



Öznitelik-Odaklı Sensor Verisi Bazlı Uçak Motorları Geriye Kalan Faydalı Ömür Tahminleme

Kemal Çağlar Girgin^{1*}, Cemil Zalluhoğlu²

¹ Hacettepe Üniversitesi, Bilişim Enstitüsü, Ankara, Türkiye, k.caglar.girgin@gmail.com

² Hacettepe Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Ankara, Türkiye (ORCID: 0000-0001-8716-6297), cemilzalluhoglu@gmail.com

(5th International Symposium on Innovative Approaches in Smart Technologies– 28-29 May 2022)

(DOI: 10.31590/ejosat.1125433)

ATIF/REFERENCE: Girgin, K. Ç. & Zalluhoğlu, C. (2022). Öznitelik-Odaklı Sensor Verisi Bazlı Uçak Motorları Geriye Kalan Faydalı Ömür Tahminleme. *Avrupa Bilim ve Teknoloji Dergisi*, (37), 21-27.

Öz

Uçak motorunun durumu, uçağın güvenliğini, uçuş kalitesini ve çalışmasını doğrudan etkiler. Uçak motorları için belirti izleme faaliyetleri, motorun kalan faydalı ömrünü tahmin etmek için bir önceden önlem alınmasını sağlayabilecek bir avantaj yaratabilir. Uçak motoru yapıları hem soyut hem de somut bileşenlerle karmaşık olduğundan, motor faaliyet bozulma sürecini göstermek oldukça zahmetlidir. Bu yazıda, kalan faydalı ömür tahmini doğruluğunu iyileştirmek için öznitelik odaklı çerçeve geliştirilmiştir. Bu çerçeve, motorlardan gelen gereksiz duyuşal girdileri ortadan kaldırır ve hesaplama maliyetlerini düşürür. Bir uygulama örneği olarak, sensör verilerine dayalı olarak uçak motorunun kalan faydalı ömrünü tahmin etmek için geliştirilmiş öznitelik odaklı çerçeve kullanılmıştır. Sonuçlar, diğer yöntemleri uygulamadan önce, birçok girdi özelliğine sahip sistemlerin, maliyeti düşürmek için özellik uyarlama prosedürlerine ihtiyaç duyduğunu, ancak kalan faydalı ömrü tahmin etmek için kesinliği artırdığını göstermektedir.

Anahtar Kelimeler: Uçak Motor, Veri Madenciliği, Makine Öğrenmesi, Destek Vektör Makineleri, Lojistik Regresyon, Doğrusal Regresyon, NASA Turbofan Veri Seti, Öznitelik Seçme, PCA

Feature-Oriented Remaining Useful Life Prediction of Aircraft Engines Based on Sensor Data

Abstract

Aircraft engine's condition straightforwardly influences the security, unwavering quality, and operation of the aircraft. Prognostics and wellbeing administration for aircraft engines can give a advantage to estimate the remaining useful life of the engine and can enable to take precautionary actions in advanced. Be that as it may, aircraft engine frameworks are complex with both intangible and dubious components, it is troublesome to demonstrate the complex degradation process. In this article, the remaining useful life estimation is developed to improve feature -oriented framework. This frame eliminates unnecessary sensory inputs from engines and reduces calculation costs. As an application example, the developed feature -oriented frame has been used to estimate the remaining use of the aircraft engine based on sensor data. The results show that before applying other methods, systems with many input characteristics need feature adaptation procedures to reduce costs, but increase the certainty to estimate the remaining useful life.

Keywords: Aircraft engine, data mining, machine learning, support vector machines, logistics regression, linear regression, nasa turbofan data set, attribute selection, PCA

* Kemal Çağlar Girgin: Hacettepe Üniversitesi, Bilişim Enstitüsü, Ankara, k.caglar.girgin@gmail.com

1. Giriş

Modern havacılık teknolojisinin hızlı gelişimi, güvenilirlik ve kalite endeksleri için daha yüksek ön koşullara sahip, giderek daha karmaşık mimarilere sahip olan uçak sistemlerine yol açmaktadır. Bu kapsamda, uçağın sahip olduğu motor birçok faktöre göre genel güvenliği, işletimi ve bakımı doğrudan etkilemektedir.

Erken teşhis ve sağlık izleme yetenekleri, sadece uçak motorlarında değil, performans gözetimlerini sağlayacak birçok özelliğe sahip tüm motorlarda giderek daha önemli hale gelmektedir. Motorlardan arıza erken teşhis ve sağlık izleme bilgilerini elde etmek için, teknik elemanların motorlarla ilgili titreşim, yağ sıcaklığı, motor sıcaklığı, likidite seviyesi vb. gibi verilere ulaşması gerekmektedir. Bu nedenle, uçak motorunun kalan faydalı ömrünün önceden tahmin edilerek arızalanmaması veya ilgili önlemlerin önceden alınması için sadece önemli sensör çıkış verilerinin alınması büyük önem arz etmektedir.

Kalan faydalı ömür (RUL) tespiti için mevcut birçok yöntemden öznitelik odaklı olan yöntemler, başarılı tahminleme konusunda en umut verici olan yöntemler olarak karşımıza çıkmaktadır. Bir uçak motorunun RUL'ünü tahmin etmeye çalışırken önemli sayılabilecek birçok faktör bulunmaktadır. Model geliştirme için ilk önemli kural verilere nasıl yaklaşılabileceği ve değerlendirileceğinin belirlenmesidir. Modeller bu ifade etrafında geliştirileceğinden, neyin hedeflendiğini anlamak çok önemli olmaktadır.

Bu çalışmada seçili bir veri setine öznitelik odaklı bir yapı kurularak, farklı makine öğrenmesi algoritmaları kullanılarak modellerin eğitilmesiyle uçak motorunun geriye kalan faydalı ömür süresi tahminleme analizinin ne oranda başarılı bir şekilde yapabildiği araştırılmaktadır. Örnek çalışmaların bazılarında kullanılan bu yapı araştırılarak, bu çalışmada daha gelişmiş ve hızlı bir yapı kurulmaya çalışılmaktadır. Kurulan yapı, çok sayıda sensör bilgisinden çıkarılabilecek önemli verilerin toplamını en üst düzeye çıkarmayı, hem deneyime dayalı hem de veriye dayalı yaklaşımları entegre etmeyi amaçlamaktadır.

