

**BAŞKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**BİLGİSAYAR MÜHENDİLİĞİ ANABİLİM DALI
BİLGİSAYAR MÜHENDİSLİĞİ TEZLİ YÜKSEK LİSANS
PROGRAMI**

**HAVACILIK SANAYİNDE KULLANILAN TAKIM VE
APARATLARIN TASARIM SÜRELERİNİN MAKİNE ÖĞRENMESİ
YÖNTEMLERİ İLE KESTİRİLMESİ**

HAZIRLAYAN

NAZMİ UMUT AKTAN

YÜKSEK LİSANS TEZİ

ANKARA - 2020

**BAŞKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**BİLGİSAYAR MÜHENDİLİĞİ ANABİLİM DALI
BİLGİSAYAR MÜHENDİSLİĞİ TEZLİ YÜKSEK LİSANS
PROGRAMI**

**HAVACILIK SANAYİNDE KULLANILAN TAKIM VE
APARATLARIN TASARIM SÜRELERİNİN MAKİNE ÖĞRENMESİ
YÖNTEMLERİ İLE KESTİRİLMESİ**

HAZIRLAYAN

NAZMİ UMUT AKTAN

YÜKSEK LİSANS TEZİ

TEZ DANIŞMANI

DR. ÖĞR. ÜYESİ MEHMET DİKMEN

ANKARA – 2020

BAŞKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Bilgisayar Mühendisliği Anabilim Dalı Makine Mühendisliği Tezli Yüksek Lisans Programı çerçevesinde Nazmi Umut AKTAN tarafından hazırlanan bu çalışma, aşağıdaki jüri tarafından Yüksek Lisans Tezi olarak kabul edilmiştir.

Tez Savunma Tarihi: 23 / 12 / 2020

Tez Adı: Havacılık Sanayinde Kullanılan Takım Ve Aparatların Tasarım Sürelerinin Makine Öğrenmesi Yöntemleri İle Kestirilmesi.

Tez Jüri Üyeleri (Unvanı, Adı - Soyadı, Kurumu)

İmza

Dr. Öğr. Üyesi Mehmet Dikmen, Başkent Üniversitesi

.....

Doç. Dr. Mustafa Sert, Başkent Üniversitesi

.....

Doç. Dr. Mehmet Serdar Güzel, Ankara Üniversitesi

.....

ONAY

Prof. Dr. Faruk ELALDI
Fen Bilimleri Enstitüsü Müdürü

Tarih: ... / ... /

BAŞKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZ ÇALIŞMASI ORJİNALLİK RAPORU

Tarih: 05 / 01 / 2021

Öğrencinin Adı, Soyadı: Nazmi Umut AKTAN

Öğrencinin Numarası: 21820153

Anabilim Dalı: Bilgisayar Mühendisliği

Programı: Bilgisayar Mühendisliği

Danışmanın Unvanı/Adı, Soyadı: Dr. Öğr. Üyesi Mehmet DİKMEN

Tez Başlığı: Havacılık Sanayinde Kullanılan Takım Ve Aparatların Tasarım Sürelerinin Makine Öğrenmesi Yöntemleri İle Kestirilmesi

Yukarıda başlığı belirtilen Yüksek Lisans/Doktora tez çalışmamın; Giriş, Ana Bölümler ve Sonuç Bölümünden oluşan, toplam 49 sayfalık kısmına ilişkin, 05 / 01 / 2021 tarihinde tez danışmanım tarafından Turnitin adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı % 1'dir. Uygulanan filtrelemeler:

1. Kaynakça hariç
2. Alıntılar hariç
3. Beş (5) kelimedenden daha az örtüşme içeren metin kısımları hariç

“Başkent Üniversitesi Enstitüleri Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Usul ve Esaslarını” inceledim ve bu uygulama esaslarında belirtilen azami benzerlik oranlarına tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Öğrenci İmzası:

ONAY

Tarih: 05 / 01 / 2021

Dr. Öğr. Üyesi Mehmet Dikmen

TEŐEKKÜR

Tez danıřmanlıđını yapan sayın hocam Dr. Öğr. Üyesi Mehmet DİKMEN'e, tüm çalışmalarım için verdiđi destek için ve çalışmakta olduđum şirketim TUSAŐ'a tez çalışmam sırasında sağladıđı imkânlardan dolayı teşekkür ederim.

ÖZET

Nazmi Umut AKTAN

**HAVACILIK SANAYİNDE KULLANILAN TAKIM VE APARATLARIN
TASARIM SÜRELERİNİN MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE
KESTİRİLMESİ**

Başkent Üniversitesi Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

2020

Özellikle büyük ölçekli tasarım projelerinde, tasarım eforunun (süresinin) doğru tahmin edilmesi, bir proje planı oluşturmada önemli bir faktördür. Sadece uzman görüşüne dayanan bu tahminler yanlış ya da yetersiz yapıldığında proje planına büyük zararlar verebilmektedir.

Belirtilen problem üzerine, havacılık endüstrisindeki takım tasarım eforunu tahmin edebilen bir makine öğrenmesi modeli geliştirilmesi amaçlanmıştır. Sayısal değerler içermeyen her bir takımın tasarım gereksinimi girdisi kurumsal veri tabanından uygun sorgular ile elde edilmiştir. İlgili verilerin sayısallaştırılabilmesi için girdi üzerinde çeşitli deneyler yapılarak en verimli kodlama yöntemi seçilmiştir. Sayısallaştırılan girdi verisinden, takım tasarım eforunu tahmin edebilmek için yaygın olarak kullanılan karar ağacı, Destek Vektör Makinesi, Doğrusal Regresyon ve Yapay Sinir Ağı makine öğrenmesi yöntemleri üzerinde deneyler yapılmıştır.

Bu çalışmada; takım tasarım eforunun en iyi şekilde kestirimi için ideal makine öğrenme modelini tespit etmek ve en uygun girdi ve parametre setini oluşturmak hedeflenmiştir.

ANAHTAR KELİMELER: takım tasarım, makine öğrenmesi, tasarım eforu kestirimi, havacılık ve savunma.

ABSTRACT

ESTIMATING DESIGN EFFORT OF TOOLS AND FIXTURES USED IN THE AVIATION INDUSTRY BY MACHINE LEARNING METHODS

AKTAN Nazmi Umut

Başkent University Institute of Science

Department of Computer Engineering

2020

Especially in large-scale design projects, the correct estimation of the design effort (time) is an important factor in creating a project plan. These estimates, based only on expert opinion, can cause great damage to the project plan when made incorrectly or insufficiently.

Based on the specified problem, it is aimed to develop a machine learning model that can predict the tool design effort in the aviation industry. The design requirement input of each tool that does not contain numerical values was obtained from the cooperate database with suitable data queries. In order to deflate the relevant data into numerical, the most efficient encoding method was selected by performing various experiments on the input data. Experiments have been carried out on commonly used decision tree, Support Vector Machine, Linear Regression and Artificial Neural Network machine learning methods to predict tool design effort from encoded input data.

In this study; it is aimed to determine the ideal machine learning model for the best estimation of tool design effort and to create the most appropriate input and parameter set.

KEYWORDS: tool design, machine learning, design effort estimation, aviation and defense.

İÇİNDEKİLER

Sayfa

TEŞEKKÜR.....	i
ÖZET	ii
ABSTRACT	iii
İÇİNDEKİLER.....	iv
EKLER	vi
TABLolar LİSTESİ	vii
ŞEKİLLER LİSTESİ	viii
SİMGELER VE KISALTMALAR LİSTESİ	x
1. GİRİŞ	1
1.1. Motivasyon.....	4
1.2. Literatür Taraması	5
2. PROBLEM TANIMI VE YÖNTEM.....	9
2.1. Problem Tanımı	9
2.2. Takım Tasarım Süreci.....	9
2.3 Çalışma Yol Haritası.....	12
3. ANALİZ	14
3.1. Kullanılan Veri Seti	14
3.2. Kategorik Verilerin Sayısallaştırılması.....	15
3.2.1 Sıralı Kodlama.....	15
3.2.2 İkili Kodlama	16
3.2.3 One Hot Kodlama	17
3.2.4 Dummy Kodlama	17
3.2.5 Efekt / Deviation Kodlama	18
3.2.6 Frekans Kodlama	18
3.2.7 Ortalama Kodlama	19
3.3. En İyi Kodlama Yönteminin Seçilmesi	19
3.3.1 Doğrusal Regresyon	20
3.3.2 Karar Ağacı	20

3.3.3	Destek Vektör Makinesi	21
3.3.4	Yapay Sinir Ağı	22
3.4.	Verilerin İncelenmesi.....	24
3.5.	Veri Ön İşlemesi ve Makine Öğrenmesi	25
3.5.1	Uç Değerlerden Arındırılmış Veri ile Makine Öğrenmesi	26
3.5.2.	Girdi Parametrelerinin Korelasyona Katkısına Göre Düzenlenmesi.....	27
3.5.3	Girdi Parametrelerinin Uzman Görüşü Alınarak Yeniden Değerlendirilmesi	30
3.6.	Gerçekleşen Çıktı Değerlerinin İncelenmesi	33
3.7.	YSA Optimizasyonu.....	33
3.8.	En Verimli Makine Öğrenme Yöntemi İçin Kodlama Deneyleri ve Sonuçları	40
3.9	Hesaplama Sonuçlarının Değerlendirilmesi	41
4.	SONUÇ VE ÖNERİLER.....	45

EKLER

EK 1: YSA OPTİMİZASYONU İÇİN MATLAB KODU

TABLolar LİSTESİ

	Sayfa
Tablo 3.1 Girdi ve çıktı parametrelerinin tanımları.....	14
Tablo 3.2 RFO parametresinin sıralı kodlanması.....	16
Tablo 3.3 RFO parametresinin ikili kodlanması	16
Tablo 3.4 RFO parametresinin One Hot Kodlama yöntemi ile kodlanması	17
Tablo 3.5 RFO parametresi için frekans kodlaması	18
Tablo 3.6 RFO parametresi için ortalama kodlaması.....	19
Tablo 3.7 Karar ağacı Matlab girdi parametreleri	21
Tablo 3.8 DVM girdi parametreleri.....	21
Tablo 3.9 YSA deneyi parametreleri	23
Tablo 3.10 Kodlama metodu ve makine öğrenme modellerinin RMSE sonuçları.....	23
Tablo 3.11 Takım tasarım eforu çeyreklik bilgileri.....	24
Tablo 3.12 Uç değerlerden arındırılmış veri üzerinde makine öğrenmesi uygulaması sonuçları.	27
Tablo 3.13 Girdi parametrelerinin korelasyona katkısına göre düzenlenmesi sonucundaki makine öğrenme performansı	30
Tablo 3.14 Girdi ve çıktı parametrelerin uzman görüşü ile temizlenmesi sonucunda yapılan makine öğrenmesi performansı sonuçları.....	33
Tablo 3.15 Tasarım eforu kestirimi için kullanılan YSA'nın başlangıç parametreleri	34
Tablo 3.16 YSA mimarisi ve kodlama yöntemi en iyi hale getirme deneyleri	41

