



Line-based process duration prediction using the vehicles features for the automotive industry

Mustafa Akpınar*^{ID}

Sakarya University, Faculty of Computer and Information Sciences, Software Engineering Department, 54187, Sakarya, Turkey

Highlights:

- Process time prediction for the automotive industry via statistics and artificial intelligence algorithms
- Prediction models based on vehicle-specific features
- Estimation of line-based process durations on vehicles that have not been manufactured previously

Keywords:

- automotive industry
- regression and classification trees
- multiple linear regression
- artificial neural networks
- process duration prediction

Article Info:

Research Article
Received: 11.02.2021
Accepted: 23.12.2021

DOI:

10.17341/gazimmfd.878469

Acknowledgement:

This study was supported by TUBITAK, the Scientific and Technological Research Council of Turkey (Grant No. 1170357) and Ford Otosan Company.

Correspondence:

Author: Mustafa Akpınar
e-mail:
akpinar@sakarya.edu.tr
phone: +90 264 295 3238

Graphical/Tabular Abstract

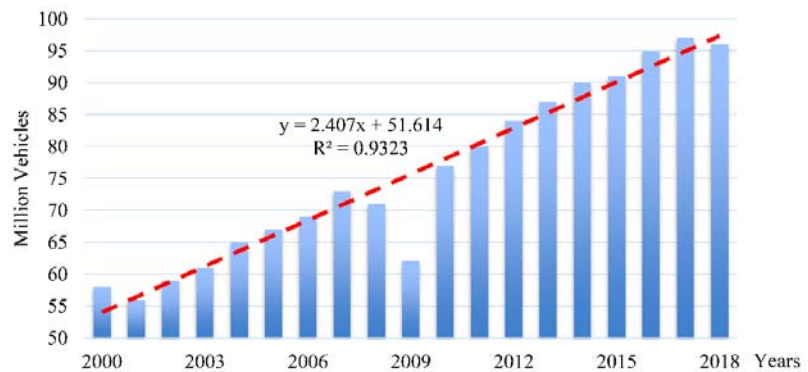


Figure A. Automobile production of World from 2000 to 2018

Purpose: In this study, unlike the traditional approach, using regression, regression and classification trees, and fully connected artificial neural networks, which are among statistical and artificial intelligence techniques, the waiting durations of the vehicle in the buffer zones, process durations of the welding, assembly, dyehouse lines were predicted.

Theory and Methods:

The techniques used in this study are Multiple Linear Regression (MLR), Classification and Regression Trees (CART) and Artificial Neural Networks (ANN). First of all, data obtained from different data sources were combined, metadata was extracted, and divided to include line-based processing times and attributes. Using the metadata obtained, the datasets were digitized in both sequential numeric and binary form. Forward selection, backward elimination, stepwise selection and coefficient of determination selection methods are the feature selection methods in MLR technique that were used to find significant features and to reduce insignificant ones in datasets. Numeric, binary, reduced numeric and reduced binary datasets obtained for 7 process lines and 3 buffer zones were used to create process time prediction models with MLR, CART and ANN techniques. In the beginning, 45 models were created, the technique with the lowest error was determined, and 16 new models were created for the technique and re-run.

Results:

Among the 45 models which are the initial models, the models with the lowest error for the process duration estimation were determined and the results were found to be acceptable. In these models, it is observed that the CART technique applies the models with the lowest error. In the CART technique, 16 new models were created, and these models improved two process lines, one buffer zone.

Conclusion:

In this study, the time estimation of 7 different lines and 3 different buffer zones was carried out with 3 different techniques over 4 different data set structures by using 43 different features of vehicles produced. As a result of the predictions, it is seen that the tree structure gave the best result for 8 lines and the regression for 2 lines. It is observed that artificial intelligence approach gives acceptable results in buffer zones other than "Welding Buffer Lower Floor", which is one of the buffer areas used only as waiting zones, where there is no production. The most important advantage of these techniques is that they can produce predictive results for the vehicle to be produced for the first time and for which there is no knowledge. In other words, when a new vehicle goes into mass production, the process duration prediction can be performed using only the attributes of 150 vehicles whose process time is unknown.



Otomotiv endüstrisi için araç özelliklerini kullanarak proses sürelerinin hat bazlı tahmini

Mustafa Akpınar*

Sakarya Üniversitesi, Bilgisayar ve Bilişim Bilimleri Fakültesi, Yazılım Mühendisliği Bölümü, 54187, Sakarya, Türkiye

Ö N E Ç İ K A N L A R

- Otomotiv sektöründe proses süre tahmini için istatistik ve yapay zeka algoritmalarının kullanılması
- Araca özgü nitelikler ile tahmin modelleri
- Hat bazlı proses sürelerinin daha önce üretilmeyen araçlar dahil tahmin edilebilmesi

Makale Bilgileri

Araştırma Makalesi

Geliş: 11.02.2021

Kabul: 23.12.2021

DOI:

10.17341/gazimmfd.878469

Anahtar Kelimeler:

Otomotiv endüstrisi,
regresyon ve sınıflandırma
ağaçları,
çoklu doğrusal regresyon,
yapay sinir ağları,
proses süre tahmini

ÖZ

Dünyadaki üretim sektörleri arasında lokomotif görevi gören sektörlerden biri de otomotivdir [1]. Bu sektöre hizmet eden fabrikalarda günlük, hatta saatlik araç çıkış sayıları önem arz etmektedir. Hattan çıkan araçların hammadde ve yarı mamullerinin tespit edilmesinde önemli etkenler araçların hatlarda geçirdikleri zaman ve üretim süreleridir. Burada hat sürelerinin belirlenmesinde klasik yaklaşım, her iş için belirli bir sürenin belirlenmesi ve bu sürelerin toplanarak hat sürelerinin bulunmasıdır. Daha önce gerçekleştirilmemiş işlemler ve arızalar nedeniyle, sürelerde sapmalar meydana gelebilmektedir. Bu çalışmada klasik yaklaşımdan farklı olarak regresyon, regresyon ve sınıflandırma ağaçları yapay sinir ağları kullanılarak kaynak, montaj, boyahane bölümlerinin araç proses ve tampon bölümlerinin bekleme sürelerinin tahmini gerçekleştirilmiştir. Yapılan tahminler için 61 model oluşturulmuş ve montaj transit ile kaynak tampon üst kat dışındaki hatlarda en düşük ortalama mutlak yüzdesel (OMYH), ortalama mutlak ve ortalama karesel hatalara sahip tekniğin sınıflandırma ağaçları olduğu görülmüştür. Montaj hatlarında süre tahmini ortalama %7,42 OMYH ile elde edilirken, boyahane hatları için %21,24 OMYH ve kaynak hatlarında için %22,49 OMYH ile belirlenmiştir. Araca mahsus bağımsız değişkenler kullanıldığından, araçların bekletildiği tampon bölgelerde süre tahmininin montaj alt kat haricinde %82,89 OMYH ile gerçekleştiği tespit edilmiştir. Montaj alt kat tampon bölgesinin %1000 OMYH değerinden daha yüksek değer belirlenmesi, bu alanın süre tahmininin uygun olmadığını göstermiştir.

Line-based process duration prediction using the vehicles features for the automotive industry

H I G H L I G H T S

- Process time prediction for the automotive industry via statistics and artificial intelligence algorithms
- Prediction models based on vehicle-specific features
- The ability to the estimation of line-based process durations on vehicles that have not been manufactured previously

Article Info

Research Article

Received: 11.02.2021

Accepted: 23.12.2021

DOI:

10.17341/gazimmfd.878469

Keywords:

Automotive industry,
regression and classification
trees,
multiple linear regression,
artificial neural networks,
process duration prediction

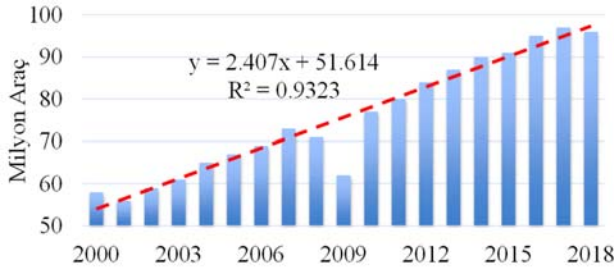
ABSTRACT

The automotive sector is an industry that acts as one of the locomotives among the manufacturing sectors in the world [1]. Daily, even hourly, vehicle production speeds are essential in these factories. The crucial factors in determining the vehicles' raw materials and semi-finished products come from the time they spend on the lines and production times of the vehicles. The traditional approach here is to determine a particular time for each job and find the line times by summing up manually. However, due to unpredictable events, breakdowns, and processes that have not been performed before, deviations in periods may occur. In this study, unlike the traditional approach, regression, regression and classification trees, and artificial neural networks; the waiting times of the vehicle, process and buffer sections of the source, assembly, dye-house sections were estimated. Sixty-one models were created for the predictions, and it is seen that the classification trees technique performed the lowest mean absolute percent (MAPE), mean absolute (MAE), and mean square errors (MSE) in the lines other than the assembly transit and the welding buffer upper floor. While the duration estimation in assembly lines was obtained with 7.42% MAPE, it was determined with MAPE 21.24% for dye-house and 22.49% MAPE for welding lines. Since the independent variables specific to the vehicle are used, it has been determined that the time estimation in the buffer zones where the vehicles are kept is realized with 82.89% MAPE except for the assembly lower floor buffer. The determination of a value higher than 1000% MAPE of the assembly lower floor buffer showed that the duration estimation of this zone was not suitable.

*Sorumlu Yazar/Yazarlar / Corresponding Author/Authors : *akpinar@sakarya.edu.tr / Tel: +90 264 295 3238

1. GİRİŞ (INTRODUCTION)

Otomotiv sektörü her geçen gün artan üretimler ile kendini geliştirmektedir. 2030'da global otomotiv sektörü piyasasının yaklaşık olarak ikiye katlayarak 5.500 milyar € olacağı öngörülmektedir [2]. Avrupa'da otomotiv sektörünün ürettiği ciro ise, Avrupa Birliği'nin gayrisafi yurtiçi hasılasının (GSYİH) yaklaşık olarak %7'sini oluşturmaktadır [2]. 2000'li yılların başından 2018 yılına kadar artan üretim trendi incelendiğinde, senede 2,4 milyon yeni araç üretildiği görülmektedir (Şekil 1). 2009 yılındaki düşüş, 2007 ve 2008 senelerini etkisi altına alan finansal krizin etkilerinden meydana gelen üretim düşmelerinden kaynaklanmaktadır. Bu düşüşe rağmen, sektör sonraki yıllarda kendisini hızlıca toparlamış ve kaldığı yerden üretim artışına devam etmiştir. Yine [2] raporunda belirtildiği gibi 1990'larda batı Avrupa'da ekonominin ağırlık merkezi olarak kabul edilen otomotiv sektörü, öncelikle Türkiye, sonrasında da Asya'ya kaymıştır. Bu açıdan, Türkiye'de 14 farklı firmanın fabrikası bulunmasıyla otomotiv üretiminde önemli bir yere sahip olduğu söylenebilir. Türkiye'nin 2020 yılı için motorlu araç üretim kapasitesi 2,056 milyondur [3]. Tüm dünya ile birlikte Türkiye'de de otomotiv üretimi hızla artmakta ve talebi karşılamak için çalışılmaktadır. Şekil 1'deki eğilimde görüldüğü gibi her sene yaklaşık olarak 2,4 milyon daha fazla araç üretilmektedir.



Şekil 1. 2000 – 2018 arası dünyadaki araç üretimi [4]
(Automobile production of World from 2000 to 2018)

Her üretim artışı beraberinde lojistik sorunları da getirmektedir. Örneğin, araç satın alan müşteriler satın aldığı araçları en kısa sürede talep etmektedir. Satış temsilcileri, olası teslim tarihini fabrikalardan alarak müşterilerini bilgilendirmektedir. Bu sürenin kısa olması üretim, kapasite ve yoğunluğa bağlıdır. Zamanında teslimatlar firmalara olan güveni ve müşteri bağlılığını arttırmaktadır. Avrupa otomotiv sektöründeki diğer önemli nokta ise %90 oranında zamanında teslimattır [2]. Bu noktada araçların lojistiği önem kazanmaktadır. Araç lojistiği iki yönlü olarak ele alınabilir: son ürünün müşteri ile buluşturulması ve araç üretilme sürecinde ilk parçanın üretilmesinden nihai ürün olarak sunulmasına kadar geçen sürede kullanılacak parça ve işlemlerin tedarikinin sağlanması. Bu çalışmada son ürünün müşteri ile buluşturulması kapsam dışıdır. Diğer taraftan otomotiv sektörü, üretim, Üretim-Geliştirme (ÜR-GE) ve Araştırma-Geliştirmeye (AR-GE) önemli katkılar raporuna göre büyük veri, yapay zeka ve endüstri 4.0 kavramları

üzerine çalışmalar yapılmasının önemi de bu durumu desteklemektedir [9]. Raporla iş süreçlerinin dijitalleştirilmesi, kestirimci bakım ve testlerde Endüstri 4.0 çözümlerinin kullanılması ilk aşama olarak yer alırken, ikinci aşamada ürün ve müşteri bilgileri kullanılarak telematik ve tahmine dayalı servisler önerilmiştir. Üçüncü aşamada mobilite ve bulut tabanlı hizmetler gibi hizmetler önerilmiştir. Bu çalışma kapsamında yapılan uygulamalar, birinci ve ikinci aşamaya hitap etmektedir.

