



Kahramanmaraş Sutcu Imam University

Journal of Engineering Sciences



Geliş Tarihi : 16.07.2024
Kabul Tarihi : 30.10.2024

Received Date : 16.07.2024
Accepted Date : 30.10.2024

MAKİNE ÖĞRENMESİ İLE ÜRETİM PERFORMANSI TAHMİNLEMESİ

PREDICT OF PRODUCTION PERFORMANCE WITH MACHINE LEARNING

Semih GÖKSU^{1*} (ORCID: 0009-0007-8158-6654)
Prof.Dr.Bülent SEZEN² (ORCID: 0000-0001-7485-3194)
Dr.Yavuz Selim BALCIOĞLU³ (ORCID: 0000-0001-7138-2972)

^{1,2,3} Gebze Teknik Üniversitesi, İşletme Fakültesi, İşletme Bölümü, Kocaeli, Türkiye

*Sorumlu Yazar / Corresponding Author: Semih GÖKSU, s.goksu2022@gtu.edu.tr

ÖZET

Bu çalışmanın amacı, makine öğrenmesi algoritmaları kullanarak üretim performansının tahmin edilmesidir. Üretim sistemleri, çeşitli makineler, parametreler ve set değerler üzerinden çalışmakta olup, her bir üretim koşulu değişkenlik gösterebilmektedir. Gelişen teknolojiler sayesinde, bu değişkenliklerin kontrol altına alınması, üretim koşullarının optimize edilmesi ve birbirini etkileyen süreçlerden çıkarım yapılması mümkün hale gelmiştir. Bu bağlamda, makine öğrenmesi, istatistiksel metotlarla veri setleri üzerinden üretim performansının tahmin edilmesine olanak tanıyan önemli bir araçtır. Çalışmada, aynı ürün grubuna ait 2 yıllık veri kullanılarak üretim performansını tahmin etmek amacıyla Karar Ağacı, Lineer Regresyon, Lasso Regresyon, XGBoost, Destek Vektör Regresyonu ve LSTM algoritmaları uygulanmıştır. Bu algoritmalar, üç farklı senaryo üzerinden değerlendirilmiş ve üretim performansını en doğru şekilde tahmin edebilecek modelin belirlenmesi hedeflenmiştir. Elde edilen sonuçlar, basit ve karmaşık modellerin performanslarını karşılaştırarak üretim süreçlerinin iyileştirilmesine yönelik pratik öneriler sunmaktadır.

Anahtar Kelimeler: makine öğrenmesi, performans, regresyon, tahminleme, üretim parametreleri

ABSTRACT

The aim of this study is to predict production performance using machine learning algorithms. Production systems operate based on various machines, parameters, and set values, with each production condition potentially exhibiting variability. Advances in technology have made it possible to control these variabilities, optimize production conditions, and derive insights from interconnected processes. In this context, machine learning serves as a valuable tool that enables the prediction of production performance through statistical methods applied to datasets. In this study, two years of data from the same product group were used to predict production performance through the application of Decision Tree, Linear Regression, Lasso Regression, XGBoost, Support Vector Regression, and LSTM algorithms. These algorithms were evaluated across three different scenarios, with the goal of identifying the model that can most accurately predict production performance. The results provide practical insights into improving production processes by comparing the performance of simple and complex models.

Keywords: machine learning, performance, regression, prediction, production parameters

GİRİŞ

Üretim sistemleri birbirini etkileyen birçok parametreden oluşmaktadır. Üretim parametreleri; üretimin hızını, kalitesini, maliyetini, fiyatını, işlem sürelerini, kapasitelerini doğrudan ya da dolaylı olarak etkilemektedir. Üretim sistemlerinin ana amacı tek seferde doğru miktarda, doğru kalitede ve doğru zamanda üretimi yapabilmektir. Üretim sistemlerini etkileyen birden fazla girdi mevcuttur. Girdiler hammadde, teknoloji, makine, ekipman, personel,

ToCite: GÖKSU, S., & SEZEN, B., & BALCIOĞLU, Y.S., (2025) MAKİNE ÖĞRENMESİ İLE ÜRETİM PERFORMANSI TAHMİNLEMESİ. *Kahramanmaraş Sütçü İmam Üniversitesi Mühendislik Bilimleri Dergisi*, 28(1), 65-79.

ergonomi, teknik gereksinimler, üretim parametreleri şeklinde sıralanabilir (Rahul vd., 2021). Operasyon yöneticileri, üretim sürecinin girdilerden etkilenme düzeyini minimum seviyeye indirmeyi amaçlarlar. Bu amaca ulaşmak her zaman mümkün olmayabilir. Bir girdide oluşan değişkenlik, diğer girdileri pozitif veya negatif anlamda etkileyebilir ve süreç bu girdiden etkilenebilir. Kontrol altında tutulabilen girdiler, üretim sürecinin performansını iyileştirmek için ele alındığında kritik öneme sahiptir. Kontrol altında tutulabilen girdiler için süreç; deney tasarımı, istatistiksel araçlar gibi endüstriyi etkileyen modeller ile çeşitli denemeler aracıyla kurgulanabilir ve sonuçları değerlendirilebilir. Üretim parametreleri, kontrol altında tutulabilen girdilere bir örnek olarak verilebilir (Sittón-Candanedo vd., 2018). Üretim sistemlerinin içerisinde yer alan üretim parametrelerinin üretimi tek seferde kaliteli şekilde üretecek net bir değer ya da belli bir aralığı ifade eden; değer alt limit ve üst limitinden oluşan aralık değerlerine ihtiyacı vardır. Parametrelerin alabildikleri değerlerde oluşan her sapma, üretimin kalitesini, işlem süresini ya da maliyetini etkileyecektir (Rimpault vd., 2018).

Geçmişten günümüze üretim sistemleri içerisinde yer alan, üretimi etkileyen parametreler incelendiğinde ustalık ve deneyimlerin kazandırmış olduğu beceri ile yapılan işlemlerin yerini günümüzde hızla otomasyon sistemlerinin aldığını gözlemlenmektedir (Angelopoulos vd., 2019). Otomasyonlar entegre edildikleri sistemlerdeki insan faktörünü azaltmaya yönelik kurgulanmış yapılardır. Otomasyonlar, daha az bağlı insan gücü ile birlikte daha verimli ve daha kaliteli üretim çıktısı için kullanılır (Cinar vd., 2020). Otomasyonlar içerisinde yer alan en temel düzeydeki sinyallerden, anlık ölçüm yapabilen cihazlara kadar çeşitlilik gösteren materyaller üretimin çıktısına etki etmektedir. Üretim çıktısı hem üretilen ürün adedi ilgili olabileceği gibi, kullanılan ekipmanlarla, ayarlanabilen parametrelerle, kullanıcı tutumlarıyla, ustalıkla, malzemeye ve makinelerle ilgili olabilir. Üretim sistemi içerisinde yer alan otomasyonlar, özellikle ayarlanabilen parametreler açısından üretim yöneticilerine operasyonunun durumu görmek için fırsat sağlamaktadır. Otomasyon sistemleri üretim ekipmanlarının, makinelerinin içerisinde gömülü ya da harici olarak yer alıp, ölçümlerle veri kaynağını oluşturur (Rahul vd., 2021). Elde edilen verilerin doğruluğu; sistemin kurgulanması ile ilgili bir konudur ve genel olarak algoritmalar üzerine kurulu bir yapıdadır. Algoritmaların oluşturduğu yapının içerisindeki veriler, belirli koşullara bağlı olarak veri ataması yapabilmektedir. Sürekliliği sağlanan veri yapısı aracılığıyla veriler elde edilir. Elde edilen verilerin anlamlı yorumlanması ve analiz edilmesi, üretim sisteminin geliştirilmesi için önemlidir (Angelopoulos vd., 2019). Verilerden sağlanan analizler, sistem ile ilgili çıkarımlar yapılmasını sağlar. Çıkarımlar tahminleme, optimizasyon, etki analizi, nedensellik analizi, kalite analizi, maliyet analizi şeklinde sınıflanabilmektedir (Rahul vd., 2021).

Endüstri 4.0 ve beraberinde getirdiği dijitalleşme süreçleri ile verilerin doğruluğu kritik önem kazanmıştır. Rekabet avantajı elde etmek isteyen firmalar üretim sistemlerine yapacakları iyileştirmeleri mümkün oldukça kapsamlı araştırmaktadırlar ve süreçlerinde uygulamak istemektedirler (Sittón-Candanedo vd., 2018). Verilerin yorumlanması, analiz edilmesi ve modellenmesi; üretim sistemlerinin iyileştirme noktalarının tespiti açısından önemlidir (Cinar vd., 2020). Tecrübeye ve geçmiş bilgi birikimine dayalı üretimin yapısını belirleyen parametre değerlerinin; üretim sistemi üzerinden alınan verilerle belli bir değer aralığına oturtulması üretim sistemi içerisinde kişi farklılıklarının getirdiği değişkenliği ve üretilen ürün birim başına oluşan kullanılmayan makine performans kaybını azaltacaktır (Angelopoulos vd., 2019).