ŞEKİLLER LİSTESİ

Sayfa

Şekil 1.1 Atak Helikopteri (sağda) ön gövdesi alt montaj takımı (solda)	2
Şekil 1.2 Bir uçağa ait kompozit motor kapağı (sağda) için tasarlanıp üretilen kompozit serme takımı (solda)	2
Şekil 1.3 Uçak kanatçığı (sağda) montajının robotik kontrol ile delinebilmesi için tasarlanıp üretilmiş bir takım (solda)	3
Şekil 1.4 Alüminyum bir hava aracı yapısalı parçasını üretmek için üç boyutlu yazıcı ile imal edilmiş bir hidroform şekillendirme kalıbı.....	3
Şekil 2.1 TUSAŞ Takım Tasarım Süreci girdi veri akış şeması blok diyagramı	10
Şekil 2.2 TUSAŞ Takım Tasarım Süreci blok diyagramı	11
Şekil 2.3 TUSAŞ Takım Tasarım Süreci sonucunda oluşan verinin saklanması.....	11
Şekil 2.4 Kurumsal veri tabanından imalat mühendisliği takım tasarım isterleri ve tasarım eforu tablosu oluşturulması	12
Şekil 2.5 Araştırma yöntemi mimarisi.....	13
Şekil 3.1 Deney için kullanılan YSA'nın genel görünümü.....	22
Şekil 3.2 Takım tasarımı için harcanan efor saatinin dağılım grafiği (0-4500 saat ve 0-150 saat arası).....	25
Şekil 3.3 Takım tasarımı için harcanan efor saatinin uç değerler çıkarıldıktan sonraki dağılım grafiği	26
Şekil 3.4 Girdi Parametrelerinin Korelasyon Katsayılarının Dağılım Grafiği	28
Şekil 3.5 Normal Dağılımın İçinde Kalan Girdi Parametrelerinin Korelasyon Katsayılarının Dağılım Grafiği	29
Şekil 3.6 Normal Dağılımın Dışındaki Girdi Parametrelerinin Korelasyon Katsayılarının Dağılım Grafiği	29
Şekil 3.7 Uç değerlerden arındırılmış veriler ile yapılan çalışmadaki karar ağacı yönteminin tahmin ve gerçekleştirmeleri grafiği (üstte), üstteki grafiğin 0 - 25 saat aralığının büyütülmüş hali (altta).....	31
Şekil 3.8 Tasarım eforu kestirimi için kullanılan YSA görseli	33
Şekil 3.9 YSA en uygun şekle sokma algoritması akışı	35

Şekil 3.10 Bir katmanlı dummy kodlama ile eğitilmiş veri setinin RMSE değerinin nöron sayısına göre değişimi	36
Şekil 3.11 Bir katmanlı dummy kodlama ile eğitilmiş veri setinin R değerinin nöron sayısına göre değişimi	37
Şekil 3.12 Birinci katmanı 202 nörondan oluşan, ikinci katmanının dummy kodlama ile eğitilmiş veri setinin RMSE değerinin nöron sayısına göre değişimi	37
Şekil 3.13 Birinci katmanı 202 nörondan oluşan, ikinci katmanının dummy kodlama ile eğitilmiş veri setinin R değerinin nöron sayısına göre değişimi	38
Şekil 3.14 Birinci katmanı 202 nörondan oluşan ve ikinci katmanı 152 nörondan oluşan, üçüncü katmanının dummy kodlama ile eğitilmiş veri setinin RMSE değerinin nöron sayısına göre değişimi	39
Şekil 3.15 Birinci katmanı 202 nörondan oluşan ve ikinci katmanı 152 nörondan oluşan, üçüncü katmanının dummy kodlama ile eğitilmiş veri setinin R değerinin nöron sayısına göre değişimi	39
Şekil 3.16 Deneyler sonucunda en iyi hale getirilmiş YSA	40
Şekil 3.17 İkili kodlama için hesaplanan en iyi YSA mimarisi.....	41
Şekil 3.18 Takım Tasarım Eforu Makine Öğrenme Modeli Hata Dağılımı.....	42
Şekil 3.19 Günlük çalışma üzerinden tahmin edilen ve gerçekleşen takım tasarım eforunun hata dağılımı	44

SİMGELER VE KISALTMALAR LİSTESİ

2D	Two dimensions
3D	Three dimensions
CAD	Computer Aided Design
CAE	Computer Aided Engineering
COCOMO	Constructive Cost Model
DFM	Design for Manufacturing
DVM	Destek Vektör Makinesi
E	matematiksel ondalık katsayı gösterimi
F	Fisher dağılımı değeri
GRNN	Generalized Regression Neural Network
μ	Yapay Sinir Ağı adaptasyon parametresi
NC	Numerical Control
p	istatistiksel ihtimal (Probablity) değeri
PCA	Principal Component Analysis
PLM	Product Lifecycle Management
RBF	Radial Basis Function
RFO	Reason for Order
RMSE	Root Mean Square Error (Kök Ortalama Kare Hata)
R	korelasyon katsayısı
SMO	Sequential Minimal Optimization
SQL	Structured Query Language
SRS	Software Requirements Specifications (Yazılım Gereksinim Özellikleri)
TD	Tool Design
TO	Tool Order
TUSAŞ	Türk Havacılık ve Uzay Sanayi Anonim Şirketi
ÜPK	Üretim Planlama ve Kontrol
YSA	Yapay Sinir Ağı

1. GİRİŞ

Takım; bir ürünün üretilebilmesi için gerekli olan çeşitli kaynak ve ekipmandan birisi olarak tanımlanabilir. Seri üretim, tekrarlanabilirlik ve değiştirilebilirlik¹ gerektirir. Bunu başarmak için üretim adımlarını kolaylaştıracak ve hata yapma olasılığını azaltacak araç ve ekipmana ihtiyaç vardır. Buna ek olarak, bazı üretim süreçleri, malzemenin aşırı basınç, kuvvet veya sıcaklıklarda tutulmasını gerektirebilir. Takım, ulaşılması gereken bu tür aşırı koşullara ulaşmak için tek çözümdür.

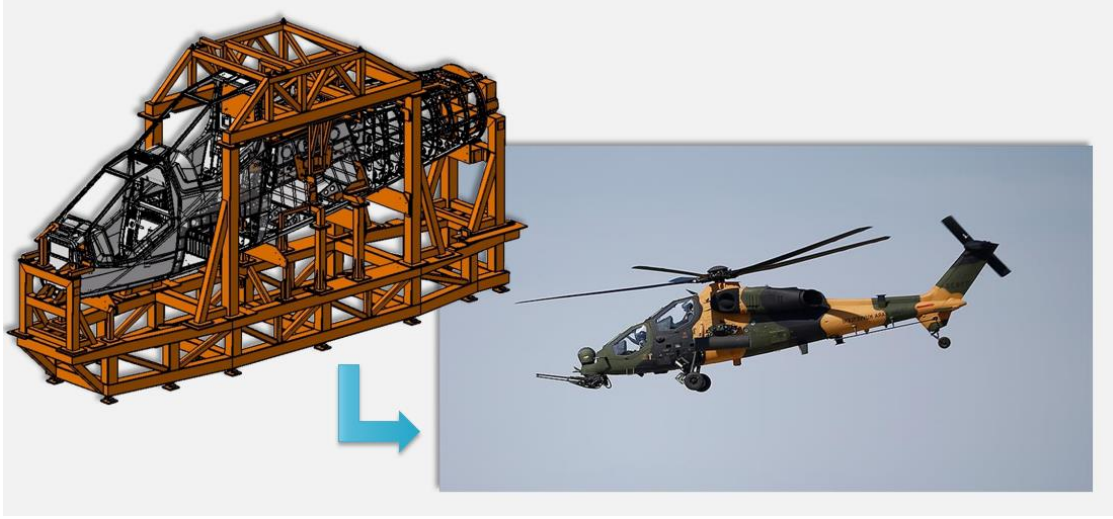
Takımlar, çeşitli amaçlar doğrultusunda ürüne özel tasarlanıp üretilirler. Örneğin, torna, freze veya robotik işlemler fikstürleri, sac metal şekillendirme kalıpları, montaj takımları, kompozit serme kalıpları, taşıma takımları vb. Diğer bir deyişle her bir ürün için bir takım ya da takım seti olmalıdır. Bahse konu bu takımlar basit bir ürün için birkaç tane olabilirken karmaşık sistemlerden oluşan bir hava aracı için takım çeşitliliği binler mertebesine rahatça çıkabilmektedir.

Havacılık sanayinde kullanılan takımların çeşitliliği yukarıda bahsedildiği gibi ürünün karmaşıklığı, üründe kullanılan materyaller, ürünün üretim adedi, bakım ve ikame edilebilirlik düzeyi, hava aracının üretim için kullanılacak standart makine ve ekipmanın türü ve çeşitliliği, takım kullanıcısının yetkinliği, iş güvenliği gereksinimleri gibi birçok parametre ile değişebilmektedir. Örnek olarak sunulan bu parametreler takım çeşitliliğine ek olarak takımın tasarım ve imalat sürecine de direkt ya da indirek şekilde etkimektedir.

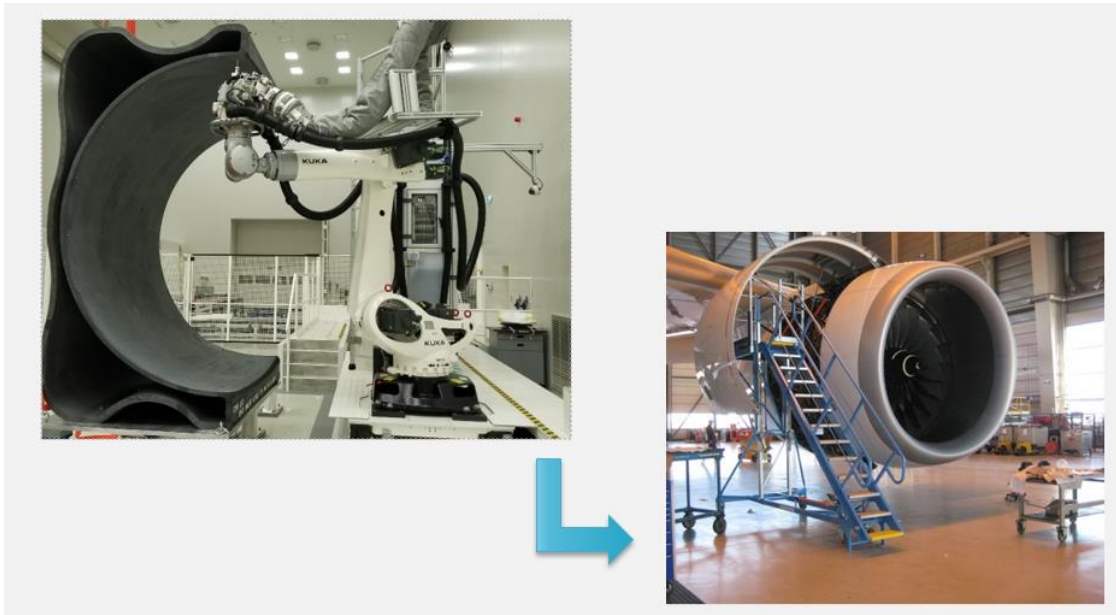
Atak helikopterin ön gövde montajını gerçekleştirmek için tasarlanıp üretilen ve ölçüleri yaklaşık olarak 8m x 4.5m x 2.5m olan bir montaj takımı Şekil 1.1’de gösterilmiş olup ilgili takımın tasarımı ortalama 8 ay sürmüştür. Benzer şekilde Airbus Firması’na ait bir uçağın motor kapağının kabuk yüzeyinin üretiminde kullanılacak bir kompozit serme

¹ İngilizce’de “interchangeably” olarak geçen “değiştirilebilirlik” kelimesi; bir ürünün içerisindeki bir parça ya da komponentin değiştirilmesi gerektiğinde yerine gelecek parçanın ürünle tam uyum sağlayabilmesi özelliğidir.

takımının (Şekil 1.2) tasarımı yaklaşık 2.5 aydır. Yaklaşık 5.5m x 2m x 1m boyutlarında olan ve Şekil 1.3’de görseli sunulan robotik delik delme takımının ise tasarımı yaklaşık 2 ay sürmüştür. Benzer şekilde Şekil 1.4’de gösterilen yaklaşık 25cm x 15cm x 7cm ölçülerindeki takım ise 3 saat içerisinde tasarlanmıştır.