Üretim planlama konusunda 1950'li yılların başlarından itibaren çalışılmaktadır [10]. Zamanla üretim planlaması başlı başına bir araştırma konu haline gelmiş, endüstri mühendisleri tarafından çalışılan önemli konular arasında görülmüştür. Üretim planlama daha sonra zamanlama ile birlikte kullanılabilir hale gelmiştir. Bunun sebebi planlamada zamanın en önemli etken olmasıdır. Düzgün bir plan zamandan bağımsız ise maliyet hesabının gerçekleştirilmesi imkansız olacaktır [11]. Gerçek dünya planlama problemleri kombinasyonlu, dinamik ve rastsaldir [12]. Dinamik çizelgeleme kaynak ve iş bağımlı olarak iki kısımda incelenebilmektedir. [13]. Günümüzde ise dinamik yaklaşımlar "tümüyle reaktif planlama" (completely reactive scheduling), "sağlam proaktif planlama" (robust pro-active scheduling) ve "tahmine duyarlı planlama" (predictive-reactive scheduling) olmak üzere üç farklı kategoride yer almaktadır [13]. Tahmine duyarlı planlama, en yaygın kullanılan yaklaşımdır. Çizelgeleme barındıran planlama yaklaşımları Polinamiyel Olmayan-Zor (Non-polynomial Hard) problem tipine girmektedir ve yüksek işlem gücü gerektirmektedir [14]. Literatürde, bu çalışmanın benzeri ya da hedefi ile örtüşen bir çalışmaya bilindiği kadarıyla rastlanmamıştır. Ancak enerji ve otomotiv sektörü üzerine yapılan yakın çalışmalar bulunmaktadır. Motorlu taşıtların üretimi/montajı üzerine yapılan farklı bölgelerdeki farklı stratejiler ile üretim yapılması olduğunu gösteren başka bir çalışmada, yalın üretim üzerine yetkinlikler araştırılmıştır [15]. Çalışmada üretim çeşitliliğinin çok hızlı değiştiği, ürün sayısının arttığı bunun da montaj üretim tesisleri için zorlayıcı bir durum olduğu belirtilmiştir. Üretimdeki karmaşıklığın artması, ürün çeşitliliği ve üretkenlikte artışın olacağını göstermiştir. Örneğin, Ford tarafından üretilen Fiesta, Focus ve Fusion modellerinin Türkiye pazarında sırasıyla 14.144, 4.992 ve 3.460 farklı çeşitliliği bulunmaktadır [16]. Bu sayılar Amerika Birleşik Devletlerinde (ABD'de) Fiesta ve Ford için sırasıyla 150.624 ve 150.624 olarak belirlenmiştir. Aslında bu değişkenlik araçların farklı anahtar, park sensörü, jant, koltuk rengi ve araç rengi gibi seçeneklerin birleşimlerinden kaynaklanmaktadır. Başka bir çalışmada otomotiv montajında enerji tüketimi üzerine sayısal bir çalışma yapılmıştır [17]. İlgili çalışmada, otomobil alt gövdesindeki kaynak ile montajın bulunduğu bölge ele alınmıştır. Farklı alt gövde tiplerindeki sonuçlar, ampirik sayısal denklemler ile bulunmuş ve gösterilmiştir. ABD için yapılan diğer bir teknik raporda, 70 farklı montaj fabrikasının bir senede 13 milyon otomobil ve büyük şase aracı (kamyon, tir) ürettiği

belirtilmiştir [18]. Çalışmada, motorlu araçların montaj süreci ve enerji kullanımı gösterilmiştir. Montaj fabrikalarında en yüksek enerji tüketiminin boyahanelerde olduğu saptanmıştır. Daha sonra aydınlatma, İHİ (Isıtma, havalandırma ve iklimlendirme) en yüksek enerji tüketimine sahipken, bunu hava basıncı takip etmektedir. Kaynak, malzeme montajı (el aletleri kullanımı) ve metal şekillendirmenin her biri %10'un altında enerji tüketimi gerçekleştirmektedir. Bu makalede anlatılan çalışmamızın veri seti hazırlanırken proses sürelerinin sırasıyla boyahane, montaj ve kaynak hatları için 584,5, 287,66 ve 263,5 dakika olduğu görülmüştür. [18] nolu çalışmada boyahanelerin en yüksek enerji tüketimine sahip olması bizim de bulduğumuz boyahane süresinin yüksek olmasını desteklemektedir. Yalın üretim yapan başka bir otomotiv fabrikasında ise montaj hattındaki değer akış analizi yapılmış, yalın üretim sonucu süredeki düşüş belirlenmeye çalışılmıştır [19]. Üretimde çevrim süresini ve toplam enerji tüketimini azaltmak için montaj hatlarına robotların eklenmesi üzerine yapılan başka bir çalışmada [20], bizim çalışmamızda olduğu gibi literatürde bu konuda kısıtlı çalışmanın yapıldığı belirtilmiş ve robot sayılarına göre performansları incelenmiştir. Görüldüğü gibi literatürde, genel olarak planlama ve çizelgeleme problemlerinin çözümüne ve enerji ihtiyacının düşürülmesine yönelik çalışmalar bulunmaktadır. Bu çalışmalar dışında, çalışmamızda kullanılan teknikler güncelliğini koruyan ve yapay zeka uygulamalarında sıklıkla kullanılan tekniklerdir. Maaouane ve arkadaşları yaptıkları çalışmada çoklu regresyon denklemi (ÇRD) kullanarak endüstriyel enerji tüketimini tahmin etmeye çalışmışlardır [21]. Liu ve arkadaşları tarafından yapılan çalışmada, Çin kırasındaki bölgesel yoksulluğu belirlemek için ÇRD kullanılmıştır [22]. Ouma ve arkadaşları tarafından 2020 yılında yapılan çalışmada ise nehirlerdeki çözünmüş oksijen miktarını ÇRD ile tahmin edilmiştir [23]. ÇRD kullanan bu çalışmalardan [23] çalışmasında yapay sinir ağı (YSA) da kullanılmıştır. Çetin ve arkadaşları ise çalışmalarında sabit mıknatıslı senkron generatör performansını YSA kullanarak belirlemeye çalışmışlardır [24]. Yerlikaya-Özkurt ve Askan tarafından yapılan çalışmada Türkiye'nin deprem hasar veritabanı kullanılarak sismik hasarların tahmini sınıflandırma ve regresyon ağaçları (SRA) ile gerçekleştirilmiştir [25]. Ebebeynlerin grip aşısı kararını açıklama SRA'nın kullanıldığı diğer bir çalışma da 2020 yılında yapılmıştır [26]. Covid-19 üzerine yapılmış bir çalışmada ise SRA kullanılarak bulaşmanın verimli tarama ve test için önerilen klinik göstergeler incelenmiştir [27]. Kullanılan tekniklerin güncelliği son yıllarda yapılan çalışmalarda görülmektedir.

Bu çalışmanın hedefi araçların üretim esnasında, yapısında farklılık gösteren özellikler kullanılarak işlem sürelerini tahmin etmektir. Literatürde, bu çalışmanın benzeri bir çalışmaya bilindiği kadarıyla rastlanamamıştır. Ancak enerji ve otomotiv sektörü üzerine yapılan enerji tüketim tahmini ve çizelgeleme çalışmaları bulunduğu görülmüştür. Bu çalışmanın öncesinde yapılan [28] çalışmasında; bir hat için çoklu regresyon denkleminin uygun olduğu görülmüş ve

çalışma genişletilmiştir. Bu çalışmanın literatüre katkısı şunlardır:

- Literatürde araç işlem sürelerinin belirlenmesi üzerine bilindiği kadarıyla ilk olacaktır.
- Araçların işlem sürelerinin belirlenmesinde araca özgü bilgiler olan araç özneliklerini kullanılmıştır.
- Bu çalışmada güncelliği gösterilmiş olan üç teknik olan çoklu regresyon denklemi, sınıflandırma ve regresyon ağaçları ve yapay sinir ağlarının kullanımının uygunluğu araştırılmıştır.

Çalışmanın devamı şu şekilde düzenlenmiştir: Proses süresi için kullanılan tahmin algoritmaları 2. bölüm olan Metotlarda, verinin elde edilişi, manipülasyon ve dönüşüm işlemleri 3. bölüm olan Modelleme Öncesinde, çalışmada hazırlanan veri setleri ve model bilgileri 4. bölümde, bulgular ve tartışma 5. bölümde, çalışmanın değerlendirmesi ise 6. Bölüm olan sonuçlarda aktarılmaktadır.

2. METOTLAR (METHODOLOGIES)

Bu çalışmada üç farklı yöntem kullanılmıştır. Bu yöntemler sırasıyla Çoklu Doğrusal Regresyon (ÇRD), Sınıflandırma ve Regresyon Ağaçları (SRA) ve Yapay Sinir Ağlarıdır. Üç yaklaşımdan MLR ve SRA yöntemleri açıklayıcı olarak tahminleri gerçekleştirebilirken, yapay sinir ağı (YSA) öğrenme yöntemi ile tahmin gerçekleştirmektedir. Çalışmamızda araç özellikleri kullanılarak fabrika içindeki işlem süreleri tahmin edilecektir. Elimizde bulunan bağımsız değişkenlerden anlamlı olanların belirlenmesi çalışmanın önceliklerinden biridir. Burada ÇRD kullanarak modelde yer alacak bağımsız değişkenlerin anlamlılık sınaması (p değeri) üzerinden modelde tutulması veya çıkartılması sağlanacaktır. Çalışmadaki veri setlerini oluşturmadaki ilk varsayım bağımsız değişkenlerin anlamlılığının belirlenmesidir. Çalışmadaki verilerde değişkenlerin kategorik olması da mümkündür. Kategorik verilerde SRA tekniği kullanışlı olmaktadır. YSA kullanımı ise verinin doğrusal olmayacağı öngörüldüğünden kullanılmıştır. Doğrusal olmayan verilerde YSA başarılı sonuçlar vermektedir. Üç teknik ile gerçekleştirilen tahminler Kolmogorov-Smirnov testi ile sınanacaktır. Kullanılan yöntemlerin temelleri kısaca bu bölümde verilmiştir.

2.1. Çoklu Doğrusal Regresyon (Multiple Linear Regression)

Çoklu regresyon denklemi (ÇRD) model oluşturma, etkin değişkenlerin belirlenmesi ve tahminleme kısımlarında kullanılan en bilinen yöntemdir [28–35]. ÇRD'de, fazla sayıdaki bağımsız değişkenlerden istatistiki olarak anlamlı ve etkin olan değişkenler belirlenebilmektedir. İlk olarak modelin anlamlı olması istenmektedir. Sonrasında ise her değişkenin modelde anlamlı ve modele fayda sağlaması beklenmektedir. Regresyonda tahmin eden ile tahmin edilen arasında neden - sonuç veya etki - tepki ilişkisi vardır. Bu ilişki Eş. 1'de gösterilmiştir. Denklemden y sütun vektörü bağımlı değişken (tahmin edilen) olarak ifade edilirken, x matrisi bağımsız değişkenleri (öznitelik - tahmin eden)

göstermektedir. Tahmin eden değişkendir bir birim değişim katsayılar yardımıyla etki eder ve bunlar B sütun vektörü ile gösterilmektedir. e sütun vektörü ise ortalaması sıfır ve varyansı normal dağılım gösteren rastgele değişken olarak gösterilmektedir.

$$[y] = [x][B] + [e] \quad (1)$$

Çoklu doğrusal regresyon (ÇDR) denkleminin çeşitli çözüm yöntemleri bulunmakla beraber en sık kullanılan sıradan en küçük kareler yöntemidir (SEKKY - ordinary least square (OLS)). SEKKY'inde matris işlemlerinden yararlanılarak B vektörü hesaplanır. Eş. 1'de e vektörünün sıfır olması beklenmektedir. Bu durumda, B vektörü yalnız bırakılırsa denklem Eş. 2'deki gibi olur.

$$[B] = ([x]^T [x])^{-1} [x]^T [y] \quad (2)$$

Burada $[x]^T [x]$ kısmının tersi alınabilir bir matris olması önemlidir. Matrisin tersinin alınamaması SEKKY'nin kullanılmamasına sebep olur. Bu durumda sezgisel algoritmalar kullanılarak en yakın çözüm bulunmaya çalışılabilir. Matrisin tersinin alınamaması, birbiriyle çok yüksek ilişkisi olan iki değişkenin var olması (çoklu doğrusal bağlantı - multicollinearity) veya tek değerden oluşan bir bağımsız değişkenden kaynaklanabilmektedir.

Modelde sadece anlamlı olan bağımsız değişkenlerin bulunması hem esnek model oluşturulmasını hem de tahminlerin yüksek doğrulukta yapılmasını sağlamaktadır [31]. Modeldeki her değişkenin anlamlılığı için belirlenen t -istatistiği, her değişkenin modeldeki birim değişim katsayısının standart hatası bölünmesiyle elde edilir. t -istatistiği, modeldeki serbestlik derecesi olarak ifade edilen ve eleman sayısına da bağlı olarak t dağılım tablosu yardımıyla p değerine dönüştürülür [36]. Anlamlılığı test etmek için ÇDR denklemi üzerinden her değişkenin p değerine bakılması gerekmektedir. Her bağımsız değişken modelden çıkartıldığında, oluşturulan yeni modelde birim değişim katsayıları (B) ve standart hatalar değişmektedir. Bu durum t -istatistiği ve p değerinin değişmesine sebep olmaktadır [31]. Modeldeki bağımsız değişkensayı azaldıkça modelin esnekliği ve değişkenlerin anlamlılığının arttırması beklenmektedir. Uygun p değeri seçilerek bağımsız değişkenlerin belirli bir güvenilirlik düzeyinde modelde yer alması sağlanabilmektedir [34, 36].

MLR tahmin için kullanılabilirken, bağımsız değişken sayısını azaltmak ve etki eden bağımsız değişkenlerin belirlenmesinde de kullanılabilir. Bu kullanım detayı bir sonraki bölümde verilmiştir.

2.2. Sınıflandırma ve Regresyon Ağaçları (Classification and Regression Tree)

Sınıflandırma ve regresyon ağaçları (SRA) veri madenciliğinin önemli tekniklerinden biridir. SRA'lar sayesinde tahmin, sınıflandırma ve ayırıştırma

yapılabilmektedir. Diğer karar ağacı algoritmalarına kıyasla regresyon ve sınıflandırma özelliklerinin bulunmasından dolayı daha sık kullanılmaktadır [37]. SRA'ların sıklıkla kullanılmasının diğer bir sebebi kategorik veri türlerinde başarılı sonuçlar elde edebilmesidir [38].

SRA'lar kök dizinden başlayarak, ikili bir ağaç üzerinde bölünerek özyinelemeli olarak derinleşmektedir. Her düğümde yer alan veriler ebeveyninin bir alt sınıfı olur. Diğer bir ifadeyle, her düğüm çocuklarının bölünmesini sağlayan kümelerdir. SRA ağaçları en iyi çözümü bulduklarını garanti etmez [38]. Regresyon ağaçları tahminde kullanılırken, sınıflandırma ağaçları bölümlemede kullanılmaktadır. Sınıflandırma ağaçlarında bağımsız değişkenlerin kurallara göre sınıflandırılması olurken, regresyon ağaçlarında önceden tanımlı sınıflar yoktur. Regresyon ağaçlarında bağımsız değişkenler kullanılarak elde edilen katsayılar döndürülmektedir [39, 40]. SRA ağaçlarında ilk aşama maksimum ağacın oluşturulması, ikinci aşama en uygun ağacın belirlenmesi, son aşama ise oluşturulan ağacın yeni veri ile sınıflandırılması veya üretilmesidir. İlk aşama olan ağacın oluşturulmasında, regresyon temelli ağaçta bölme ve oluşturma yapılır. Burada toplam karesel hatanın en düşük olacağı yer belirlenmektedir [37]. Son gözlem incelenene kadar ağaçta bölünme devam etmekte ve büyük bir ağaç elde edilmektedir. İkinci aşamada elde edilen büyük ağaçta budama (pruning) yapılmaktadır. Bu aşama ağacın karmaşıklığının azaltılmasında önemlidir. Budama tekniği sayesinde anlamsız olan düğümler kırılır. Ön budama (Prepruning) ve son budama (postpruning) olarak iki yaklaşım vardır. Son budama daha yaygın olarak kullanılmaktadır [41]. Budama algoritması iki türlü olabilmektedir. Bunlar çapraz geçerlilik (cross-validation) ve yapraktaki maksimum eleman sayısının sınırlandırılmasıdır. Bölme işlemi yapılırken, yapraktaki veri sayısı önceden tanımlanmış olan yapraktaki minimum veri sayısından düşük duruma gelindiğinde optimizasyon durur. Çapraz geçerlilik tekniği ile budama sınıflandırma hatası ve ağacın karmaşıklığı kullanılarak sonlandırılmaktadır. Çapraz geçerlilik kısmında test veri setleri de olduğundan, her yeni durum için farklı ağaçlar olabilmektedir.

