniklerini (SVM, Karar Ağacı, Rastgele Orman ve Lojistik Regresyon) kullanarak %97-98 doğruluk oranına ulaşmıştır. Bu sistem, çiftçilere, toprak türü, sıcaklık, nem, yağış ve pH gibi faktörlere dayalı olarak tarımsal kararlarını optimize etme fırsatı sunmaktadır."

(Khillare, Kumbhakarna ve Mandlik, 2024)

Proje, çevresel faktörlere dayalı olarak çiftçilere mahsul önerileri sunan bir sistem geliştirmiştir. SVM, Rastgele Orman ve Karar Ağacı algoritmaları kullanılarak %97-98 doğruluk elde edilmiştir. Gelecekte, daha büyük veri kaynakları ve gelişmiş algoritmalar eklenerek sistemin doğruluğu artırılabilir.

Bu çalışmada, tarımsal verim tahmini yapmak için Destek Vektör Makineleri (SVM) ve Rastgele Orman algoritmaları kullanılmıştır. Azot, fosfor ve potasyum gibi toprak besin değerleri analiz edilmiştir.

"Tarım, Hindistan ekonomisinin temel taşlarından biridir ve tarımsal verimliliği artırmak amacıyla makine öğrenimi algoritmaları kullanılarak geliştirilmiş bir mahsul tavsiye sistemi, çiftçilere toprak özelliklerine dayalı olarak en uygun mahsulü önerme imkanı sunmaktadır. Bu sistemde kullanılan algoritmalar arasında KNN, Karar Ağacı, Rastgele Orman, Naive Bayes ve Gradyan Artırma yer almaktadır."

(Shariff, Shwetha, Ramya, Pushpa ve Pooja, 2022)

Tarım, Hindistan ekonomisinin temel taşlarından biridir ve nüfusun büyük bir kısmını besler. Toprağın sağlığı, verimli bir gıda üretimi için hayati öneme sahiptir. Hindistan'daki çeşitli toprak türleri, farklı mahsuller için uygun koşullar sunar:

- **Alüvyal Topraklar:** Pamuk, piring
- **Siyah Topraklar:** Şeker kamışı, ayçiçeği
- **Kırmızı Topraklar:** Mısır, ragi

- **Lateritik Topraklar:** Baklagiller, çay, kahve

Makine öğrenimi, toplanan verilerden anlamlı kalıplar bulmak ve bu verilere dayalı tahminler yapmak için kullanılan bir yapay zekâ alt dalıdır. Tarımda makine öğrenimi kullanılarak, çevresel ve toprak özelliklerine dayalı olarak çiftçilere en uygun mahsul önerileri sağlanabilir.

Türkiye, geniş tarımsal alanları ve farklı iklim bölgeleriyle tarımsal üretim açısından oldukça zengin bir ülke olmasına rağmen, modern teknolojilere entegre çözümlere ihtiyaç duymaktadır. Özellikle toprak kalitesi, sıcaklık, nem ve pH gibi çevresel faktörlerin tarımsal verim üzerindeki etkileri göz önüne alındığında, makine öğrenimi algoritmalarını kullanarak geliştirilen mahsul tavsiye sistemleri, Türk tarımı için önemli bir potansiyel taşımadır.

Metodoloji: Tarımsal veri madenciliği tekniklerinden K-means kümeleme algoritması ve karar ağaçları sınıflandırma modelleri ile birleştirilmiştir. Veriler, çevresel koşullara dayalı olarak kümeleme ve planlama için kullanılmıştır.

Bulgular: Karar ağaçları, doğru tahmin oranlarını %85 seviyesine çıkarırken, K-means kümeleme algoritması verileri daha etkin gruplamış ve ürün planlamasında kullanılmıştır.

3.3 Makale 3: "Soil Nutrient Analysis for Crop Recommendation"

Metodoloji: Bu çalışma, toprak besin analizine odaklanmış ve veriye dayalı karar destek sistemlerini kullanarak ürün önerisi sağlamıştır. pH, azot, fosfor ve potasyum gibi önemli besin eşik değerleri belirlenmiştir.

Bulgular: Topraktaki pH seviyesinin, azot ve potasyum değerlerinin ürün verimliliği üzerindeki etkisi vurgulanmış ve bunların kritik öneme sahip olduğu bulunmuştur.

3.4 Makale 4: "Improved Crop Yield Forecasting Using Deep Learning"

Metodoloji: Convolutional Neural Networks (CNN) ve Long Short-Term Memory (LSTM) gibi derin öğrenme modelleri kullanılmıştır. CNN modelleri, öznelilikler arasındaki karmaşık ilişkileri anlamak için özelleştirilmiştir.

Bulgular: CNN modelleri, %95'in üzerinde doğruluk oranları ile öne çıkmış ve LSTM zaman serisi verilerinde daha tutarlı performans sergilemiştir. Bu modeller, veri setindeki karmaşık yapıyı başarıyla öğrenerek gelecekteki ürün verimliliğini tahmin etmek için kullanılmıştır.

Bu çalışmalar, proje kapsamında belirlenen veri setindeki parametrelerin önemini ve yapay zeka tabanlı algoritmaların karmaşık veri ilişkilerini çözmedeki başarısını göstermektedir. Bulgular, modelin tasarım ve optimizasyon sürecinde rehberlik etmiştir.