Şekil 1.1 Atak Helikopteri (sağda) ön gövdesi alt montaj takımı (solda)



Şekil 1.2 Bir uçağa ait kompozit motor kapağı (sağda) için tasarlanıp üretilen kompozit serme takımı (solda)



Şekil 1.3 Uçak kanatçığı (sağda) montajının robotik kontrol ile delinebilmesi için tasarlanıp üretilmiş bir takım (solda)



Şekil 1.4 Alüminyum bir hava aracı yapısalı parçasını üretmek için üç boyutlu yazıcı ile imal edilmiş bir hidroform² şekillendirme kalıbı

² Çelik, titanyum ve alüminyum gibi metallerin şekillendirilmesine olanak sağlayan bir imalat yöntemidir.

1.1. Motivasyon

Yukarıda verilen örneklerden de anlaşılacağı üzere takım kavramı içerisinde tasarlanıp üretilen ürünler çok geniş bir özellik yelpazesine sahiptir. O nedenle takımların tasarım sürelerinin gerçeğe yakın olarak kestirilmesi karmaşık bir problem haline gelebilmektedir.

Havacılık ve savunma sanayinde geriye doğru takip edilebilirliğin garanti altına alınabilmesi için ürünle ilgili oluşturulan tüm veri kayıt altında tutulur. Tasarım girdileri, çıktıları ve belirli bir iş için harcanan efor da bu kayıt altında tutulan veriler tarafından kapsamaktadır.

Klasik yöntemde ilgili proje planının tasarım süresi ya da tasarım işinin büyüklüğü tahmin edilirken uzman görüşüne başvurulur. Uzman, öncesinde yaptığı tasarım işlerinin gerçekleşen eforu ile kendi deneyimini kullanarak proje planı için en iyi kestirimi yapmaya çalışır. Fakat yapılan bu kestirimin belirli bir metodu yoktur ve proje planına girdi olacak veri tamamen kestirimi yapan tasarım mühendisinin inisiyatifindedir. O nedenle yapılan kestirimin doğruluğu ancak projenin sonunda belirlenebilir.

Diğer yandan herhangi bir kurumsal firmanın bu kestirimi yeterince doğru yapabilecek deneyime sahip uzman mühendisi bulunmayabilmektedir. Havacılık ve savunma sanayindeki projeler göreceli olarak uzun projeler olup on yıllarca sürebilmektedir. Belirtilen süre zarfında bir deneyimin oluşması da en az bir proje süresinin tamamlanması ile olabildiğinden dolayı tasarım tahminlerinin kirşlerden ziyade kurumların yetkinliğinde olması şirketlerin kalıcılığı bakımından büyük önem arz etmektedir.

Bu tez çalışmasında kurumsal firmaların sahip olduğu kurumsal bilgi varlıkları ve yukarıda belirtilen kısıtların ışığında takım tasarım sürecinin efor tahmini için uygun olabilecek en iyi makine öğrenmesi yönteminin belirlenmesi ve bu yöntemin uygulanabilmesi için oluşturulabilecek optimal sürecin belirlenmesi hedeflenmiştir.

1.2. Literatür Taraması

Her bir tasarım sürecinde tasarımın gerçekleştirilmesi için yeterli teknik girdi ve gereksinimler net tanımlı olmalıdır. Gereksinimlerin karmaşıklığı ve belirsizliği tasarım eforunu arttıran etmenlerdendir. Tasarım eforu kestirimi için yapılan çalışmaların araştırılmasının sonucunda dikkat çeken bilimsel yayınlar aşağıda özetlenmeye çalışılmıştır;

Poli ve ark., "Tasarım"ı, tasarlanan bir nesne hakkında bilinen ve kaydedilen bilgilerin eklendiği, iyileştirildiği (yani daha ayrıntılı hale getirildiği), değiştirildiği veya az çok kesinleştirildiği faaliyetler dizisi olarak tanımlamaktadır [1].

Tasarım sürecinde bilgi akışının önemine değinen A.J. Dentsoras Tasarım gereksinimleri ve efor arasındaki bir kasnaklı konveyör sistemi üzerinden örneklendirerek tanımlamıştır [2]. Tasarım sürecinde bulunan anahtar terimler, tasarım döngüsü ve tasarımı en yüksek oranda etkileyen bağımlı değişkenleri incelenmiştir. Verileri belirtilen bu parametreler ile sayısallaştırarak bir matris haline getirmişlerdir. Böylelikle tasarım eforu ve parametrelerin önemi arasında bir ilişki kurmayı başarmışlardır.

Mark V. Martin ve ark. yaptığı çalışmada "Design For Variety" adını verdiği yöntemde çeşitli araçlar kullanarak tasarım projelerinin farklılıklarını belirleyerek projenin tasarım maliyetini çıkartmaya çalışmıştır [3]. Otomobil pencere motorları, sıcaklık ölçüm kablosu konektörleri ve sabit disk sürücülerini projelerinin kullandığı bu çalışmada; kaba olarak maliyet kalemlerini belirleyip düzenleyebilmişlerdir.

Ürün geliştirme sürecinde, ürünün tamamlanma ihtimalini zaman ile bağımlı olarak tahmin etmek için M. Carrascosa ve ark. bir model önermişlerdir [4]. Bu modelde görevler arası girdi ve çıktıyı değerlendirerek bir tasarım yapısı matrisi oluşturmuşlardır. Bu girdi matrisini kullanarak sürecin değişiklik analizini yapmışlardır.

Benzer olarak ürün ailelerinin geliştirme ve değişiklik maliyetlerini tahmin edebilmek için; problem tanımı, gereksinimler ve kapsam içeriği üzerinde geliştirilen bir modeli ise J. Park ve ark. sunmuştur [5]. Üretim adetleri, işçilik ve makine maliyetleri şirket

giderleri gibi kalemlerin hesaplanmasını da içeren bu modelde tasarım aşamasında hesaplama için gereken girdi parametrelerinin yeterli olmayabileceğinden bahsetmiştir.

Rajkumar Roy ve ark. otomotiv endüstrisinde yeni teknolojiye sahip ürünlerin maliyetlerinin tahmininin önemli olduğuna değinmiştir [6]. Yeni teknoloji içeren bir otomobil güç aktarma organı tasarımı için maliyet çıkarılması üzerine olan bu çalışmada; parametrik, analoji ve ayrıntılı tahmin tekniklerini kullanan bir yöntem önermiştir. Girdilerin çoklu regresyon modeli ile maliyete dönüştürüldüğü bu çalışmanın kapsamının sadece yeni teknoloji içeren ürünler ile sınırlı olduğu belirtilmiştir.

Bilgisayar mühendisliğinde Yazılım Gereksinim Özellikleri (SRS) dokümanı ile takım tasarım gereksinimleri tanımlı farklı bilimlere ait olmasına karşın benzer girdileri içermektedir. Her ikisi de projenin yapılabilmesi ya da tasarlanabilmesi için gerekli olan gereksinimleri içermektedir.

Sumeet Kaur Sehra ve ark. 1996-2017 arasında yayınlanan yazılım efor tespit araştırmalarını doğal dil işleme yolu ile inceleyerek anlamsal olarak gruplandırmıştır [7]. 1178 makalenin incelenmesi sonucunda içerik olarak 15 ana konu başlığında yayınlanan makale sayıları hesaplanmıştır. Hesaplama sonucunda son 15 yılda yazılım eforu tespitinde makine öğrenmesi metotlarının kullanımının arttığı gözlemlenmiştir.

Çeşitli regresyon modellerini kullanarak çalışma yapan J. Mesa ve ark. Constructive Cost Model (COCOMO) ve Function Points gibi tasarım eforu kestirim yöntemlerine benzer şekilde mikroişlemcilerin tasarım eforlarını kestirebilmek için bir yöntem geliştirmişlerdir [8] [9] [10].

Benedetto ve Silva ise tasarım kestirimine destek olacak bir yapı önermiştir [11]. Bu yapıyı geliştirmek için 13 uzman tasarımcıdan oluşan bir ekibe çeşitli sorular yönlendirilmiş ve sonuçlarını değerlendirmiş ve dış süreç girdilerinin proje maliyetlerini etkileyen en büyük etmen olduğunu tespit etmişlerdir.

Genel tasarım eforu kestirimi hakkında birden fazla çalışmaları olan Bashir ve Thomson ise ürün karmaşıklığını tasarımla ilişkilendirecek “Sezgi”, “Duyarlılık”,

“Tutarlılık”, “Genellik”, “Basitlik” parametreleri ile ürün karmaşıklığını ölçmeyi hedeflemiştir [12].

Bashir ve ark. diğer bir çalışmada 15 hidroelektrik santrali projesi üzerinde çeşitli parametrik yöntemler kullanarak tasarım eforlarını kestirmeye çalışmıştır [13]. Yine hidroelektrik santrali sektöründe Bashir, çoklu regresyon analizi ile tasarım geliştirme süresini tahmin etmeye çalışmıştır [14]. Araştırmasının sonucunda veri setinin homojenliğinin verimi arttıracığına değinen yazar, sunduğu modelin diğer tasarım sektörleri için de kullanılabilirliğini belirtmiştir.

Bashir ve ark. başka bir çalışmada analoji tabanlı bir model matris yapısı oluşturarak Kanada’da bulunan şarj cihazı ve iletişim cihazı üreten iki şirketin tasarım eforları üzerinden bir tahmin çalışması yapmıştır [15].

Diğer yandan tasarımın ardından yapılacak üretim eforunun kestirimi eldeki girdilerin daha detaylı olduğundan dolayı daha kolaydır. Bu konuda ise C. Ou-Yang ve T. S. Lin Bilgisayar Destekli Tasarım (CAD) programlarındaki parametreleri yorumlayarak bir Nümerik Kontrol (NC) tezgâhta işlenecek bir iş parçasının üretim süresini kestirebilecek bir yöntem tanıtmıştır [16].

Tasarım ve üretim üzerine çalışmalara benzer olarak bakım ve sürdürülebilirlik eforu konusunda da çalışmalar bulunmaktadır. Raj Sodhi, Manuela Sonnenberg & Sanchoy Das ürünlerin de montesinin eforunu hesaplamak için U-Effort adında bir çalışma yapmışlardır. Diğer çalışmalardan farklı olarak bu çalışma deney sonucuna dayanmaktadır.

Yapılan literatür araştırması sonucunda tasarım efor tespiti konusunda makine öğrenmesi ile ilgili bir uygulanmaya rastlanmamıştır. Yalnızca Salam ve Bhuiyan [16] çalışmada, Pratt & Whitney³ Şirketi’ndeki kompresör fan tasarımlarında doğrusal ve doğrusal olmayan regresyon modelleri ile tasarım süresi kestirimi yapmışlardır. Aynı

³ Merkezi Hartford, Connecticut'ta olan bir havacılık ve uzay şirkettir

çalışmada tasarım sürecinin kestirimi için Yapay Sinir Ağları (YSA) da kullanılabileceğini önermişlerdir.

Bir sonraki bölümde tanıtılacak problemde girdi verileri kategorik olarak tanımlanmıştır. Bu konudaki çalışmalara bakıldığında; K. Potdar ve ark. “Car Evaluation” veri seti üzerinde çalışmışlardır [17]. Çeşitli kodlama yöntemleri ve YSA kullanılarak çalışmalar yapmış ve bu çalışmalar sonucunda hangi kodlama yönteminin daha başarılı olduğunu değerlendirmişlerdir. Ayrıca kullandıkları kodlama yöntemlerinin olumlu ve olumsuz yanlarını da sıralamışlardır.

Yüksek parametre çeşitliliğine (kardinalite) sahip veri setleri üzerinde çalışma yapan Patricio Cerda ve ark. Gamma-Poisson Matrix Factorization ve Min-hash kodlama yöntemleri üzerinde yeni bir yaklaşım tanıtmışlardır [18]. Ayrıca bu çalışma verideki ölçeklenebilirliğin durumuna göre uygun olan kodlama yönteminin seçilebilmesi konusunda önerilerde bulunmuştur. One Hot kodlamanın pratik ve güvenilir olduğuna değinmesine karşın yüksek kardinaliteye sahip verilerde yazarların önerdiği yaklaşım alan karmaşıklığını düşürerek One Hot kodlamaya yakın sonuçlar ulaşılmasını sağlamıştır.