2. Materyal ve Metot

2.1. İlgili Çalışmalar

Uçak motor RUL tahminleme konusunda gerçekleştirilen çok sayıda çalışma bulunmaktadır. Bunlardan bazıları doğru tahminlemeye ulaşmak için öznitelik özelliklerine [1-3] odaklanırken, bazıları çalışan ve arızalı motorlar arasındaki geçiş durumlarına odaklanmaktadır [4-6]. Bu çalışmaların amacı, arızalar hakkında önceden uyarı sağlamak, sistem ömrünü uzatmak ve arızalı motorları önceden teşhis etmektir.

Uçak motorları kalan faydalı ömür üzerine literatürde yapılmış çalışmaların bazıları raporun bu kısmında listelenmektedir:

Xu vd. [1] çalışmasında, sistem durumu tahmin doğruluğunu geliştirmek için prognostik ve sağlık yönetimi odaklı bir entegre füzyon prognostik çerçevesi geliştirilmektedir. Bu çerçeve, izleme sensörü verilerini stratejik olarak birleştirmektedir ve ilgili sınırlamalarını azaltırken veriye dayalı prognostik yaklaşımın ve deneyime dayalı yaklaşımın güçlü yanlarını bütünleştirmektedir.

Javed vd. [4] çalışmasında, başarısızlık prognostiğini iyileştirmek için yeni bir yaklaşım önermektedir. Özetle, önerilen prognostik model, yıpranma davranışı tahmin etmek, yıpranan makinelerin durumlarını otomatik olarak belirlemek için bir Toplama Dalgacığı Aşırı Öğrenme Makinesi (SWELM)

ve Çıkarıcı-Maksimum Entropi Bulanık Kümeleme (S-MEFC) olmak üzere iki yeni algoritmayı bütünleştirmektedir.

Wang vd. [7] çalışmasında, Kalan Faydalı Ömür(RUL) tahmin etmek için benzerliğe dayalı bir yaklaşım sunulmaktadır. Yaklaşımında, özellikle tasarlanmış bir sistem için bol miktarda çalışmadan-arızaya verinin mevcut olduğu durumlar için uygunluğu belirtilmiştir. Aynı sistemin birden çok biriminden gelen veriler, çalışmanın bozulması şeklinde oluşturulan kalıpların kitaplığını oluşturmak için kullanılmaktadır. Bir test biriminin RUL'u tahmin edilirken, test biriminden gelen veriler kitaplıktaki bu kalıplarla eşleştirilmekte ve bu eşleşen birimlerin kalan gerçek ömrü tahminin temeli olarak kullanılabilir. Sonuçlar, benzerlik temelli yaklaşımın RUL tahminini gerçekleştirmede çok etkili olduğunu göstermektedir.

Liu vd. [8] çalışmasında, sensör teknolojisinin hızlı gelişimi ve bir mühendislik sisteminin durumunu izlemek için birden fazla sensörün kullanılmasına değinerek kalan kullanım ömrünü tahmin etmedeki doğruluğunu geliştirmek amacıyla birden fazla sensörden gelen veriyi entegre edebilen metodolojiler geliştirmenin önemini vurgulamaktadır. Bu çalışmada, çoklu bozulma tabanlı sensör verilerinin birleştirilmesi yoluyla sistemin performansını karakterize etmek için birleşik bir sağlık endeksi oluşturmak için bir metodoloji sunulmaktadır. Bu metodoloji, gelişmiş bir bozulmaya dayalı modele yol

açan veri seçimi, veri işleme ve veri birleştirme adımlarını içermektedir.

Moghaddass vd. [9] çalışmasında, etkin prognostik ve sistem durumu yönetimi (PHM) çerçevesi aracılığıyla arızaların ve arızaların zamanında tespitinin, yıkıcı arızaları önlemek ve sistemlerin genel bakım maliyetini en aza indirmek için proaktif olarak programlanabilecek uygun bakım eylemlerine yol açabileceğine değinilmektedir. Çalışmada, gözlemlenemeyen bozulma altında mekanik sistemlerin çevrimiçi teşhisi ve prognostiğinin pratik zorluklarına değinilmektedir. İlk olarak, çok durumlu bir bozulma yapısının unsurları gözden geçirilmektedir ve ardından bir model seçim çerçevesi tanıtılmaktadır. Çevrimiçi teşhis ve prognostik için kullanılabilir önemli dinamik performans ölçüleri tanıtılmaktadır.

Tamilselvan vd. [10] çalışmasında, derin inanç ağı (DBN) kullanan yeni bir "çok sensörlü sağlık tanı yöntemi" sunulmaktadır. DBN, hızlı çıkarım, daha zengin ve daha yüksek dereceli ağ yapılarını kodlama yeteneği gibi vaat ettiği avantajları nedeniyle son zamanlarda makine öğreniminde popüler bir yaklaşım haline gelmektedir. DBN, birden çok yığılmış kısıtlı Boltzmann makinesiyle hiyerarşik bir yapı kullanılmaktadır ve katmanlar ardışık öğrenme süreci boyunca çalışmaktadır. DBN tabanlı durum sınıflandırmasını kullanan önerilen çok sensörlü sağlık tanı metodolojisi birbirini takip eden üç aşamada yapılandırılabilir: ilki, sağlık durumlarının tanımlanması ve DBN eğitimi ve testi için duyuşal verilerin ön işlenmesi; ikincisi, önceden tanımlanmış sağlık durumlarının teşhisi için DBN tabanlı sınıflandırma modelleri geliştirmek; üçüncüsü, duyuşal veri setini test ederek DBN sınıflandırma modellerini doğrulanmasıdır.

Javed vd. [11] çalışmasında, öğrenme süresini sınırlandırırken ve rastgele başlatma prosedürünün etkisini azaltırken, doğruluk ve genelleme performansları sağlayan yeni bir bağlantı ağ yapısı olan Summation Wavelet Extreme Learning Machine (SW-ELM) önerilmektedir. SW-ELM, hızlı toplu öğrenme için Extreme Learning Machine (ELM) algoritmasını temel almaktadır, ancak gizli katman

düğümlelerinde ikili etkinleştirme işlevleri bulunmaktadır. Bu, doğrusal olmayan durumlarla verimli bir şekilde başa çıkmayı geliştirmektedir. Dalgacıkların (gizli düğümlerin) ve sinir ağı parametrelerinin (giriş-gizli katmanın) başlatma aşaması, veriler modele sunulmadan önce gerçekleştirilmektedir. Tüm önerme, zaman serisi uygulamasıyla ilgili üç konuda testler gerçekleştirilerek gösterilmektedir ve tartışılmaktadır.