Makine öğrenmesi endüstride görüntü olarak ve statü olarak kalite problemlerini tespit etme, talep ve pazar tahminleri, ürün gruplandırma/sınıflandırma gibi alanlarda kullanılmaktadır (Angelopoulos vd., 2019). Makine öğrenmesinin endüstrilere entegrasyonu ile insan işgücüne bağlı ve hataya açık operasyonlardaki kaliteli çıktısı ve yüksek verimlilik beklentisi oluşmuştur (Rahul vd., 2021). Makine öğrenmesi modelleri çeşitli algoritmalar kullanarak ileri istatistik ile beraber süreçleri modelleme ve modele uygun şekilde optimizasyon fırsatı sunmaktadır (Cinar vd., 2020).

Çalışmamızda iki yıllık veri üzerinden, üretim performansı üç farklı senaryoya dayalı olarak tahmin edilmiştir. Bu, mevcut literatürde kullanılan tek veri seti veya belirli parametreler üzerine odaklanan çalışmalardan önemli bir farklılık sunmaktadır. Örneğin, literatürde yaygın olan Cinar vd. (2020) çalışmasında makine öğrenmesi ve tahmin edilebilir bakım üzerine odaklanılırken, bizim çalışmamızda daha geniş kapsamlı üretim parametreleri üzerine odaklanılmıştır. Bununla birlikte, Karar Ağacı, Lineer Regresyon, Lasso Regresyon, XGBoost, Destek Vektör Regresyonu ve LSTM gibi çeşitli makine öğrenmesi algoritmalarını kullanarak geniş bir tahminleme analizi yapılmıştır. Literatürdeki çalışmalarda genellikle birkaç algoritmanın karşılaştırılmasıyla sınırlı kalan analizler varken (Rahul vd., 2021), bu çalışma kapsamlı bir model performansı değerlendirmesi sunmaktadır. Son olarak, üç farklı senaryo kullanılarak performans ölçümü yapılması, bu çalışmayı literatürdeki benzer çalışmalarla kıyaslandığında daha özgün ve uygulamaya yönelik kılmaktadır. Mevcut çalışmalar genellikle tek bir model veya

senaryo üzerinden ilerlerken, bu çalışmada parametrelerin farklı senaryolarda nasıl değiştiği derinlemesine incelenmiştir (Sittón-Candanedo vd., 2018). Sonraki bölümlerde, çalışmada kullanılan makine öğrenmesi üzerine üretim süreçlerinde yapılan çalışmalar literatür incelemesi olarak, materyal ve metot bölümünde kullanılan algoritmaların açıklamaları ve kullanım alanları, analiz bölümünde yapılan çalışmanın detaylarını içeren analizler, bulgular bölümünde çalışmaya ait bulgulara ve sonuçlar kısmında da çalışmanın sonuçları detaylı olarak aktarılmıştır.

LİTERATÜR İNCELEMESİ

Makine öğrenmesi (ML) ve ilgili yöntemlerin üretim süreçlerindeki kullanımı, son yıllarda üretim performansını optimize etmek ve süreç verimliliğini artırmak amacıyla önemli bir alan haline gelmiştir. Chen vd. (2020), iş birliğine dayalı üretim planlamasında ML ve Model Öngörücü Kontrol (MPC) yöntemlerini birleştirerek, üretim parametrelerindeki belirsizliklerin giderilmesine ve üretim verimliliğinin artırılmasına katkı sağlamışlardır. Benzer şekilde, küçük ölçekli sanayilere yönelik geliştirilen ML tabanlı modeller, hava durumu ve satış geçmişi gibi dış parametreleri kullanarak üretim planlamasında belirsizlikleri azaltmayı hedeflemektedir (Batra vd., 2023).

Denetimsiz öğrenme algoritmalarıyla üretim hatlarının performansının tahmin edilmesi ise arıza oranlarını azaltarak üretim süreçlerinde karar almayı optimize etmektedir (Adesiyan, 2021). Mokrova vd. (2018), k-En Yakın Komşu ve Rastgele Ormanlar algoritmalarını kullanarak üretimle ilgili olası tehlikeleri önceden tespit etmenin iş güvenliği ve üretim maliyetlerini düşürmedeki önemini vurgulamışlardır.

Petrol endüstrisinde ML kullanımı ise üretim tahminlerinde doğruluğu artırarak maliyetlerin düşürülmesine katkıda bulunmaktadır (Noshi vd., 2019). Ayrıca, genetik algoritmalarla üretim parametrelerinin optimize edilmesi, yüksek hacimli üretim süreçlerinde maliyetleri düşürerek verimliliği artırmaktadır (Al-Aomar & Al-Okaily, 2006).

Katmanlı üretim ve Bernoulli hatları gibi farklı üretim sistemlerinde makine öğrenimi algoritmalarının kullanımı da literatürde dikkat çekmektedir. Meng vd. (2020), ML'in katmanlı üretim süreçlerinde parametre optimizasyonu ve anomali tespitinde nasıl kullanıldığını incelemişlerdir. Sun ve Zhang (2020) ise Bernoulli hatlarında ML tabanlı parametrik tahmin yöntemleri geliştirerek üretim süreçlerinin doğruluğunu artırmayı hedeflemişlerdir.

Sinir ağları (CNN ve yapay sinir ağları) ise üretim süreçlerinde verimliliği artırmak ve ürün deformasyonlarını tahmin etmek amacıyla yaygın olarak kullanılmaktadır (Khdoudi vd., 2019; Mezentsev & Yasnitsky, 2022). Bu yöntemler, endüstriyel süreçlerde kaliteyi artırmak ve maliyetleri azaltmak için önemli araçlar haline gelmiştir. Bu çalışmalar, makine öğrenimi ve ilgili algoritmaların üretim süreçlerinde maliyetleri azaltma, verimliliği artırma ve kaliteyi iyileştirme konusundaki potansiyelini ortaya koymaktadır. Literatürdeki çeşitli yöntemler ve uygulamalar, ML'in üretim süreçlerine entegrasyonunun önemli faydalar sağladığını göstermektedir.

MATERYAL ve METOT

Materyal

Çalışmamızda, üretim performansının tahmin edilmesi amacıyla Karar Ağacı, Lineer Regresyon, Lasso Regresyon, XGBoost, Destek Vektör Regresyonu (DVR) ve LSTM algoritmaları kullanılmıştır. Her bir algoritmanın uygulanış şekli ve katkıları aşağıda açıklanmıştır:

1. Karar Ağacı (Decision Tree): Karar Ağacı Regresyon Modeli, bir hedef değişkeni tahmin etmek için kullanılan bir makine öğrenmesi yöntemidir. Veriyi dallara ayırarak kararlar oluşturan bir yapıdır. Karar Ağacı algoritmasının temel çalışma prensibi, bağımsız değişkenlerin değerlerine göre veri setini belirli kurallara dayanarak dallara ayırmaktır. Ağaç yapısı şeklinde organize edilir ve veri seti içindeki örnekleri bölerek çalışmaktadır. Temel çalışma mekanizması sırasıyla düğümler ve bölünmeler, karar kuralları, yaprak düğümler ve tahmin şeklindedir (Klusowski & Tian, 2021). Karar ağacı regresyon modelleri, mühendislik, finans, sağlık, pazarlama gibi birçok alanda kullanılmaktadır. Özellikle karmaşık veri yapılarında, açıklanabilirliğin önemli olduğu durumlarda tercih edilir. Bununla birlikte, genellikle diğer modelleme teknikleriyle (örneğin, rastgele ormanlar, gradyan artırılmalı ağaçlar) birlikte kullanılarak performans artırılabilir ve dezavantajları minimize edilebilir (Sevvanthi vd., 2023). Çalışmada kullanımı üretim parametreleri arasındaki ilişkileri hiyerarşik bir şekilde inceleyerek, her adımda en uygun kararı belirlemektedir. Bu algoritma, özellikle kararların görsel olarak anlaşılmasını kolaylaştırarak, üretim

süreçlerinde hangi parametrelerin daha etkili olduğunu belirlemeye katkı sağlar. Çalışmamızda, üretim performansını etkileyen en önemli parametreleri belirlemek amacıyla kullanılmıştır.