2.3. Tam Bağlı Yapay Sinir Ağları (Fully Connected Artificial Neural Network)

Yapay zeka tekniklerinden biri olan YSA, en çok kullanılan ve iyi bilinen bir tekniklerin başında gelmektedir [42, 43]. Bu teknikte insan sinir sistemi benzetilmeye çalışılmış ve enerji-elektrik tüketim tahmini [44, 45], kanser sınıflandırması [46], çizelgeleme [47] gibi karmaşık problemlerin çözümünde kullanılabilirdiği gösterilmiştir. YSA tekniğinin diğer kullanım alanı da doğrusal olmayan ve yüksek belirsizlik içeren problemlerdir. Yapay sinir ağları, tam bağlı YSA (fully connected neural network), zaman gecikmeli YSA (time delay neural network), evrişimli YSA (convolutional neural network), tekrarlayan YSA (recurrent neural network) gibi farklı yapılarda bulunabilmektedir. Bu çalışmada tam bağlı YSA modelleri kullanılmıştır. Tek katmanlı YSA yapısı ÇRD'ye benzemektedir fakat genellikle

tek katman olmayıp, en az bir gizli katman bulunur. Katmanlar arası düğümler (neron) ağırlık (w) olarak adlandırılan yapılar ile birbirine bağlanmaktadır [48–50]. Bir YSA; giriş, gizli ve çıkış olmak üzere en az üç katmandan oluşmaktadır. Gizli katman sayısı artırılarak daha karmaşık problemler çözülebilmektedir. Her girişin ağırlıklar ile çarpılarak toplanması ile bir değer elde edilir. Bu değer bir sonraki katmana geçip geçmeyeceğine, geçecekse ne derecede geçeceğine aktivasyon fonksiyonu karar verir. Yine ağırlıklardan dolayı eşiği geçemeyen sinyallerin ötelenmesi için de eşik (bias) değeri her ağırlık toplaması işleminde modele eklenir. Yapay sinir ağlarında geri yayılım (backpropagation) en iyi bilinen eğitim algoritmasıdır.

3. MODELLEME ÖNCESİ (PRE-MODELING)

Çalışmada modelleme öncesinde verinin hazırlanması gerekmektedir. Bu kapsamda üretimi yapılacak her aracın hatlar üzerinde işlem süresinin tahmini, verinin hazırlanmasından sonra gerçekleşmektedir. Fabrika içerisinde üretimi gerçekleştiren toplamda 7 alt hat bulunmaktadır (Tablo 1). Bu alanlar dışında, üretimin devamlılığını sağlamak için akışı sürekli olacak biçimde besleyen tampon bölgeleri de fabrika içindeki kaynak-boyahane arasında 2 adet ve boyahane-montaj arasında bir adet yer almaktadır. Üretim yapılan üç farklı ana hat sırasıyla kaynak, boyahane ve montajdır. Ana hatlarda kendi içinde birbirine paralel alt hatlar bulunmaktadır. Üç ana hattın biri olan kaynak hattı üç alt hattan oluşmaktadır. Bunlar *Vbk*, *Custom* ve *Cab* olarak ifade edilmektedir. Boyahane hattında *Eski Kabin* ve *Yeni Kabin* olarak iki alt hat bulunmaktadır. Montaj hattında da *Custom* ve *Transit* olarak iki alt hat bulunmaktadır.

Üretimdeki her araç yukarıda belirtilen üç ana hattın geçmekte olup hatların alt kısımlarında birbirinden farklı yollar izleyebilmektedir. Araçları fabrika ortamında birbirinden ayırmak için 80 haneye sahip Avrupa Sipariş Kartı (ASK - European Order Card (EOC)) adında bir seri kodu bulunmaktadır. Bu ifade içinde benzersiz araç değeri, aracın üretim yeri, donanım ve yazılımsal kendine özgü yapısı bulunmaktadır. Bu özellikler gövde tipi, sürücü tarafı (sol-sağ), motor tipi, vites türü, sunulan market (ülke), kilit türü, anahtar türü ve rengi gibi birçok parametreyi içermektedir.

Her aracın birbirinden ayırımında bu bağımsız değişkenler (öznitelikler) kullanılmıştır (Tablo 2). Bu bağımsız 80 haneli koddan, 43 adet bağımsız değişken elde edilmiştir. Öncelikli olarak Ford Otosan veritabanlarından üç farklı bölgeye ait süreleri içeren veri setleri elde edilmiştir. Bu veri setleri "Boyahane", "Kaynak & Kaynak Tampon", "Montaj & Montaj Tampon" için hazırlanmıştır. Araçların ASK kodları

bu veri kaynağında bulunmamaktadır. Onun yerine 17 karakterden oluşan araç tanımlama numarası (ATN - vehicle identification number (VIN)) kodları bulunmaktadır. Süreleri belirlenen araçların VIN kodları kullanılarak EOC kodlarına erişilmiş ve elde edilen ASK kodları ayrıştırılmıştır. Bir sonraki aşamada ise iki farklı kaynaktan alınan bu araç bilgileri birleştirilmiştir. Bu işlem sonrasında her araç tek bir satırda olacak biçimde alt hat bazlı kırılımda araçların harcadıkları süreler çeşitli dönüşüm ve özetleme işlemleri ile elde edilmiştir. Alt hat bazlı araç sürelerinin elde edilmesinde molalar, öğle araları, arızalar ve tatil günleri çıkartılmıştır. Bu aşamadaki diğer önemli bir işlem sayısal olmayan ve 2 karakter içerebilen bağımsız değişkenlerin sayısallaştırılmasıdır. Tüm ASK listesinden bağımsız değişkenlerin tekil olacağı bir meta liste elde edilmiştir. Bir sonraki aşamada bu meta tablosu sayısallaştırılmıştır. Sayısallaştırılmış meta tablosuna göre her araç için sayısal olarak yeni tablo elde edilmiştir. Ham veriden sayısallaştırmaya dönüşüm için Meta tablosu oluşturulmuştur. Bu tablonun oluşturulmasında 100.954 araç kullanılmış olup, 1056 adet meta bilgisi bulunmaktadır. 140 meta türü bulunan görünüm paketi (*visibility pack*), en çok meta türüne sahip bağımsız değişkendir. Görünüm paketi içerisinde ısıtmalı ön cam, ısıtmalı aynalar ve cam suyu seviye lambası, geri görüş kamerası, yıkama tankı uyarı seviyesi, ön sis lambaları, şeritten ayrılma uyarısı, Ice Pack 21 (Dijital Ses yayını (DAB)/AM/FM radyo, 2 adet USB, 8 inç dokunmak ekran, acil durum asistanı gibi özellikleri olan multimedya sistemi), ısıtmalı ön cam, ön cam suyu sensörü, elektrikli kapanır ve ısıtmalı aynalar gibi özellikler bulunmaktadır. Özel araç işlemleri; siparişte yer alan aracın itfaiye, ambulans, hava kirliliği ölçüm aracı gibi özel araç olması halinde gerçekleştirilen işlemlerdir. Şov aracı değişkeninin 6 türü bulunmaktadır. Proje aracı ve şov araç değişkenleri bağımsız değişken olarak ASK içinde bulunmaktadır fakat bu çalışmada kullanılmamıştır. Bu işlemlerden sonra toplamda 61.570 adet araç verisinin değeri, hem sayısal hem de alfanümerik biçimde analitik baz tablo olarak elde edilmiştir.

Bir sonraki aşamada üç ana hatta (kaynak, boyahane, montaj) bulunan alt hatlar (kaynak: *vbk*, *cab*, *custom*; boyahane: *eski kabin*, *yeni kabin*; montaj: *custom*, *transit*; kaynak tampon: *alt kat*, *üst kat*; montaj tampon) ayrı veri seti olacak biçimde bölünmüştür. Böylece 10 farklı veri seti her kısım için kendi işlem sürelerini kapsayacak biçimde hazırlanmıştır. Her kısım için ilk haldeki araç sayıları Tablo 3' deki ilk sütunda gösterilmiştir. Veri setlerinde bazı bağımsız değişkenlerin boş olduğu (örneğin gövde tipi) görülmüştür ve bu veriler kaldırılmıştır. Daha sonra kalan veriler proses süresi bazında incelenmiştir. Şekil 2'de yedi alt hat ve üç tampon bölgesinin proses süreleri iç içe iki grafik üzerinde verilmiştir. Grafiklerdeki yatay eksen,

Tablo 1. Ford Otosan Üretim ve Tampon Hatları (Ford Otosan process and buffer lines)

Ana Hat ve Ana Tampon Bölgeleri	Kaynak	Kaynak Tampon	Boyahane	Montaj Tampon	Montaj
Alt Hat ve Alt Tampon	Vbk Custom Cab	Ust Kat Alt Kat	Eski Kabin Yeni Kabin	Montaj	Transit Custom

süreleri düşükten yükseğe sıralanmış olan araçların sırasıdır. İç içe grafiklerde dışta bulunan grafik, tüm verileri göstermektedir. Örneğin, Boyahane Eski Kabinde yer alan 31.130 aracın, bu hattın 17 – 36.590 dakikalık süre aralığında geçtiği Şekil 2(a)'da görülmektedir. Bu grafiklerde, yükselen noktalarda uç değerler olduğu görülmektedir. Uç değerler hattın genel süresini göstermesini engellemekte ve veriyi bozmaktadır. Bu elde edilen değerlerde aşırı uç değerler olduğu Ford Otosan ile yapılan görüşmelerde onaylanmış ve temizlenmesi gerektiği ortaya çıkmıştır. Öncelikle çeyreklik aralığı kullanılmak istenmiş fakat verinin %50'sinin atılacağı görülmüştür. Bu oran çok yüksek olduğu için, hat işlem süresine göre küçükten büyüğe sıralanmış serinin alttan ve üstten sırasıyla %20, %15, %10 ve %5'lik kısımları kırılarak veri setleri incelenmiştir. Altan ve üstten %20'lik ve %15'lik (toplamda %40 ve %30) veri kırıldığında da çok fazla verinin modelde yer almadığı görülmüştür. Altan ve üstten %5'lik silinme durumunda ise yine uç değerlerin veri setlerinde bulunduğu görülmüştür.

Sonuç olarak, uç değerler için alt ve üst değerlerde %10'luk orandaki araçların silinmesi hem veriyi, hem de tahmini iyileştireceği öngörülmüştür. Böylece Tablo 3'de gösterilmiş "İlk Hal" değerlerin yerine "Temiz" değerler kullanılmış olacaktır. Böylece Boyahane Eski Kabin hattında temizlenmiş veriler ile elde edilmiş grafik Şekil 2(a)'da iç grafik olarak göstermiş ve bu grafiğin süre eksenini 300–1.200

dakika arasında gösterilmiştir. Böylece 17 dakika olan en düşük değer 385 dakikaya çıkmış, 36.590 dakika olan en yüksek değer de 1.207 dakikaya düşmüş ve zaman aralığı %97,79 azalmıştır. Benzer durum diğer hatlar için de geçerlidir. Şekil 2'de yer alan (b), (c), (d), (e), (f), (g), (h), (i) ve (j) grafiklerinde gösterilmiş olan Boyahane Yeni Kabin, Kaynak Custom, Kaynak Vbk, Kaynak Cab, Kaynak Tampon Alt Kat, Kaynak Tampon Üst Kat, Montaj Custom, Montaj Transit ve Montaj Tampon hatlarında indirgeme sonrası zaman aralıkları sırasıyla %98,47, %99,69, %98,76, %94,68, %95,38, %98,9, %99,72, %99,73 ve %99,43 düşmüştür. Veri setinde yapılan indirgeme ne kadar uç değerleri temizlese de, hat işlem sürelerindeki doğrusal olmama durumu devam etmektedir. Böylelikle, bu değerlerin gerçekten uç değer olduğu rahatlıkla görülmüş, yapılan çalışmanın doğruluğundan emin olunmuştur. Tablo 3'deki tanımlayıcı istatistikler temizlenmiş veri setlerine aittir ve en yüksek proses süresi olan hatlar sırasıyla boyahane, kaynak ve montaj hatlarıdır. En kısa sürede araçları prosese gönderen tampon bölgeleri ise sırasıyla, montaj tampon, üst kat kaynak tampon ve alt kat kaynak tampon olmuştur.

4. MODELLEME (MODELING)

Çalışmanın modelleme aşamasında ÇDR hem veri setindeki bağımsız değişkenlerin belirlenmesi hem de tahmin modellerinin oluşturulmasında kullanılmıştır. Verilerin

Tablo 2. Tahmininde kullanılan araç özellikleri (Vehicle properties used in the estimate)

Değişkenler					
Seri	Motor	Arka Cam	Trim Panel	Sıralı Koltuk	Görünüm Paketi
Çatı	Tavan	Şanzıman	Güneşlikler	Sürücü Tarafı	Motor Governörü
Renk	Bölme	Yıkayıcı	Çeki Demiri	Lastik Boyutu	Market Seviyesi
Marş	Kilit	Arka Aks	Yakıt Tankı	Koltuk Paketi	Yangın Söndürücü
Ayna	Modül	Arka Kapı	Şov Arabası	Proje Arabası	Gövde Geliştirmesi
Teker	Market	Ses Paketi	Park Desteği	Sıralı Pencere	Özel Araç İşlemleri
Klima	Anahtar	Gövde Tipi	Egzoz Çıkışı	Yükleme Tabanı	Brüt Araç Kütle Puanı
					Hava Deflektörü, Yük Bölmesi

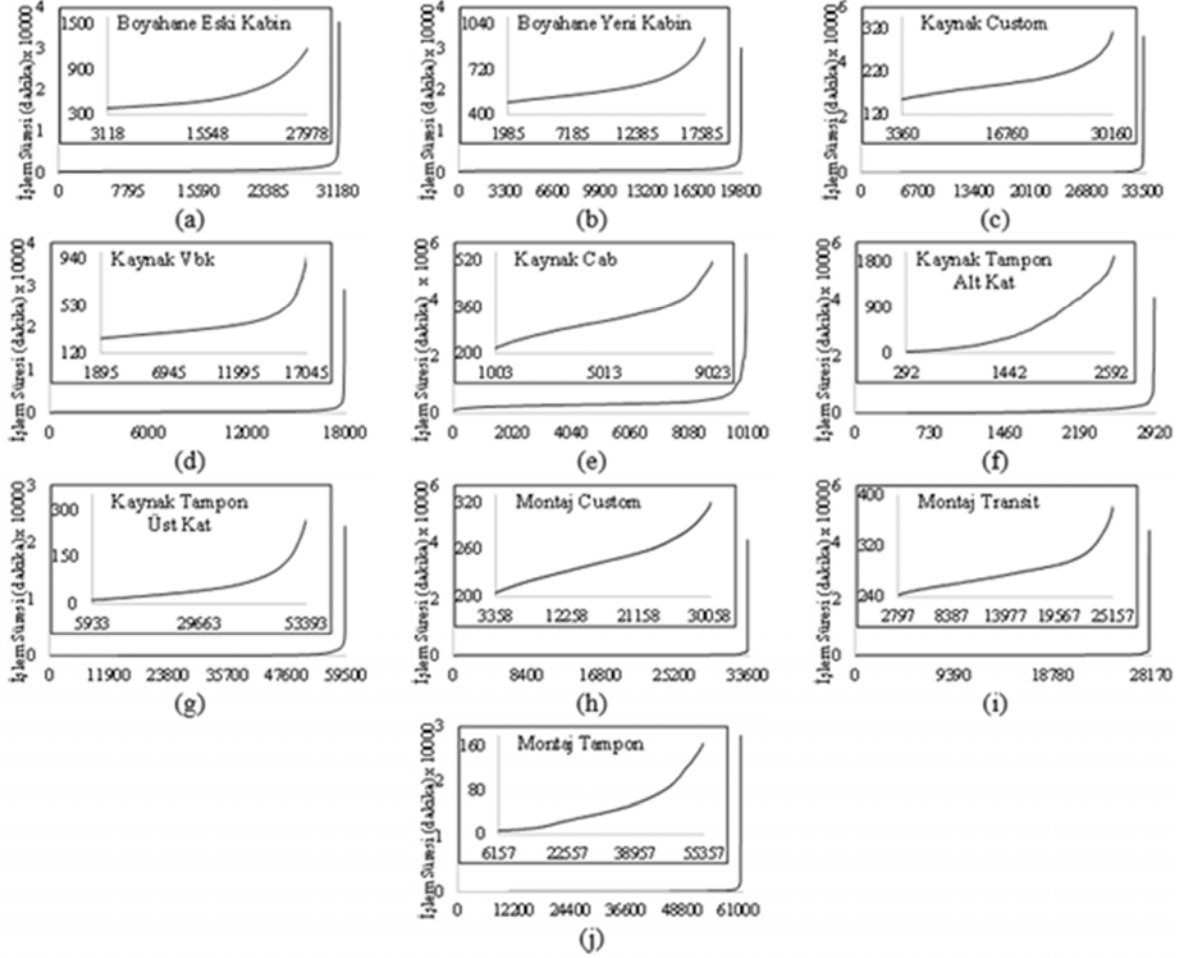
Tablo 3. Alt hat proses sürelerinin dakika cinsinden tanımlayıcı istatistikleri (Descriptive statistics of the line process durations in minutes)