2.2. Öznitelik Odaklı Yaklaşım Yapısı

Birçok veri uygulamasında, özniteliklerin önemi ve katkıları model geliştirmede çok önemli bir rol oynamaktadır. Makine öğreniminde, veri kümesinin karmaşıklığının üstesinden gelmek ve etkili modeller oluşturmak için en iyi öznitelikleri analiz etmek ve seçmek için çeşitli yöntem türleri bulunurken, derin öğrenme gibi daha gelişmiş uygulamalarda oluşturulan yapı, sağlanan veri kümesinde en uygun özellikleri bulmaktadır. Yine de derin öğrenme modelinin gereksiz girdilerle beslenmesi, maliyet sorunları yaratabilmektedir ve yüksek hesaplama durumundan dolayı gereksiz zaman harcanmasına neden olabilmektedir.

Herhangi yararlı bir bilgi içermeyen öznitelikler ile başa çıkmanın birden fazla metodu bulunmaktadır. Bu metodlardan biri genellikle öznitelik seçimi diğeri ise boyut azaltmadır. Öznitelik seçme metodolojisinde, makine öğreniminde değerli bilgiler içeren en yararlı öznitelikleri seçmeye yönelik en çok bilinen yöntemler korelasyon matrisleri, varyans, ki-kare ve bilgi kazancı yöntemleridir.

Korelasyon yöntemi, değişkenler arasındaki korelasyon katsayılarını gösteren bir tablodur. Tablodaki her hücre, iki değişken arasındaki korelasyonu göstermektedir. Verileri özetlemek için, daha gelişmiş bir analize girdi olarak ve gelişmiş analizler için bir teşhis olarak bir korelasyon matrisi kullanılmaktadır. Varyans, bir veri setindeki sayılar arasındaki yayılmanın istatistiksel bir ölçümünü ifade etmektedir. Daha spesifik olarak, varyans, kümedeki her sayının ortalamadan ve dolayısıyla kümedeki diğer her sayıdan ne kadar uzakta olduğunu ölçmektedir. Ki-kare, iki değişkenin bağımsız olup olmadığını incelemesinde kullanılmaktadır. Bilgi kazanımı, bir veri kümesini bir şekilde dönüştürmenin entropisindeki azalmayı hesaplamaktadır.

Boyut azaltma metodolojisinde, makine öğrenmesinde en çok bilinen yöntemler genel olarak PCA, SVD ve LASSO'dur. PCA, verilere dayalı hiyerarşik bir koordinat sistemi kullanarak veri kümelerindeki istatistiksel varyasyonları temsil etmeye olanak tanımaktadır. SVD, veriye dayalı bir Fourier dönüşümü genellemesidir. LASSO ise öznitelik seçimi için temsili bir yöntemdir.

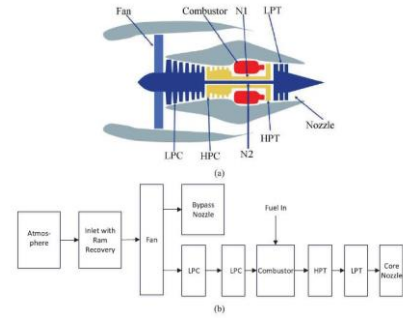
Bilgi içeren özniteliklerin analiz edilmesi ile daha hızlı sonuç ve maliyet azaltma imkânları karşımıza çıkmaktadır. Fakat her zaman görünürde bilgi içeren özniteliklerin kullanımı en optimal sonucu vermeyebilir. Örneğin 2 farklı öznitelğin fazla bilgi içermediği analiz sonucunda belli olduğunu varsayalım. Fakat bu 2 öznitelğin birleştirilmesi ile oluşturulan sentez öznitelği önemli bilgi içeriyor olabilir. Bu yüzden bu tip işlemlerin uygulanabilirliğinden de haberdar olmak gereklidir.

Yukarıda belirtilen yöntemler tekil olarak kullanıldığı gibi, entegre yöntemler olarak da kullanılabilir. Bu araştırmada belirtilen yöntemlerden seçilenler tekil olarak uygulanarak etkileri gözlemlenmeye çalışılmaktadır. Kurulan yapının değerlendirilmesi için ise aşağıdaki modeller kullanılmakta olup, skor olarak MAE ve R2 değerleri ele alınmaktadır:

- Doğrusal Regresyon,
- Optimize Edilmiş Doğrusal Regresyon,
- Doğrusal Support Vector Regression (SVR),
- Optimize Edilmiş Doğrusal Support Vector Regression (SVR),
- Optimize Edilmiş Poly Support Vector Regression (SVR),
- Optimize Edilmiş RGF Support Vector Regression (SVR).

2.3. Veri Seti Özellikleri

Makine öğrenme tekniklerini kullanarak uçak motor Kalan Faydalı Ömür tahmini yapılabilmesi için güvenilir kaynaktan edilen bir veri seti olması gerekmektedir. NASA'nın Turbofan motor bozulma veri seti, Ticari Modüler Hava Tahrik Sistemindeki model Simülasyon (C-MAPSS) test yatağı [12] simülasyonu tarafından birçok arıza ile sonuçlanan yörüngelerden oluşmaktadır. Farklı sorunlardan ve 4 alt veri kümesinden oluşan C-MAPSS veri kümesi, veriye dayalı prognostiklerin geliştirilmesi için iyi bir şekilde uyarlanmış olduğu kanıtlanmıştır. Her alt veri kümesi, kendi içlerinde belirli motor sayısı içeren eğitim ve test setinden oluşmaktadır. Her bir motor basınçlı kompressörler, türbinler, vb. gibi komponentlerden oluşmaktadır. Motor modelinin ana bileşenleri ve simülasyonda çeşitli alt rutinlerin nasıl birleştirildiğine ilişkin akış şeması Şekil 1'de sunulmaktadır.