2. **Lineer Regresyon (Linear Regression):** Lineer regresyon modeli, bağımlı bir değişkeni (hedef değişken) bir veya daha fazla bağımsız değişkene (girdi değişkenleri) dayalı olarak tahmin etmek için kullanılan istatistiksel bir tekniktir. Lineer regresyon, bağımsız ve bağımlı değişkenler arasında doğrusal bir ilişki olduğunu varsayar. Bu varsayım her zaman geçerli olmayabilir ve doğrusal olmayan ilişkileri modellemek için yetersiz kalabilir (Ramachandran & Tsokos, 2021). Lineer regresyon, ekonomiden mühendisliğe, sosyal bilimlerden biyolojiye kadar pek çok alanda kullanılmaktadır (James vd., 2021). Çalışmada üretim performansını etkileyen parametreler arasındaki doğrusal ilişkiyi bulmak ve bu ilişkiyi kullanarak gelecekteki üretim performansını tahmin etmek amacıyla uygulanmıştır. Model, basit ve etkili tahminler sunarak üretim süreçlerinin optimizasyonuna katkıda bulunmuştur.
3. **Lasso Regresyon (Lasso Regression):** Lasso regresyon modeli, lineer regresyonun bir türüdür ve değişken seçimi ve düzenleme amacıyla kullanılır. Temel amacı, modelin karmaşıklığını azaltarak aşırı uyumu önlemektir. Lasso regresyon, klasik lineer regresyon modeline benzer, yani $y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$ formülüne sahiptir. Ancak, lasso regresyon, katsayıları belirlerken bir ek kısıtlayıcı terim ekler (Lutay & Khusainov, 2021). Lasso regresyonu gereksiz veya anlamsız değişkenleri otomatik olarak sıfıra indirerek, modelin sadeliğini ve yorumlanabilirliğini artırmaktadır. Bu, yüksek boyutlu veri setlerinde özellikle faydalıdır. Lasso regresyonun düzenleme terimi, modelin aşırı öğrenmesini engeller ve genelleme yeteneğini artırır. Bu, modelin yeni veriler üzerinde daha iyi performans göstermesini sağlar (Ben-Moshe, 2020). Lasso regresyon, standart doğrusal regresyon tekniklerine benzer şekilde uygulanabilir ve mevcut istatistiksel yazılım paketlerinde kolayca kullanılabilir. Lasso regresyon, yüksek korelasyonlu bağımsız değişkenlerin bir kısmını sıfıra indirerek modelin stabilitesini artırabilir.
4. **XGBoost (Extreme Gradient Boosting):** XGBoost, denetimli öğrenme problemleri için kullanılan güçlü ve esnek bir makine öğrenmesi algoritmasıdır. Ayrıca XGBoost, karar ağaçları topluluğu oluşturarak hataları minimize eden bir gradyan artırma algoritmasıdır. Genellikle sınıflandırma ve regresyon görevlerinde kullanılır. XGBoost, bir dizi zayıf öğreniciyi, genellikle karar ağaçları olmak üzere ardışık olarak eğiterek güçlü bir öğrenici oluşturmaktadır. Her yeni ağaç, önceki ağaçların hatalarını düzeltmeye çalışır. Bu işlem, hataları minimize edecek şekilde ağırlıklandırılarak yapılır (Zhang vd., 2022). XGBoost'un avantajları arasında yüksek tahmin doğruluğu, hızlı hesaplama, esneklik ve düzenleme özellikleri sayılabilir. Özellikle yüksek doğruluk sunan ve hesaplama açısından verimli olan bu algoritma, büyük veri setleriyle iyi performans gösterir. Dezavantajları ise, hiperparametre ayarlarının karmaşıklığı ve büyük veri setlerinde hafıza tüketiminin yüksek olabilmesidir (Chen & Fan, 2021). Çalışmada, çok sayıda parametre içeren üretim süreçlerinin performans tahmin edilmesinde etkili olmuştur.
5. **Destek Vektör Regresyonu (Support Vector Regression - SVR):** Destek Vektör Regresyonu (DVR), makine öğrenmesinde kullanılan ve doğrusal ya da doğrusal olmayan bir modelle veri noktaları arasındaki ilişkileri tahmin etmeye çalışan bir yöntemdir. DVR'nin amacı, belirli bir hata toleransı içinde en geniş karar sınırını oluşturup yalnızca bu sınırın dışında kalan veri noktalarına odaklanarak tahminlerin doğruluğunu artırmaktır. DVR'de model, tahmin edilen değer ile gerçek değer arasındaki fark belirlenmiş bir hata toleransı (ϵ) sınırında kaldığında bu hatayı göz ardı eder ve yalnızca daha büyük sapmalara dikkat eder; böylece aşırı öğrenmenin önüne geçilerek genelleme kapasitesi artırılır (Vapnik, 1995). Bu yaklaşım, doğrusal bir fonksiyon ile verilerin arasındaki ilişkiyi tahmin etmek yerine, destek vektörleri olarak adlandırılan belirli veri noktalarına dayanarak karar sınırı belirler ve modelin daha hızlı işlem yapmasını sağlar (Cortes & Vapnik, 1995). Doğrusal verilerde DVR basit bir doğrusal fonksiyon kullanırken, doğrusal olmayan verilerde kernel (çekirdek) yöntemleri ile karmaşık ilişkiler öğrenilebilir. Kernel fonksiyonları, DVR'nin doğrusal olmayan problemlerdeki başarısını artıran en önemli bileşenlerden biridir ve bu sayede polinom, RBF (Radial Basis Function) ve sigmoid fonksiyonları gibi yaygın kernel fonksiyonları kullanılabilir (Bishop, 2006). Cortes ve Vapnik'in çalışmalarında doğrusal olmayan veri yapılarına da başarıyla uygulanabilen kernel yöntemleri, DVR'nin geniş bir uygulama alanında kullanılmasını sağlamıştır (Cortes & Vapnik, 1995). DVR, özellikle yüksek boyutlu veri setlerinde etkili bir şekilde çalışır ve finansal zaman serisi tahminleri gibi doğruluğun kritik olduğu alanlarda tercih edilir. Ancak, veri setinin boyutu ve seçilen kernel fonksiyonuna göre hesaplama maliyeti artabileceğinden, DVR büyük veri setlerinde uzun eğitim sürelerine ihtiyaç duyabilir (Vapnik, 1995).

6. LSTM (Long Short-Term Memory): LSTM, zaman serisi verileri ve geçmiş bilgilere dayalı olarak gelecekteki değerleri tahmin eden bir yapay sinir ağı modelidir. Üretim performansını etkileyen parametrelerin zamana dayalı değişimlerini öğrenerek gelecekteki performansı tahmin etmek amacıyla kullanılmıştır. Bu model, üretim süreçlerindeki değişiklikleri geçmiş verilere dayalı olarak daha doğru bir şekilde tahmin etme yeteneğiyle çalışmamıza katkı sağlamıştır. LSTM, tekrarlayan sinir ağlarının bir türüdür ve özellikle zaman serisi verileri veya metin, ses gibi dizilim verileriyle çalışmak için tasarlanmıştır. LSTM'ler, uzun vadeli bağımlılıkları yakalayabilme yeteneği ile bilinir. LSTM'ler, zaman serisi verileri, dil modelleme, konuşma tanıma gibi sıralı veri problemlerinde çok etkilidir.

Metot

Üretim sistemleri birbirini etkileyen birçok parametreden oluşmaktadır. Haddeleme işlemi metalurji sektöründe sıklıkla kullanılan bir işlem türüdür. Genellikle kalınlık inceltme ve yüzey düzgünlüğü sağlamak amacıyla haddeleme operasyonları yapılmaktadır. Haddeleme operasyonlarını etkileyen çeşitli parametreler bulunmaktadır. Hadde hızı, makinelerin uyguladığı yük, gergi kuvvetleri parametrelere örnek olarak verilebilir. Bu çalışma 2021-2023 yılları arasında makineden alınan sinyallerle oluşturulan ham veri seti üzerinde yapılmıştır. Veriler aynı ürün grubuna ait verilerdir. Veri seti içerisindeki girdi statüsündeki veriler aracılığıyla üretim performansını tahminlemek amaçlanmıştır. Veri setinde yer alan girdiler parametre olarak adlandırılabilir. Parametrelere ait veriler üretim hatlarında yer alan sinyaller ile toplanan verilerdir. Her bir veri, sinyaller üzerinden zaman ve tarih sınırlamaları ile üretim kayıtları ile eşleştirilmiştir. Üretim veri kayıtları SAP üzerinden geri bildirim esaslı olarak tutulmaktadır. 2 yıllık üretim verisinde yer alan her satır bir üretilen ürünü ifade etmektedir. Her bir kolon ise üretim sonucunda o üretimin oluşmasını sağlayan parametrelerdir. SAP verisi içerisinde yer alan tarih ve saat bazlı zaman bildirimleri ile anlık alınan üretim parametrelerine ait sinyal verileri zaman bazında eşleştirilerek toplamda 30000*39 yapısında bir matris elde edilmiştir. Sütunlar içerisinde yer alan veriler numerik verilerdir. Bu sebeple metinsel ifadelerin numerik ifadelere dönüştürülmesine gerek kalmamıştır. Veriler kamuya açık değildir ve şirket içi verilerdir. Bu sebeple parametreler detaylıca açıklanamamıştır. Korelasyon analizi yapılarak ihtiyaç duyulan veri ön işleme işlemi yapılmıştır. Veri ön işleme aşamasında birbirleri ile yüksek korelasyon derecesine sahip verilerin modele girmemesi, eksik ya da uç verilerin modelden çıkartılması, birbirlerinden hesaplanan verilerin ayrıştırılması amaçlanmıştır. Veri ön işleme öncesindeki veri setinin boyutu 30000*39 iken, veri ön işleme sonrası 30000*22 matris yapısına sahip veri, ana veriyi oluşturmaktadır. Ana veride yer alan kategorik veri yer almamaktadır. Makine performansının bağlı olduğu iki ana kriter vardır. Kriterler birbirlerini etkiledikleri için ayrıştırılarak da tahminleme performansları ölçülmüştür. 3 farklı senaryo için makine öğrenmesi uygulanmıştır. Tahminleme uygulamalarında tolerans değeri %10 olarak belirlenmiştir. Destek Vektör Uygulaması tahminleme uygulamasında kernel ve lineer olarak ele alınmıştır. Veriler 3 farklı şekilde incelenmiştir. İncelenen koşullar senaryolar şeklinde ifade edilmiştir. Performans sütunu, akım oranına bağlı olarak formülize edildiği için ilk senaryoda akım sütunu model dışında tutulmuştur. İkinci ve üçüncü senaryolarda ise performans kolonu model dışında tutularak korelasyon etkisi indirmek amaçlanmıştır. İlk senaryo olarak üretim verilerinin herhangi bir kısıtlayıcıya bağlı olmadan değerlendirilmesi koşuldur. Bu koşul üzerinden yapılan hesaplamalar için toplam veri boyutu 30000*22 olarak kalmıştır. Performans sütunu hedef sütun olarak seçildiğinde 30000*21 boyutlu matris sırasıyla 0,8 eğitim ve 0,2 test olmak üzere ayrıştırılmıştır. İlk senaryo için oluşan matris boyutu eğitim ve test veri setleri için sırasıyla; 24000*21 ve 6000*21 olarak oluşmuştur. İkinci senaryo için akım performansı kısıtlayıcı olarak kabul edilmiştir. Akım makinenin ne ölçüde güç kullandığının ölçütüdür. Bu kapsamda belirlenen akım kullanım kapasitesi %80'dir. İkinci senaryo ve üçüncü senaryo akım oranı %80 oranına bağlı olarak oluşmuştur. İkinci senaryo akım oranının %80 altında olduğu üretim koşullarını içerir. Üçüncü senaryo ise akım oranının %80 üzerinde olduğu üretim koşullarını içerir. Bu kapsamda ikinci ve üçüncü senaryo için veri boyutları 30000*20 olarak oluşturulmuştur. İkinci senaryo için %80 oranından yüksek üretim koşulları için toplam veri seti boyutu 20500*20 ve %80 oranından düşük üretim koşulları için toplam veri seti boyutu 9500*20 olarak oluşmuştur. Numerik verilerden oluşan veri setinde MinMaxScaler uygulanarak veriler 0-1 arasında ölçeklendirilmiştir. Her bir satırın bir ürünün üretimini temsil ettiği veri setinde yer alan sütunlar üretilen ürünün üretim koşullarını niteler. Satırlarda yer alan ürün ve üretim parametreleri göz önüne alarak hangi parametrelerin üretimin performansına etki ettiğini saptamak amaçlanmıştır. Bu çalışma gelecek dönemlerde yapılacak olan parametre aralıklarının belirlenmesi çalışmalarına verimsiz üretimlerin oluştuğu koşulları kategorize ederek katkıda bulunacaktır. Verilerin tamamı güncel üretim verisi olduğu için sürekli kontrole tabi tutulmaktadır. Bu sebeple verilerde herhangi bir boş hücre yer almamaktadır.