2. PROBLEM TANIMI VE YÖNTEM

2.1. Problem Tanımı

Bu çalışmada takım tasarımı süresini kestirebilmek için bilinen makine öğrenmesi yöntemlerinin kullanılması ve en iyi sonuç vereninin yöntemin hâlihazırda olan yöntem ile karşılaştırılması sunulmuştur.

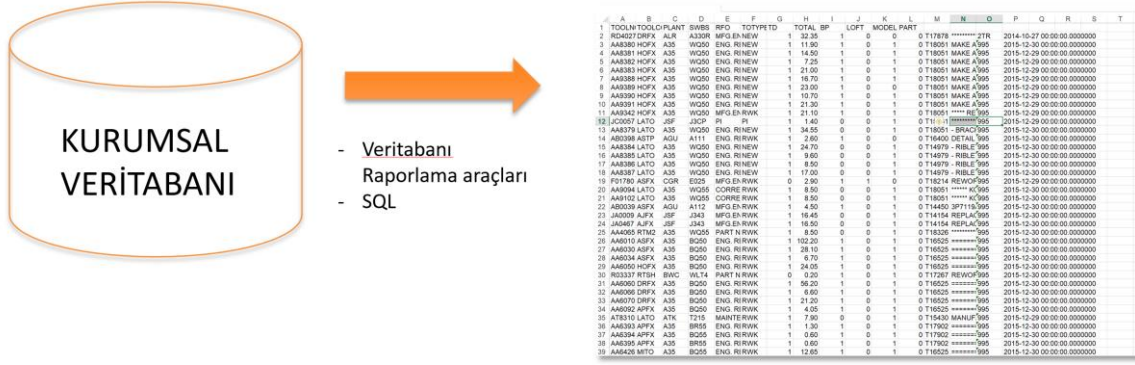
Havacılık ve uzay sanayinde çalışılan projeler diğer sanayi projelerinin pek çoğundan göreceli olarak uzun sürmektedir. Bunun nedeni hava araçlarının pek çok alt sistemden oluşması ve güvenlik faktörünün diğer ürünlere göre çok sıkı olmasından kaynaklanmaktadır. Bu nedenle üretilen tasarım verileri de ürün gereksinimleri ile orantılı olarak fazladır. Bu boyut ve çeşitlilikteki verileri geleneksel yöntemler ile saklamak oldukça maliyetli ve risklidir. Product Lifecycle Management (PLM) yazılımları bu noktada bu tür projelerin idamesi için gerekli çözümleri sağlayabilmekle beraber pek çok çeşit veriyi de tablolar halinde kullanıcılara sunabilmektedir. Teamcenter PLM, ENOVIA, SAP SE, ORACLE, Üretim için Tasarım (DFM) sürecinde kullanılan PLM çözümlerinin örnekleridir [19] [20] [21] [22].

Bu noktada girdi verilerin bir kısmı ve harcanan efor bilindiğinden dolayı makine öğrenmesi yöntemleri ile gelecek projelerin tasarım eforlarının kestirilmesi teorik olarak mümkün görünmektedir.

2.2. Takım Tasarım Süreci

Türk Havacılık ve Uzay Sanayi Anonim Şirketi (TUSAŞ) PLM sistemine gömülü olarak çalışabilen Siemens NX ve CATIA olmak üzere iki ayrı CAD sistemi bulunmaktadır [23] [24]. Takım tasarım sürecinde ise yaygın olarak CATIA yazılımı kullanılmaktadır. Her bir takımın tasarım gereksinimleri ve eforu benzer şekilde PLM destekli TUSAŞ kurumsal veri tabanında tutulmaktadır. Şekil 2.1’de sunulan blok diyagram takım tasarım süreci için gereken veri akışını özetlemektedir.

Ürüne ait olan mühendislik ve tasarım verileri imalat mühendisliği tarafından incelenir. Ürünün üretim rotasını oluşturan imalat mühendisi aynı zamanda ilgili ürün için



Şekil 2.4 Kurumsal veri tabanından imalat mühendisliği takım tasarım isterleri ve tasarım eforu tablosu oluşturulması

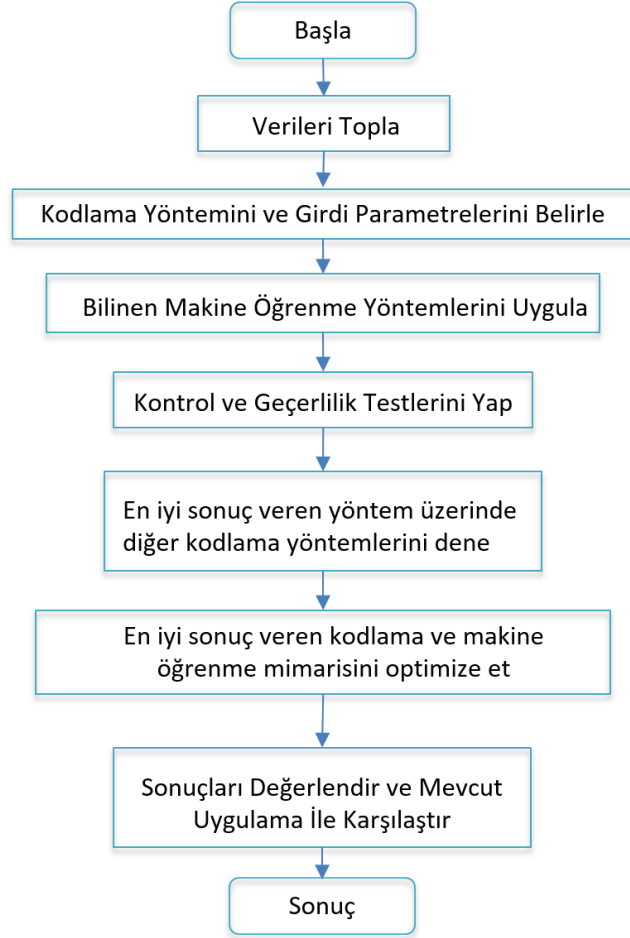
2.3 Çalışma Yol Haritası

İmalat mühendisliğinin tarif ettiği takım tasarım gereksinimleri ve tasarımcının harcadığı süre arasında bir korelasyon olduğundan bu veri ile makine öğrenmesi yöntemleri denenerek gelecekteki projeleri kestirmek mümkün olabilir. Girdiler ile çıktılar arasındaki korelasyonu belirleyebilmek için Şekil 2.5’de sunulan süreç planlanmıştır. Öncelikle edinilebilen tüm veriler kurumsal veri tabanı üzerinden bir Structured Query Language (SQL) sorgusu ile çekilerek tek bir tablo haline getirilmiştir.

İmalat mühendisliğinin hazırladığı takım tasarım gereksinimleri tablosu tamamen kategorik değerler içerdiğinden öncelikle bu veriler kodlanmalıdır. Literatürde kategorik değerler için çeşitli kodlama yöntemleri bulunduğundan en iyi kodlama yöntemini belirleyebilmek için bir çalışma yapılmalıdır. Bu çalışma yapılırken aynı zamanda girdileri oluşturacak imalat mühendisliği takım emri verileri içerisinde uygun parametreler de seçilmiş olur.

En iyi kodlama yönteminin belirlenmesinin ardından regresyon için uygun olan makine öğrenme teknikleri denenir. Yapılan kontrol ve geçerlilik testlerinin ardından en iyi sonuç veren yöntemi bulabilmek için korelasyon içerisindeki Kök Ortalama Kare Hata

(RMSE) ve Korelasyon Katsayısı (R) deęerleri karřılařtırılır. Bu noktada RMSE ve R deęeri en uygun olan makine öğrenme yöntemi üzerinde dięer kodlama yöntemleri de denenerek, en iyi RMSE ve R deęerine sahip olan kodlama ve makine öğrenme yöntemi üzerinde bir iyileřtirme de yapılır.



řekil 2.5 Arařtırma yöntemi mimarisi

Son olarak, elde edilen sonuçlar takım tasarım proje süresi kestiriminde hâlihazırda uygulanan yöntem ile karřılařtırılarak sonuçlar deęerlendirilmiřtir.

3. ANALİZ

Bu bölümde girdi ve çıktı verileri tanımlanarak, yöntem mimarisi üzerinden yapılan çalışma sunulmuştur.

3.1. Kullanılan Veri Seti

TUSAŞ Kurumsal veri tabanından son 10 yıl içerisinde tasarımı tamamlanmış tüm projelerin tasarım eforu saatleri bir SQL sorgusu ile raporlanmış ve 52,215 takım tasarım projesinin verisi elde edilmiştir.

Takım Emri'nde de geçen bu verinin içeriğindeki alanlar aşağıda özetlenmiştir;

- Takımın kimlik numarası olarak tanımlanabilen takım ve seri numaraları
- Proje müşterisine ait proje kodu (PLANT)
- Ürünün kullanılacağı yere göre tanımlanan komponent kodu
- Tasarım gerekçesini belirten RFO) kodu
- Tasarımın yeni ya da değişiklik olduğunu belirten, ayrıca değişikliğin türünü tanımlayan TOTYPE kodu
- Takımın türünü belirleyen TOOLCODE
- Takımın kayıtlı tasarımının yapıp yapılmadığını belirleyen Tool Design (TD) kodu
- Tool Order (TO) hazırlayan imalat mühendisinin kullanıcı kodu
- Takım üretiminin tamamlandığını belirten istasyon kodu
- Saat cinsinden tasarım eforu (TOTAL)

Yukarıda belirtilen alanlardan takım tasarım süresine etken olan parametreler seçilerek Tablo 3.1'de tanımları ile gösterilmiştir;

Tablo 3.1 Girdi ve çıktı parametrelerinin tanımları

Parametre	Türü	Veri Türü	Açıklama
TOOLCODE	Girdi	Kategorik	Takım Kodu
PLANT	Girdi	Kategorik	Proje Kodu
RFO	Girdi	Kategorik	Takım Emri Hazırlanma Gerekçesi
TOTYPE	Girdi	Kategorik	Takım Türü
TOTAL	Çıktı	Reel Sayı	Toplan Harcanan Tasarım Süresi

TOOL CODE: Takımın türünü ve işlevini belirten parametredir ve genellikle dört karakterden oluşur. Her bir takım türünün tasarım eforu farklılık gösterdiğinden girdi parametresi olarak kullanılmıştır.

PLANT: Takıma ait projenin adını tanımlar. Projelerin çeşitleri genel olarak tasarım standardı kullanımında ve zorluğunda belirleyici bir faktördür.

RFO: Takım emrinin imalat tarafından ne amaç ile yazıldığını belirleyen dokümandır. Örneğin bir mühendislik değişimi olduğunda ya da bir tasarımın düzeltilmesi gerektiğinde bu parametre belirleyicidir.

TO TYPE: Takım emrinin türünü belirler. Örneğin takım yeni olması ya da bir değişikliğe ihtiyacı olması durumu bu parametre ile tanımlanır.

TOTAL: Takım tasarım mühendisinin takım tasarımı için harcadığı saat cinsinden süredir.

3.2. Kategorik Verilerin Sayısallaştırılması

Bu çalışmadaki tüm girdiler kategorik olduğundan makine öğrenmesi yöntemlerini uygulayabilmek için öncelikle kategorik verilerin sayısal olarak kodlanması gerekmektedir. Bu bölümde literatürde bulunan, uygulanacak kodlama yöntemleri tanıtılarak en başarılı kodlama yönteminin belirlenmesi için yapılmış olan çalışma tanıtılacaktır.