Şekil 1 – (a)C-MAPSS'de simüle edilen motorun basitleştirilmiş diyagramı.(b)Simulasyondaki çeşitli modüller arasındaki bağlantıları gösteren diagram [13]

Dataset	FD001	FD002	FD003	FD004
Fault mode	1	1	2	2
Operational condition	1	6	1	6
Engines in train set	100	260	100	249
Engines in test set	100	259	100	248
Train minimum lifespan(cycles)	128	128	145	128
Train maximum lifespan(cycles)	362	378	525	543
Minimum test length(cycles)	31	21	38	19
Maximum test length(cycles)	303	367	475	486

Şekil 2 – C-MAPSS veri kümesinin temel bilgileri[13]

Her bir motorun bozulma davranışları, bir operasyonel döngü periyodu boyunca 21 sensör çıkışından örneklenen çok değişkenli zamansal veri noktalarına karşılık gelen yörüngeler sağlanarak temsil edilmektedir. Ayrıntılı olarak, her veri noktası, yalnızca sensör bileşenlerinin çalışmasını kaydeden 21 sensör sinyali değil, aynı zamanda çalışma ayarlarını, motor kimliklerini ve çalışma döngüsünü de içeren bir anlık görüntüdür. Şekil 2, hem tek operasyonel koşulu hem de birden fazla operasyonel koşulu içeren dört alt veri kümesinin bazı temel bilgilerini özetlemektedir. Her motor sağlıklı bir hal ile başlamaktadır ve sistem arızası vermeden önce belirli bir

zamanda motor arızası oluşmaya başlamaktadır. Sistemin çeşitli bölümlerinden toplanan veriler, bozulma zamanında basınç, sıcaklık, hız gibi sensör ölçümleri üzerindeki etkilerini kaydetmektedir. Daha fazla ayrıntı için okuyucular [14]'e başvurabilir. Sensör verileri doğrusal olmadığı, yüksek düzeyde gürültü içerdiği ve farklı geniş zaman aralıklarına dağıldığı için, RUL tahmini oldukça karmaşık ve zor olmaktadır.

Şekil-2'de belirtilen 4 farklı veri alt kümesinden, çalışma kondisyonu olarak daha az, üzerinde analiz gerçekleştirilmesi daha uygun ve veri setini daha rahat tanıma olanağı sunduğundan RUL analiz ve model çalışmaları için FD001 alt veri seti seçilmektedir. FD001 alt veri setine ait olan .csv uzantılı bu dosyalar aşağıda belirtilmektedir:

- train_FD001
- test_FD001
- RUL_FD001

Eğitim seti olarak train_FD001 ve test seti olarak test_FD001 kullanılmaktadır. Eğitim ve test setleri ile bilgi Şekil-2'de verilmektedir. Eğitim setinde RUL değeri rasgele bir değerden başlamaktadır ve 0'a kadar birer olarak azalmaktadır. Değer 0'a ulaşınca ilgili motorun arızaya girdiğini göstermektedir. Test setinde ise RUL değeri 0'a kadar gitmeyip, belirli rasgele bir değere kadar gitmektedir. Amaç ise test setindeki RUL'un kaç adım sonra arızaya geçeceğini, yani 0'a geleceğini bulmaktır. Bu kesin değer ise RUL_FD001 setinde bulunması gereken değer şeklinde sunulmaktadır.

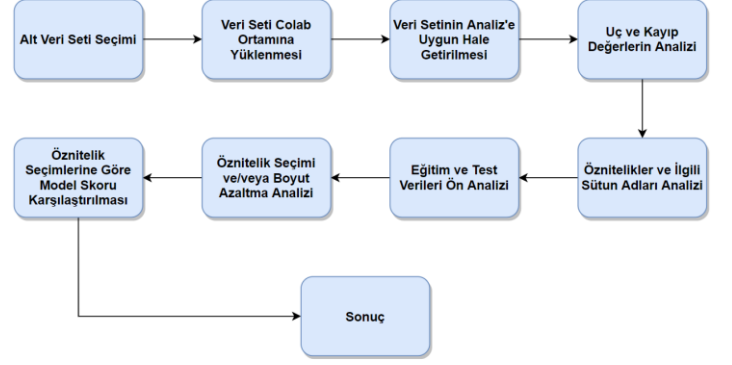
Veri kümesine NASA sitesinden (NASA n.d.) [15] ulaşılabilir.

2.4. Yaklaşım ve Uygulama

NASA Turbofan Motor Veri Setini analiz edebilmek ve RUL tahminlemesi yapabilmek için Python programlama dili kullanılarak proje hazırlanmaktadır. Bu programın yazımı için Google Colab ortamı Python 3.7.13 versiyonu kullanılmış olup, alt bölümlerde bahsedilecek Python kütüphaneleri de gerektiğince kullanılmaktadır.

Projenin özet niteliğini taşıyan aşamaları aşağıda belirtilmektedir:

- İlk adım hangi alt veri setinin kullanılacağını seçmek olmaktadır. Belirtildiği üzere sadelik ve anlaşılabilirlik adına FD001 alt veri seti kullanılmaktadır.
 - Alt veri kümesi analiz ortamına yüklenmektedir.
 - Veri kümesi veri analizine uygun yapıda değilse uygun yapıya getirilmektedir.
 - Alt veri kümesi uç değerler ve kayıp değerler için analiz edilmektedir.
 - Öznitelik sütunlarda sütun adları kontrol edilmektedir. Yoksa eklenmektedir.
 - Eğitim ve test verileri uygun formata getirilmektedir. Bunun için ulaşılmaya çalışılan değerler göz önünde bulundurulacaktır.
 - Öznitelik seçimleri ve boyut azaltma metodları uygulanmaktadır.
 - Daha sonra veri setindeki eğitim verileri SciKit-Learn kütüphanesindeki makine öğrenmesi algoritmaları titizlikle uygulanıp, farklı parametrelerle iterasyonlara sokularak modellerin sonuçları elde edilmektedir.
 - Seçilen algoritmalar test verileri ile karşılaştırılarak, mutlak sapma(MAE) ve model uygunluğu(R2) parametrelerine bakılarak en uygun öznitelik-boyut indirgeme yöntemi ile model belirlenmektedir.
- Bu aşamalar Şekil 3'te görülmektedir.



Şekil 3 – Proje Aşamaları

2.4.1. Python Geliştirme Ortamının Belirlenmesi ve Hazırlanması

Analiz için gerekli kodlar Google COLAB platformu üzerinde yazılmıştır. COLAB üzerinde mevcut Python sürümü 3.7.13'tür.