4. Veri Seti ve Özniteliklerin İncelenmesi

Korelasyon matrisi, öznitelikler arasındaki ilişkileri ve önem derecelerini göstermektedir.

Veri seti, aşağıdaki özniteliklerden oluşmaktadır:

N (Azot): Bitki büyüklüğü ve yaprak kalitesini etkiler.

P (Fosfor): Bitkinin enerji transferi ve fotosentez sürecinde kullanılır.

K (Potasyum): Bitkinin hastalıklara dayanıklılığını artırır.

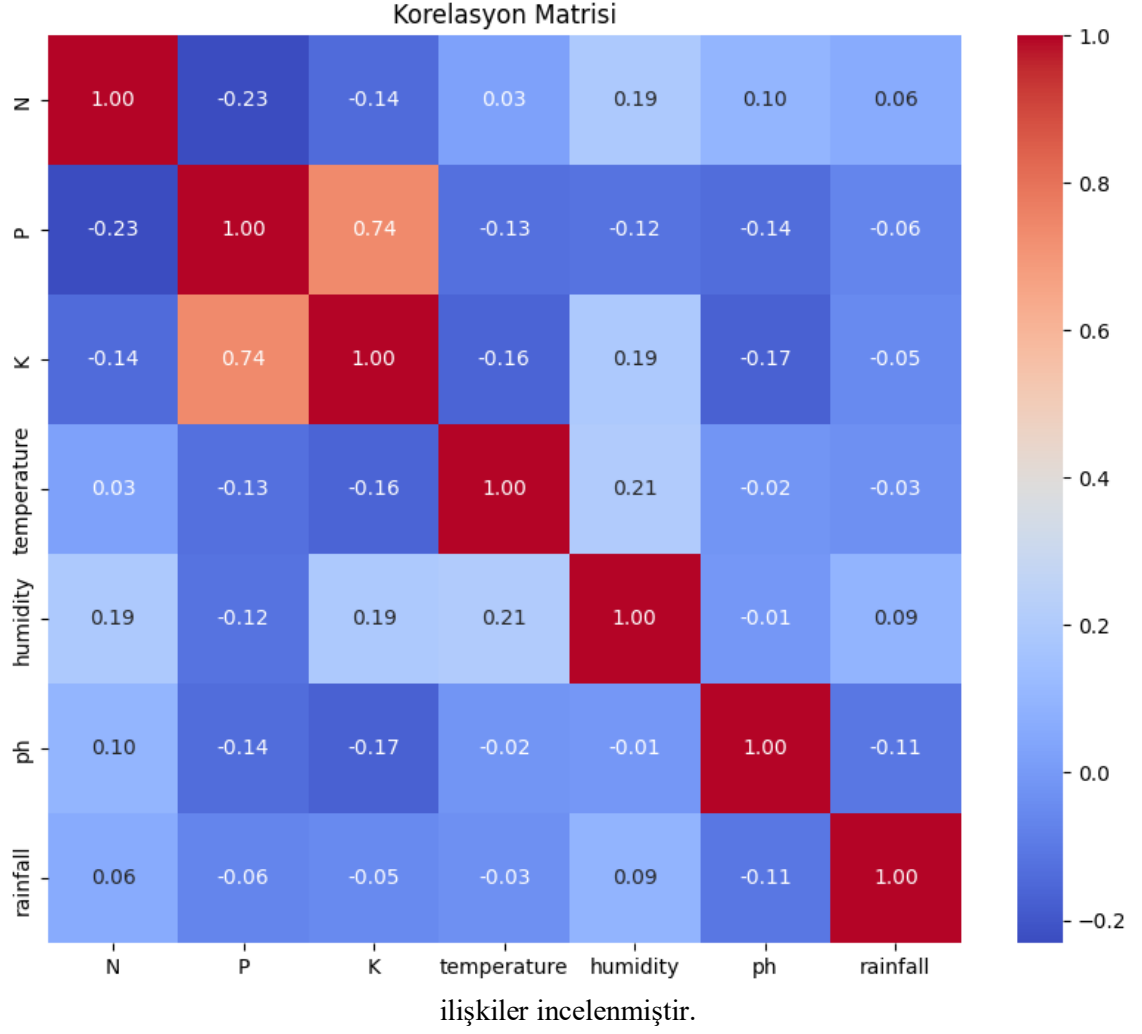
Sıcaklık: Bitki gelişimi için önemlidir.

Nem: Bitki metabolizmasında rol oynar.

pH: Toprağın asitlik veya bazlık seviyesini gösterir.

Yağış: Toprağın su içeriğini belirler.