3.2.1 Sıralı Kodlama

Bu kodlama türünde girdi parametresinin içeriğinde bulunan değerlere veri sırası rakam verilerek girdi sayısallaştırılır. Tablo 3.2'de girdi parametrelerinden RFO için oluşturulmuş kodlama görülebilir.

Tablo 3.2 RFO parametresinin sıralı kodlanması

RFO	KOD
CORRECT TOOL / DESIGN	1
CORRECT TOOL / MFG	2
CORRECT TOOL / PLAN	3
CORRECT TOOL / VENDOR	4
ENG. REL.	5
MAINTENANCE	6
MFG.ENG.REQ.	7
P/N REVISION LEVEL CHANGE	8
PART NO CHANGE	9
PI	10
PROD.IMPVT	11
QUALITY	12
RATE TOOL	13
RATE TOOLS	14
REPL LOST TOOL	15
REPLACE LOST TOOL	16
REPLACE LOST TOOL DETAIL	17
TOOL LABEL CHANGE	18

3.2.2 İkili Kodlama

Tablo 3.3 RFO parametresinin ikili kodlanması

RFO	İKİLİ KODLANMIŞ RFO				
	Rfo_01	Rfo_02	Rfo_03	Rfo_04	Rfo_05
ENG. REL.	0	0	1	0	1
ENG. REL.	0	0	1	0	1
ENG. REL.	0	0	1	0	1
TOOL LABEL CHANGE	1	0	0	1	0
TOOL LABEL CHANGE	1	0	0	1	0
ENG. REL.	0	0	1	0	1
ENG. REL.	0	0	1	0	1
ENG. REL.	0	0	1	0	1
ENG. REL.	0	0	1	0	1
MFG.ENG.REQ.	0	0	1	1	1
CORRECT TOOL / DESIGN	0	0	0	0	1
CORRECT TOOL / DESIGN	0	0	0	0	1

İkili kodlamada sıralı kodlamada olduğu gibi önce parametrelerin türlerine sıra ile numara verilir. Örneğin Tablo 3.2’de RFO değeri için 1-18 arasındaki değerleri vardır. En büyük değer olan “18” değeri ikili değere dönüştürüldüğünde 5 bit uzunluğunda olacaktır. Bu da RFO değeri için “5” adet parametre olarak tanımlanır. (Tablo 3.3)

3.2.3 One Hot Kodlama

Bu kodlama türünde tüm girdi parametrelerinin her bir değeri için ikili değerden oluşan bir parametre seti oluşturulur. Bellek karmaşıklığı yüksek olan bu yöntem için örnek bir kodlama Tablo 3.4’de gösterilmiştir.

Tablo 3.4 RFO parametresinin One Hot Kodlama yöntemi ile kodlanması

ENG. REL.	TOOL LABEL CHANGE	MFG.ENG. REQ.	CORRECT TOOL / DESIGN	CORRECT TOOL / MFG	CORRECT TOOL / PLAN	MAINTENANCE	P/N REVISION LEVEL	RFO PARAMETRESİNİN DEĞERLERİ								CORRECT TOOL / VENDOR	
								PART NO CHANGE	PI	PROD.IMP VT	QUALITY	RATE TOOLS	REPL LOST TOOL	REPLACE LOST TOOL	REPLACE LOST TOOL		RATE TOOL
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

3.2.4 Dummy Kodlama

Dummy kodlama, One Hot kodlamaya çok benzer olup bellek karmaşıklığı bakımından biraz iyileştirilmiş olarak tanımlanabilir. Aralarındaki tek fark, her bir parametrenin içerdiği değerlerden birinin kukla değer olarak seçilmesi ve sıfırlardan oluşturulmasıdır. Diğer bir deyişle One Hot kodlamada türetilen N adet girdi parametresine karşın Dummy kodlamada N-1 tane parametre olacaktır.

3.2.5 Efekt / Deviation Kodlama

Bu kodlama türü ise Dummy kodlamaya çok benzemektedir. Aralarındaki tek fark Dummy kodlamada sıfırlardan oluşan bir parametre bu kodlama türünde -1'lerden oluşturulur. H Alkharusi yaptığı çalışmada Dummy ve Efekt kodlamayı karşılaştırmış ve regresyon denklemindeki değişiklik dışında R^2 ve F değerlerinde anlamlı bir fark bulamamıştır [25].

3.2.6 Frekans Kodlama

Frekans kodlaması her bir parametrenin içeriğinde bulunan değerlerin verinin tamamında bulunma adedinin sayılması ile kodlanır. Tablo 3.5'de RFO parametresi frekans kodlaması örnek olarak verilmiştir.

Tablo 3.5 RFO parametresi için frekans kodlaması

RFO	Değerin veri içerisindeki frekansı
CORRECT TOOL / DESIGN	233
CORRECT TOOL / MFG	188
CORRECT TOOL / PLAN	36
CORRECT TOOL / VENDOR	8
ENG. REL.	845
MAINTENANCE	231
MFG.ENG.REQ.	641
P/N REVISION LEVEL CHANGE	20
PART NO CHANGE	120
PI	188
PROD.IMPVT	254
QUALITY	130
RATE TOOL	8
RATE TOOLS	92
REPL LOST TOOL	24
REPLACE LOST TOOL	10
REPLACE LOST TOOL DETAIL	11
TOOL LABEL CHANGE	33

3.2.7 Ortalama Kodlama

Ortalama kodlama frekans kodlama değerlerinin toplam veri sayısına bölünmesi ile oluşturulur. Değerler “0” ile “1” arasındaki pozitif reel sayılardır. (Tablo 3.6)

Tablo 3.6 RFO parametresi için ortalama kodlaması

RFO	Değerin veri içerisindeki bulunma oranı
CORRECT TOOL / DESIGN	0.0758
CORRECT TOOL / MFG	0.0612
CORRECT TOOL / PLAN	0.0117
CORRECT TOOL / VENDOR	0.0026
ENG. REL.	0.2751
MAINTENANCE	0.0752
MFG.ENG.REQ.	0.2087
P/N REVISION LEVEL CHANGE	0.0065
PART NO CHANGE	0.0391
PI	0.0612
PROD.IMPVT	0.0827
QUALITY	0.0423
RATE TOOL	0.0026
RATE TOOLS	0.0299
REPL LOST TOOL	0.0078
REPLACE LOST TOOL	0.0033
REPLACE LOST TOOL DETAIL	0.0036
TOOL LABEL CHANGE	0.0107

3.3. En İyi Kodlama Yönteminin Seçilmesi

En başarılı kodlama yöntemini seçebilmek için doğrusal regresyon, karar ağacı, Destek Vektör Makinesi (DVM) ve YSA'dan oluşan dört kategoride makine öğrenmesi deneyleri yapılmıştır. İlgili deneyler için Matlab⁴ yazılımı kullanılmış olup veriler üzerinde hiçbir değişiklik yapılmadan makine öğrenmesi tekniklerine verilmiştir. Kullanılan öğrenme modelleri ile özet bilgi aşağıda verilmiştir:

⁴ MathWorks tarafından geliştirilmiş dördüncü nesil programlama diline sahip çok amaçlı bir sayısal hesaplama yazılımıdır

3.3.1 Doğrusal Regresyon

Bu model, bir bağımlı değişken ve bir veya daha fazla bağımsız değişken X arasındaki doğrusal ilişkiyi tanımlayan modeldir [26]. Öğrenme modelleri arasında en basit olan bu yöntem diğer istatistikî makine öğrenme yöntemlerine referans teşkil etmesi için seçilen model şu şekilde tanımlanır;

$$y_i = \beta_0 + \sum_{k=1}^K \beta_k f_k(X_{i1}, X_{i2}, \dots, X_{ip}) + \varepsilon_i, \quad i = 1, \dots, n, \quad (3.1)$$

y_i : i . sonuç,

β_k : k . katsayı,

β_0 : modelin sabit katsayısı,

$f(\cdot)$: bağımsız değişkenin skalar değerli fonksiyonu,

X_{ij} : $j = 1, \dots, p$ iken girdi verisi içerisindeki i . Gözlemi,

ε_i : rastgele hata gürültü terimi

3.3.2 Karar Ağacı

Genellikle sınıflandırma için kullanılan bir makine öğrenmesi türüdür. Regresyon için kullanılması ilk defa Breiman, L. ve ark. tarafından sunulmuştur [27].

Bu çalışmada ikili olarak dallanan bir karar ağacı yapısı kullanılmıştır. Kullanılan karar ağacının parametreleri Tablo 3.7'de belirtilmiştir

Tablo 3.7 Karar ağacı Matlab girdi parametreleri

PARAMETRE ADI	PARAMETRE DEĞERİ
RegressionTree.ModelParameters.SplitCriterion	'mse'
RegressionTree.ModelParameters.MinParent	24
RegressionTree.ModelParameters.MinLeaf	12
RegressionTree.ModelParameters.NVarToSample	'all'
RegressionTree.ModelParameters.MergeLeaves	'on'
RegressionTree.ModelParameters.Prune	'on'
RegressionTree.ModelParameters.PruneCriterion	'mse'
RegressionTree.ModelParameters.QEToler	1.00E-06
RegressionTree.ModelParameters.NSurrogate	'off'
RegressionTree.ModelParameters.MaxCat	10
RegressionTree.ModelParameters.AlgCat	'auto'
RegressionTree.ModelParameters.PredictorSelection	'allsplits'
RegressionTree.ModelParameters.UseChisqTest	1

3.3.3 Destek Vektör Makinesi

DVM regresyon modeli düşük ila orta boyutlu girdiler kümesinde kullanılabilen bir yöntemdir. DVM deneyinde Fan, R ve ark. tarafından geliştirilen Sequential Minimal Optimization (SMO) algoritması kullanılmıştır [28]. Tablo 3.8’de oluşturulan kübik DVM modelinin detay parametreleri sunulmuştur.

Tablo 3.8 DVM girdi parametreleri

PARAMETRE ADI	PARAMETRE DEĞERİ
RegressionSVM.ModelParameters.Solver	'SMO'
RegressionSVM.CacheInfo.Size	1000
RegressionSVM.CacheInfo.Algorithm	'Quene'
RegressionSVM.ModelParameters.ClipAlphas	1
RegressionSVM.ModelParameters.DeltaGradientTolerance	0
RegressionSVM.ModelParameters.Epsilon	0.861989191
RegressionSVM.ModelParameters.GapTolerance	1.00E-03
RegressionSVM.ModelParameters.KKTTolerance	0
RegressionSVM.ModelParameters.IterationLimit	1.00E+06
RegressionSVM.ModelParameters.KernelFunction	'polynomial'
RegressionSVM.ModelParameters.KernelScale	'auto'
RegressionSVM.ModelParameters.KernelOffset	0
RegressionSVM.ModelParameters.KernelPolynomialOrder	3 (kübik)
RegressionSVM.ModelParameters.NumPrint	0
RegressionSVM.ModelParameters.StandardizeData	1

3.3.4 Yapay Sinir Ağı

D.F. Specht tarafından önerilen; girdi parametreleri ve çıktı arasında doğrusal olmayan bir ilişkinin oluşturulmasında kullanılan Generalized Regression Neural Network (GRNN) modeli genel olarak aşağıdaki şekilde tanımlanabilir [29];