Alt Hat	İlk Hal	Temiz	En Küçük	En Büyük	Aralık	Ortalama	Medyan	Mod	Standart Sapma	Toplam
Boyahane Eski Kabin	31.310	24.191	385	1.203	818	569	488	397	194	13.770.613
Boyahane Yeni Kabin	19.846	15.340	487	947	460	600	569	513	100	9.210.891
Kaynak Cab	10.032	7.938	217	521	304	321	311	302	65	2.548.049
Kaynak Vbk	17.951	13.964	246	606	360	342	322	281	77	4.777.796
Kaynak Custom	33.586	26.060	156	311	155	200	192	189	32	5.203.341
Kaynak Tampon Alt Kat	2.918	2.245	22	1.903	1.881	510	307	31	494	1.145.317
Kaynak Tampon Üst Kat	59.327	46.184	12	265	253	60	42	13	51	2.760.784
Montaj Transit	27.969	21.858	243	378	135	280	275	259	27	6.113.843
Montaj Custom	33.577	26.041	205	319	114	247	244	250	26	6.439.945
Montaj Tampon	61.569	47.759	7	167	160	49	38	7	40	2.345.513

sayısallaştırılmasında iki yaklaşım önerilmiştir (Tablo 4). İlk yaklaşımda alfanümerik ifadeler sayıya dönüştürülmüştür. *Body Type* (gövde tipi) özneliği A, B, C gibi alfanümerik ifadelerden oluşmaktadır. Sayıya dönüşümde “A”, “B”, “C” ifadelerine sırasıyla 1, 2, 3 denilmiştir. Tüm veriler meta bilgisi kullanılarak dönüştürülmüştür. İkinci yaklaşımda her bağımsız değişkenden meta ifadesi ayrı bir sütun olarak alınmıştır. Buna göre *Body Type* özneliğindeki 17 tipten sadece 1 tanesi “1”, diğer 16 tip “0” olacaktır. Diğer bir

ifadeyle ikinci yaklaşımda bağımsız değişkenler ikili hale getirilmiştir.

Veri seti bağımsız değişkenlerinin seçimi için farklı yaklaşımlar vardır. Bağımsız değişken sayısını azaltmak için için dört farklı yapı bulunmaktadır [51]. Bu çalışmada ileri doğru seçim (forward selection), geriye doğru eleme (backward elimination), aşamalı seçim (stepwise selection) ve belirlilik katsayısı (R^2) seçim yöntemleri kullanılmıştır



Şekil 2. Yedi alt hat ve üç tampon bölgesinin proses süreleri (Process durations of the seven line and three buffers)

Tablo 4. Veri setlerinin (a) alfa numerik, (b) sayısal, (c) ikili hale dönüştürülmesi (Converting data sets to (a) alfa numeric, (b) numeric and (c) binary)

(a)		(b)			
Ham - Alfa numerik		Sayısal			
Body Type	Color	Body Type	Color		
A	W	1	2		
B	Z	2	3		
C	Q	3	1		
(c)					
Body Type A	Body Type B	Body Type C	Color Q	Color W	Color Z
1	0	0	0	1	0
0	1	0	0	0	1
0	0	1	1	0	0

[51]. İleri doğru seçim algoritmasında sabit değer modele eklenerek başlanır ve her iterasyonda bir bağımsız değişken modele eklenir. İşlem eklenen bağımsız değişken anlamlı oldukça sonuna kadar devam eder. Modeldeki değişkenin anlamlılığı hipotez testi ile sağlanır ve p değeri ile gösterilir. Özniteliğin modele girişi için anlamlılık sınırı ileri seçim algoritmasında $0,5 \geq p$ olarak belirlenmiştir. Geriye doğru eleme algoritmasında ise tüm bağımsız değişkenler ile model oluşturulur. Elde edilen modeldeki bağımsız değişkenlerin anlamlılığı incelenir, en yüksek p değerine sahip bağımsız değişken modelden çıkartılır ve yeni model kurulur. Bu işlem iteratif olarak tüm bağımsız değişkenler anlamlı olana kadar devam eder. Modelde kalması için anlamlılık sınırı $0,2 \geq p$ olarak belirlenmiştir. Aşamalı seçim algoritması, ileri doğru seçim ve geriye doğru eleme yöntemlerinin birleştirilmiş halidir. İleri seçim algoritmasında olduğu gibi bağımsız değişkenin modele girişi için anlamlılık sınırı $0,5 \geq p$ olarak belirlenmiştir. Her yeni bağımsız değişken eklendikçe diğer bağımsız değişkenlerin modeldeki anlamlılıkları değişmektedir. Bu yüzden, bağımsız değişkenlerin modelde kalması için de $0,2 \geq p$ anlamlılık sınırı belirlenmiştir. Model oluşturmadaki diğer bir yaklaşım ise R^2 seçim yöntemidir [51]. Bu yöntemde bağımsız değişkenler içinden alt kümeler oluşturularak en doğru tahmini veren model belirlenmeye çalışılmaktadır. Bu yöntemde olabilecek tüm alt kümeler değerlendirilir ve elde edilen modellerin istatistiksel olarak doğru olup olmadığına bakılmaz. Bu yaklaşımdaki ana amaç farklı modellerde bulunan ortak bağımsız değişkenlerin belirlenmesidir. Ortak bağımsız değişkenler modelde yer alması gereken değişkenleri gösterecektir. Bu çalışmada her alt hat için modeller oluşturulmuş, her alt hat için sayısal - ikili veri setlerinde ayrı ayrı modeller sınanarak anlamlı bağımsız değişkenler belirlenmiştir. Tablo 5’de örnek bir bağımsız değişken azaltımı ve etki eden bağımsız değişkenlerin belirlenmesi gösterilmiş olup “Motor”, “Anahtar”, “Klima”,

“Marş”, “Arka Kapı”, “Tekere”, “Çatı”, “Kilit”, “Yıkayıcı”, “Görünüm Paketi”, “Arka Aks” gibi değişkenlerin anlamsız olduğu ($p \geq 0,2$) görülmüştür. Bu değişkenler modelden çıkartıldığında diğer bağımsız değişkenlerin anlamlılıklarının arttığı (p değerlerinin düştüğü) görülmektedir. Tablo 6’de, 7 alt hat ve 3 tampon alanı için 6 farklı bağımsız değişkenin modeldeki durumları gösterilmiştir. Bağımsız değişken belirlenmesinde kullanılan dört yapıya göre oluşturulmuş bu tabloda, değişken hücresinin doğru işareti olması modelde bulunması gerektiğini, çarpı işareti olması anlamsız olduğunu göstermektedir. Örneğin Kaynak Vbk kısmı için 3 bağımsız değişken anlamlı, 3 bağımsız değişken anlamsızdır.

Tablo 5’de gösterilmiş veri seti boyahane eski kabin için tüm sayısal veri kullanıldığında süresi 552 dakika ile başlarken, azaltılmış sayısal veride bu süre 562 dakika olmuştur. Tablo 4’de verilmiş olan Boyahane Eski Kabin için ortalamasının 569 dakika olduğu dikkate alındığında, bağımsız değişkenlerin belirlendiği model sabitinin ortalamaya yaklaştığı görülmektedir. Bu durum bağımsız değişken belirlenmenin faydalı olduğunu göstermektedir. Tablo 7’de ise ikili dönüşüm yapılmış veri seti üzerine uygulanmış olan örnek bağımsız değişken belirleme kısmı olarak gösterilmiştir. Boyahane eski kabin veri setinde bulunan 43 bağımsız değişken, ikili veri setinde 933 adet olarak elde edilmiştir. İkili bağımsız değişken sayısı her alt hat için farklılık göstermektedir (Tablo 8). Benzer şekilde bağımsız değişkenleri belirlenmiş veri setlerinde de her alt hat için anlamlı bağımsız değişkenler farklıdır. İkili veri setleri sayısal veri setlerine göre daha fazla özniteliğe sahiptir. Sayısal veri setlerinde “Gövde Tipi” özniteliğinin anlamsız olması halinde tümü modelden çıkartılmaktadır. İkili veri setlerinde ise bağımsız değişken tanımı (Tablo 7 – ÖT) “Gövde Tipi” içinde bulunan detaylı bağımsız değişken tanımlarında (Tablo 7 – AÖT) görülmektedir.

Tablo 5. Boyahane Eski Kabin için sayısal veri setinin bağımsız değişkenlerinin belirlenmesi
(Determination of the independent variables of the numerical data set for dyehouse old cabinet)

Bağımsız Değişken	Tüm Sayısal Veri				Belirlenmiş Bağımsız Değişken Sayısal Veri			
	Katsayı	Std Hata	t	p değeri	Katsayı	Std Hata	t	p değeri
Sabit	552,6553	20,03162	27,59	<0,0001	562,3466	14,1484	39,75	<0,0001
Gövde Tipi	2,2323	0,69177	3,23	0,0013	2,10154	0,41213	5,1	<0,0001
Motor	0,08087	0,20311	0,4	0,6905				
Çeki Demiri	-3,89552	0,73394	-5,31	<0,0001	-3,67064	0,70559	-5,2	<0,0001
Şanzıman	-1,96927	0,7525	-2,62	0,0089	-2,38105	0,67651	-3,52	0,0004
Anahtar	0,04615	0,2061	0,22	0,8228				
Ses Paketi	0,2891	0,20212	1,43	0,1526	0,33365	0,19167	1,74	0,0817
Klima	-0,4748	0,88009	-0,54	0,5896				
Marş	-0,40319	0,36686	-1,1	0,2718				
Sıralı Pencere	2,14006	0,85396	2,51	0,0122	2,77841	0,78646	3,53	0,0004
Sıralı Koltuk	1,35931	0,38615	3,52	0,0004	1,15778	0,35578	3,25	0,0011
Arka Kapı	0,5761	0,437	1,11	0,2670				
Arka Cam	-2,37764	0,85282	-2,79	0,0053	-2,69134	0,77122	-3,49	0,0005
Teker	0,49819	0,44999	1,11	0,2683				
Ayna	0,54315	0,26612	2,04	0,0413	0,46647	0,25624	1,82	0,0687
Çatı	-0,04056	0,58512	-0,07	0,9447				
Yakıt Tankı	-0,84077	0,40201	-2,09	0,0365	-0,97238	0,38122	-2,55	0,0108
Kilit	0,3543	1,3519	0,26	0,7933				
Görünüm Paketi	-0,04384	0,0336	-1,237	0,2161				
Lastik Boyutu	1,02209	0,43292	2,36	0,0182	1,01939	0,41407	2,46	0,0138
Arka Aks	0,25177	0,3104	0,81	0,4173				
Renk	-0,39315	0,0495	-7,94	<0,0001	-0,39636	0,04886	-8,11	<0,0001

Tablo 6. Farklı alt hatlar için örnek seçilmiş 6 özneliğin modeldeki durumları
(Status in the model of selected 6 sample attributes for the different lines)

Bağımsız Değişken	Boyahane										
	Boyahane Eski Kabin	Yeni Kabin	Kaynak Cab	Kaynak Custom	Kaynak Vbk	Montaj Custom	Montaj Transit	Kaynak Üst Kat Tampon	Kaynak Alt Kat Tampon	Montaj Tampon	
Gövde Tipi	✓	✓	x	✓	✓	✓	x	x	✓	✓	
Arka Cam	✓	✓	✓	✓	x	x	✓	✓	x	✓	
Kilit	x	✓	x	x	x	x	x	x	x	✓	
Renk	✓	✓	x	✓	✓	✓	x	✓	x	✓	
Motor	x	x	✓	x	x	x	x	✓	x	✓	
Çeki Demiri	✓	x	✓	✓	✓	x	✓	✓	x	✓	

Tablo 7. İkili veri setinin bağımsız değişkenlerinin belirlenmesi (Determination of the independent variables of the binary data set)

Bağımsız Değ.	ÖT	AÖT	Tüm Sayısal Veri				Belirlenmiş Bağımsız Değişken Sayısal Veri			
			Katsayı	Std Hata	t	p değeri	Katsayı	Std Hata	t	p değeri
Sabit			1.372,311	766,5639	1,79	0,0734	532,4365	56,8920	9,36	<0,0001
Col2	Gövde Tipi	1	-205,614	371,7378	-0,55	0,5802	-75,8534	29,9009	-2,54	0,0112
Col3	Gövde Tipi	2	-223,126	340,562	-0,66	0,5124	-88,6346	34,6314	-2,56	0,0105
Col4	Gövde Tipi	3	273,7304	569,6869	0,48	0,6309	-	-	-	-
Col5	Gövde Tipi	9	429,2567	588,4615	0,73	0,4657	-	-	-	-
Col6	Gövde Tipi	D	-620,988	726,7096	-0,85	0,3928	-15,7686	29,3997	-0,54	0,5917
Col7	Gövde Tipi	E	-1.068,12	676,5718	-1,58	0,1144	-	-	-	-
Col8	Gövde Tipi	F	-349,306	644,1003	-0,54	0,5876	-	-	-	-
Col9	Gövde Tipi	G	-512,564	659,0583	-0,78	0,4367	-67,7134	53,6458	-1,26	0,2069
Col10	Gövde Tipi	H	-579,414	677,0354	-0,86	0,3921	-	-	-	-
Col11	Gövde Tipi	J	16,04719	735,5363	0,02	0,9826	-	-	-	-
Col12	Gövde Tipi	K	-247,448	610,0473	-0,41	0,6850	37,1451	44,1984	0,84	0,4007
Col13	Gövde Tipi	M	-13,8299	296,0943	-0,05	0,9627	66,82507	43,1852	1,55	0,1218
Col14	Gövde Tipi	P	-122,751	290,8039	-0,42	0,6729	22,61559	40,7301	0,56	0,5787
Col15	Gövde Tipi	V	-408,55	594,0354	-0,69	0,4916	51,09767	76,6454	0,67	0,505
Col16	Gövde Tipi	X	-193,791	657,7117	-0,29	0,7683	-	-	-	-
Col17	Gövde Tipi	Y	0	0	0	0	-	-	-	-

Tablo 8. Alt hat veri setlerindeki bağımsız değişken sayıları (Numbers of attribute in line data sets)

Alt Hat	Sayısal Veri Seti	İndirgenmiş Sayısal Veri Seti	İkili Veri Seti	İndirgenmiş İkili Veri Seti
Boyahane Eski Kabin	43	24	933	374
Boyahane Yeni Kabin	43	23	831	399
Kaynak Cab	43	22	345	133
Kaynak Custom	43	24	642	340
Kaynak Vbk	43	24	539	257
Montaj Custom	43	26	645	304
Montaj Transit	43	16	626	308
Kaynak Tampon Üst Kat	43	30	996	540
Kaynak Tampon Alt Kat	43	12	743	272
Montaj Tampon	43	30	1.008	440