2.4.2. Veri Setinin Hazırlanması

Geliştirme ortamı hazırlandıktan sonra, alt veri seti seçilerek COLAB ortamına .csv uzantılı dosyalar olarak eklenmektedir. FD001 alt veri setine ait olan .csv uzantılı bu dosyalar aşağıda belirtilmektedir:

- train_FD001
- test_FD001
- RUL_FD001

Dosyalar COLAB ortamına yüklendikten sonra veri yapısı şekline ilgili kütüphaneler ile getirilmektedir. Bu işlem sonrasında özniteliklere ilişkin belirleyici sütun başlıkları olmadığından dolayı sütun adlandırma işlemi yapılmaktadır.

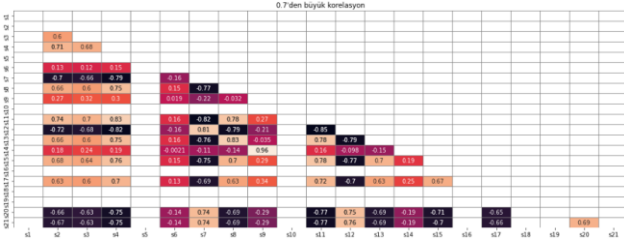
2.4.3. Veri Ön Analizi Gerçekleştirilmesi

Öznitelik sütun tanımlamasından sonra değeri olmayan satır ve sütunlar analiz edilmektedir. CSV formatındaki verinin son sütununun boş veri içerdiği tespit edilmektedir ve son sütun kaldırılmaktadır.

Ardından eğitim seti aşama(cycle) sayısı'nın en büyük değeri, başlangıçtan başlayacak şekilde ve 0'a inene kadar azalacak şekilde her bir motor numarası için yeniden düzenlenmektedir ve RUL sütunu oluşturularak buraya eklenmektedir. Test verisinde gerçek sonuç bilgisi RUL_FD001 veri dosyasında olduğundan, buradaki hedef sonuç, motor numarası ile ilişkilendirilerek RUL sütunu oluşturularak buraya eklenmektedir.

Eğitim veri setinin aynı zamanda karactersitikliklerini görmek adına *describe()* fonksiyonu çağırılmıştır. *Describe* fonksiyonu veri setinde seçilen her bir öznitelik ortalamasını, standart sapmasını, yüzdelik dağılım ağırlığını ve öznitelik veri setindeki minimum ve maksimum değerlerini gösteren açıklayıcı bir fonksiyondur.

Yukarıdaki görsellerden görüldüğü üzere net bir şekilde *Setting3*, *S1*, *S5*, *S10* ve *S16*'da veriler değişmemekte ve standart sapma(std)'leri çok düşük gözükmektedir. Bunun yanında özniteliklerin birbirleriyle olan korelasyonuna bakmak uygun gözükmektedir. Şekil 5'de 0.7'den büyük korelasyonlar gösterilmektedir.



Şekil 5–0.7’den büyük korelasyonlar

Yukarıdaki şekilden görüldüğü üzere 0.8’den büyük olarak yansyan korelasyonlar (s14-s9), (s11- s4), (s11-s7), (s11-s12), (s4-s12), (s8-s13) ve (s7-s12)’dir. Bu bilgi öznelik seçiminde kullanılmaktadır.

2.4.4. Makine Öğrenmesi Regresyon Algoritmaları ve Öznelik İşlemleri

Sağlanan veri dosyalarında eğitim ve test veri setleri hazır olduğundan herhangi bir ayırma yapılmamaktadır. Belirlenen öznelikler ilgili verilerden çıkartılıp veya boyut indirgeme yapıp, farklı modellere sokulmaktadır. Algoritmalar bazda doğrusal tahminleyici algoritmalar seçilmektedir. Bundaki temel neden verinin sürekli bir yapıya sahip olması ve net bir değer bulma amacından kaynaklanmaktadır. Kullanılan modellemelerde aşağıdaki algoritmalar tabanlı modeller kullanılmaktadır:

- Doğrusal Regresyon,
- Optimize Edilmiş Doğrusal Regresyon,
- Doğrusal Support Vector Regression (SVR),
- Optimize Edilmiş Doğrusal Support Vector Regression (SVR),
- Optimize Edilmiş Poly Support Vector Regression (SVR),
- Optimize Edilmiş RGF Support Vector Regression (SVR).

Doğrusal Regresyon

Doğrusal regresyon analizi, bir değişkenin değerini başka bir değişkenin değerine dayalı olarak tahmin etmek için kullanılmaktadır. Tahmin etmek istenilen değişkene bağımlı değişken denirken diğer değişkenin değerini tahmin etmek için kullanılan değişkene bağımsız değişken denmektedir. Bu analiz biçimi, bağımlı değişkenin değerini en iyi tahmin eden bir veya daha fazla bağımsız değişkeni içeren doğrusal denklemin katsayılarını tahmin etmektedir. Doğrusal regresyon, tahmin edilen ve gerçek çıktı değerleri arasındaki tutarsızlıkları en aza indiren düz bir çizgiye veya yüzeye uymaktadır. Bir dizi eşleştirilmiş veri için en uygun doğruyu bulmak için "en küçük kareler" yöntemini kullanan basit doğrusal regresyon hesaplayıcıları bulunmaktadır.

SVR

Destek Vektör Makinesi, algoritmayı karakterize eden tüm ana özellikleri (maksimum marj) koruyarak bir regresyon yöntemi olarak kullanılabilir. Destek Vektör Regresyonu (SVR), sınıflandırma için SVM ile aynı prensipleri kullanmaktadır, sadece birkaç küçük farklılık bulunmaktadır. Her şeyden önce, çıktı gerçek bir sayı olduğundan, sonsuz olasılıklara sahip olan eldeki bilgiyi tahmin etmek çok zor hale gelmektedir. Gerileme durumunda, problemde zaten talep etmiş

olacak olan SVM'ye yaklaşık olarak bir tolerans marjı (epsilon) belirlenmektedir. Ana fikir her zaman aynı olmaktadır: hatayı en aza indirmek, marjı en üst düzeye çıkaran hiperdüzlemi bireyselleştirmek, hatanın bir kısmının tolere edildiğini akılda tutmaktır. SVR algoritması, doğrusal(doğrusal), polinomiyal çekirdek(poly)ve radyal tabanlı fonksiyon çekirdeği(rbf) yöntemlerini uygulayabilmektedir.