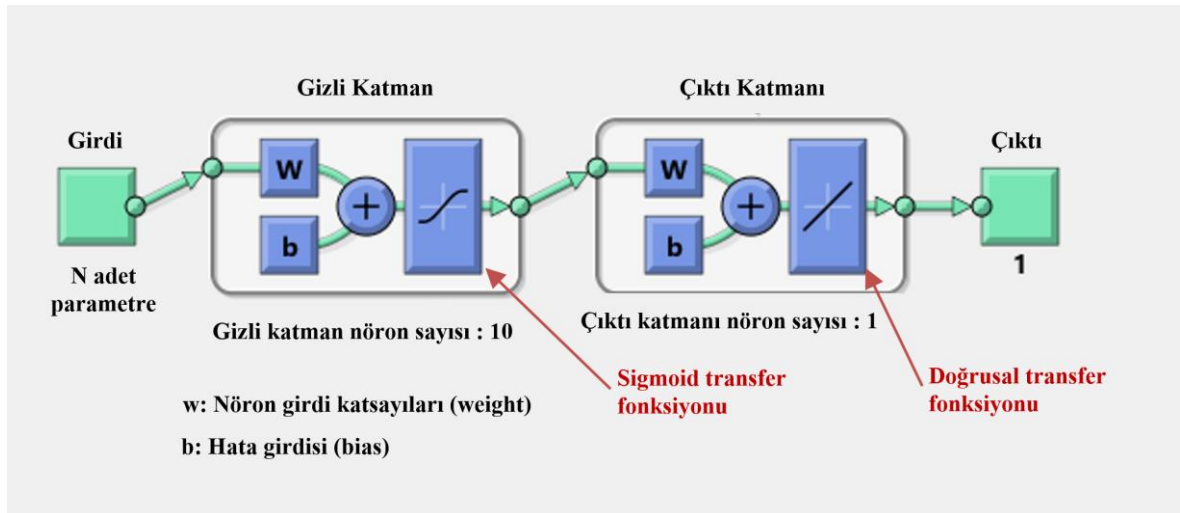
$$Y(x) = \frac{\sum_{k=1}^N y_k K(x, x_k)}{\sum_{k=1}^N K(x, x_k)} \quad (3.2)$$

$Y(x)$: x girdisinin tahmin değeri

y_k : k nöron katmanının aktivasyon katsayısı

$K(x, x_k)$: Radyal Tabanlı Fonksiyon (RBF) çekirdeği;

Deney için oluşturulan YSA Şekil 3.1’de verilmiştir. 10 adet gizli katman sigmoid transfer fonksiyonu kullanılarak tanımlanmıştır. Çıktı katmanında ise 1 nörondan oluşan doğrusal aktivasyonu olan bir katman vardır.



Şekil 3.1 Deney için kullanılan YSA'nın genel görünümü

Detaylı parametre bilgileri Tablo 3.9’da belirtilen YSA’nın kullanımında Levenberg ve Kenneth tarafından sunulmuş Damped Least-Squares olarak ta bilinen Levenberg–Marquardt algoritması kullanılmıştır [30].

Tablo 3.9 YSA deneyi parametreleri

PARAMETRE ADI	PARAMETRE DEĞERİ
net.trainParam.epochs	1000
net.trainParam.goal	0
net.trainParam.max_fail	6
net.trainParam.min_grad	0.0000001
net.trainParam.mu	0.001
net.trainParam.mu_dec	0.1
net.trainParam.mu_inc	10
net.trainParam.mu_max	10000000000

Yukarıda anlatılan makine öğrenme yöntemleri ile yapılan deneylerde ortak olarak 10 katlı çapraz doğrulama yapılmıştır. Kullanılan her yöntem için RMSE değerleri hesaplanarak Tablo 3.10’da özetlenmiştir:

Tablo 3.10 Kodlama metodu ve makine öğrenme modellerinin RMSE sonuçları

Kodlama Metotları	RMSE Değerleri				
	Doğrusal Regression	Karar Ağacı	Destek Vektör Makinesi	Yapay Sinir Ağı	Yöntemlerin Ortalaması
Sıralı	82.77	80.22	76.67	76.52	79.89
İkili	72.00	73.74	72.71	63.98	72.82
One Hot	65.81	63.26	55.51	68.03	61.53
Dummy	64.31	63.05	55.29	67.12	60.88
Efekt	71.90	70.01	73.04	68.18	71.65
Frekans	77.76	72.06	74.67	66.27	74.83
Ortalama	91.53	67.65	77.39	88.26	78.86

Yapılan deneyin sonucunda dummy kodlama doğrusal regresyon, karar ağacı ve DVM’de en iyi sonucu verirken YSA ise ikili kodlama ile daha düşük bir RMSE değerine ulaşılabilmiştir.

Yine Tablo 3.10’da makine öğrenme yöntemlerinin ortalamasına bakıldığında; dummy kodlama iyi sonuç vermiştir. Dummy kodlamanın, one hot kodlama arasındaki yapısal fark; dummy kodlamanın one hot kodlamaya göre N-1 parametreye sahip olması nedeni ile one hot kodlama da gözden geçirildiğinde birbirleri arasında %5 den az fark olduğu gözlenmektedir. Bahsedilen iki kodlama türü arasında düşürülen parametrenin korelasyona gürültü ekleyebileceği düşüncesini de göz önüne alarak ileride sunulacak olan makine öğrenme mimarisinin ve parametrelerinin oluşturulması sırasında belirtilen durumun da göz önünde bulundurulması daha doğru bir sonuca ulaşmakta faydalı görünmektedir.

Bu bölümde sunulan sonuçlar göz önünde bulunarak çalışmanın bundan sonraki aşamalarında tüm makine öğrenme yöntemlerinde, ortalamada en düşük sonucu veren dummy kodlama ile çalışma ilerletilmiştir.

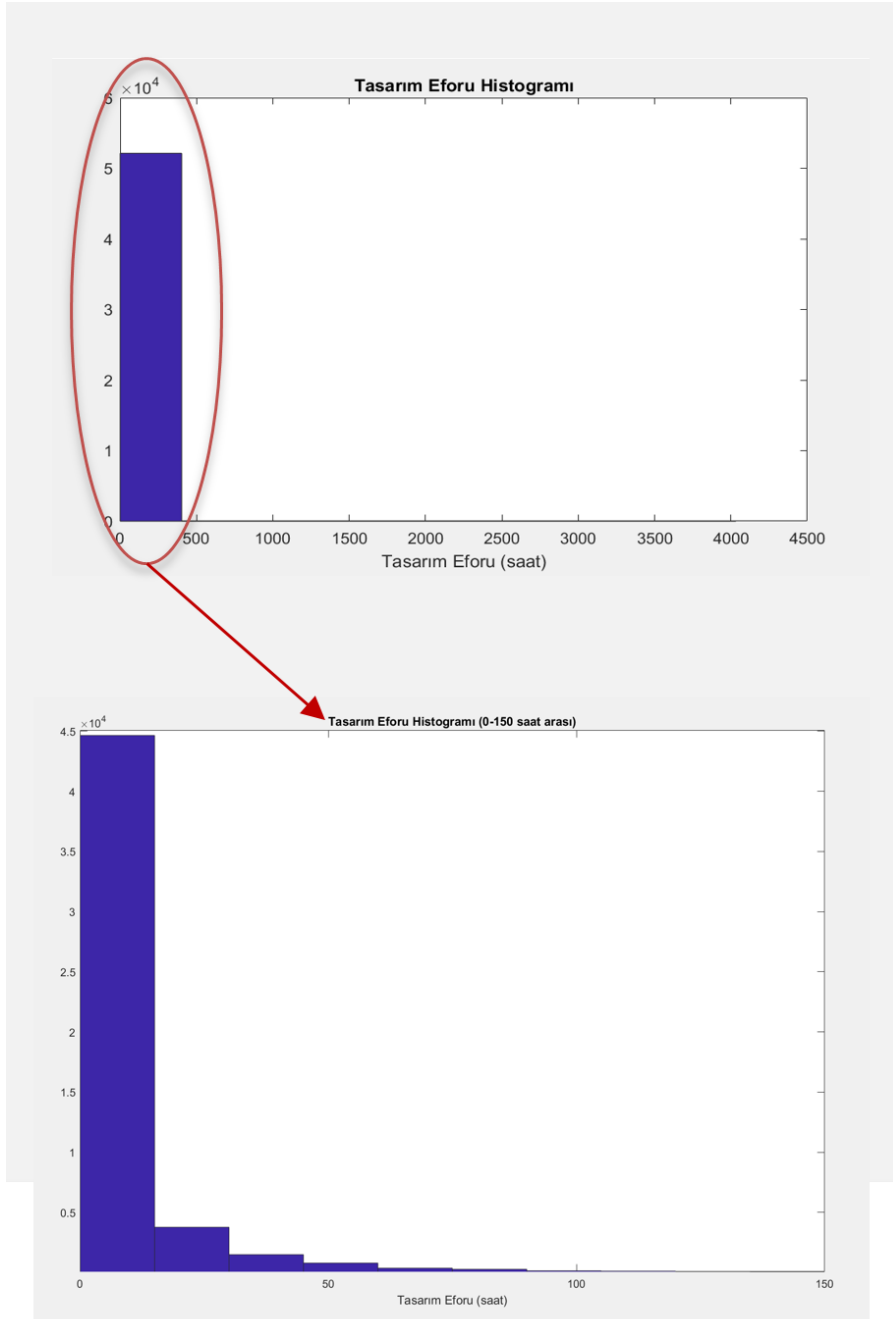
3.4. Verilerin İncelenmesi

52213 örneklem boyutuna sahip veride çıktı parametresinin ortalaması 10.62 olarak hesaplanmıştır. Verinin içerisindeki tasarım eforu saati en düşük; 0 ve en yüksek; 4031.14 saat olarak görülmüştür. Çeyreklik bilgileri Tablo 3.11’de verilen verinin tel örneklem Kolmogorov-Smirnov normallik testi sonucunda $2.2e-16$ p değeri ile %95 güven aralığında normal dağılım göstermediği görülmüştür.

Tablo 3.11 Takım tasarım eforu çeyreklik bilgileri

0%	25%	50%	75%	100%
0.05	0.8	2.2	8	4031.14

Ek olarak makine öğrenmesi ile tahmin edilecek sonuçlar Şekil 3.2’deki dağılım grafiğinde gösterilmiştir. Ayrıca ilgili verinin %95’inin 0-50 saat arasında toplandığı hesaplanmıştır.



Şekil 3.2 Takım tasarımı için harcanan efor saatinin dağılım grafiği (0-4500 saat ve 0-150 saat arası)

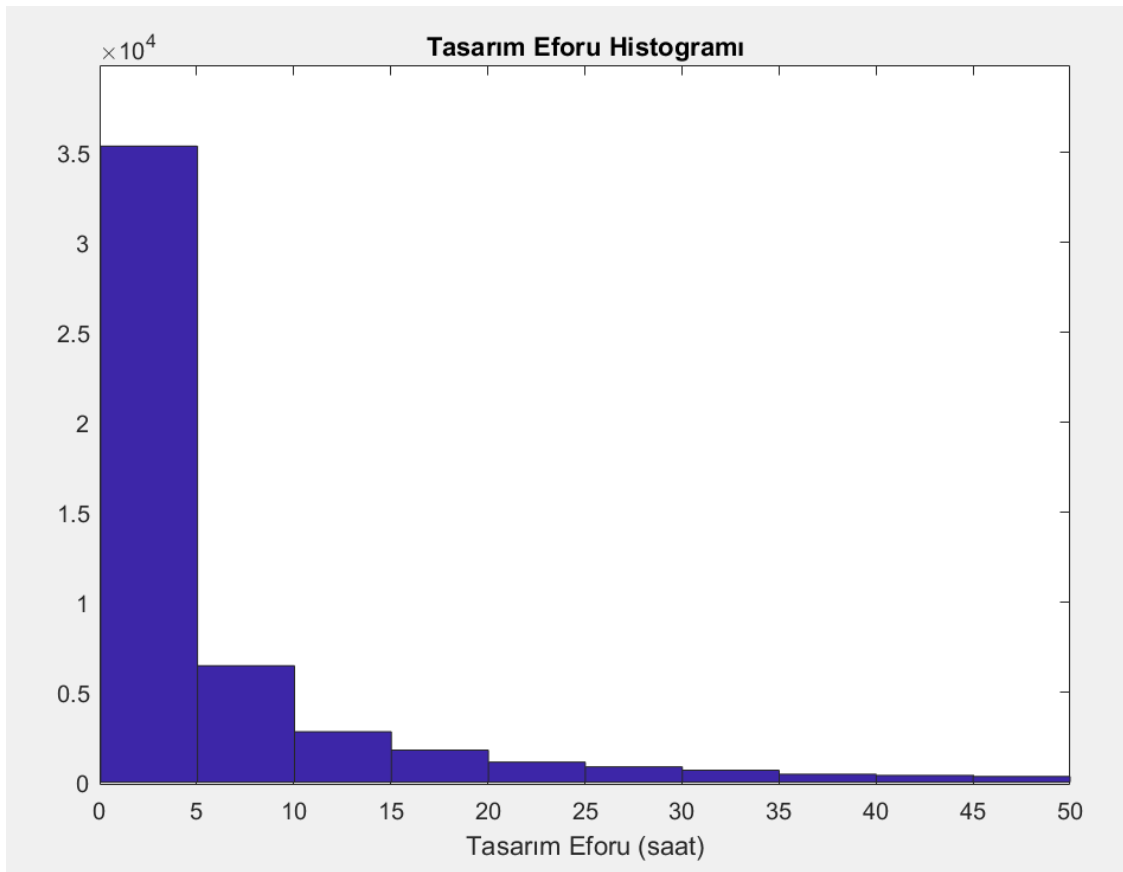
3.5. Veri Ön İşlemesi ve Makine Öğrenmesi

Bu bölümde verilerin çeşitli yöntemler ile işlenerek makine öğrenmesindeki doğru kestirim performansının artırılması için yapılan çalışmalar sunulacaktır.