AÖT değerlerinden “3”, “9”, “E”, “H”, “J” ve “X” gibi tanımlar p değerinin yüksek olmasından dolayı modelden çıkartılabilmektedir. Böylece tüm değişken yerine AÖT’ler modelden çıkartılmakta ve anlamlı olanlar modelde tutulabilmektedir. İkili veri setleri bize “bağımsız değişkenden bağımsız - AÖT’ye bağımlı” model kurgulama imkanı vermektedir. Sayısal bağımsız değişkenlerin belirlendiği veride tüm değişken modelden kaldırıldığı için değişkenin içinde bulunduğu faydalı AÖT’ler kaldırılmaktadır. Fakat ikili yaklaşımda bunun önüne geçilmektedir. Yine benzer şekilde bağımsız değişkenleri belirlenmiş ikili veri setinde sabit terimi 532 dakika olarak bulunmuş ve ortalamaya yakın olarak elde edilmiştir. Tablo 5’deki her iki modelin R² değerleri sırasıyla 0,732 ve 0,694 olmuştur. Yine örnekteki modellerin hata terimlerinin normallik sınamalarında Kolmogorov-Smirnov testi (KS Testi) kullanılmış olup test istatistiği olan D değerleri 0,00767 ve 0,00743, p değerleri ise 0,1315 ve 0,1093 olarak belirlenmiştir. Hata; KS test istatistiği üzerinden

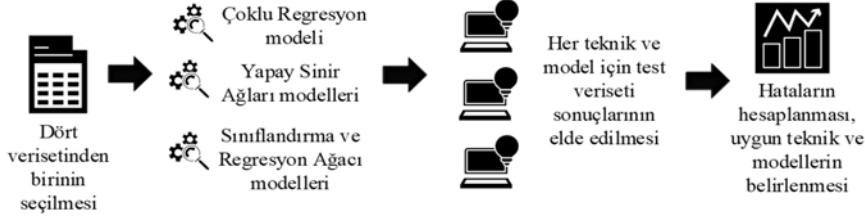
sınandığında, tüm sayısal veri içeren modelin normal dağılıma daha yakın olduğu ortaya çıkmaktadır. Her ne kadar tüm sayısal veriyi içeren modelin KS testi p değeri tüm bağımsız değişkenleri içeren modele göre daha yüksek hem de R² değeri tüm bağımsız değişkenleri içeren modele göre düşük olsa da, daha esnek bir model elde edilmesini sağlamıştır.

4.1. Çalışmanın Yol Haritası (Roadmap of the Study)

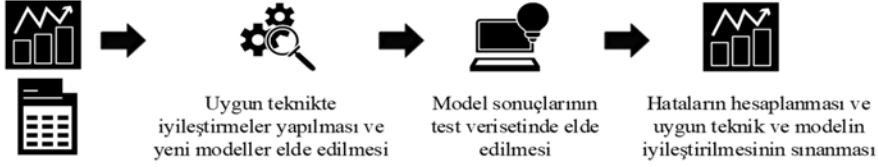
Bu bölümde çalışmanın süreçleri ve işlemlerinde nasıl bir yol izlendiği anlatılacaktır. Burada Şekil 3, Şekil 4 ve Şekil 5’deki gösterimler sadece bir hat için yapılan işlemlerdir ve tüm hatlar için aynı işlemler gerçekleştirilmiştir. Öncelikle Şekil 3’de görüldüğü gibi farklı kaynaklardan elde edilen veriler birleştirilerek analitik baz tablo (veriseti) elde edilmiştir. Bu verisetleri öncelikle Tablo 4’de gösterildiği gibi ikili veri biçimine dönüştürülmüştür, ÇRD kullanılarak sayısal ve ikili verisetlerinde bağımsız değişken sayıları



Şekil 3. Çalışmanın yol haritası – bölüm 1 (Roadmap of the study – part 1)



Şekil 4. Çalışmanın yol haritası – bölüm 2 (Roadmap of the study – part 2)



Şekil 5. Çalışmanın yol haritası – bölüm 3 (Roadmap of the study – part 3)

azaltılmıştır. Çalışmanın sonraki aşaması modelleme ve model tahmin sonuçlarının elde edilmesidir. Burada Şekil 4'de gösterilen ilk işlem olan dört verisetinden birinin seçilmesidir. Şekil 3'ün son adımındaki dört veriseti için Şekil 4 ve Şekil 5'deki işlemler gerçekleştirilmiştir. Seçilen veriseti için farklı modeller üç farklı teknik ile oluşturulmuştur. Bu modeller Tablo 8'de gösterilmiştir. Eğitilen modeller Tablo 10'da gösterilmiş olan test veriseti ile sınanmıştır. Test verisetlerinin de Şekil 4'deki gibi işlemler uygulanarak dönüşüm işlemleri gerçekleştirilmiştir. Elde edilen sonuçlar üç hata terimi ile değerlendirilmiştir ve uygun teknik ve modelin belirlenmesi sağlanmıştır.

Şekil 4'in son adımında belirlenen uygun teknik ve modeller ile bu modellerinin eğitiminde kullanılan verisetleri alınarak çalışmanın üçüncü bölümü başlamıştır (Şekil 5). Belirlenmiş olan uygun teknik ve modelin farklı parametre ve özellikler ile tekrar eğitilmesi ve iyileştirilmesi sağlanarak test veriseti ile test edilmiştir. Son aşama olarak da Şekil 4'de elde edilen modeller ile karşılaştırılarak tahmin hatalarında düşme ve tahminlerde iyileşmeler belirlenmeye çalışılmıştır.

5. SONUÇLAR VE TARTIŞMALAR (RESULTS AND DISCUSSIONS)

Yapılan çalışmada 4 farklı veri seti kullanılarak 3 farklı teknikle tahminler gerçekleştirilmiştir. Tablo 9'de farklı modeller açıklamaları ile verilmiştir. Buradaki ilk 4 model ÇDR sonuçlarıdır. Sırasıyla sayısal tüm değişkenlerin kullanıldığı (Model 1), anlamlılığına bakılarak belirlenen

sayısal değişkenlerin kullanıldığı (Model 2), ikili değişkenlerin tamamının kullanıldığı (Model 3) ve anlamlılığına bakılarak belirlenen ikili değişkenlerin kullanıldığı (Model 4) veri setlerinin modelleri gösterilmiştir. Sonraki 16 model SRA yapılarıdır. Devamında 25 model YSA yapısındaki modelleri göstermektedir. Tablo 9'da her teknik bir sütunda gösterilmiştir ve model oluşturulan her hücrede tekniğe ilişkin açıklamalar verilmiştir. Örneğin regresyon tekniği Model 1 ile Model 4 arasında değişmektedir ve katsayıları belirlemek için SEKKY kullanılmıştır. Opt parametresi 30 iterasyonlu Bayes optimizasyon modellerini gösterirken, Standart parametresi optimizasyonsuz ağaç yapısını göstermektedir. YSA tarafında ise köşeli parantez içinde yer alan değerler gizli katmanlardaki nöron sayılarını, bu değerlerin kaç adet olduğu ise gizli katman sayısını vermektedir.

Çoklu regresyonda 4 farklı model çalışılmıştır. Model 1 ve Model 2 değişkenlerinin bazılarının katsayıları, Boyahane Eski Kabin için Tablo 5'de verilmiştir. İkili değişken kullanan Model 3 ve Model 4'ün oluşturduğu katsayıların bazıları ise Boyahane Eski Kabin Tablo 7'de gösterilmiştir. Sayısal belirlenmiş veride tüm değişken modelden kaldırıldığı için değişkenin içinde bulunduğu faydalı tiplerde kaldırılırken, ikili yaklaşımda bu durum oluşmamaktadır.

ÇDR çalışması hem süre tahmin modelinde temel tahmin oluşturulmasında, hem de anlamsız verilerin temizlenmesinde kullanılmıştır. Böylece çoklu regresyon,

Tablo 9. Model tanımları ve parametreleri (Models definitions and parameters)

Model Adı	Veri Seti	Regresyon	Tam Bağlı Yapay Sinir Ağları	SRA
Model 1	Sayısal tüm veri	SEKKY		
Model 2	Sayısal belirlenmiş veri	SEKKY		
Model 3	İkili tüm veri	SEKKY		
Model 4	İkili belirlenmiş veri	SEKKY		
Model 5	Sayısal tüm veri			Regresyon Standart
Model 6	Sayısal tüm veri			Regresyon Opt 30
Model 7	Sayısal tüm veri			Sınıflandırma Standart
Model 8	Sayısal tüm veri			Sınıflandırma Opt 30
Model 9	Sayısal belirlenmiş veri			Regresyon Standart
Model 10	Sayısal belirlenmiş veri			Regresyon Opt 30
Model 11	Sayısal belirlenmiş veri			Sınıflandırma Standart
Model 12	Sayısal belirlenmiş veri			Sınıflandırma Opt 30
Model 13	İkili tüm veri			Regresyon Standart
Model 14	İkili tüm veri			Regresyon Opt 30
Model 15	İkili tüm veri			Sınıflandırma Standart
Model 16	İkili tüm veri			Sınıflandırma Opt 30
Model 17	İkili belirlenmiş veri			Regresyon Standart
Model 18	İkili belirlenmiş veri			Regresyon Opt 30
Model 19	İkili belirlenmiş veri			Sınıflandırma Standart
Model 20	İkili belirlenmiş veri			Sınıflandırma Opt 30
Model 21	İkili tüm veri		[500 750 1500 700 550] 6000 iter.	
Model 22	İkili tüm veri		[500 750 1500 700 550] 6000 iter.	
Model 23	İkili tüm veri		[250 350 500 300 150] 6000 iter.	
Model 24	İkili tüm veri		[750 2500 700] 6000 iter.	
Model 25	İkili tüm veri		Ortalama	
Model 26	Sayısal tüm veri		[500 750 1500 700 550] 6000 iter.	
Model 27	Sayısal tüm veri		[500 750 1500 700 550] 6000 iter.	
Model 28	Sayısal tüm veri		[250 350 500 300 150] 6000 iter.	
Model 29	Sayısal tüm veri		[750 2500 700] 6000 iter.	
Model 30	Sayısal tüm veri		Ortalama	
Model 31	Sayısal belirlenmiş veri		[500 750 1500 700 550] 100 iter.	
Model 32	Sayısal belirlenmiş veri		[500 750 1500 700 550] 100 iter.	
Model 33	Sayısal belirlenmiş veri		[250 350 500 300 150] 100 iter.	
Model 34	Sayısal belirlenmiş veri		[750 2500 700] 100 iter.	
Model 35	Sayısal belirlenmiş veri		Ortalama	
Model 36	İkili belirlenmiş veri		[500 750 1500 700 550] 50 iter.	
Model 37	İkili belirlenmiş veri		[500 750 1500 700 550] 50 iter.	
Model 38	İkili belirlenmiş veri		[250 350 500 300 150] 50 iter.	
Model 39	İkili belirlenmiş veri		[750 2500 700] 50 iter.	
Model 40	İkili belirlenmiş veri		Ortalama	
Model 41	İkili belirlenmiş veri		[500 750 1500 700 550] 10000 iter.	
Model 42	İkili belirlenmiş veri		[500 750 1500 700 550] 10000 iter.	
Model 43	İkili belirlenmiş veri		[250 350 500 300 150] 10000 iter.	
Model 44	İkili belirlenmiş veri		[750 2500 700] 10000 iter.	
Model 45	İkili belirlenmiş veri		Ortalama	

eğitim temelli veri madenciliği ve yapay zeka teknikleri olan SRA ve YSA modellerine karar aşamasında yardımcı tahmin olarak gösterilmekte olup, veriyi değişken bazlı azaltma ve anlamlı verinin elde edilmesinde kullanılmıştır.

Ağaç yapısındaki tahminlerde 4 farklı ağaç modeli elde edilmiştir. Bu ağaçlar ilk olarak sırasıyla sınıflandırma ve regresyon ağaçları olarak ayrılırken, ikinci olarak da en fazla bölünme sayısını vererek standart ağaç oluşturulması ve Bayes optimizasyonunu kullanan ağaç olarak ayrılmaktadır. Diğer bir ifadeyle 4 farklı veri seti için 4 farklı tahmin oluşturulmuştur. Bayes optimizasyonunda sınıflandırma ve regresyon ağacının oluşturulmasında 30 iterasyon ile uygun ağaç bulunmaktadır.

YSA tahminlerinde 4 farklı ağ yapısı kullanılmıştır. Bu ağ yapılarından ilki ileri beslemeli ağ yapısıdır ve 5 gizli katmanlı [500 750 1500 700 550] nöron sayısına sahiptir. İkinci ağ yapısı ilk ağ yapısı ile aynı nöron sayısına sahiptir 2270

fakat ölçekli eşlenik gradyan geri yayılım (ÖEGGY - scaled conjugate gradient backpropagation) algoritması kullanılmıştır. ÖEGGY algoritması üçüncü ve dördüncü ağ yapılarında da kullanılmış olup bu ağ yapılarında sırasıyla [250 350 500 300 150] ve [750 2500 700] nöron kullanılmıştır. Bu 4 modelin tahmin sonuçlarının ortalaması alınarak, beşinci tahmin sonucu oluşturulmuştur. Transfer fonksiyonu olarak daha yüksek doğruluk [52] ile, üstel veya trigonometrik fonksiyon gerektirmedikinden ($f(x) = 1/(1+|x|)$) basit hesaplama sayesinde daha az işlem süresi [53] gerektirdiği için Elliot simetrik sigmoid kullanılmıştır. Bu ağların eğitiminde NVIDIA GeForce 1080 GPU ve Intel i7 6700K 4Ghz donanım kullanılmıştır. Burada da 4 farklı veri seti için her hatta 5 tahmin sonucu üretilmiştir. YSA modellerindeki ağın girdilerinin sayısı Tablo 8'de gösterilmiştir ve bu girdilere göre alt hat işlem süreleri model çıktısı olarak tahmin edilmiştir. Eğitim ve test veri setlerinde, her alt hat için farklı sayıda araç verisi bulunmaktadır. Eğitim veri setleri sırasıyla Boyahane Eski Kabin hattı için

24.191, Boyahane Yeni Kabin hattı için 15.340, Kaynak Cab hattı için 7.938, Kaynak Vbk hattı için 13.964, Kaynak Custom hattı için 26.060, Kaynak Tampon Alt Kat için 2.245, Kaynak Tampon Üst Kat için 46.184, Montaj Transit hattı için 21.858, Montaj Custom hattı için 26.041 ve Montaj Tampon için 47.759 araç eğitimde kullanılmıştır.

5.1. Test Veri Seti ve Hataların Ölçülmesi (Test Data Set and Errors Measurement)

Bu bölümde alt hat bazlı sonuçlar gösterilmiştir. Sonuçların yorumlanmasında 3 farklı hata terimi kullanılmıştır. Bunlar sırasıyla OMYH (Mean Absolute Percent Error – Ortalama Mutlak Yüzdesele Hata - Eş. 3), OMH (Mean Absolute Error – Ortalama Mutlak Hata - Eş. 4) ve OKH (Mean Square Error – Ortalama Karesel Hata - Eş. 5)' dir. Bu hata terimlerinin formülleri aşağıda sırasıyla verilmiştir.