Çalışmada kullanılan modellerin değerlendirilmesi için birden fazla ölçüt kullanılmıştır. Kullanılan ölçütler sırasıyla maddeler halinde açıklanmıştır.

1. MAE (Mean Absolute Error): MAE, tahmin edilen değerler ile gerçek değerler arasındaki farkların mutlak değerlerinin ortalamasıdır. Modelin tahminlerinin ne kadar sapma gösterdiğini ölçer ve hataların büyüklüğünü anlamak için kullanılmaktadır. Küçük MAE değerleri, modelin daha doğru tahminler yaptığını göstermektedir.
2. MSE (Mean Squared Error): MSE, tahmin edilen değerler ile gerçek değerler arasındaki farkların karesinin ortalamasıdır. MSE, hataları karelediği için büyük sapmaların etkisini daha fazla vurgular. Küçük MSE değerleri, modelin daha iyi performans gösterdiğini göstermektedir.
3. RMSE (Root Mean Squared Error): RMSE, MSE'nin kareköküdür ve tahmin hatalarının ortalama büyüklüğünü ifade eder. RMSE, MSE'den farklı olarak hataları gerçek birimlerle ifade eder ve hataların büyüklüğünü daha kolay yorumlamayı sağlar. Küçük RMSE değerleri daha iyi bir model performansına işaret etmektedir.
4. R^2 (R-Squared): R^2 , modelin veri üzerindeki açıklayıcılık oranını gösterir. 0 ile 1 arasında bir değer alır ve 1'e yakın bir değer modelin tahminlerinin mükemmel olduğunu gösterirken, 0'a yakın bir değer modelin veriyi iyi açıklayamadığını göstermektedir.
5. MAPE (Mean Absolute Percentage Error): MAPE, tahmin edilen değerlerin gerçek değerlere göre yüzdesel hata oranını ölçmektedir. Hataların yüzde olarak ifadesi, tahmin edilen değer ne kadar doğru olduğunu anlamak için kullanılır. Küçük MAPE değerleri, modelin daha iyi bir performans gösterdiğini belirtmektedir. Bununla birlikte, MAPE değerinin teorik olarak alt sınırı 0'dır. Bu, tahmin edilen değerlerin tam olarak gerçek değerlerle örtüştüğünü göstermektedir; yani hata yoktur. MAPE'nin üst sınırı ise belirli bir değerle sınırlı değildir ve 100'ün üzerine çıkabilir. Yüksek MAPE değerleri, tahminlerin gerçek değerlere kıyasla çok büyük hatalar içerdiğini gösterir. MAPE metriği büyük hatalar durumunda çok büyük değerlere ulaşabilir, bu yüzden uç değerler göz önünde bulundurulmalıdır. Ancak genellikle pratikte çok yüksek MAPE değerleri (örneğin 100'den büyük) kötü performans anlamına gelmektedir.

Çalışmada üretimi etkileyen parametreler bilgi gizliliği kapsamında paylaşılmamaktadır. Çalışmada kullanılan algoritmalar sırasıyla; Karar Ağacı, Lineer Regresyon, Lasso Regresyon, XGBoost, Destek Vektör Regresyonu ve LSTM'dir. Algoritmalar oluşturulurken kullanılan hiperparametre değerleri maddeler halinde verilmiştir.

1. Karar Ağacı (Decision Tree):
 - DecisionTreeRegressor, max_depth:2, min_samples_split: 2, min_samples_leaf:1, criterion: "mse"
2. Lineer Regresyon (Linear Regression):
 - Fit_intercept: True, normalize: False
3. Lasso Regresyon (Lasso Regression):
 - Alpha:2, max_iter: 1000, tol:0.0001
4. XGBoost (Extreme Gradient Boosting):
 - XGBRegressor, Booster="gblinear", learning_rate:0.1, n_estimators: 100, max_depth: 3, subsample: 0.8, colsample_bytree: 0.8
5. Destek Vektör Regresyonu (Support Vector Regression - SVR):
 - SVR1: SVR, kernel: "RBF", C: 1.0, epsilon: 0.1, gamma: "scale" (RBF kernel için)
 - SVR2: SVR, kernel:"linear", C: 1.0, epsilon: 0.1
6. LSTM (Long Short-Term Memory):
 - Sequential, nöron=50, return_sequences=True, optimizer="adam", loss="mean_squared_error", Dense=1

Tüm analiz senaryoları için eğitim aşamasında veriler; 0,8 ve 0,2 oranında eğitim ve test verisi olarak ikiye ayrılmıştır. Ardından veriler MinMaxScaler metodu ile normalleştirilmiştir. Veri normalleştirilmesinin ardından veri seti üzerinde makine öğrenmesi algoritmaları Karar Ağacı, Lineer Regresyon, Lasso Regresyon, XGBoost, Destek Vektör Makinesi ve LSTM sıralı olarak uygulanmıştır. Senaryo açıklamaları sırasıyla maddeler halinde açıklanmıştır. Ayrıca, DVR1 ve DVR2, Destek Vektör Makinesi algoritmasının farklı çekirdek fonksiyonları (kernel) ile uygulanan varyantlarını temsil etmektedir. Bu doğrultuda, DVR1, RBF (Radial Basis Function) çekirdeğini, DVR2 ise lineer çekirdek fonksiyonunu kullanmaktadır. Bu iki varyant arasındaki farklar, çekirdek fonksiyonlarının veri üzerinde gösterdiği performansa dayanmaktadır.

- Senaryo 1:
 - Makine hız performansı ve akım performansı yüzde oranının büyük olanının kabul edildiği ortak performans tahminlemesi
- Senaryo 2:
 - Akım performansının performans kısıtı olarak kabul edildiği %80'nin altındaki senaryoya ait tahminleme
- Senaryo 3:
 - Akım performansının performans kısıtı olarak kabul edildiği %80'nin üzerindeki senaryoya ait tahminleme

ANALİZLER

Senaryo 1 Analizleri:

Veri setinde yer alan Ortak Performans değişkeni hedef değişken olarak belirlenmiştir.