Aşağıda sunulacak olan çalışmalarda yapılan her bir makine öğrenme yöntemi değerlendirilirken verideki var olabilecek hatalar göz ardı edilerek her bir çalışmanın girdi verisi kendi içerisinde “doğru” olarak kabul edilmiştir.

3.5.1 Uç Değerlerden Arındırılmış Veri ile Makine Öğrenmesi

Önceki bölümde tasarım eforu saati üzerinde yapılan inceleme sonucunda, öncelikle veri, yaklaşık %5’ini oluşturan uç değerlerden arındırılmıştır. 50 saat ve üzerine denk gelen bu %5’lik veri ana veriden çıkartıldığında Şekil 3.3’de de sunulan dağılım gözlemlenmiştir.



Şekil 3.3 Takım tasarımı için harcanan efor saatinin uç değerler çıkarıldıktan sonraki dağılım grafiği

Düzenlenen bu veri üzerinde, dummy kodlama kullanarak 10 katlı çapraz doğrulama ile makine öğrenmesi yapıldığında ise Tablo 3.12’deki sonuçlar gözlemlenmiştir. Genel olarak, uç değer elemesi işleminin hata değerlerinin düşmesini sağladığı gözlemlenmiştir.

Tablo 3.12 Uç değerlerden arındırılmış veri üzerinde makine öğrenmesi uygulaması sonuçları.

	SONUÇLAR	
	RMSE	R
Doğrusal Regression	45.3285	0.2646
Karar Ağacı	41.4070	0.4583
Destek Vektör Makinesi	44.7683	0.3464
Yapay Sinir Ağı	51.4925	0.5740

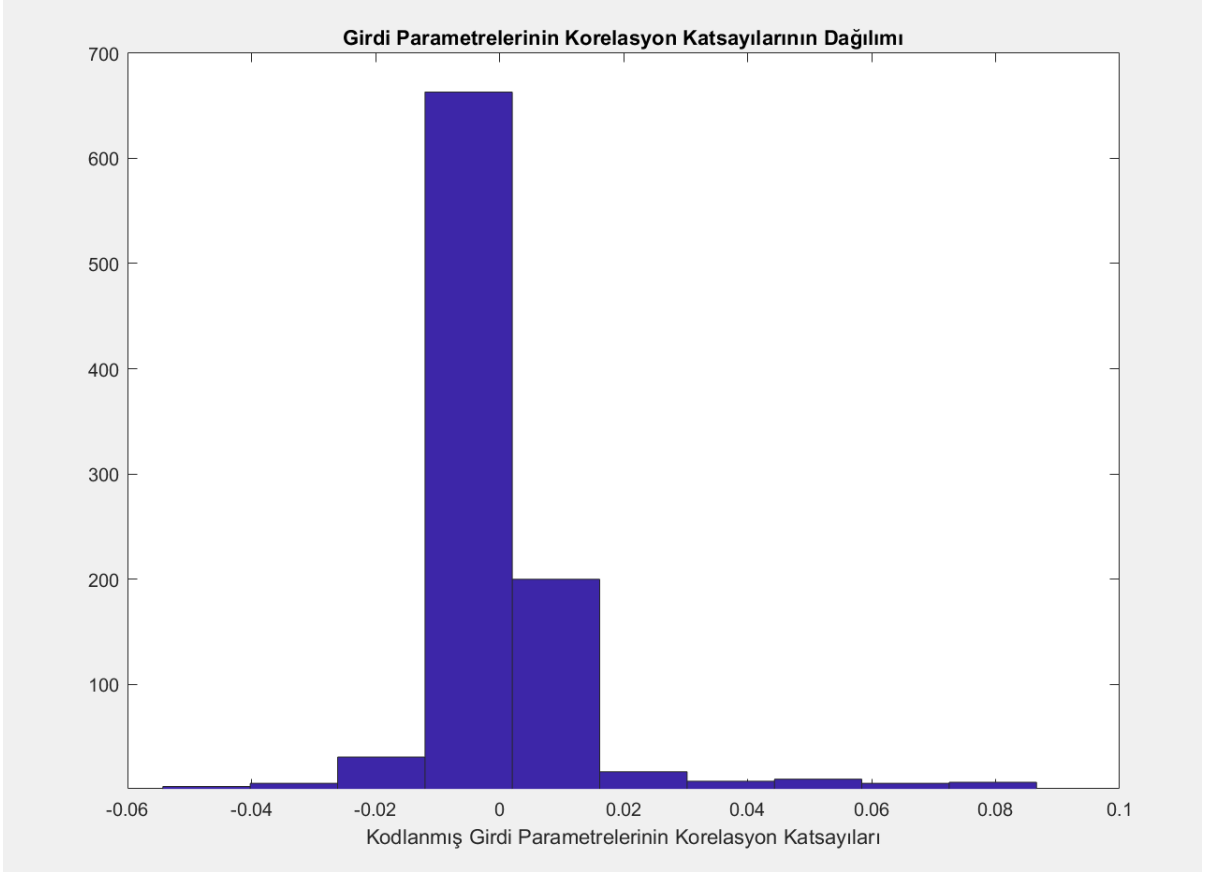
Ayrıca yine Tablo 3.12’de hesaplanan R değerleri incelendiğinde; Doğrusal Regresyon ve DVM’de girdi ile sonuçlar arasında zayıf bir ilişkiyi işaret etmektedir. Diğer yandan YSA ve Karar Ağacı’nın sırasıyla 0.574 ve 0.4583 olan R değeri ile orta şiddette bir ilişki göstermiştir. Bununla beraber en yüksek RMSE’yi veren fakat en yüksek R değerini veren YSA ile öğrenme yöntemi de gözden kaçırılmamalıdır.

Bu deneylerin sonucunda tasarım eforu gerçekleşme saatlerini kestirmede, girdi parametrelerinin doğrudan kullanımının yeterli olmadığı ortaya çıkmıştır. Bu nedenle, bir sonraki adımda girdi parametreleri incelenecek ve çeşitli düzenlemeler gerçekleştirilerek sonuçlar karşılaştırılacaktır.

3.5.2. Girdi Parametrelerinin Korelasyona Katkısına Göre Düzenlenmesi

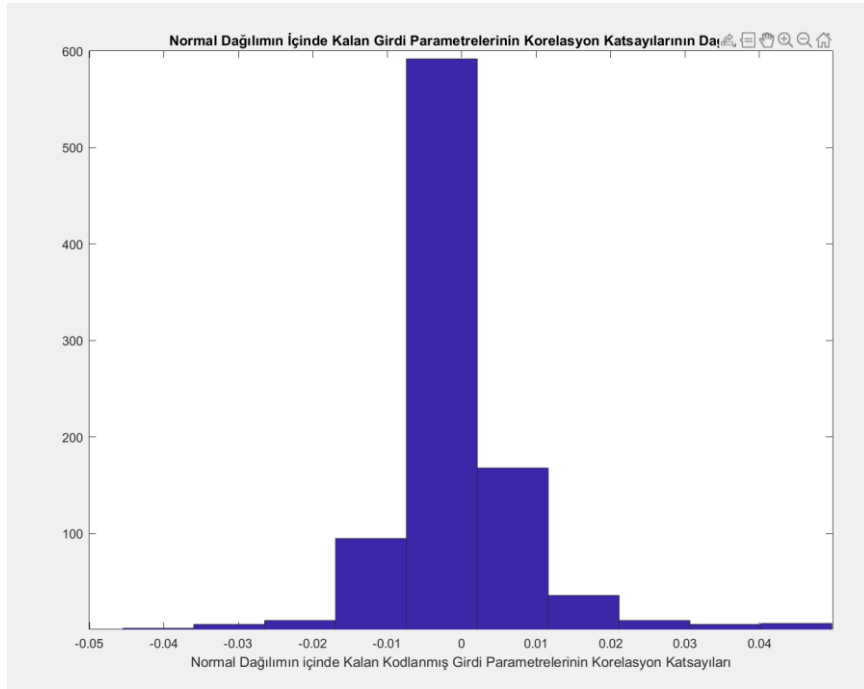
Bu bölümde dummy kodlama ile elde edilmiş parametrelerin çıktıya katkısını değerlendirebilmek için Pearson Korelasyonu kullanılarak her bir parametrenin sonuca etkisi değerlendirilmiştir.

Şekil 3.4’de dummy kodlama yöntemi ile kodlanmış ikili parametrelerin Pearson R değeri dağılımı verilmiş olup bu dağılım, veri içerisinde bulunan ve -0.05 ile 0.05 aralığında bulunan dummy kodlanmış girdi parametrelerinin girdi verisinden çıkarılması ile elde edilmiştir.

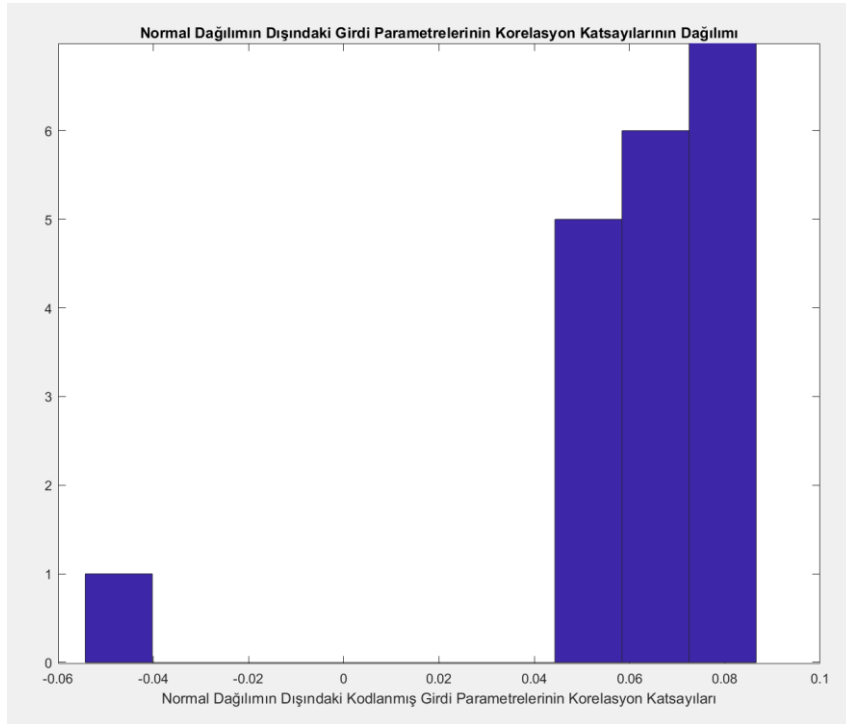


Şekil 3.4 Girdi Parametrelerinin Korelasyon Katsayılarının Dağılım Grafiği

Şekil 3.5’de ise Pearson R değerlerinin normal dağılım içerisinde olanları gösterilmiştir. Şekil 3.4 ve Şekil 3.5 arasındaki fark ise Şekil 3.6’da gösterilmiş olan veri parametrelerinin Pearson R değerlerinin normal dağılım göstermeyen katsayılarının dağılımıdır. Düşük korelasyon katsayısı, bu katsayıya karşılık gelen girdi parametresi ile çıktı parametresi arasındaki zayıf ilişkiyi ifade ettiğinden, bu katsayıları analize dahil etmek tahmin sırasında sapma oranını arttıracaktır. Bu nedenle, tahmin sonucunda iyileşme sağlamak amacı ile bu katsayılara karşılık gelen girdi parametreleri çıkarılarak, yalnızca Şekil 3.6’da gösterilen yüksek R değerine sahip katsayıların sahibi olan parametreler kullanılacaktır.