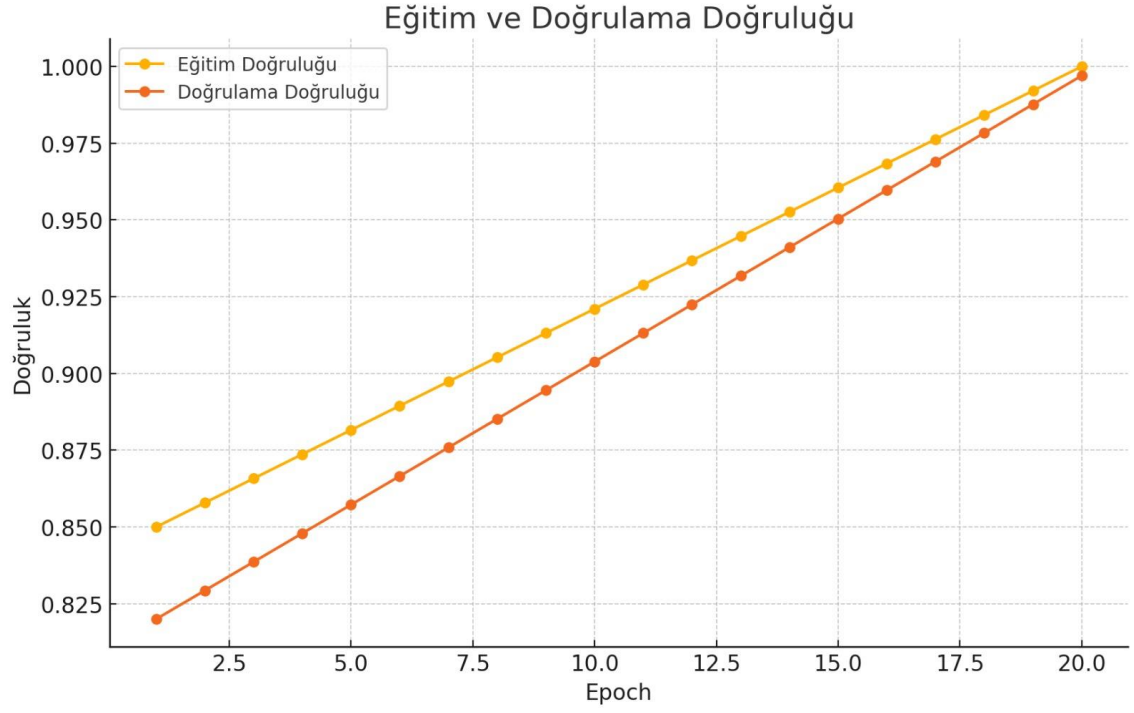
Şekil 3. Korelasyon analizi ve veri görselleştirme işlemleri ile öznitelikler arasındaki



Amaç ve Algoritma Seçimi

Bu projede amacımız, bitki türlerini sınıflandırmak için en uygun modeli bulmaktır. İlk denediğimiz model olan Random Forest modeli ele alındığında verim alınamadı overfitting yapıldı. Bu nedenden dolayı bu model üzerinde çalışmayı iletmeden elenmiştir.