Yukarıda belirtilen algoritmalar temel alınarak özneliklerin 10 farklı seçim oluşturulmaktadır. Oluşturulan 9 seçim, bölüm 5.3 öznelik analizleri ve bazı yayınlar temel olarak oluşturulmuş olup sırası ile belirtilen 6 farklı algoritmaya sokulmuş olup sonuçları karşılaştırılmaktadır. Geriye kalan 1 seçim ise tüm veri setine uygulanan boyut indirgeme yöntemi olup, PCA metodu kullanılmaktadır. Özneliklerin metotlara göre seçimi Şekil 6’da gösterilmektedir.

ID	Metot-1	Metot-2	Metot-3	Metot-4	Metot-5	Metot-6	Metot-7	Metot-8	Metot-9	Metot-10
	Feature	Feature	Feature	PCA	Feature	Feature	Feature	Feature	Feature	Feature
Cycle	X	X	X	X	X	X	X	X	X	X
Setting-1	X	X	X	X	X	X	X	X	X	X
Setting-2	X	X	X	X	X	X	X	X	X	X
Setting-3	X	X	X	X	X	X	X	X	X	X
S1										
S2	X	X	X	X	X	X	X	X	X	X
S3	X	X	X	X	X	X	X	X	X	X
S4	X	X	X	X	X	X	X	X	X	X
S5										
S6										
S7	X	X	X	X	X	X	X	X	X	X
S8	X	X	X	X	X	X	X	X	X	X
S9	X	X	X	X	X	X	X	X	X	X
S10										
S11	X	X	X	X	X	X	X	X	X	X
S12	X	X	X	X	X	X	X	X	X	X
S13										
S14	X	X	X	X	X	X	X	X	X	X
S15	X	X	X	X	X	X	X	X	X	X
S16										
S17	X	X	X	X	X	X	X	X	X	X
S18										
S19										
S20	X	X	X	X	X	X	X	X	X	X
S21	X	X	X	X	X	X	X	X	X	X

Şekil 6– Değerlendirecek öznelik temelli metotlar (X’ler seçilen öznelikleri göstermektedir)

Yukarıda belirtilen metotlardaki seçimi etkileyen faktörler aşağıda belirtilmektedir:

- **Metot-1:** Tercihen seçilen sadece değişmeyen-önemsiz veri içeren özneliklerin çıkartıldığı metot,
- **Metot-2:** Metot-1’in farklı bir versiyonu olarak düzenlenen metot,
- **Metot-3:** “PHM-Oriented Integrated Fusion Prognostics for Aircraft Engines Based on Sensor Data” araştırmasının temel aldığı öznelikleri içeren metot [1],
- **Metot-4:** Tüm veriler ile yapılan boyut indirgeme (PCA) içeren metot,
- **Metot-5:** “Making use of partial knowledge about hidden states in HMMs: an approach based on belief functions.” araştırmasının temel aldığı öznelikleri içeren metot [2] (Değerlendirme kriteri bulunmamaktadır),
- **Metot-6:** Metot-5’in Setting-2 ve Setting-3’ün dâhil olmadığı olarak düzenlenen metot,
- **Metot-7:** “A Similarity-Based Prognostics Approach for Remaining Useful Life Estimation of Engineered Systems” araştırmasının temel aldığı öznelikleri içeren metot [7] (Değerlendirme kriteri bulunmamaktadır),
- **Metot-8:** Metot-7’nin Setting-2 ve Setting-3’ün dâhil olmadığı olarak düzenlenen metot,
- **Metot-9:** Öznelik analizleri sonucu kişisel öznelik seçimlerini içeren metot,
- **Metot-10:** Metot-9’un Setting-2 ve Setting-3’ün dâhil olmadığı olarak düzenlenen metot.

Veri setlerinin farklı formasyonlarda analizi için sete model analizi sırasında bir takım işlemler uygulanmaktadır. Bunlardan

biri verinin dağılımının homojen olmasını sağlayan *Minmax* ölçeklendirmesidir. *Minmax* ölçeklendirmesinde değerler 0 ile 1 arasına, aralarındaki ölçek bozulmadan dağıtılmaktadır. Bir diğer işlem ise RUL değerlerinin kırpma işlemi olmaktadır. Kırpma işleminde RUL değerlerinin dağılımına bakıldığında değerlerin çoğunun belirli bir değerden fazla olmayacağı benimsenmektedir. Bu nedenden dolayı üst değer olarak döngü değeri 125 olarak alınmaktadır ve kırpma işlemi olması durumunda bu değer üstünde belirtilen değerler 125 değerine eşitlenerek işlem yapılmaktadır [16].

Metotlar algoritmalara 3 farklı şekilde sokulmaktadır:

- Ölçeklendirme olmadan, kırpma olmadan
- Ölçeklendirme olmadan, kırpma ile
- Ölçeklendirme olarak, kırpma ile

Sonuç Değerlendirme-Karşılaştırma Parametresi Seçimi

Eğitim ve test verilerine seçilen öznelik metotları uygulandıktan sonra, veri setleri algoritma modellerine sokularak algoritmanın ne kadar başarılı olduğunun ölçülmesi gerekmektedir. Veriler sürekli veri olduğundan ve tahmin etmeye çalışılan değer kümesi bir tam sayı kümesi olduğundan buna uygun sonuç karşılaştırma, yani skor parametresi seçimi olmaktadır. Bu projede uygun görülen sonuç karşılaştırma-değerlendirme parametreleri olarak Ortalama Mutlak Hata (MAE) ve Ortalama Kare Hatası (R2) olarak belirlenmiştir.

Ortalama Mutlak Hata (MAE)

Ortalama Mutlak Hata, regresyon modelleriyle birlikte kullanılan bir model değerlendirme metriğidir. Bir test setine göre bir modelin ortalama mutlak hatası, test setindeki tüm örnekler üzerindeki bireysel tahmin hatalarının mutlak değerlerinin ortalaması olmaktadır. Her tahmin hatası, örneğin gerçek değeri ile tahmin edilen değeri arasındaki farktır. Modelin ortalama mutlak hatası, test setindeki tüm örnekler üzerindeki bireysel tahmin hatalarının mutlak değerlerinin ortalamasıdır.