OMYH birimsiz olarak oran döndürmektedir. Genel olarak zaman serilerinde, zamana bağlı sonuçlarda ve birbirine bağlı olan olayların hatasını ölçmede kullanılır [34, 54, 55]. OMYH'da \hat{y}_i ; i . tahmini, y_i ; i . gerçekleşeni, n ; eleman sayısını göstermektedir (Eş. 3). Negatif ve pozitif hataların birbirini götürmesi mutlak değer ile engellenmiş olur.

$$OMYH = \left(\frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \right) \cdot 100\% \quad (3)$$

OMH ise ortalama hatayı yönüzsüz olarak bulur [57, 59]. Birimi; hesaplanan değer kendisidir (Eş. 4). Bu çalışmada dakikadır. Diğer bir ifadeyle ortalama kaç dakikalık hata ile sonuçların bulunduğunu göstermektedir. Mutlak değer ifadesi de yönün etkisini ortadan kaldırmak için kullanılır.

$$OMH = \left[\frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \right] \quad (4)$$

OKH ise OMH' den farklı olarak yönünü karesel olarak pozitif tarafa döndürür (Eş. 5). Birimi kareseldir [31,58,59]. Yani dakika kare olarak ifade edilir. Burada büyük hataların etkisi kendini daha fazla hissettirmektedir.

$$OKH = \left[\frac{1}{n} \left(\sum_{i=1}^n (\hat{y}_i - y_i)^2 \right) \right] \quad (5)$$

Hazırlanan modelleri test etmek için her alt hat için yeni veri setleri hazırlanmıştır. Test veri setinin tanımlayıcı istatistikleri Tablo 9' da gösterilmiştir. Bazı araçların işlemi bitmediği veya iptal edilmesinden dolayı sistemden süre gelemediği için en küçük değerlerinin sıfır, en büyük değerlerinin 11.000 dakikayı geçmekte olduğu görülmüştür. Bu durum aslında veri setinde eğitim veri setinde de olduğu gibi uç değerler olduğu göstermiştir.

Test veri setinde de alt ve üst kısımdan %10 oranında veri kırılmıştır (Tablo 11). Test veri setinde uç değerler çıkartıldığında tüm hatlar için son duruma göre ortalama %29'luk değişim meydana gelmiştir. Bu değer kırmanın etkisinin fazla olduğunu göstermektedir. Mod durumları incelendiğinde ise en çok tekrar eden değerlerin sıfır olduğu görülürken, temizleme sonrası mod değerleri değişmiştir. Uç değerlerin çıkartıldığı veri setinde ise medyanın boyahane eski kabin, kaynak tampon alt kat ve montaj tampon hatları dışında ortalamaya yakın değerler olduğu görülmüştür. Eski durumdaki aralık yeni durum aralığı ile karşılaştırıldığında ise ortalama 25 katlık fark görülmektedir.

Tablo 10. Test veri seti ham halinin tanımlayıcı istatistikleri (Descriptive statistics of test data set in the raw form)

Alt Hat	Araç Sayısı	Ortalama (dk)	Medyan	Mod	Standart Sapma (dk)	En Küçük	En Büyük	Aralık	Toplam
Boyahane Eski Kabin	2.880	498	388	0	407	0	5.141	5.141	1.435.335
Boyahane Yeni Kabin	13.250	592	502	0	628	0	11.241	11.241	7.839.994
Kaynak Cab	5.559	376	296	293	372	26	3.834	3.807	2.089.126
Kaynak Vbk	11.250	341	255	0	421	0	9.715	9.715	3.838.253
Kaynak Custom	2.839	134	132	0	121	0	3.533	3.533	381.485
Montaj Transit	16.809	401	296	0	441	0	3.617	3.617	6.736.170
Montaj Custom	3.123	228	231	0	95	0	1.484	1.484	712.332
Kaynak Tampon Alt Kat	935	504	145	0	887	0	9.572	9.572	471.337
Kaynak Tampon Ust Kat	3.080	91	21	0	319	0	4.563	4.563	280.682
Montaj Tampon	3.311	115	53	0	220	0	3.818	3.818	382.404

Tablo 11. Test veri seti, uç değerlerden temizlenmiş halinin tanımlayıcı istatistikleri (Descriptive statistics of test data set in cleared from outliers)

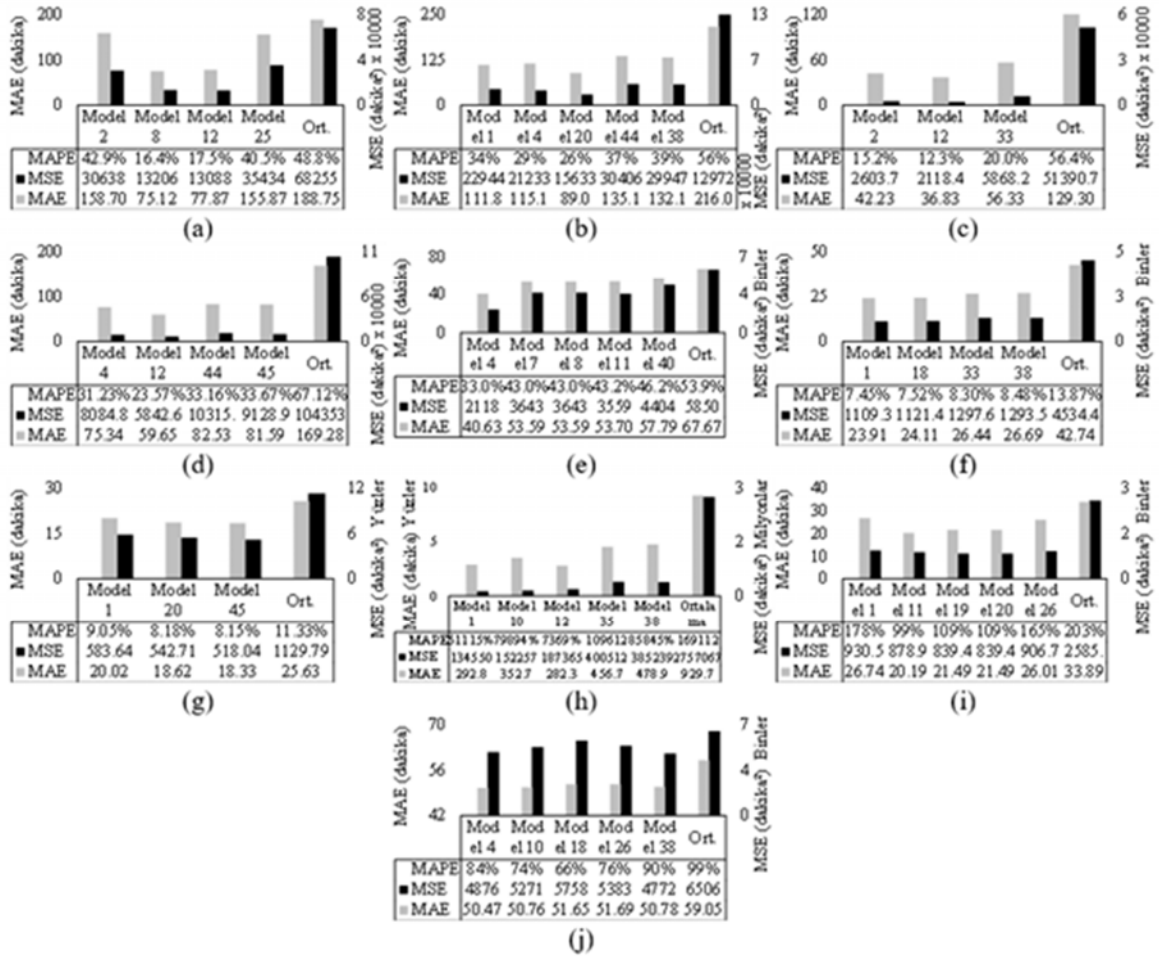
Alt Hat	Araç Sayısı	Ortalama (dk)	Medyan	Mod	Standart Sapma (dk)	En Küçük	En Büyük	Aralık	Toplam
Boyahane Eski Kabin	2.307	423	388	330	112	212	845	633	976.249
Boyahane Yeni Kabin	10.602	486	502	469	123	182	735	553	5.155.406
Kaynak Cab	4.448	299	296	293	45	219	443	224	1.329.536
Kaynak Vbk	9.003	268	255	244	51	197	397	200	2.413.421
Kaynak Custom	2.273	132	132	129	16	93	166	73	298.926
Montaj Transit	13.451	300	296	285	28	257	401	143	4.040.173
Montaj Custom	2.499	232	231	223	16	193	263	70	579.799
Kaynak Tampon Alt Kat	749	304	146	0	345	0	1.288	1.288	227.793
Kaynak Tampon Ust Kat	2.467	29	21	7	22	7	110	103	71.948
Montaj Tampon	2.651	80	53	30	65	11	221	210	212.645

Araçların işlem süresi ise bağımsız değişkenleri belirlenmiş veri setinde %50 oranında azalmıştır. Burada en yüksek işlem süresi boyahane, montaj ve kaynak olarak sıralanmıştır. Bu sıralama [18] numaralı referansta bulunan çalışmaya paralellik göstermektedir. Yapılan kırpmanın doğruluğu böylece görülmüştür. Bu aşamadan sonraki tüm test sonuçları uç değerler arındırılmış veri seti üzerinden gösterilmiştir.

5.2. Alt Hat Bazlı Modellerin Proses Sürelerinin Tahmini (Process Duration Results of Line Based Models)

Bu bölümde her alt hat için tahmin modellerinin sonuçları gösterilecek ve en uygun tekniğin belirlenmesi sağlanacaktır. Regresyon, SRA ve YSA modellerinde OMYH, OKH ve OMH değerlerinde en düşük sonuç veren modeller Şekil 6'de gösterilmiştir. Grafiklerde sol eksen OMH değerini, sağ eksen OKH değerlerini göstermektedir. Boyahane Eski Kabinde 45 modelin ortalama OMH değeri 188,75 dakika, OKH değeri 68.255,84 dakika², OMYH değeri %48,84 olmuştur (Şekil 6a). Regresyon modellerinde Model 2 OMYH, OKH ve OMH değerlerinde en düşük sonucu

vermiştir. SRA modellerinde ise Model 8 OMYH ve OMH'de en düşük sonuçları verirken Model 12 OKH'de en düşük sonuçları vermiştir. YSA'da ise Model 25 en düşük OMYH, OKH ve OMH değerlerini vermiştir. Boyahane yeni kabinde modellerin ortalama OMH değeri 216 dakika olmuştur (Şekil 6b). Yapay zeka yaklaşımlarında ise bu süre Model 8, Model 16 ve Model 20'da 89 dakika olmuştur. Bu üç model de Bayes optimizasyonun kullanıldığı sınıflandırma ağaçlarıdır. Regresyon modelleri YSA modellerine göre daha iyi sonuç verirken, 45 modelin OMYH, OMH ve OKH ortalaması her üç tekniğin en iyi sonuç veren modellerinin gerisinde kalmıştır. Kaynak Cab hattında modellerin ortalama OMH değeri 129 dakika olmuştur (Şekil 6c). Yapay zeka yaklaşımlarında ise bu süre Model 12'de 37 dakika olmuştur. En düşük OMYH ile OKH değeri de aynı modelde gerçekleşmiştir. Bu model dışında OMH ve OMYH'de 31 model ortalama değerlerin altında hata değerleri gösterirken, OKH'de 34 model ortalamanın altındadır. Her üç teknikteki en düşük hatalı modeller OMYH, OKH ve OMH için de aynı bulunmuştur. Bu üç model de OMH'de ortalamanın yarısından az, OKH'de 1/20, OMYH'de ise yaklaşık 1/6 oranında hata ile sonuçları



Şekil 6. Yedi alt hat ve üç tampon bölgesinin araç proses sürelerinin tahminleri
(Process durations estimations of seven line and three buffer area)

göstermektedir. Kaynak Vbk hattında modellerin ortalama OMH değeri 169 dakika olmuştur (Şekil 6d). Yapay zeka yaklaşımlarında ise bu süre Model 12' de 60 dakika olmuştur. YSA modelleri regresyon ve ağaç modellerine göre daha yüksek hata değerleri bulurken, ağaç modelleri en düşük hataya sahip sonuçları vermiştir. OMYH ve OMH için 34, OKH için 39 model ortalama hata değerlerinden daha düşük gerçekleşmiştir. En iyi modellerin ortalama OKH değerine göre 1/10'luk düşüşü, sadece düşük proses süreli (197 – 250 dakika) araçların değil, yüksek süreli (300 – 397 dakika) araçların da tahmininde hataların düşük olduğunu göstermektedir. Kaynak Custom hattında modellerin ortalama OMH değeri 68 dakika olmuştur (Şekil 6e). Bu hat üzerinde en düşük hataya sahip tahminler regresyon modeli olan Model 4' de gerçekleşmiş ve 40,63 dakikalık ortalama sapma olmuştur. Ağaç yapısında ise OMYH, OMH ve OKH değerleri için Model 7 ve Model 8 aynı hata oranlarına sahip olmuştur. Model 11 ise ağaç modelleri arasında en düşük OKH değerine sahiptir. Genelde regresyon ve ağaçlara göre daha yüksek OMYH, OMH ve OKH veren YSA modelleri, bu hattaki süre tahminlerinde üç hata terimi göz önüne alındığında diğer iki tekniğe yaklaşmıştır. Yapay zeka tekniklerinde OKH değerinin birbirine yakın olması, serilerde paralel tahminlerin olduğunu göstermektedir. Montaj transit hattında modellerin ortalama OMH değeri 42,74 dakika olmuştur (Şekil 6f). Tahmin için hazırlanan 45 modelden 33 model OMYH hatasında ortalamadan düşük olurken, OKH'de 38, OMH'de 32 model daha düşük hataya sahip olmuştur. Model 1 ve Model 18 26 dakikalık hataya sahipken, YSA'daki Model 33 ve Model 38 26,5 dakika civarında hataya sahip olmuştur. Burada da regresyon ve ağaç yapılarının daha düşük OMYH, OMH ve OKH ile tahmin yaptıkları görülmektedir fakat YSA modelleri de ÇDR ve SRA tekniklerine yaklaşmıştır. Montaj Custom hattında modellerin ortalama OMH değeri 25,63 dakika olmuştur (Şekil 6g). Diğer hatlardan farklı olarak Montaj Custom hattında YSA modellerinden biri olan Model 45 (ikili veride bağımsız değişkenleri belirlenmiş veri setinde dört YSA modelinin tahmin ortalamalarının alındığı model) tüm hata değerlerinde en düşük hatalı sonuçları vermiştir. YSA modelini sırasıyla ağaç modeli olan Model 20 izlemiştir. Ağaç modellerinde, ortalama süre hatası olan OMH 22 dakika ve OMYH %9,8 değerine sahiptir. Regresyonda ise 4 modelin ortalama OMYH değeri %10,05 olurken, OMH değeri 22,47 dakika olmuştur. Tampon hatlarının üretim hatlarından farklı bir durumu bulunmaktadır. Üretim hatlarında şasenin farklılığı kaynak biçimi ve bölgelerini etkileyebilmektedir. Benzer şekilde boya rengi üretimin yapılacağı kabinde farklılıklar gösterebilir. Ekstra eklenecek parçalar montaj süresini değiştirebilir. Üretim hatlarında gerçekleştirilen her işlemin araçların özellikleri ile bağlantısı bulunmaktadır. Tampon hatlarında ise bu durum sipariş durumu, termin tarihi, özel üretimlerde parça tedariki gibi süreçlerden etkilenebilmektedir. Bu durumda aracın özelliklerine göre süre tahmininden çok iki üretim hattı arasındaki bekleme, başka değişkenlere bağlı olmaktadır. Bu sebeple tampon hatlarının süre tahmin değerlendirmesi buna uygun olarak değerlendirilmelidir. Tampon hatlarından ilk kısım kaynak hatları ve boyahane hatları arasında bulunanlar kaynak tamponudur. Bu tamponda üst ve alt kat olmak üzere iki