Tablo 1. Senaryo 1 Model Doğruluk Ölçütleri

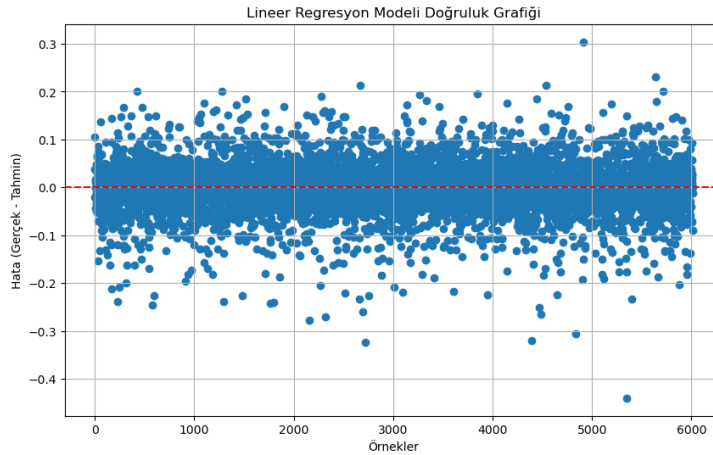
Algoritma	MAE	MSE	RMSE	R2	MAPE
Karar Ağacı	0,035	0,003	0,059	0,515	4,135
Lineer Regresyon	0,038	0,003	0,053	0,598	4,338
Lasso Regresyon	0,056	0,007	0,084	0,000	6,735
XGBoost	0,055	0,005	0,072	0,263	6,225
DVR1	0,070	0,006	0,079	0,109	7,324
DVR2	0,045	0,003	0,058	0,517	4,964
LSTM	0,860	0,747	0,865	-104,825	90,141

Model performanslarına ait veriler Tablo-1'de gösterilmiştir. Karar Ağacı düşük hata değerlerine ve makul bir R^2 değerine sahiptir. Bu da modelin veriyi iyi açıklayabildiğini göstermektedir. Lineer Regresyon, Karar Ağacı modelinden biraz daha yüksek MAE ve MAPE değerlerine sahip olsa da, en yüksek R^2 değerini elde ederek veriyi en iyi şekilde açıklayan model olarak öne çıkmaktadır. Lasso Regresyon, yüksek hata değerleri ile R^2 değerinin sıfır olması nedeniyle diğer modellerden çok daha kötü performans göstermektedir. XGBoost ise Lasso Regresyon'a benzer şekilde yüksek hata değerlerine sahiptir ve R^2 değeri daha düşüktür. Bu değer modelin yeterince iyi performans sağlamadığını göstermektedir. DVR1, yüksek hata değerleri ve düşük R^2 değeri ile diğer modellerden daha kötü performans göstermektedir. DVR2 ise SVM1'den daha iyi performans göstermektedir ve Karar Ağacı ile benzer sonuçlar elde etmektedir. LSTM modeli diğer modellerden çok daha kötü performans göstermektedir. Özellikle negatif R^2 değeri analiz sonuçlarında elde edilmiştir. Modelin tahminlerinin, ortalama tahminlerden daha kötü olduğunu göstermektedir. Özetle, Lineer Regresyon, en yüksek R^2 değeri ve makul hata metrikleriyle en iyi performans gösteren model olduğu tespit edilmiştir. Karar Ağacı ve DVR2, iyi performans gösteren diğer modellerdir. Lasso Regresyon, XGBoost, ve DVR1 daha düşük performans sergilerken, LSTM modeli belirgin şekilde en kötü performans sergilmektedir. Sonuç olarak, basit modeller (Lineer Regresyon ve Karar Ağacı) bu veri setinde daha karmaşık modellerden (LSTM, DVR1, XGBoost) daha iyi performans göstermiştir.

Tablo 2. Senaryo 1 Doğru Tahmin Yüzdesi

Algoritma	Doğru Tahmin Yüzdesi
Karar Ağacı	%93
Lineer Regresyon	%94
Lasso Regresyon	%90
XGBoost	%88
DVR1	%77
DVR2	%92
LSTM	%0

Model doğru tahmin yüzdesine ait veriler Tablo-2’de gösterilmiştir. Tablo-2’den yapılan çıkarım ile; Lineer Regresyon, en yüksek doğruluk oranı ile en iyi performansı göstermektedir. Karar Ağacı ve DVR2 de yüksek doğruluk oranları ile iyi performans göstermektedir. Lasso Regresyon ve XGBoost, diğer modellerden biraz daha düşük ancak kabul edilebilir performans sergilemektedir. DVR1 düşük doğruluk oranıyla diğer modellerden daha kötü performans göstermektedir. LSTM modelinin doğruluk oranı sıfır tespit edilmiştir, bu da modelin analizlerde en başarısız model olduğunu göstermektedir.

**Şekil 1.** Senaryo 1 Lineer Regresyon Doğruluk Grafiği

Şekil 1’de modelin tahmin performans değerlendirmesine ait grafik görülmektedir. Y ekseninde yer alan rakamlar, modelin tahmin ettiği değerler ile gerçek değerler arasındaki farkı göstermektedir. X ekseninde ise veri noktalarını temsil etmektedir. Her bir mavi nokta bir örneği temsil etmektedir. Bu noktalar modelin tahmin hatalarını göstermektedir. Mavi noktaların sıfır (0) değerine yakın olması modelin doğruluk oranının yüksek olmasını gösterir. Grafikte yer alan kırmızı çizgi, tahmin ile gerçek durumun örtüştüğü durumu niteler. Grafik genel olarak tahmin hatalarının dağılımını göstermektedir.

Şekil 1 incelendiğinde, %94 oranında doğruluk gösteren modelin hatalarının 0 değerinin altında ve üstünde dağılım gösterdiğini söylemek mümkündür. Hatalar sıfır çevresine yoğunlaştığı görülmektedir ve bu durum model açısından olumlu bir durumdur.

Model performanslarına ait veriler Tablo-3’te gösterilmiştir. Bu tabloda farklı makine öğrenimi modellerinin performans ölçümleri olumsuz sonuçlar göstermektedir. Genel olarak, tüm modeller negatif R^2 değerlerine sahip, bu da modellerin veriye iyi uyum sağlayamadığını ve tahminlerde başarısız olduğunu göstermektedir. Özellikle, LSTM modelinin R^2 değeri (-3186,204) ile en düşük performansı sergilediği, aynı zamanda en yüksek MAE (0,990) ve MAPE (100,001) değerlerine sahip olduğu gözlemlenmektedir. Bu, LSTM'nin bu verisetinde çok kötü bir sonuç verdiğini ortaya koymaktadır. Lasso Regresyon, en düşük MAE (0,318) ve MAPE (32,121) değerlerine sahip olmasına rağmen, R^2 değeri (-329,207) hala oldukça kötü, bu da modelin tahmin doğruluğunun düşük olduğunu

göstermektedir. Karar Ağacı ve DVR1 modelleri de benzer MAE ve MAPE değerleri sunuyor, ancak onların da R^2 değerleri oldukça negatif (-436,04 ve -459,651) olduğundan, başarı oranları düşük kalmaktadır. XGBoost ise en yüksek MSE (0,330) ve RMSE (0,574) değerleriyle birlikte, en kötü R^2 değerlerinden biri olan (-1072,142) ile tahminlerde oldukça başarısız bir performans sergilemektedir. Lineer Regresyon ve DVR2 modelleri de yüksek hata oranlarına sahip ve negatif R^2 değerleri (-764,418 ve -716,682) ile başarısızlıklarını göstermektedir. Genel olarak, bu modellerin hiçbirinin bu verisetinde iyi performans göstermediği ve veriyi yeterince iyi tahmin edemedikleri söylenebilir. Özellikle, negatif R^2 değerleri modellerin düşük uyum yeteneğini ve zayıf genelleme yapabildiğini işaret etmektedir.

Senaryo 2 Analizleri:

Veri setinde yer alan Kriter-1 kolonu hedef olarak belirlenmiştir. (Akım performansı < %80)

Tablo 3. Senaryo 2 Model Doğruluk Ölçütleri

Algoritma	MAE	MSE	RMSE	R2	MAPE
Karar Ağacı	0,337	0,134	0,366	-436,04	34,048
Lineer Regresyon	0,461	0,235	0,485	-764,418	46,491
Lasso Regresyon	0,318	0,101	0,318	-329,207	32,121
XGBoost	0,550	0,330	0,574	-1072,142	55,459
DVR1	0,334	0,141	0,376	-459,651	33,598
DVR2	0,446	0,220	0,470	-716,682	44,920
LSTM	0,990	0,981	0,990	-3186,204	100,001

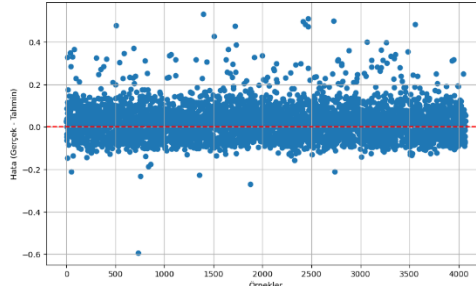
Tablo 4. Senaryo 2 Doğru Tahmin Yüzdesi

Algoritma	Doğru Tahmin Yüzdesi
Karar Ağacı	%48,95
Lineer Regresyon	%32,94
Lasso Regresyon	%27,43
XGBoost	%16,9
DVR1	%85,34
DVR2	%39,32
LSTM	%44,93

Model doğru tahmin yüzdesine ait veriler Tablo-4'te gösterilmiştir. Tablo-4'ten yapılan çıkarım ile; DVR1, en yüksek doğruluk oranı ile en iyi performansı sergileyen modeldir. Bu, modelin veriyi çok iyi öğrendiğini ve güvenilir tahminler sağladığını gösterir. Karar Ağacı ve LSTM, orta düzeyde performans sergileyen modellerdir. Bu modellerin doğruluk oranları %40-50 arasında değişmektedir. DVR2, doğruluk oranı açısından DVR1'den çok daha düşük performans sergiliyor ve %40 civarında doğruluk sağlıyor. Lineer Regresyon ve Lasso Regresyon, düşük doğruluk oranları ile diğer modellerden daha kötü performans gösteriyor. XGBoost modeli, %16,9 doğruluk oranı ile başarısız olmuştur. Bu, modelin yapılandırmasında veya eğitiminde ciddi sorunlar olduğunu göstermektedir.