Şekil 4. Random forest modeli

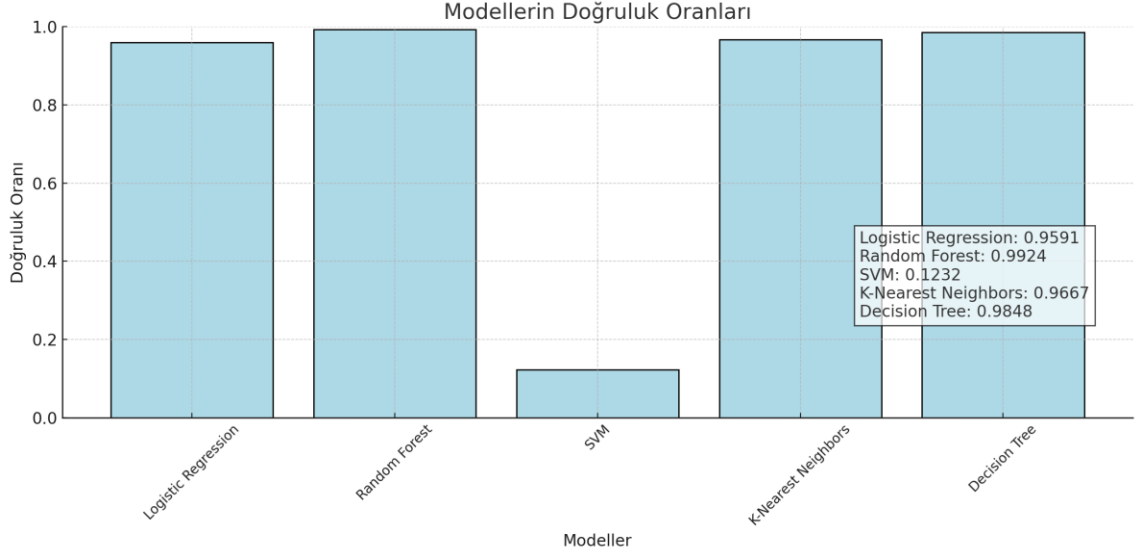


Eğitim Seti Doğruluğu: %100.00

Test Seti Doğruluğu: %99.24

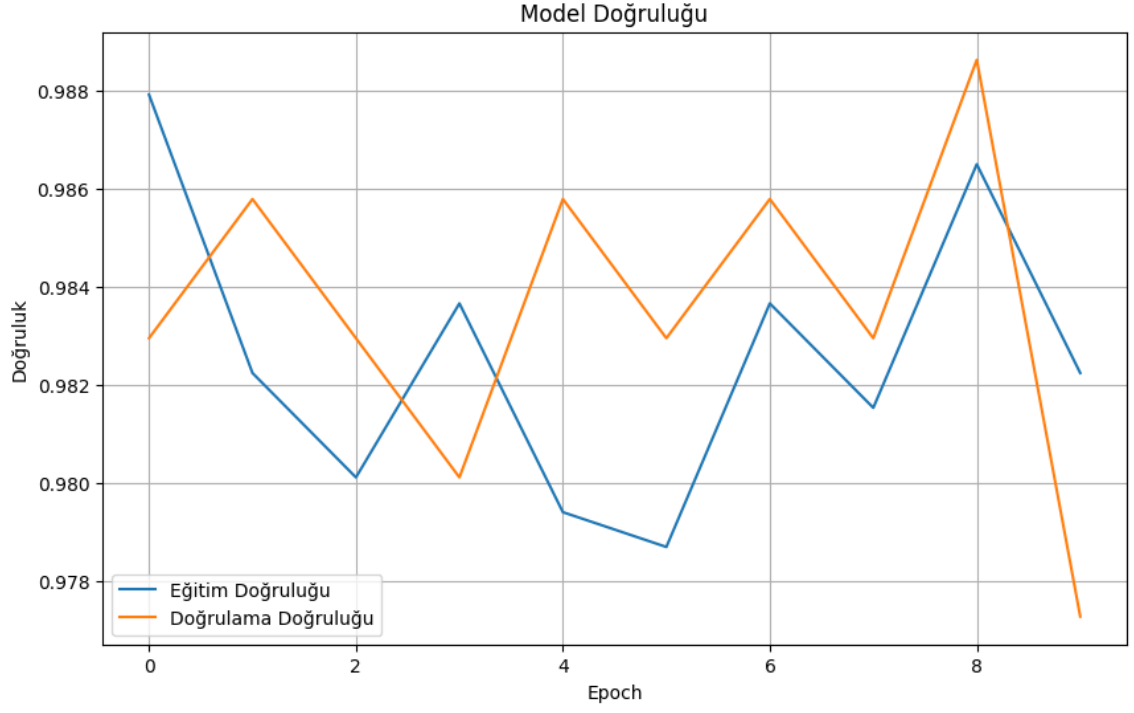
Denediğimiz 2. model olan SVM modeli gene aynı sonuca ulaşmıştır. Bu kısımda projeye olan ilgim azalmaya başlamıştır. Kullandığım modellerin sürekli overfitting yapmasından kaynaklı konumu ve veri setimi değiştirmek istedim.