kısım vardır. Üst kat tampon bölgesi, alt kata göre daha yoğun kullanılmaktadır. Bunun sebebi kaynak ve boyahanenin farklı binalarda olmasıdır. İki bina (kaynak – boyahane) arasında oluşturulmuş hat yolu kesmeyecek biçimde üst kattan gitmektedir. Parça tedariki ve benzeri durumlarda da araç, kaynak üst kat tampon bölgesine girmeden alt katta bekletilmektedir. Kaynak alt kat tampon bölgesinde araçların geçirdiği süre tahmininde modellerin ortalama OMH değeri 930 dakika olmuştur (Şekil 6h). Bu tampon bölgesinde en iyi sonuç veren yaklaşım Model 12 olmuştur ve OMH değeri 282 dakika olmuştur. Model 1'de ise OMH değeri 292 dakika olmasına rağmen, OKH değeri Model 12'den daha düşük gerçekleşmiştir. Bu durum, tampon bölgesinde yüksek bekleme süresi olan araçlarda Model 12'nin daha fazla hata yaptığını göstermektedir. YSA modellerinde OMYH ve OKH için Model 38 en düşük değerleri verirken, OMH değerinde Model 35 en düşük değeri vermiştir. Ağaç modellerinde de en düşük OMYH ve OMH Model 12'de gerçekleşirken, en düşük OMH değeri Model 10'da gerçekleşmiştir. Regresyon modelinde ise üç hatanın da en düşük olduğu yaklaşım Model 1 olmuştur. Kaynak Üst Kat tampon hattı ve bölgesinde daha düşük bekleme süreleri bulunmaktadır. Bu sürelere uygun olarak da tüm modellerin ortalama OMH değeri 34 dakika olmuştur (Şekil 6i). Yapay zeka yaklaşımlarında ise en düşük OMH değerine sahip yaklaşım OMH değeri 20 dakika ile Model 11 olup OMYH değeri de en düşük modeldir. Model 1 regresyon modelleri arasında en düşük hata terimlerine sahip yaklaşım olurken, Model 26 YSA modelleri arasında en düşük hata terimlerine sahip yaklaşım olmuştur. Her üç teknikte de tahminlerin hataları düşük düzeyde kalmıştır. Montaj tampon bölgesi boyahane ile montaj arasında yer almaktadır. Boyahanedan çıkan araçlar montaj hattını beslerken bekleme gerektiğinde bu bölgede bulunmaktadır.

Bu bölgenin tüm modellerdeki ortalama bekleme süresi (OMH değeri) 59 dakika olmuştur (Şekil 6j). Üç hata terimi için ağaç modellerinde iki, YSA modellerinde iki, regresyon modellerinde bir yaklaşım en düşük hata terimlerine sahiptir. En düşük OMH değeri regresyon yaklaşımındaki Model 1'de 50,47 dakika ile gerçekleşirken, en düşük OMYH değeri Model 10 ile ağaç yaklaşımında gerçekleşmiştir. OKH değeri en düşük olan YSA yaklaşımı olmuş ve Model 38 ile gerçekleşmiştir. Her üç terim de farklı algoritmalar ile en düşük değerde tahminde bulunmuştur. Genel olarak bakıldığında üç teknikte de OMH ve OKH değerleri birbirine çok yakın gerçekleşirken, OMYH değerinde farklılıklar görülmektedir. Bu durum, düşük sürelerin tahminlerinin daha yüksek hataya sebep olduğunu göstermektedir.

5.3. Alt Hat Bazlı Model Sonuçlarının İyileştirilmesi (Improvement of Line Based Results)

Sonuçlarda en iyi araç bazlı süre tahmininin SRA modellerinde olduğu görülmüştür. Bu aşamada SRA modellerinde özelleştirilmiş biçimlerde yeni tahmin modelleri elde edilmiştir ve Tablo 12'da gösterilmiştir. Bu model sonuçları ile önceki tahmin sonuçları karşılaştırılmıştır. Bu yeni modellerde "Opt" ifadesi Bayes

optimizasyonunu, 500 değeri iterasyon sayılarını göstermektedir. Standart ağaç modellerinde minimum yaprak sayısının belirtilmesi de iyileştirme çalışmalarına eklenmiştir. Böylece, eğer belirtilmişse, modeller en düşük yaprak sayısı 100 olacak şekilde oluşturulmuştur. Bayes optimizasyonu ile ağaç oluşturulurken yaprak sayısında alt veya üst limit konulamamaktadır. Ağaç yaklaşımında önce 16 yeni model eğitilmiş, sonrasında test veri seti ile sınanmıştır.

İlk bulunan 45 model ve ikinci aşamada oluşturulan 16 modeldeki sonuçlarda en düşük değerler Tablo 13’de sırasıyla “Başlangıç Modelleri” ve “Geliştirilmiş Modeller” olarak verilmiştir. Başlangıç modelleri ile geliştirilmiş modeller karşılaştırıldığında yatık alanlı hücreler sonuçlarda iyileşme olan kısımları, altı çizili olan hücreler aynı kalan kısımları, diğer hücreler ise sonuçların iyileşmediği kısımları göstermektedir. Buna göre Boyahane yeni kabin hattı ve Montaj tampon bölgesinde herhangi bir değişiklik olmamıştır. Boyahane Yeni Kabinde Geliştirilmiş modeller içerisinde yer alan Model 55 ve 59 da, Model 20’de olduğu gibi Bayes optimizasyonu ile sınıflandırma ağacı oluşturan belirlenmiş ikili veri seti kullanan modellerdir. Model 59’un Model 20’den tek farkı iterasyon sayısının 500 olmasıdır. Benzer şekilde Model 55 de 500 iterasyonlu olup, ikili değişkenlerin tümü modellemede kullanılmıştır. Model 20 ile Model 59’un aynı değerleri bulması ilk 30 iterasyon içinde optimum çözümün bulunduğunu göstermiştir. Model 55’de de aynı değerlerin bulunması bağımsız değişken sayısı artmasa da iterasyon sayısını artırılarak optimum çözüme ulaşabileceğini göstermiştir. Benzer durum Boyahane Eski Kabin’de de bulunmaktadır.

Bu alt hat üzerinde kullanan model olan Model 8; aynı veri seti ve yaklaşımı kullanarak 500 iterasyona sahip Model 47 ile üç hata teriminde de aynı sonuçları vermiştir. OKH değerinde, Model 8 ve 47 en düşük değerleri elde edemedikleri ($OKH_{Model\ 8\ ve\ 47} = 13.206$) için tabloda bu durum görülememektedir. Boyahane Eski Kabinde Model 59’un 30 iterasyona sahip yaklaşımındaki karşılığı Model 20’dir ve iterasyon sayısını arttırmak sadece OKH sonucunu

iyileştirirken, OMYH ve OMH değerlerini yükseltmiştir. Kaynak Cab hattında ise başlangıç yaklaşımlarından Model 12 her üç hatada da en düşük değerlere sahip olmuştur. Geliştirilmiş modeller içinde OMYH değeri en düşük yaklaşım olan Model 46, Model 12’den farklı olarak tüm bağımsız değişkenlerin olduğu sayısal veri seti ile regresyon ağacı kullanmakta ve 500 iterasyon gerçekleştirmektedir. Model 46 ile Model 12 arasındaki OMYH farkı %3,2’dir. OKH ve OMH değerlerinde ise Model 50 en düşük geliştirilmiş model sonuçlarını vermiştir. Model 46’dan farklı olarak belirlenmiş veri seti kullanılmakta olup, OMYH değeri %15,56 olmuştur. Model 50’nin hata değerleri de Model 12’den yüksek gerçekleşmiştir. Kaynak Vbk hattında başlangıç modellerindeki en iyi performans Model 12’de gerçekleşmiştir. Model 49, hem OMYH hem de OMH değerlerinde hatayı düşürmüştür ($OKH_{Model\ 49} = 5.026$). OKH değerinde ise Model 57, en düşük hataya sahiptir. Kaynak Vbk hattında minimum yaprak sayısının tanımlandığı modeller fayda sağlamıştır. Kaynak Custom hattında ise belirlenmiş ikili veri setinde regresyon yöntemi en düşük hatalı değerleri vermiştir. Geliştirilmiş modellerde ise Model 57 en düşük hata değerlerine sahip olmuş, başlangıç modellerinde yer alan tüm ağaç modellerinden daha iyi sonuç vermiş fakat Model 4’ün performansını geliştirememiştir. Transit Montaj hattında ise Model 58, geliştirilmiş modellerde en iyi sonucu vermiştir. Bu model, tüm veriyi içeren regresyon yaklaşımına sahip Model 1 ile neredeyse aynı değerlere sahiptir. Başlangıç modellerinden olan Model 18 ile Model 58 aynı hata değerlerine sahiptir. Model 58 ile Model 18’in tek farkı iterasyon sayılarıdır. Bu durum en uygun modelin ilk 30 iterasyonda bulunduğunu göstermektedir. Montaj Custom hattında ise Model 47, en düşük hata değerlerini vermiştir. Burada başlangıç modellerinde en düşük hatalara sahip yaklaşım Model 45’dir. Başlangıç modellerinde Model 45’e en yakın yaklaşım Model 20 olarak görülmektedir ve neredeyse hata değerleri aynıdır. Model 20 ile Model 47’nin ortak yönü Bayes optimizasyonu kullanmaları ve sınıflandırma ağacı olmalarıdır. Tampon bölgeleri daha önce de belirtildiği gibi farklı yaklaşımlar gerektirmektedir. Fakat sonuç alınabildiği için gösterilmiştir. Kaynak Alt Kat Tampon bölgesinde OKH

Tablo 12. İyileştirilmiş modeller ve parametreleri (Improved models and parameters)

Model Ad	Veri Seti	SRA	Model Ad	Veri Seti	SRA
Model 46	Sayısal tüm veri	Regresyon Opt 500	Model 54	İkili tüm veri	Regresyon Opt 500
Model 47	Sayısal tüm veri	Sınıflandırma Opt 500	Model 55	İkili tüm veri	Sınıflandırma Opt 500
Model 48	Sayısal tüm veri	Regresyon Std. Min. Yaprak Sayısı	Model 56	İkili tüm veri	Regresyon Std. Min. Yaprak Sayısı
Model 49	Sayısal tüm veri	Sınıflandırma Std. Min. Yaprak Sayısı	Model 57	İkili tüm veri	Sınıflandırma Std. Min. Yaprak Sayısı
Model 50	Sayısal belirlenmiş veri	Regresyon Opt 500	Model 58	İkili belirlenmiş veri	Regresyon Opt 500
Model 51	Sayısal belirlenmiş veri	Sınıflandırma Opt 500	Model 59	İkili belirlenmiş veri	Sınıflandırma Opt 500
Model 52	Sayısal belirlenmiş veri	Regresyon Std. Min. Yaprak Sayısı	Model 60	İkili belirlenmiş veri	Regresyon Std. Min. Yaprak Sayısı
Model 53	Sayısal belirlenmiş veri	Sınıflandırma Std. Min. Yaprak Sayısı	Model 61	İkili belirlenmiş veri	Sınıflandırma Std. Min. Yaprak Sayısı

Tablo 13. Başlangıç ve geliştirilmiş modellerin karşılaştırılması (Comparison of the starter and improved models)

Hata Terimi		OMYH (%)		OKH (dakika ²)		OMH (dakika)	
Bölüm	Alt Hat	Geliştirilmiş Modeller	Başlangıç Modelleri	Geliştirilmiş Modeller	Başlangıç Modelleri	Geliştirilmiş Modeller	Başlangıç Modelleri
Boyahane	Eski Kabin	Model 47 %16,40	Model 8 %16,40	Model 59 12.974	Model 12 13.088	Model 47 75,12	Model 8 75,12
	Yeni Kabin	Model 55, 59 %26,07	Model 20 %26,07	Model 55, 59 15.633	Model 20 15.633	Model 55, 59 88,98	Model 20 88,98
	Cab	Model 46 %15,55	Model 12 %12,33	Model 50 2.749	Model 12 2.118	Model 50 43,38	Model 12 36,83
	Kaynak	Vbk %22,15	Model 12 %23,57	Model 57 4.736	Model 12 5.842	Model 49 55,98	Model 12 59,65
Montaj	Custom	Model 57 %41,12	Model 4 %32,99	Model 57 2.992	Model 4 2.118	Model 57 51,09	Model 4 40,63
	Transit	Model 58 %7,52	Model 1 %7,45	Model 58 1.121	Model 1 1.109	Model 58 24,11	Model 1 23,91
	Custom	Model 47 %7,39	Model 45 %8,15	Model 47 413	Model 45 518	Model 47 16,54	Model 45 18,33
Tampon	Kaynak	Model 49 %9352	Model 12 %7369	Model 56 131.361	Model 1 134.550	Model 49 283,85	Model 12 282,31
	Alt Kat	Model 57 %75,19	Model 11 %99,71	Model 53 584	Model 19,20 839	Model 53 17,51	Model 19,20 20,19
	Üst Kat	Model 58 %66,07	Model 18 %66,07	Model 50 5.291	Model 38 4.772	Model 50 50,96	Model 4 50,47
	Montaj						

değeri geliştirilmiş modelde daha düşükken, OMH değeri, başlangıç ve geliştirilmiş modellerde neredeyse aynı olmuştur. OMYH değerinde geliştirilmiş modeller içerisinde yer alan modelin yüksek başlangıç modeline göre çıkması düşük bekleme sürelerindeki hatanın daha fazla olduğunu göstermektedir.

Kaynak Üst Kat Tampon bölgesinde Model 53, her üç hata teriminde de başlangıç modellerinden daha iyi sonuç vermiştir. Montaj tampon bölgesinde OMYH değerinde Model 58 ile Model 18 aynı sonuçları vermiştir. İki model arasındaki fark iterasyon sayılarıdır ve optimum ağacın 30 iterasyonda belirlendiği tespit edilmiştir. OMH değerinde ise Model 50'nin başlangıç modellerine yakın sonuç verdiği görülmüştür.

Sonuç olarak 7 üretim alt hattı ve 3 alt tampon bölgesinin sonuçları incelendiğinde bir kaynak, bir montaj ve bir tampon alt hattında geliştirilmiş modellerin tüm hata terimlerinde daha iyi sonuçlar verdiği görülmüştür. OKH hata teriminde de geliştirilmiş modellerin bir boyahane ve bir tampon alt hattında sonuçları iyileştirdiği görülmüştür. Bir boyahane hattında ise geliştirilmiş ve başlangıç modellerinin aynı sonuç verdiği belirlenmiştir. OMYH değerinin, iki boyahane ve bir tampon bölgesinde geliştirilmiş modeller ile başlangıç modellerinin aynı sonuç verdiği görülürken, OMH değerinde iki boyahanenin aynı sonuç verdiği belirlenmiştir. İki kaynak alt hattı, bir montaj alt hattında başlangıç modellerinin tüm hata terimleri için daha iyi sonuç verdiği görülmüştür.