En yüksek doğru tahmin yüzdesine sahip olan DVR1 Modeli için doğruluk grafiği Şekil-2'de verilmiştir. Şekil-2'den yapılan çıkarımla, hatalar genelde -0,4 ile 0,4 arasında yoğunlaşmıştır. Bu, modelin genel olarak tahminlerinde büyük sapmalar yapmadığını, hataların çoğunlukla küçük olduğunu göstermektedir. Hataların çoğu sıfır etrafında yoğunlaşmıştır, bu da modelin tahminlerinin genel olarak doğruya yakın olduğunu işaret etmektedir. Grafikte, belirgin şekilde sıfırdan sapmış bazı aykırı değerler mevcuttur. Bunlar, modelin belirli örneklerde büyük hatalar

yaptığını göstermektedir. Ancak bu aykırı değerler genel dağılıma kıyasla az sayıda görünmektedir. X ekseninde boyunca hatalar geniş bir yelpazede dağılmıştır. Örnekler arasında belirgin bir desen ya da sistematik hata gözükmemektedir. Hatalar örnek sayısına göre rastgele dağılım göstermektedir. Bu tarz bir hata grafiği, modelin genellikle doğru tahminler yaptığını, ancak zaman zaman aykırı değerler gördüğünü ve hatalarının rastgele dağıldığını göstermektedir.



Şekil 2. Senaryo 2 DVR1 Doğruluk Grafiği

Senaryo 3 Analizleri:

Veri setinde yer alan Kriter-1 kolonu hedef olarak belirlenmiştir. (Akım performansı > %80)

Tablo 5. Senaryo 3 Model Doğruluk Ölçütleri

Algoritma	MAE	MSE	RMSE	R2	MAPE
Karar Ağacı	0,083	0,012	0,111	0,595	13,547
Lineer Regresyon	0,052	0,004	0,070	0,840	8,545
Lasso Regresyon	0,148	0,030	0,174	0,000	25,787
XGBoost	0,081	0,009	0,095	0,703	12,695
DVR1	0,064	0,006	0,083	0,772	9,108
DVR2	0,049	0,004	0,065	0,859	7,949
LSTM	0,646	0,449	0,670	-13,757	94,699

Model performanslarına ait veriler Tablo-5'te gösterilmiştir. Tablo, farklı makine öğrenimi algoritmalarının performansını karşılaştırmak için çeşitli hata metrikleriyle değerlendirilmiş sonuçları içermektedir. Modeller, MAE (Ortalama Mutlak Hata), MSE (Ortalama Kare Hatası), RMSE (Kök Ortalama Kare Hatası), R² (Doğruluk) ve MAPE (Ortalama Yüzde Hatası) gibi ölçütlerle analiz edilmiştir. Genel olarak, DVR2 ve Lineer Regresyon gibi bazı modeller iyi sonuçlar verirken, Lasso Regresyon ve LSTM gibi diğer modeller beklenenden çok daha düşük performans sergilemiştir. En iyi sonuçları veren model DVR2'dir. En düşük hata oranına (MAE: 0,049) ve en yüksek doğruluk oranına (R²: 0,859) sahiptir. Ayrıca MAPE değeri de en düşük olan modeldir (%7,949). DVR2, veri üzerinde oldukça iyi uyum sağlamış ve başarılı tahminler yapmıştır. Lineer Regresyon da DVR2'ye yakın bir performans sergilemiştir. Hata oranları düşük (MAE: 0,052) ve R² değeri oldukça yüksektir (0,840). Ayrıca MAPE değeri %8,545 ile oldukça başarılıdır. Bu iki model, düşük hata oranları ve yüksek doğrulukları ile veri üzerinde en iyi performansları sergilemişlerdir. DVR1 modeli de başarılı modeller arasında yer alır. MAE değeri 0,064, R² değeri ise 0,772'dir. Hata oranları biraz daha yüksek olmasına rağmen, doğruluk oranı tatmin edicidir. XGBoost ise orta seviye bir performans sergilemektedir. MAE değeri 0,081 ve R² değeri 0,703 ile veriye kabul edilebilir düzeyde uyum sağlamıştır. Ancak hata oranları diğer başarılı modellere göre biraz daha yüksektir. Karar Ağacı modeli ise yine orta düzeyde bir performans sergilemektedir. MAE değeri 0,083 ve R² değeri 0,595 ile DVR1 ve XGBoost'un gerisinde kalmış, ancak yine de makul sonuçlar üretmiştir. Diğer yandan, Lasso Regresyon modeli çok kötü bir performans sergilemiştir. MAE değeri 0,148 ile en yüksek değerlerden biridir ve R² değeri sıfırdır, yani model veriye hiç uyum

sağlayamamıştır. Yüzdesel hata oranı da oldukça yüksektir (MAPE: %25,787). LSTM ise en düşük performansı gösteren modeldir. Çok yüksek bir MAE değeri (0,646) ve negatif R² değeri (-13,757) ile veriye tamamen uyumsuz kalmıştır. Yüzdesel hata oranı da oldukça yüksektir (%94,699), bu da modelin veri üzerinde ciddi şekilde başarısız olduğunu göstermektedir. Sonuç olarak, DVR2, Lineer Regresyon ve DVR1 modelleri en iyi performansı sergileyen modellerdir. XGBoost ve Karar Ağacı ise orta seviye performans göstermektedir. Ancak Lasso Regresyon ve özellikle LSTM modeli oldukça kötü sonuçlar vermiştir.

Tablo 6. Senaryo 3 Doğru Tahmin Yüzdesi

Algoritma	Doğru Tahmin Yüzdesi
Karar Ağacı	%73,80
Lineer Regresyon	%86,97
Lasso Regresyon	%32,94
XGBoost	%67
DVR1	%77,61
DVR2	%88,80
LSTM	%0,00

Model doğru tahmin yüzdesine ait veriler Tablo-6'da gösterilmiştir. Tablo-6'dan yapılan çıkarım ile; DVR2 ve Lineer Regresyon en yüksek doğruluk oranları ile en iyi performansı sergileyen modellerdir. Bu, modellerin veriyi çok iyi öğrendiğini ve doğru tahminler sağladığını göstermektedir. DVR1, Karar Ağacı ve XGBoost modelleri de yüksek doğruluk oranları ile iyi performans sergilemektedir. Lasso Regresyon modeli düşük doğruluk oranı ile yetersiz performans sergilemektedir. LSTM modelleri ise neredeyse hiç doğru tahmin yapmamış ve tamamen başarısız olmuş. Bu, modellerin yapılandırmasında veya eğitiminde ciddi sorunlar olduğunu göstermektedir.

Özellikle Lineer Regresyon ve Karar Ağacı modellerinin düşük hata metrikleri ile iyi performans sergilemesinin, üretim süreçlerinde önemli avantajlar sunmaktadır (Gültepe, 2019). Düşük hata oranları, üretim süreçlerinde daha doğru tahminlemelere olanak tanıyarak üretim hattındaki aksaklıkların ve hataların önüne geçilmesini sağlamaktadır (Aktaş & Aydın, 2018). Bu da hem maliyetleri düşürmekte hem de zaman kayıplarını en aza indirmektedir. Örneğin, doğru bir şekilde tahmin edilen üretim parametreleri sayesinde, fazla ya da yetersiz üretimden kaynaklanan stok maliyetleri minimize edilebilmektedir. Ayrıca, üretim hatlarındaki işlem sürelerinin optimize edilmesi, makinelerin duruş sürelerini azaltarak verimliliği arttırmaktadır. Bu durum, özellikle hızlı üretim süreçlerinin kritik olduğu sektörlerde, zamanında teslimat ve müşteri memnuniyeti açısından da büyük faydalar sağlamaktadır. Ayrıca, bulgularımızın literatürdeki benzer çalışmalarla karşılaştırılması da yapılmıştır. Moura vd. (2011) çalışmasında da benzer şekilde Lineer Regresyon'un iyi performans gösterdiği, ancak bizim çalışmamızda geniş veri seti ve farklı senaryoların kullanılmasıyla daha kapsamlı bir analiz yapıldığı vurgulanmıştır. Ayrıca, DVR2 modelinin diğer modellere göre daha yüksek doğruluk oranlarıyla öne çıktığı ve bu sonucun literatürdeki diğer çalışmalarda da benzer şekilde rapor edildiği belirtilmiştir (G. Angayarkanni & S. Hemalatha, 2023). Sonuç olarak, çalışmamızda elde edilen bulguların, üretim süreçlerini optimize etmek için makine öğrenmesi modellerinin nasıl kullanılabileceği konusunda önemli katkılar sunduğu vurgulanmıştır.