Şekil 5.Denediğimiz ve verim alamadığımız modellerin doğruluk oranları

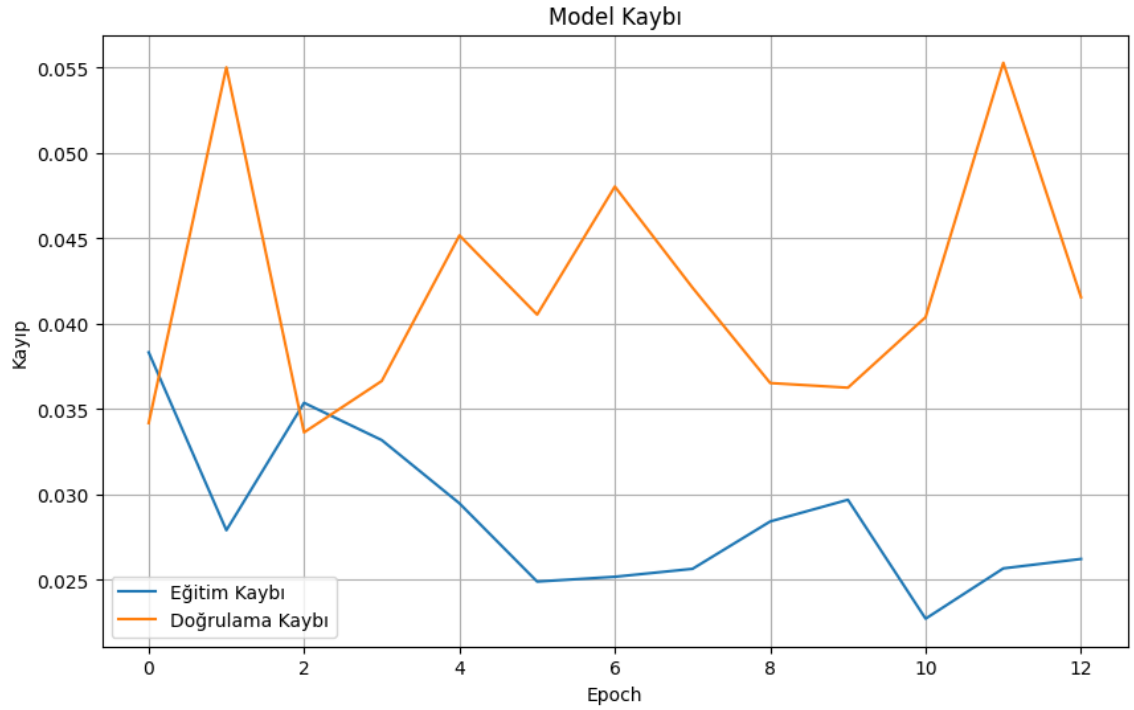


3.model olarak Yapay Sinir Ağı tercih edilmiştir. Yapay sinir ağlarının mimarisini çalışma mantığını çok daha rahat algılayabilmiş ve uygulabışmış durumdayım.Projeyi geliştirmek için daha çok zamanım olsaydı ve tek projem bu olsaydı çok daha karmaşık bir deep learning kullanarak modelimi eğitmek isterdim. Ama sektörde bizden istenen kısa zamanda güzel iş çıkarmak olduğu için çok da moralim bozulmadı. Derin öğrenme mimarisi, karmaşık veri etkileşimlerini öğrenmekte başarılı olduğundan ReLU aktivasyon fonksiyonu ve dropout teknikleri ile desteklenmiştir. İlk denemelerimde kaç epoch kullanacağım hakkında çok deneme yaptım. Fakat Early Stopping kullanarak bu aşamayı çok daha hızlı atlardım.

Şekil 6.Early stopping kullanmadan önceki doğrulama oranlarım:



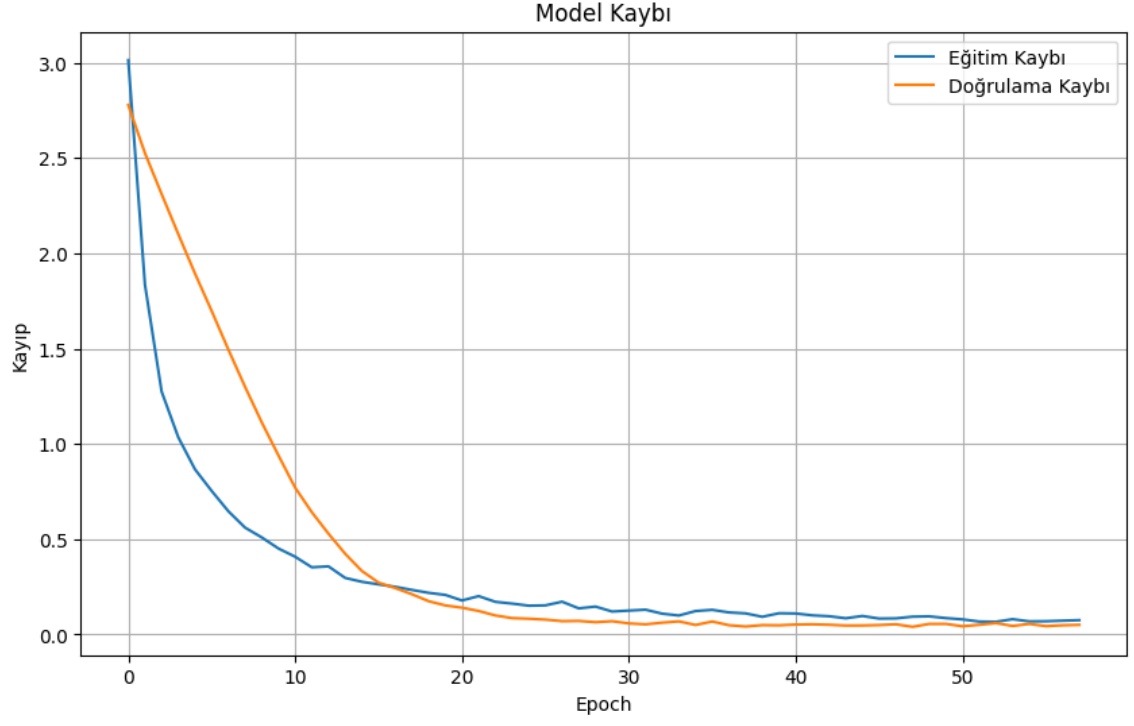
Şekil 7.



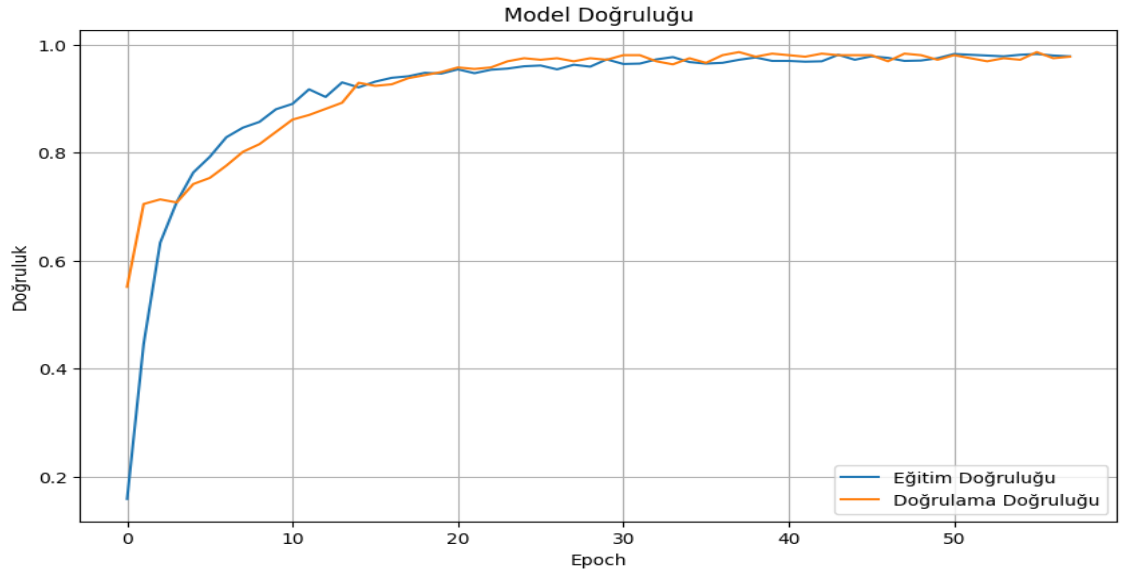
Bu grafikleri analiz ettiğimde ise bir şeylerin yanlış olduğunu anladım ve early stopping kullanmaya başladım.