Elde edilen sonuçlar üretilen araçların özelliklerinin kullanılarak alt hat işlem sürelerinin tahmin edilebileceğini göstermiştir. Burada kullanılan eğitim ve test veri setleri birbiri ardına gelen zamanlarda üretilmiş olan taşıtlardan elde edilmiştir. Böylece çalışmanın gerçek fabrika ortamında da kullanılabilirliği sınanmıştır. Test veri setinde elde edilen

sonuçlar süre tahminlerinin kabul edilebilir sınırlarda olduğunu göstermiş, herhangi bir süre tahmininin yapılmadığı bir ortamda kullanılabilirliği belirlenmiştir. Yine burada hazırlanan modellerin o anki yakın geçmişteki verileri alınarak eğitilmesi sağlanarak, eğitim yapılan zaman diliminde üretimde bulunan veya üretimi başlayacak araçların alt hat bazlı süre tahminleri yapılabilecektir. İndirgeme işlemi otomatize edilerek sıralı görev haline getirildiğinde çok fazla zaman almasa da, eğitim gerektiren modellerin öğrenme sürelerinin uzun olması fabrika ortamında her gün eğitim yapılmasına izin vermeyecektir. Haftalık modellerin eğitimi de tahminlerin yapılması için yeterli olmaktadır. Eğitim yapılan zaman diliminde üretimde bulunan veya üretimi başlayacak araçların alt hat bazlı süre tahminleri toplam 5 dakikalık bir süre içerisinde gösterilebilmektedir. Bu tahmin sonuçları üretimi yapılan araçların ham madde ve yarı mamullerinin üretim ortamına verilmesi, temini ve sevkiyatının takibinde olumlu yönde katkı yapması sağlanabilir. Böylece depo, parça siparişi ve üretim planlama maliyetleri düşürülebilecektir.

6. SONUÇLAR (CONCLUSIONS)

Yapılan bu çalışmada üretilen motorlu taşıtların 43 adet birbirinden farklı özelliği kullanılarak 4 farklı veri seti yapısı üzerinden 3 farklı teknik ile 7 farklı alt hat, 3 farklı tampon bölgesinin süre tahmini gerçekleştirilmiştir. Üç teknik ve çeşitli parametre değişiklikleri ile 61 farklı model oluşturulmuş ve bu modeller kullanılarak araçların hatlardaki işlem süreleri tahmin edilmiştir. Yapılan tahminler sonucunda, ağaç yapısının 8 alt hat, regresyonun 2 alt hat için en iyi sonucu verdiği görülmüştür. Üretimin olmadığı, sadece bekleme alanı olarak kullanılan tampon alanlarından Kaynak Alt Kat harici alanların yapay zeka yaklaşımının kabul edilebilir sonuçlar verdiği görülmüştür. Kullanılan bu tekniklerin en önemli üstünlüğü, ilk kez

üretilecek araç için daha önce bilgi olmasa dahi tahmin sonucu üretebilmesidir. Diğer bir ifadeyle yeni bir araç seri üretime geçtiğinde içeride işlem süresi bilinmeyen 150 aracın sadece bağımsız değişkenleri kullanılarak proses süresi tahmini gerçekleştirilebilecektir. Çalışmanın sonraki aşamalarında modellerinin daha iyi sonuç vermesi için araç bilgileri dışında hatlarda araç üretilirken çalışan personel sayıları, araçların hangi vardiyalarda üretildikleri, hatların proses hızları, araçların üretildikleri günler (pazartesi, salı, vb.), hatlarda arızaların olup olmadığı, tatil günleri gibi bağımsız değişkenlerin eklenmesi faydalı olacaktır. Bunun dışında diğer bir öneri, araçların sipariş ve nihai teslim tarihlerinin de modele eklenmesidir. Böylece aslında öncelikli araçların daha hızlı hattan çıkmalarının etkisi de model üzerinde görülecektir. Yine çalışmadaki üç teknik dışında güncel tekniklerden evrimsel sinir ağları gibi yaklaşımlar ile süre tahmini gerçekleştirmek faydalı olabilir. Bu üç önerinin sonraki çalışmalarda gerçekleştirilmesi hedeflenmektedir.

TEŞEKKÜR (ACKNOWLEDGEMENT)

Bu çalışma Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (Proje No. 1170357) ve Ford Otosan A.Ş. tarafından desteklenmiştir.

KAYNAKLAR (REFERENCES)

1. T.C. Kalkınma Bakanlığı, Onbirinci Kalkınma Planı - Otomotiv Sanayii Çalışma Grubu Raporu (2019 - 2023), Ankara, 2018.
2. McKinsey&Company, RACE 2050 – a vision for the European automotive industry, 2019.
3. Turkish Automotive Manufacturer Association, General and Statistical Information Bulletin Of Automotive Manufacturers, 2020.
4. Manufacturers IO of MV, 2018 World Motor Vehicle Production Statistics. <http://www.oica.net/category/production-statistics/2018-statistics/> , 2019, Erişim tarihi: 05.08.2020.
5. Pavlínek, P., The Internationalization of Corporate R&D and the Automotive Industry R&D of East-Central Europe. *Economic Geography*, 88 (1), 279–310, 2012.
6. Badillo, E.R., Galera, F.L., Moreno Serrano, R., Cooperation in R&D, firm size and type of partnership. *European Journal of Management and Business Economics*, 26 (1), 123–143, 2017.
7. Chamsuk, W., Fongsuwan, W., Takala, J., The Effects of R&D and Innovation Capabilities on the Thai Automotive Industry Part's Competitive Advantage: A SEM Approach, *Management and Production Engineering Review*, 8 (1), 101–112, 2017.
8. Mordue, G., Karmally, D., Frontier Technologies in Non-Core Automotive Regions: Autonomous Vehicle R&D in Canada. *Canadian Public Policy*, 46 (1), 73–93, 2020.
9. Lazard, Roland Berger, Global Automotive Supplier Study, Munich, 2019.
10. Salveson, M.E., On a Quantitative Method in Production Planning and Scheduling, *Econometrica*, 20 (4), 554–590, 1952.
11. Dautère-Péres, S., Lasserre, J.B., An Integrated Approach in Production Planning and Scheduling, Springer-Verlag, Berlin, 1994.
12. Terekhov, D., Tran, T.T., Down, D.G., Beck, J.C., Integrating Queueing Theory and Scheduling for Dynamic Scheduling Problems, *Journal of Artificial Intelligence Research*, 50 (1), 535–572, 2014.
13. Bukkur, K.M.M.A., Shukri, M.I., Elmardi, O.M., A Review for Dynamic Scheduling in Manufacturing, *Global Journal of Researches in Engineering*, 18 (5), 25–37, 2018.
14. Vlk, M., Bartak, R., Replanning in Predictive-reactive Scheduling, *International Conference on Automated Planning and Scheduling*, Jarusalem, Israel, June 7-11, 2015.
15. MacDuffie, J.P., Sethuraman, K., Fisher, M.L., Product Variety and Manufacturing Performance: Evidence from the International Automotive Assembly Plant Study, *Manage Sci*, 42 (3), 350-369, 1996.
16. Scavarda, L.F., Schaffer, J., Scavarda, A.J., Reis, A.C., Schleich, H., Product variety: an auto industry analysis and a benchmarking study, *Benchmarking: An International Journal*, 16 (3), 387–400, 2009 .
17. Fysikopoulos, A., Anagnostakis, D., Salonitis, K., Chryssolouris, G., An Empirical Study of the Energy Consumption in Automotive Assembly, *Procedia CIRP*, 3 (1), 477–482, 2012.
18. Galitsky, C., Worrell, E., Energy Efficiency Improvement and Cost Saving Opportunities for the Vehicle Assembly Industry: An ENERGY STAR® Guide for Energy and Plant Managers, Ernest Orlando Lawrence Berkeley National Laboratory, 2008.
19. Ar, R., Al-Ashraf, M., Production Flow Analysis through Value Stream Mapping: A Lean Manufacturing Process Case Study, *Procedia Engineering*, 41 (1), 1727–1734, 2012.
20. Nilakantan, J.M., Huang, G.Q., Ponnambalam, S.G., An investigation on minimizing cycle time and total energy consumption in robotic assembly line systems, *Journal of Cleaner Production*, 90 (1), 311–325, 2015.
21. Maaouane, M., Zouggar, S., Krajačić, G., Zahboune, H., Modelling industry energy demand using multiple linear regression analysis based on consumed quantity of goods, *Energy*, 225 (c), 120270, 2021.
22. Liu, M., Hu, S., Ge, Y., Heuvelink, G.B.M., Ren, Z., Huang, X., Using multiple linear regression and random forests to identify spatial poverty determinants in rural China, *Spatial Statistics*, 42, 100461, 2021.
23. Ouma, Y., Okuku, C.O., Njau, E.N., Use of Artificial Neural Networks and Multiple Linear Regression Model for the Prediction of Dissolved Oxygen in Rivers: Case Study of Hydrographic Basin of River Nyando, Kenya, *Complexity*, 2020 (8), 1–23, 2020.
24. Çetin, O., Dalcalı, A., Temurtaş, F., Estimation of Permanent Magnet Synchronous Generator Performance with Artificial Neural Network Models,

- Sakarya University Journal of Computer and Information Sciences, 3 (1), 59–72, 2020.
25. Yerlikaya-Özkurt, F., Askan, A., Prediction of potential seismic damage using classification and regression trees: a case study on earthquake damage databases from Turkey, *Nat. Hazard.*, 103 (1), 3163–3180, 2020.
 26. Lama, Y., Hancock, G.R., Freimuth, V.S., Jamison, A.M., Quinn, S.C., Using classification and regression tree analysis to explore parental influenza vaccine decisions, *Vaccine*, 38 (5), 1032–1039, 2020.
 27. Zimmerman, R.K., Nowalk, M.P., Bear, T., Taber, R., Clarke, K.S., Sax, T.M., Eng, H., Clarke, L.G., Balasubramani, G.K., Proposed clinical indicators for efficient screening and testing for COVID-19 infection using Classification and Regression Trees (CART) analysis, *Human Vaccines & Immunotherapeutics*, 17 (4), 1109–1112, 2021.
 28. Akpınar, M., Ovaz Akpınar, K., Aktas, E., Senocak, O.F., Donmez, H., Cavdar, V., Sarioglu, O.S., Duration Estimation of Vehicle Based Production: Case Study of Assembly Line, *International Conference on Data Analytics for Business and Industry*, Sakhir, Bahrain, 1-4, 2020.
 29. Walpole, R.E., Myers, R.H., Myers, S.L., Ye, K., *Probability & Statistics for Engineers & Scientists*, Global Edition, Pearson Education Limited, 2017.
 30. Akpınar, M., Yumusak, N., Estimating household natural gas consumption with multiple regression: Effect of cycle, *International Conference on Electronics, Computer and Computation*, Ankara, 188–191, 2013.
 31. Ross, S.M., *Introduction to Probability and Statistics for Engineers and Scientists*, Academic Press, 2014.
 32. Adak, M.F., Akpınar, M., A Hybrid Artificial Bee Colony Algorithm using Multiple Linear Regression on Time Series Datasets, *International Conference on Smart and Sustainable Technologies*, Split, 1–5, 2018.
 33. Makridakis, S., Wheelwright, S.C., Hyndman, R.J., *Forecasting: methods and applications*, John Wiley, New York, 2008.
 34. Delurgio, S., Bhamé, C., *Forecasting Systems for Operations Management*, Irwin Professional Pub, New York, 1991.
 35. Yılmaz A., Prediction of base and subbase resilient modulus (Mr) using regression methodology, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 35 (1), 507-518, 2020.
 36. O'Rourke, N., Hatcher, L., Stepanski, E.J., *A Step-by-Step Approach to Using SAS for Univariate & Multivariate Statistics*, SAS Institute Inc., North Carolina, 2005.
 37. Rokach, L., Maimon, O., *Data Mining with Decision Trees: Theory and Applications*, World Scientific Publishing, MA, 2015.
 38. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., *Classification And Regression Trees*, Routledge, New York, 1984.
 39. Choubin, B., Zehtabian, G., Azareh, A., Rafiei-Sardooi, E., Sajedi-Hosseini, F., Kişi, Ö., Precipitation forecasting using classification and regression trees (CART) model: a comparative study of different approaches, *Environmental Earth Sciences*, 77 (1), 314, 2018.
 40. Timofeev, R., *Classification and Regression Trees (CART) Theory and Applications*, Humboldt University, 2004.
 41. Han, J., Kamber, M., Pei, J., *Data Mining: Concepts and Techniques*. Morgan Kaufmann, MA, 2012.
 42. Ahmad, A.S., Hassan, M.Y., Abdullah, M.P., Rahman, H.A., Hussin, F., Abdullah, H., Saidur, R., A review on applications of ANN and SVM for building electrical energy consumption forecasting, *Renewable and Sustainable Energy Reviews*, 33 (1), 102–109, 2014.
 43. Zhao, H., Magoulès, F., A review on the prediction of building energy consumption, *Renewable and Sustainable Energy Reviews*, 16 (6), 3586–3592, 2012.
 44. Reşat H.G., Design and development of hybrid forecasting model using artificial neural networks and ARIMA methods for sustainable energy management systems: A case study in tobacco industry, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 35 (3), 1129–1140, 2020.
 45. Bulut M., Başoğlu B., Development of a hybrid system based on neural networks and expert systems for shortterm electricity demand forecasting, *J Journal of the Faculty of Engineering and Architecture of Gazi University*, 32 (2), 575-583, 2017.
 46. Toraman S., Türkoğlu İ., A new method for classifying colon cancer patients and healthy people from FTIR signals using wavelet transform and machine learning techniques, *Journal of the Faculty of Engineering and Architecture of Gazi University.*, 35 (2), 933–942, 2020.
 47. Özcan E., Danişan T., Eren T., A hybrid model proposal for maintenance scheduling in hydropower plants, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 35 (4), 1815–1827, 2020.
 48. Schalkoff, R.J., *Artificial Neural Networks*, McGraw Hill, New York, 1997.
 49. Haykin, S., *Neural Networks: A Comprehensive Foundation*, Prentice Hall, New Jersey, 1998.
 50. Swingler, K., *Applying Neural Networks: A Practical Guide*, Academic Press, New York, 1996.
 51. SAS Institute Inc., *SAS/ETS® 13.2 User's Guide*. SAS Institute Inc., North Carolina, 2014.
 52. Salmeron, J., Ruiz-Celma, A., *Elliot and Symmetric Elliot Extreme Learning Machines for Gaussian Noisy Industrial Thermal Modelling*, *Energies*, 12 (1), 90, 2019.
 53. Elliott, D.L., *A Better Activation Function for Artificial Neural Networks*. Maryland, 1993.
 54. Akpınar M., Yumuşak N., Daily basis mid-term demand forecast of city natural gas using univariate statistical techniques, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 35 (2), 725–742, 2020.
 55. Özger, Y. E., Akpınar, M., Musayev, Z., Yaz, M., *Electrical Load Forecasting Using Genetic Algorithm Based Holt-Winters Exponential Smoothing Method*, *Sakarya University Journal of Computer and Information Sciences*, 2 (2), 108–123, 2019.

56. Lewis, C.D., Green, B., Sevenoaks, K., Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting, Butterworth-Heinemann, London, 1982.
57. Bulut M., Başoğlu B., Development of a Hybrid System Based on Neural Networks and Expert Systems for Short-Term Electricity Demand Forecasting, Journal of the Faculty of Engineering and Architecture of Gazi University, 32 (2), 575-583, 2017.
58. Acı M., Avcı M., Acı Ç., Reducing simulation duration of carbon nanotube using support vector regression method, Journal of the Faculty of Engineering and Architecture of Gazi University, 32 (3), 901-907, 2017.
59. Akpınar, M., Adak, M.F., Yumuşak, N., Time series forecasting using artificial bee colony based neural networks, International Conference on Computer Science and Engineering, Antalya, 554-558, 2017.