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{x}|$$

Şekil 7– MAE Formül [17]

R-Kare (R2)

R-kare (R2), bir bağımsız değişken veya bir regresyon modelindeki değişkenler tarafından açıklanan bir bağımlı değişken için varyansın oranını temsil eden istatistiksel bir ölçüdür. R-kare değerleri 0 ile 1 arasında değişmektedir ve genellikle %0 ile %100 arasındaki yüzdeler olarak belirtilmektedir. Dolayısıyla, bir modelin R2'si 0,50 ise, gözlemlenen varyasyonun yaklaşık yarısı modelin girdileriyle açıklanabilmektedir. %100'lük bir R-kare, bağımlı bir değişkenin tüm hareketlerinin, ilgilenilen bağımsız değişkenin hareketler tarafından tamamen açıklandığını belirtmektedir.

$$R^2 = 1 - \frac{\text{Unexplained Variation}}{\text{Total Variation}}$$

Şekil 8– R-Kare Formül [18]

3. Araştırma Sonuçları ve Tartışma

Öznelik odaklı sensör verisi bazlı uçak motorları geriye kalan faydalı ömür tahminleme projesi kapsamında; eğitim ve test verilerinin analiz ortamına aktarılması, eğitim ve test verilerinin analiz için uygun formata getirilmesi, verilerin eksik değer ve diğer ön analizlerinin yapılması, kullanılacak özneliklerin seçilmesi ve belirlenmesi ve oluşturulan farklı öznelik metotları üzerinde farklı algoritmaların etkinliğinin nasıl değiştiğinin gözlemeleme işlemi yapılmaktadır. Bu algoritmaların skorunun R-kare ve Ortalama Mutlak Hata ile ölçülerek veri setine en uygun model algoritması ve öznelik seçimi değerlendirilmiş olup, sonuç tablosu Şekil 9'de sunulmaktadır.

	Lineer Regr.		Lineer Regr.		Lineer SVR		Lineer SVR		Poly SVR		RBF SVR	
	Ölç.	Kırp.	Ölç.	Kırp.	Ölç.	Kırp.	Ölç.	Kırp.	Ölç.	Kırp.	Ölç.	Kırp.
Metot-1 Eğitim	30,99	0,65	16,27	0,77	19,72	0,67	17,74	0,71	13,76	0,81	11,93	0,83
Metot-1 Test	25,46	0,45	17,61	0,70	20,62	0,66	19,92	0,66	16,01	0,76	14,48	0,77
Metot-2 Eğitim	31,20	0,64	16,92	0,74	NA	NA	16,96	0,74	14,94	0,79	12,88	0,80
Metot-2 Test	26,15	0,42	18,84	0,67	NA	NA	18,92	0,67	17,66	0,72	15,94	0,72
Metot-3 Eğitim	31,29	0,64	17,05	0,74	17,18	0,73	16,99	0,74	15,05	0,78	13,07	0,80
Metot-3 Test	26,58	0,40	19,25	0,67	19,90	0,65	19,96	0,65	17,98	0,71	16,64	0,72
Metot-4 Eğitim	32,45	0,62	17,71	0,71	18,69	0,68	17,64	0,71	18,28	0,68	14,17	0,77
Metot-4 Test	29,12	0,27	22,00	0,56	23,16	0,52	22,44	0,55	23,93	0,47	20,02	0,61
Metot-5 Eğitim	32,46	0,62	17,71	0,71	20,85	0,60	17,64	0,71	19,71	0,64	14,18	0,77
Metot-5 Test	29,24	0,27	22,00	0,56	23,09	0,50	22,44	0,55	24,18	0,77	20,01	0,61
Metot-6 Eğitim	33,04	0,78	17,04	0,74	17,07	0,74	16,97	0,74	15,02	0,78	13,05	0,80
Metot-6 Test	18,13	0,71	19,07	0,67	19,29	0,67	19,16	0,66	18,18	0,71	16,36	0,72
Metot-8 Eğitim	31,27	0,64	17,05	0,74	17,05	0,74	16,98	0,74	15,04	0,78	13,08	0,80
Metot-8 Test	26,64	0,41	19,16	0,67	19,27	0,67	19,14	0,66	18,13	0,71	16,35	0,72
Metot-9 Eğitim	31,08	0,65	16,30	0,76	18,75	0,76	16,46	0,76	13,94	0,81	12,10	0,83
Metot-9 Test	26,00	0,42	18,20	0,68	18,20	0,68	18,59	0,67	15,98	0,74	14,95	0,74
Metot-10 Eğitim	31,08	0,65	16,31	0,76	21,44	0,56	16,26	0,76	13,95	0,81	12,13	0,83
Metot-10 Test	25,88	0,42	18,17	0,68	21,17	0,41	18,56	0,67	15,85	0,75	14,95	0,75
Tüm Öznelikler												
PCA, Korelasyon>0,7			PCA, Korelasyon>0,8			PCA, Korelasyon>0,9			PCA, Korelasyon>0,95			
Öznelik Dahil			Öznelik Dahil			Öznelik Dahil			Öznelik Dahil			
En İyi Sonuç Poly SVR			En İyi Sonuç Poly SVR			En İyi Sonuç RBF SVR			En İyi Sonuç RBF SVR			
Metot-4 Eğitim	12,62	0,81	12,54	0,81	11,87	0,83	11,79	0,83				
Metot-4 Test	18,80	0,65	15,10	0,74	14,41	0,76	14,05	0,77				

Şekil 9– Metot-Model Algoritma Sonuç Karşılaştırması (Ölçeklendirme: Ölç. ; Kırpma: Kırp. ; X: Mevcut)

Şekil 9 sonuç tablosuna bakıldığında, gerçek sonuç veri kümesine en az hata ile yakınlık gösteren ilk 5 metot ve algoritma: Metot-1/RBF SVR, Metot-4/PCA>0,9, Metot-4/PCA>0,95, Metot-10/RBF SVR olarak görülmektedir. Burada ayrıca Metot-3 ile elde edilen sonuç olan MAE=17,98 ile referans verilen çalışmada elde edilen DSR, SVM ve RNN algoritmalarının performanslarından (sırası ile 26,19 / 31,14 / 29,00) daha iyi sonuç verdiği görülmektedir. Ayrıca yine Metot-3'teki referans araştırma ile karşılaştırdığında, PCA bazlı boyut indirgeme işleminin RBF SVR ile elde edilen sonuç olan MAE=14,05'in, referansta verilen Comentropy-based fusion algoritmasından elde edilen sonuç olan MAE=14,1997'den daha yüksek olduğu gözlemlenmiştir. Burada öznelikler aynı olmasına rağmen, uygulanan kırpma ve ölçeklendirme işleminin elde edilen performansa katkısının yüksek olduğu çıkarılabilmektedir.