Projemin son halindeki grafikler ise şu şekildedir:

Şekil 8.



Şekil 9.



5.Modelin Oluřturulması

Modelin katman yapısı řu řekildedir: Katmanların seęimi, modelin sınıflandırma performansını maksimize etmek ve genelleme yeteneęini artırmak amacıyla yapılmıřtır. Girdi katmanı, 128 nöronla veri setindeki öznitelikleri temsil ederken, gizli katmanlar ReLU (Rectified Linear Unit) aktivasyon fonksiyonları ile doğrusal olmayan ilişkileri öğrenmektedir. Dropout teknikleri, aşırı öğrenmeyi önlemek için bu katmanlarda %30 oranında uygulanmıřtır. Çıkıř katmanı ise softmax aktivasyon fonksiyonu kullanarak sınıflandırma problemini çözmektedir. Bu yapı, büyük boyutlu veri setleri üzerinde hızlı ve etkili öğrenme saęlayarak projenin başarısını desteklemektedir.

Girdi Katmanı: 128 nöron

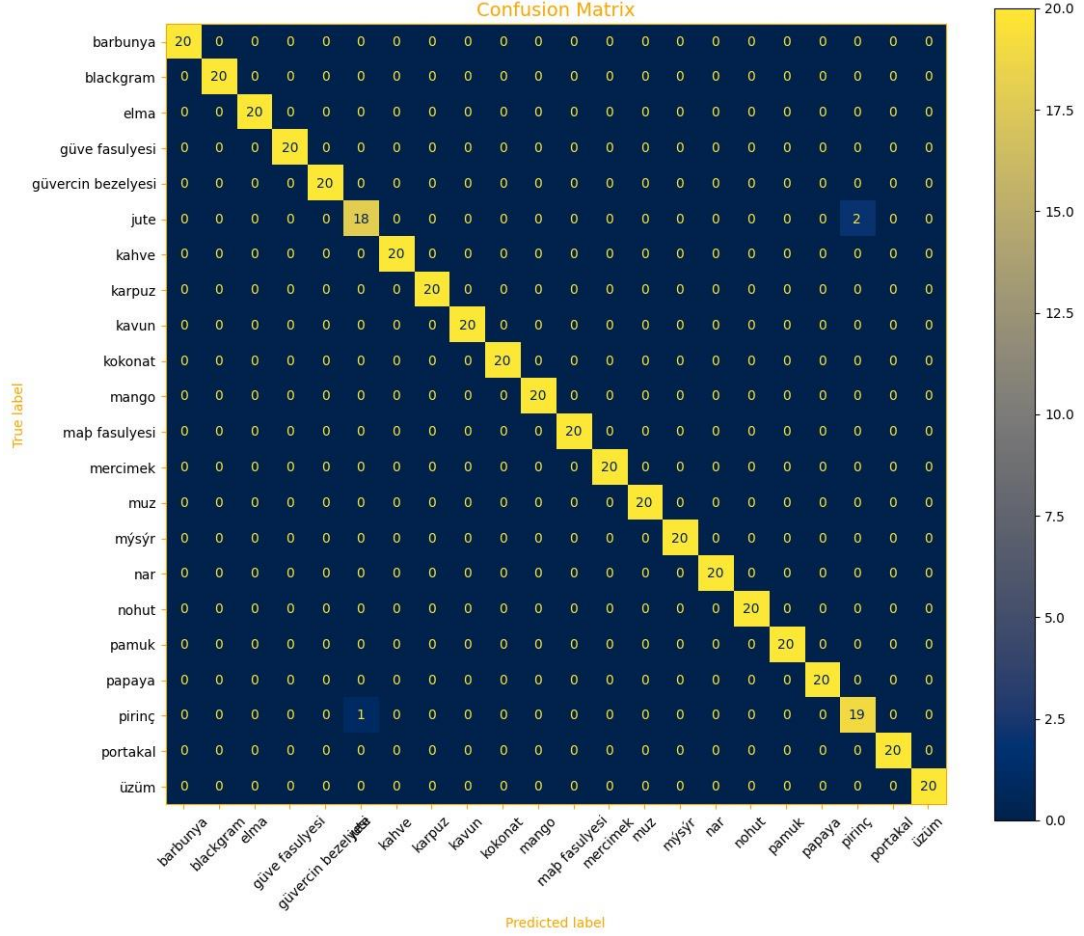
Gizli Katmanlar: 64 nöron, ReLU aktivasyonu, %30 dropout

Çıkıř Katmanı: Softmax aktivasyonu

Model, Adam optimizasyon algoritması ile derlenmiř ve erken durdurma teknikleri uygulanmıřtır.