BULGULAR

Üretim performansı için kullanılan geliştirme yöntemlerinden biri olan veri analizi yoluyla makine öğrenmesi algoritmaları kullanılarak yapılan sonuçlara uygulama kısmında yer verilmiştir. Senaryo 1 için; Lineer Regresyon ve Karar Ağacı modelleri, düşük MAE ve RMSE değerleri ile iyi performans sergilemiştir. LSTM modeli ise çok yüksek hata metrikleri ve negatif R² değeri ile kötü performans göstermiştir. DVR1 ve DVR2 modelleri arasında DVR2, daha düşük hata metrikleri ile daha iyi performans sergilemiştir. Senaryo 2 için; R² değerinin negatif olması, modelin veriyi tahmin etmede başarısız olduğunu göstermektedir. Bu durum, modellerin çoğunun veri setindeki varyasyonu yakalayamadığını ve doğruluk oranlarının oldukça düşük olduğunu göstermektedir. Lasso Regresyon ve Karar Ağacı, görece en iyi sonuçları verse de, R² değerlerinin negatif olması ve hataların oldukça yüksek olması, modellerin genel performansını sınırlandırmaktadır. Özellikle LSTM'nin çok düşük performansı, bu modelin bu veri seti için uygun

bir seçenek olmadığını işaret etmektedir. Senaryo 3 için; DVR2 ve Lineer Regresyon modelleri en iyi sonuçları sunmakta olup, düşük hata oranları ve yüksek R^2 değerleri ile dikkat çekmektedir. Özellikle DVR2 modeli, hata oranlarının düşük olması ve yüksek doğruluğuyla veri setine çok iyi uyum sağlamaktadır. DVM1 modeli de tatmin edici bir performans sergilemektedir ve veriyi oldukça iyi açıklayabilmektedir. Bu üç model, düşük hata oranları ve yüksek R^2 değerleri ile güvenilir sonuçlar sağlamaktadır. Orta seviye performans gösteren XGBoost ve Karar Ağacı, farklı hiperparametre optimizasyonları ile daha başarılı sonuçlar verebilir. Lasso Regresyon ve LSTM, düşük uyum ve yüksek hata oranları nedeniyle veri seti için uygun modeller gibi görünmemektedir. Özellikle LSTM, bu veri setine uygun bir model olmadığını net bir şekilde göstermektedir.

Çalışmada Lineer Regresyon ve Karar Ağacı modellerinin düşük MAE ve RMSE değerleriyle iyi performans gösterdiği tespit edilmiştir. Bu sonuçlar, Moura vd. (2011) makinesel komponentlerin başarısızlık zamanlarını tahmin etmeye yönelik çalışmalarıyla uyum içindedir. Moura vd. SVM modelini kullandıkları çalışmada, DVR modeli de yüksek performans göstermiştir, ancak Lineer Regresyon'un bu bağlamda daha iyi sonuçlar verdiği bazı uygulamalarda rapor edilmiştir. Ayrıca, çalışmamızda LSTM yüksek hata metrikleri ve negatif R^2 değeri ile düşük performans gösterdiği gözlemlenmiştir. Bu sonuçlar, Hitam ve Ismail'in (2018) zaman serisi verisiyle kripto para birimi tahmini üzerine yaptıkları çalışmayla da paralellik göstermektedir. Bu çalışmada DVR'in, LSTM ve XGBoost modellerine göre daha doğru sonuçlar verdiği gözlemlenmiştir. Bununla birlikte, çalışmamızda DVR modelleri, özellikle DVR2'nin, diğer modellere kıyasla daha düşük hata metriklerine sahip olduğu ve yüksek performans sergilediği bulunmuştur. Bu sonuç, G. Angayarkanni ve S. Hemalatha'nın (2023) çeşitli hastalıkları tahmin etmek için farklı makine öğrenmesi algoritmalarını karşılaştırdıkları çalışmayla da örtüşmektedir. Onların çalışmasında da DVR'in yüksek doğruluk ve hassasiyet değerleri elde ettiği gözlemlenmiştir. Son olarak, çalışmamızda Karar Ağacı'nın da birçok durumda iyi performans gösterdiği vurgulanmıştır. Bu sonuç, Kathiresan vd. (2023) makine öğrenmesi algoritmalarının çeşitli veri setlerinde performanslarını karşılaştırdıkları çalışmalarla benzerlik taşımaktadır. Onların çalışmasında Karar Ağacı, diğer algoritmalar arasında ortalama bir performans sergilemiş, ancak belirli durumlarda daha etkili olmuştur. Bu çalışmada kullanılan üç senaryo üzerinde yapılan analizler sonucunda, Lineer Regresyon ve DVR2 modelleri genellikle en iyi performansı göstermiştir. Karar Ağacı ve DVR1 modelleri de çoğu durumda iyi performans sergilemiştir. Lasso Regresyon, XGBoost ve LSTM modelleri ise genel olarak düşük performans göstermiştir. Özellikle XGBoost ve LSTM, bazı veri setlerinde çok yüksek hata metrikleri ve negatif R^2 değerleri ile en kötü performansı sergilemiştir. Bu sonuçlar, model seçiminde dikkat edilmesi gereken önemli faktörleri ve her veri seti için en uygun modelin dikkatle belirlenmesi gerektiğini göstermektedir.

Model seçiminde dikkate alınması gereken çeşitli kriterler bulunmaktadır. İlk olarak, veri setinin yapısı, model seçiminde en önemli unsurlardan biridir. Veri setinin boyutu, özellik sayısı, verilerin dağılımı ve varsa sınırlamalar, hangi modelin daha iyi performans göstereceğini doğrudan etkilemektedir. Basit modeller, genellikle daha küçük ve daha düzenli veri setlerinde karmaşık modellere göre daha iyi performans göstermektedir. Bu çalışmada da görüldüğü üzere, Lineer Regresyon ve Karar Ağacı gibi daha basit modeller, karmaşık modellere (LSTM, XGBoost) kıyasla daha düşük hata oranları ve daha yüksek doğruluk oranları ile öne çıkmıştır. Bu durum, veri setinde fazla karmaşıklık olmaması halinde basit modellerin tercih edilmesinin performans açısından avantajlı olabileceğini göstermektedir. Karmaşık modeller, genellikle büyük ve karmaşık veri setlerinde daha etkili olsa da, bu çalışmada kullanılan veri seti, daha basit modellerle daha iyi sonuçlar vermiştir. Bunun temel nedeni, karmaşık modellerin yüksek miktarda veriye ve dikkatli hiperparametre ayarlamalarına ihtiyaç duymasındır. Ayrıca, karmaşık modellerin hesaplama maliyetleri daha yüksek olabilir ve fazla öğrenme (overfitting) riski taşıyabilmektedirler. Bu bağlamda, üretim süreçlerinde maliyetlerin düşürülmesi ve süreçlerin optimize edilmesi açısından, veri setine uygun basit modellerin tercih edilmesi hem pratik hem de maliyet açısından daha uygun olabilmektedir. Doğru model seçiminin, üretim süreçlerinde sadece tahmin doğruluğunu artırmakla kalmayıp, aynı zamanda süreç iyileştirmelerini daha etkin hale getirdiği tespit edilmiştir. Daha düşük hata oranları ile tahmin edilen üretim parametreleri, yanlış üretim kararlarının önüne geçilmesine, üretim maliyetlerinin düşürülmesine ve süreçlerin daha hızlı ve verimli bir şekilde optimize edilmesine olanak tanımaktadır. Bu nedenle, doğru model seçimi, yalnızca teknik bir karar değil, aynı zamanda operasyonel verimliliği artıran stratejik bir unsurdur.

SONUÇ

Bu çalışmada, iki yıllık veri seti kullanılarak üretim performansının tahmin edilmesi amacıyla farklı makine öğrenmesi algoritmaları (Karar Ağacı, Lineer Regresyon, Lasso Regresyon, XGBoost, Destek Vektör Makinesi ve LSTM) incelenmiştir. Üç farklı senaryo altında gerçekleştirilen analizler sonucunda, Lineer Regresyon ve Karar Ağacı modelleri genel olarak en düşük hata oranlarına ve en yüksek doğruluk seviyelerine ulaşarak veri setine uyum

sağlama konusunda güvenilir sonuçlar vermiştir. Özellikle Lineer Regresyon modeli, basit yapısı ile yüksek tahmin doğruluğu sağlamış, DVR2 modeli ise düşük hata oranlarıyla öne çıkmıştır. Bu çalışmada daha karmaşık modeller olan LSTM ve XGBoost, veri setinin yapısına uyum sağlamakta zorlanmış ve negatif R² değerleriyle düşük performans göstermiştir. LSTM modelinin özellikle üretim parametrelerini tahmin etmede başarısız olması, veri setinin karmaşık zaman serisi yapılarına uygun olmamasından kaynaklanabilir. Bununla birlikte, Lineer Regresyon ve Karar Ağacı gibi daha basit modeller, veri setinde düşük hata oranları ve yüksek doğruluk oranları sunarak özellikle küçük ve daha düzenli veri setleri için uygun modeller olduğunu ortaya koymuştur. Çalışmanın bulguları, literatürdeki benzer çalışmalarla karşılaştırıldığında, Lineer Regresyon ve Karar Ağacı modellerinin özellikle düşük hata oranları ve yüksek performansı ile öne çıktığı görülmektedir (Moura & arkadaşları, 2011; Hitam & Ismail, 2018).