4. Sonuç

Bu çalışma kapsamında NASA Turbofan veri seti ele alınarak öznelik odaklı makine öğrenmesi modelleri ile RUL tahminleme çalışması sunulmaktadır. RUL tahminlemesinin yapılmasının nedeni hataya açık olan uçak motorlarının hatanın daa fazla ilerleyip, motorun çalışmasına engel olacak aşamaya gelmeden önce önleyici aksiyonlar alınmasını sağlayabilmektir.

Çalışma kapsamında öznelik seçimi ve boyut indirgeme işlemleri gerçekleştirilerek, takibinde farklı makine öğrenmesi modelleri eğitilmiş olup sonuçları karşılaştırılmıştır. Karşılaştırma sonucunda öznelik seçiminin ve boyut indirgemenin önemi bir kez daha anlaşılmış olup, RBF SVR modellerinin genel anlamda bu veri setinde sunulan diğer modellere nazaran daha iyi performans gösterdiği gözlemlenmiştir.

İlerleyen çalışmalarda öznelik seçimi ve boyut indirgeme uygulanarak veriler derin öğrenme modellerine sokulacak ve

derin öğrenme modellerinin uygun eğitimi parametrelerini bulması beklencektir.

5. Teşekkür

Çalışma kapsamında önce kendime sonra her zorlukta yanımda olan ailem Mustafa Girgin, Candan Girgin ve İremso Girgin'e ve son olarak çalışmaya olan tüm yardımlarından dolayı Cemil Zalluhoğlu'na teşekkürlerimi sunarım.

Kaynakça

- [1] J. Xu, Y. Wang and L. Xu. (April 2014). *PHM-Oriented Integrated Fusion Prognostics for Aircraft Engines Based on Sensor Data*. (2014 IEEE Sensors Journal, vol. 14, no. 4, pp. 1124-1132). <https://doi.org/10.1109/JSEN.2013.2293517>.
- [2] E. Ramasso and T. Denoeux. (April 2014). *Making Use of Partial Knowledge About Hidden States in HMMs: An Approach Based on Belief Functions*. (2014 IEEE Transactions on Fuzzy Systems, vol. 22, no. 2, pp. 395-405). <https://doi.org/10.1109/TFUZZ.2013.2259496>.
- [3] P. Tamilselvan, Y. Wang and P. Wang. (2012). *Deep Belief Network based state classification for structural health diagnosis*. (2012 IEEE Aerospace Conference, pp. 1-11). <https://doi.org/10.1109/AERO.2012.6187366>.
- [4] K. Javed, R. Gouriveau and N. Zerhouni. (2013). *Novel failure prognostics approach with dynamic thresholds for machine degradation*. (IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Societ, pp. 4404-4409). <https://doi.org/10.1109/IECON.2013.6699844>.
- [5] Chao Hu, B. D. Youn and Pingfeng Wang. (2011). *Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life*. (2011 IEEE Conference on Prognostics and Health Management, pp. 1-10). <https://doi.org/10.1109/ICPHM.2011.6024361>.
- [6] Gouriveau, Rafael et al. (2013). *Strategies to Face Imbalanced and Unlabelled Data in Phm Applications*. Chemical engineering transactions 33: 115-120.
- [7] T. Wang, Jianbo Yu, D. Siegel and J. Lee. (2008). *A similarity-based prognostics approach for Remaining Useful Life estimation of engineered systems*". (2008 International Conference on Prognostics and Health Management, pp. 1-6). <https://doi.org/10.1109/PHM.2008.4711421>.
- [8] K. Liu, N. Z. Gebraeel and J. Shi. (July 2013). *A Data-Level Fusion Model for Developing Composite Health Indices for Degradation Modeling and Prognostic Analysis*. (2013 IEEE Transactions on Automation Science and Engineering, vol. 10, no. 3, pp. 652-664). <https://doi.org/10.1109/TASE.2013.2250282>.
- [9] Ramin Moghaddass, Ming J. Zuo. (2014). *An integrated framework for online diagnostic and prognostic health monitoring using a multistate deterioration process*, *Reliability Engineering & System Safety*. (Volume 124, Pages 92-104, ISSN 0951-8320). <https://doi.org/10.1016/j.ress.2013.11.006>.
- [10] Tamilselvan, P & Wang. (2013). *Failure diagnosis using deep belief learning based health state classification*. (P 2013, Reliability Engineering and System Safety, vol. 115, pp. 124-135). <https://doi.org/10.1016/j.ress.2013.02.022>
- [11] Javed, Kamran & Gouriveau, Rafael & Zerhouni, Nouredine. (2013). *SW-ELM : A summation wavelet extreme learning machine algorithm with a priori initialization*. (2014, Neurocomputing, 123). <https://doi.org/10.1016/j.neucom.2013.07.021>.
- [12] D. K. Frederick, J. A. Decastro, and J. S. Litt. (2007). *Users guide for the commercial modular aero-propulsion system simulation (c-mapss)*". (Tech. Rep. NASA/TM2007-215026).
- [13] C. Liu, L. Zhang, Y. Liao, C. Wu and G. Peng. (2019). *Multiple Sensors Based Prognostics With Prediction Interval Optimization via Echo State Gaussian Process*. (2019 IEEE Access, vol. 7, pp. 112397-112409). <https://doi.org/10.1109/ACCESS.2019.2925634>.
- [14] A. Saxena, G. Kai, D. Simon, and N. Eklund. (Oct. 2008). *Damage propagation modeling for aircraft engine run-to-failure simulation*. (Proc. Int. Conf. Prognostics Health Manage, pp. 1-9).
- [15] National Aeronautics and Space Administration. (Mayıs,2022). PCoE Datasets. <https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#turbofan>
- [16] Fitzgerald E. (Ağustos, 2021) <https://4cda.com/intro-to-predictive-maintenance-on-nasa-turbofan-engine-dataset-using-machine-learning/> .(Mayıs, 2022)
- [17] Stephanie. (Ekim, 2016). <https://www.statisticshowto.com/absolute-error/> .(Mayıs,2022)
- [18] Fernando J. (Eylül, 2021). <https://www.investopedia.com/terms/r/r-squared.asp> .(Mayıs, 2022)