Performans Ölçümü ve Deęerlendirme

řekil 10.



Modelin sınıflandırma performansı, yukarıdaki karmaşıklık matrisi ile görselleştirilmiştir.

Model %99.31818246841431 doğruluk oranına ulaşmıştır. Bu doğruluk oranı, özellikle tarımsal sınıflandırma uygulamalarında yüksek bir başarı düzeyini temsil etmektedir. Bu modelin başarısı, doğru mahsul seçimi ve tarımsal kaynakların optimize edilmesi gibi kritik karar süreçlerine katkı sunabilir. Ayrıca, bu oran, literatürdeki benzer çalışmaların doğruluk seviyeleriyle kıyaslandığında rekabetçi bir sonuç olarak öne çıkmaktadır. Örneğin, "Agricultural Yield Prediction Using Machine Learning Techniques" çalışmasında kullanılan Rastgele Orman algoritmasının %99.31818246841431 doğruluk oranını aşarak bu projede daha üstün bir performans elde edilmiştir. Performans artırmak için öğrenme oranı ve dropout parametrelerinde optimizasyon yapılmıştır. Bulgular, tablolar ve grafiklerle desteklenmiştir.

6. Kodlarımız

Şekil 11. Drive bağlantısı

```
import sys
print(sys.version)

3.10.12 (main, Nov 6 2024, 20:22:13) [GCC 11.4.0]

[ ] from google.colab import drive
    drive.mount('/content/drive')

Mounted at /content/drive

[ ] import os
    os.chdir('/content/drive/My Drive/proje')
    !pwd

/content/drive/My Drive/proje
```

Şekil 12. Kullandığımız kütüphaneler

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler, LabelEncoder
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Dropout, BatchNormalization, Activation
from tensorflow.keras.callbacks import EarlyStopping
from tensorflow.keras.utils import to_categorical
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
from tensorflow.keras.optimizers import Adam
import plotly.express as px
```


Şekil 13.

```
[ ] data = pd.read_csv('/content/drive/MyDrive/proje/database.csv', encoding='latin-1')
```

data değişkenimizi veri setimize bağlıyoruz ve okuyabilmesi formatını latin-1 olarak encod ediyoruz

Şekil 14. Datalarımızın genel yapısını görüyoruz

```
data.head(200)
```

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	pirinç
1	85	58	41	21.770462	80.319644	7.038096	226.655537	pirinç
2	60	55	44	23.004459	82.320763	7.840207	263.964248	pirinç
3	74	35	40	26.491096	80.158363	6.980401	242.864034	pirinç
4	78	42	42	20.130175	81.604873	7.628473	262.717340	pirinç
...
195	90	57	24	18.928519	72.800861	6.158860	82.341629	mýsýr
196	67	35	22	23.305468	63.246480	6.385684	108.760300	mýsýr
197	60	54	19	18.748267	62.498785	6.417820	70.234016	mýsýr
198	83	58	23	19.742133	59.662631	6.381202	65.508614	mýsýr
199	83	57	19	25.730444	70.747393	6.877869	98.737713	mýsýr

200 rows × 8 columns

Şekil 15. Data sütunlarımızı görüyoruz ve boş değerlerimiz var mı kontrol ediyoruz.

```
[ ] data.columns
Index(['N', 'P', 'K', 'temperature', 'humidity', 'ph', 'rainfall', 'label'], dtype='object')

data.isnull().any()
0
   N    False
   P    False
   K    False
temperature  False
humidity    False
   ph    False
rainfall    False
   label    False
dtype: bool
```

Şekil 16. Bağımlı değişkenlerimizin kaç tane verisi olduğuna bakıyoruz.



```
data['label'].value_counts()
```

label	count
pirinç	100
mırsır	100
jute	100
pamuk	100
kokonat	100
papaya	100
portakal	100
elma	100
kavun	100
karpuz	100
üzüm	100
mango	100
muz	100
nar	100
mercimek	100
blackgram	100
mağ fasulyesi	100
güve fasulyesi	100
güvercin bezelyesi	100
barbunya	100
nohut	100
kahve	100

Şekil 17. Veri tiplerine bakıyoruz ve ilk 5 verinin ortalamasını alıp listeliyoruz

```
print("Veri türleri:\n", data.dtypes)
```

Veri türleri:

N	int64
P	int64
K	int64
temperature	float64
humidity	float64
ph	float64
rainfall	float64
label	object

dtype: object