Bu çalışma, belirli bir ürün grubuna ve iki yıllık bir zaman dilimine dayalı olarak yapılmış olup, sonuçların genellenebilirliği sınırlıdır. Farklı sektörlerde ve daha uzun zaman dilimlerinde yapılacak benzer çalışmalar, bulguların geçerliliğini ve genellenebilirliğini artırmaktadır. Ayrıca, çalışmamızda yalnızca statik veri kullanılmış olup, gelecekteki araştırmalarda gerçek zamanlı verilerin kullanımıyla daha dinamik analizler yapılabilir. Endüstri 4.0 ve dijitalleşme süreciyle birlikte, makine öğrenmesi algoritmalarının üretim süreçlerine entegrasyonu giderek daha önemli hale gelmektedir. Gelecekte, yapay zeka algoritmalarının daha geniş çapta kullanımını araştırmak ve bu algoritmaların çevresel ve sosyal etkilerini incelemek, sürdürülebilir üretim süreçlerine katkı sağlayabilecektir. Bu nedenle, gelecekteki araştırmalar, farklı veri setleri ve gelişmiş algoritmalar kullanılarak daha kapsamlı ve genellenebilir sonuçlar elde etmeye yönelik olmalıdır.

KAYNAKÇA

- Adesiyan, A. (2021). Performance Prediction Of Production Lines Using Machine Learning Algorithm. <https://doi.org/10.14293/s2199-1006.1.sor-.ppa7be8.v1>.
- Aktaş, B., & Aydın, C. (2018). Talaşlı imalat sektöründe zaman serileri kullanarak üretim etkililiğinin tahmini. *Bilişim Teknolojileri Dergisi*, 11(4), 407-416.
- Al-Aomar, R., & Al-Okaily, A. (2006). A GA-based parameter design for single machine turning process with high-volume production. *Comput. Ind. Eng.*, 50, 317-337. <https://doi.org/10.1016/j.cie.2006.02.003>.
- Angayarkanni, G., & Hemalatha, S. (2023). Evaluating the performance of supervised machine learning algorithms for predicting multiple diseases: A comparative study. 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS).
- Angelopoulos, A., Michailidis, E. T., Nomikos, N., Trakadas, P., Hatziefremidis, A., Voliotis, S., & Zahariadis, T. (2019). Tackling faults in the Industry 4.0 era—a survey of machine-learning solutions and key aspects. *Sensors* (Basel, Switzerland).
- Batra, R., Abbi, P., Sharma, R., Agarwal, H., & Bhulania, P. (2023). Production prediction using machine learning. 2023 10th International Conference on Signal Processing and Integrated Networks (SPIN), 399-404. <https://doi.org/10.1109/SPIN57001.2023.10116936>.
- Ben-Moshe, D. (2021). IDENTIFICATION OF LINEAR REGRESSIONS WITH ERRORS IN ALL VARIABLES. *Econometric Theory*, 37(4), 633–663. <https://doi.org/10.1017/S0266466620000250>
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer. <https://doi.org/10.5555/1162264>
- Chandra Prabha, S., & Lakshmi, S. (2023). Data analysis and machine learning-based modeling for real-time production. *The Scientific Temper*. <https://doi.org/10.58414/scientifictemper.2023.14.2.22>.
- Chen, Y., Zhou, Y., & Zhang, Y. (2020). Collaborative Production Planning with Unknown Parameters using Model Predictive Control and Machine Learning. 2020 Chinese Automation Congress (CAC), 2185-2190. <https://doi.org/10.1109/CAC51589.2020.9326614>.
- Chen, Y., Zhou, Y., & Zhang, Y. (2021). Machine Learning-Based Model Predictive Control for Collaborative Production Planning Problem with Unknown Information. *Electronics*. <https://doi.org/10.3390/electronics10151818>.
- Chen, Z., & Fan, W. (2021). A freeway travel time prediction method based on an XGBoost model. *Sustainability*, 13(15), 8577. <https://doi.org/10.3390/su13158577>

- Cinar, Z., Abdussalam Nuhu, A., Zeeshan, Q., Korhan, O., Asmael, M. B. A., & Safaei, B. (2020). Machine learning in predictive maintenance towards sustainable smart manufacturing in Industry 4.0. *Sustainability*.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297. <https://doi.org/10.1007/BF00994018>
- Gültepe, Y. (2019). Makine öğrenmesi algoritmaları ile hava kirliliği tahmini üzerine karşılaştırmalı bir değerlendirme. *Avrupa Bilim ve Teknoloji Dergisi*, (16), 8-15.
- Hitam, N. A., & Ismail, A. R. (2018). Comparative performance of machine learning algorithms for cryptocurrency forecasting. *Indonesian Journal of Electrical Engineering and Computer Science*.
- James, G., Witten, D., Hastie, T., Tibshirani, R., James, G., Witten, D., ... & Tibshirani, R. (2021). Linear model selection and regularization. *An introduction to statistical learning: with applications in R*, 225-288.
- Jun, Z. (2021). The development and application of support vector machine. In *Journal of Physics: Conference Series* (Vol. 1748, No. 5, p. 052006). IOP Publishing.
- Kathiresan, V., Dinesh, G., Sarveshwaran, V., Jayakanth, J. J., & Kiruthika, M. (2023). Comparative analysis of diverse classification algorithms of machine learning by using various quality metrics. *2023 IEEE 5th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA)*.
- Khdoudi, A., Masrour, T., & Mazgualdi, C. (2019). Using Machine Learning Algorithms for the Prediction of Industrial Process Parameters Based on Product Design. , 728-749. https://doi.org/10.1007/978-3-030-36671-1_67.
- Klusowski, J. M., & Tian, P. M. (2024). Large scale prediction with decision trees. *Journal of the American Statistical Association*, 119(545), 525-537.
- Leha, A., Pangercic, D., Rühr, T., & Beetz, M. (2009). Optimization of simulated production process performance using machine learning. *2009 IEEE Conference on Emerging Technologies & Factory Automation*, 1-5. <https://doi.org/10.1109/ETFA.2009.5347229>.
- Lutay, V. N., & Khusainov, N. S. (2021, November). The selective regularization of a linear regression model. In *Journal of Physics: Conference Series* (Vol. 2099, No. 1, p. 012024). IOP Publishing
- Meng, L., McWilliams, B., Jarosinski, W., Park, H., Jung, Y., Lee, J., & Zhang, J. (2020). Machine Learning in Additive Manufacturing: A Review. *JOM*, 72, 2363 - 2377. <https://doi.org/10.1007/s11837-020-04155-y>.
- Mezentsev, A., & Yasnitsky, L. (2022). Neural network model for determining the regulations parameters in the technological process of ore raw materials processing. *Journal Of Applied Informatics*. <https://doi.org/10.37791/2687-0649-2022-17-6-56-67>.
- Mokrova, N., Mokrov, A., Safonova, A., & Vishnyakov, I. (2018). Machine Learning Methods for Production Cases Analysis. *IOP Conference Series: Materials Science and Engineering*, 317. <https://doi.org/10.1088/1757-899X/317/1/012044>.
- Moura, M., Zio, E., Lins, I. D., & Droguett, E. (2011). Failure and reliability prediction by support vector machines regression of time series data. *Reliability Engineering & System Safety*, 96(12), 1527–1534.
- Noshi, C., Eissa, M., & Abdalla, R. (2019). An Intelligent Data Driven Approach for Production Prediction. Day 4 Thu, May 09, 2019. <https://doi.org/10.4043/29243-MS>.
- Rahul, R., Tiwari, M., Ivanov, D., & Dolgui, A. (2021). Machine learning in manufacturing and Industry 4.0 applications. *International Journal of Production Research*, 59(18), 4773-4778.
- Ramachandran, K. M., & Tsokos, C. P. (2020). *Mathematical statistics with applications in R*. Academic Press.
- Rimpault, X., Balazinski, M., & Chatelain, J. (2018). Fractal analysis application outlook for improving process monitoring and machine maintenance in Manufacturing 4.0. *Journal of Manufacturing and Materials Processing*.
- Sevvanthi, K., Ganapathy, S., Penumadu, P., & Harichandrakumar, K. T. (2023). Comparing the predictive performance of a decision tree with logistic regression for oral cavity cancer mortality: A retrospective study. *Cancer Research, Statistics, and Treatment*, 6(1), 103-110.

Sittón-Candanedo, I., Hernández Nieves, E., Rodríguez, S., Santos-Martín, M., & González-Briones, A. (2018). Machine learning predictive model for Industry 4.0. In Proceedings of the 16th International Conference on Information Systems, Management, and Automation (pp. 501-510).

Strasser, S., Tripathi, S., & Kerschbaumer, R. (2018). An Approach for Adaptive Parameter Setting in Manufacturing Processes. , 24-32. <https://doi.org/10.5220/0006894600240032>.

Sun, Y., & Zhang, L. (2020). Parameter Identification for Multiple-Machine Bernoulli Lines using Statistical Learning Methods. 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE), 810-815. <https://doi.org/10.1109/CASE48305.2020.9216960>.

Vapnik, V. (1995). The Nature of Statistical Learning Theory. Springer. <https://doi.org/10.1007/978-1-4757-2440-0>

Zhang, P., Jia, Y., & Shang, Y. (2022). Research and application of XGBoost in imbalanced data. International Journal of Distributed Sensor Networks, 18(6), 15501329221106935.