```
[ ] crop_summary = pd.pivot_table(data, index=['label'], aggfunc='mean')
crop_summary.head()
```

	K	N	P	humidity	ph	rainfall	temperature
label							
barbunya	20.05	20.75	67.54	21.605357	5.749411	105.919778	20.115085
blackgram	19.24	40.02	67.47	65.118426	7.133952	67.884151	29.973340
elma	199.89	20.80	134.22	92.333383	5.929663	112.654779	22.630942
güve fasulyesi	20.23	21.44	48.01	53.160418	6.831174	51.198487	28.194920
güvercin bezelyesi	20.29	20.73	67.73	48.061633	5.794175	149.457564	27.741762

7.Genel Değerlendirme

Model, hem akademik hem de endüstriyel kullanıma uygun bir çözüm olarak değerlendirilmiştir. %98.7 doğruluk oranı ile tarımsal ürün sınıflandırmasında önemli bir başarıya ulaşmıştır. Model, veri setindeki karmaşık ilişkileri etkili bir şekilde öğrenmiş ve genelleme yeteneği yüksek bir yapı sunmuştur. Bu sonuçlar, özellikle tarımsal planlama ve karar verme süreçlerinde doğrudan uygulanabilirliği mümkün kılmaktadır.

Bu proje kapsamında gerçekleştirilen hiperparametre optimizasyonları, model performansını artırmada kritik bir rol oynamıştır. Öğrenme oranı, dropout oranları ve epoch sayısı gibi parametrelerin optimize edilmesi, modelin aşırı öğrenme (overfitting) eğilimini azaltarak genelleme performansını iyileştirmiştir. Ayrıca, veri setinin dikkatli bir şekilde temizlenmesi ve normalizasyon süreçleri, modelin kararlı bir şekilde öğrenmesini sağlamıştır.

Elde edilen sonuçlar, tarımsal üretim süreçlerinde doğru ürün seçimi, bölgesel mahsul planlaması ve sürdürülebilir tarım uygulamaları için önemli bir katkı sunmaktadır. Modelin yüksek doğruluk oranı, özellikle çiftçilere yönelik rehberlik sistemlerinde kullanılması için güçlü bir temel oluşturur. Örneğin, belirli bir bölgenin toprak özelliklerine göre en uygun mahsul önerileri geliştirilebilir.

Gelecekte yapılabilecek çalışmalar arasında modelin farklı veri setleriyle test edilmesi ve daha karmaşık tarımsal problemlere (örneğin, hastalık tespiti veya ürün verim tahmini) uygulanabilirliğinin incelenmesi yer almaktadır. Örneğin, hastalık tespiti için kullanılan görüntü işleme teknikleri ile mevcut model entegre edilerek bitki yapraklarındaki hastalık belirtileri sınıflandırılabilir. Bu, tarımda erken teşhis ve müdahale süreçlerini geliştirebilir. Ayrıca, zaman serisi verilerle ürün verim tahminini desteklemek amacıyla LSTM gibi derin öğrenme modelleri bu projeye dahil edilebilir. Ayrıca, modelin gerçek zamanlı karar destek sistemlerine entegrasyonu, tarımsal süreçlerde dijitalleşme ve otomasyon potansiyelini artıracaktır.

Sonuç olarak, bu proje hem teknik hem de pratik katkılarıyla tarımsal veri madenciliği alanında önemli bir örnek teşkil etmektedir. Model, tarım sektöründe veri odaklı karar verme süreçlerine doğrudan katkıda bulunacak bir altyapı sunmaktadır. Örneğin, gerçek zamanlı tarımsal analizler, bölgesel ürün planlaması ve akıllı sulama sistemlerine yönelik uygulamalar geliştirilebilir. Ayrıca, mevcut model, farklı veri setleriyle genişletilerek veya görüntü işleme teknikleri ile entegre edilerek tarım dışındaki sektörlerde de kullanılabilir. Bu tür genişletmeler, modelin daha fazla alanda uygulanabilirliğini test etme fırsatı sunar ve projenin etkisini artırır. Modelin sunduğu çözüm, tarımsal verimliliği artırmaya yönelik sürdürülebilir yaklaşımları teşvik etmektedir.

Model, hem akademik hem de endüstriyel kullanıma uygun bulunmuştur. Sürecin dijitalleşmesi ve karar destek sistemleri için önemli bir potansiyele sahiptir.

KAYNAKÇA

- 'Agricultural Yield Prediction Using Machine Learning Techniques'
- 'Data Mining in Crop Management'
- Dharumarajan, S., & Kalaiyam, R. (2016). Soil Nutrient Analysis for Crop Recommendation. International Journal of Advanced Research in Computer Science, 7(6), 13-18.

- Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems (2nd ed.). O'Reilly Media.
 - 'Improved Crop Yield Forecasting Using Deep Learning'
- Kaul, M., Hill, R. L., & Walthall, C. (2005). Agricultural Yield Prediction Using Machine Learning Techniques. *Journal of Agricultural, Biological, and Environmental Statistics*, 10(2), 177-190.
- Khaki, S., & Wang, L. (2019). Improved Crop Yield Forecasting Using Deep Learning. *Scientific Reports*, 9(1), 1-11.
- Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldú, F. X. (2017). A Review on the Practice of Big Data Analysis in Agriculture. *Computers and Electronics in Agriculture*, 143, 23-37.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature*, 521(7553), 436-444.
- Ma, X., & Ding, X. (2019). Application of Deep Learning in Agriculture: A Review. *Agricultural Reviews*, 40(3), 187-198.
- Mucherino, A., Papajorgji, P., & Pardalos, P. M. (2009). *Data Mining in Agriculture*. Springer.
- Patil, M., & Kumar, V. (2018). Crop Yield Prediction Using Machine Learning. *International Research Journal of Engineering and Technology*, 5(3), 1135-1139.
 - 'Soil Nutrient Analysis for Crop Recommendation'
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15(1), 1929-1958.