Şekil 3.5 Normal Dağılımın İçinde Kalan Girdi Parametrelerinin Korelasyon Katsayılarının Dağılım Grafiği



Şekil 3.6 Normal Dağılımın Dışındaki Girdi Parametrelerinin Korelasyon Katsayılarının Dağılım Grafiği

Yukarıda anlatılan yöntemle belirlenen parametreler üzerinde aynı makine öğrenme yöntemleri denendiğinde Tablo 3.13'deki sonuçlara ulaşılmıştır. Bu yöntem ile bulunan sonuçlar Tablo 3.12'de bulunan sonuçlar ile karşılaştırıldığında hata miktarında bir miktar azalma görülmesine karşın R değeri daha da düşmüştür.

RMSE değerindeki iyileşmeye karşın, R değerindeki düşüş sonuçların bir önceki yönteme göre daha da tesadüfi olduğunu düşündürmektedir. Bu nedenle, farklı girdi verisi düzenleme yöntemleri ile deneyler yapılmasının gerektiği sonucuna varılmıştır.

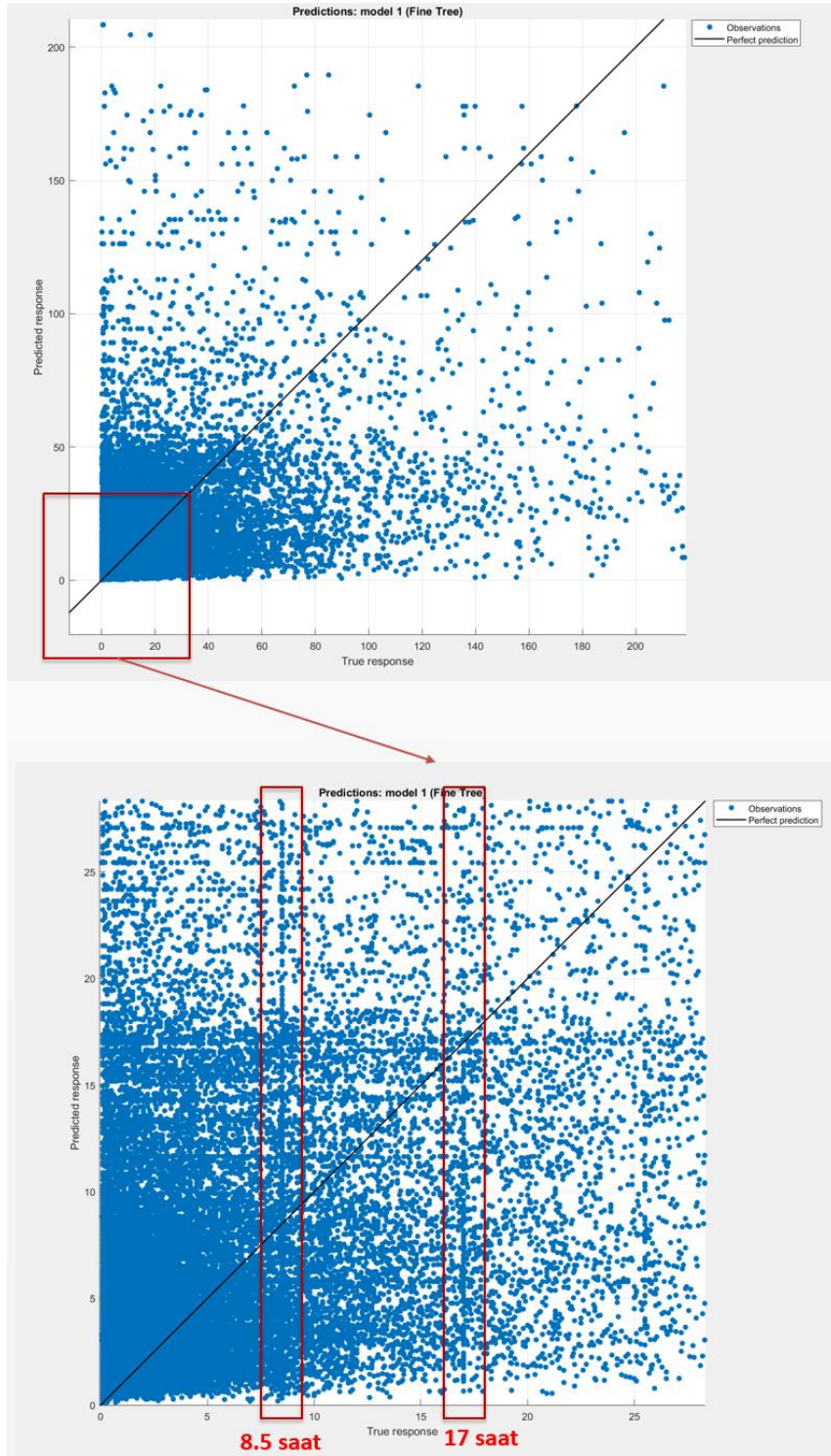
Tablo 3.13 Girdi parametrelerinin korelasyona katkısına göre düzenlenmesi sonucundaki makine öğrenme performansı

	SONUÇLAR	
	RMSE	R
Doğrusal Regression	37.0740	0.2236
Karar Ağacı	36.3043	0.3162
Destek Vektör Makinesi	37.1847	0.2000
Yapay Sinir Ağı	49.2679	0.3165

3.5.3 Girdi Parametrelerinin Uzman Görüşü Alınarak Yeniden Değerlendirilmesi

Girdi parametrelerinin incelenerek istatistiksel yöntemlerle yapılan çalışmaların sonucu başarılı sayılabilecek anlamlı bir sonuç verememesi nedeniyle yukarıdaki alt başlıklarda sunulan her iki deneyde de en iyi sonucu veren karar ağacı verileri yeniden değerlendirilmiştir.

Uç değerlerden arındırılmış veri ile yapılan çalışmada karar ağacı ile eğitilmiş modelin hesapladığı efor süreleri ile gerçekleşen efor süreleri Şekil 3.7'de karşılaştırıldığında Tablo 3.12'deki hesaplanan R değeri ile de örtüşen tesadüfi sonuçlar çıktığı gözlemlenmiştir.



Şekil 3.7 Uç değerlerden arındırılmış veriler ile yapılan çalışmadaki karar ağacı yönteminin tahmin ve gerçekleştirmeleri grafiği (üstte), üstteki grafiğin 0 - 25 saat aralığının büyütülmüş hali (altta)

İlgili grafiğin 0 – 25 saat ölçeği yakından incelendiğinde yatay ekseninde tahmin edilen değerlere paralel olarak iki alan göze çarpmaktadır. (Şekil 3.7, altta). 8.5 ve 17 değerine sahip bu iki alan veri setinin %2.022'sine denk gelmekte olup TUSAŞ takım tasarım mühendisinin bir ve iki günlük çalışma süresine denk gelmektedir. Bu durumun çalışanın işe başlarken çalışma kaydı oluşturup gün içinde başka bir işe geçtiğinde yeni çalışma kaydını oluşturmayı unutmasından kaynaklanabileceği düşünülerek ilgili veriler eğitim verisinden çıkarılmıştır.

Ayrıca ilgili girdi verisinde aynı girdi parametrelerine ait fakat farklı gerçekleşme sürelerine sahip veriler olduğu fark edilmiştir. Veride belirtilen bu tutarsızlığın önüne geçmek için ortak değere sahip verilerin gerçekleşen efor saatleri için ortalama değer alınmıştır. Ortalamaların belirlenme sürecinde ise bazı verilerin tasarım mühendisliği uzmanlığına göre anlamsız olarak yorumlanan değerler de veriden çıkartılmıştır. Örnek olarak; tasarım işlemi olmayan, shop aid⁵ olarak tanımlanan takımlar için hatalı efor girişleri gözlemlenmiş ve bu değerler veri içerisinden çıkartılmıştır. İşlem yapılan bu verilerin toplam verinin % 93.795'ini içerdiği görülmüştür.

Yapılan bu gözlemler sonucunda girdi verilerinin tesadüfi olması nedenlerinin bir kısmı da netlik kazanmıştır.

Temizlenmiş veri üzerinde yine 10 katmanlı çapraz doğrulama ile aynı makine öğrenme yöntemleri denenerek Tablo 3.14'deki sonuçlara ulaşılmıştır. Yapılan diğer iki deneyden farklı olarak YSA ile öğrenmenin diğerlerinden daha başarılı olduğu gözlemlenmiştir. Ayrıca R değerinin de artmış olması, sonuçların tesadüfi olma durumunu azaltmıştır.

⁵ Üretim sürecinde bir işe ya da işleme yardımcı olan üretim alanlarının kendi ihtiyaçları ve inisiyatifleri doğrultusunda hazırlanan tasarımı ve konfigürasyon kaydı olmayan aparatlara verilen ortak isimdir.

Tablo 3.14 Girdi ve çıktı parametrelerin uzman görüşü ile temizlenmesi sonucunda yapılan makine öğrenmesi performansı sonuçları

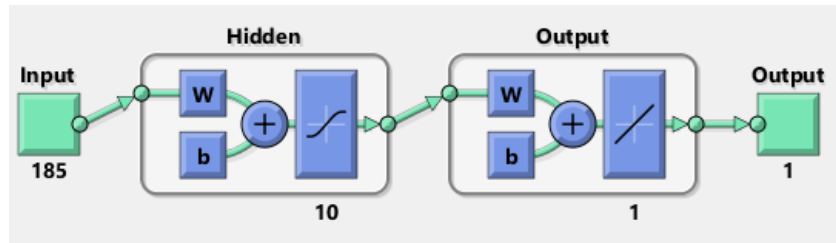
	SONUÇLAR	
	RMSE	R
Doğrusal Regression	8.7635	0.4416
Karar Ağacı	8.8633	0.4796
Destek Vektör Makinesi	9.0507	0.3894
Yapay Sinir Ağı	8.5359	0.4819

3.6. Gerçekleşen Çıktı Değerlerinin İncelenmesi

Bir önceki bölümde yapılan incelemeler ve düzenlemeler kodlama yönteminin ve makine öğrenmesi yönteminin belirlenmesi üzerinedir. Bu noktadan sonra ise en iyi sonucu veren YSA yönteminin daha iyi sonuçları verebilmesi için yapılabilecek iyileştirme çalışmalarına yer verilecektir. Bunun yanında, üzerinde çalışılmış olan kodlama yöntemleri YSA üzerinde yeniden denenerek, YSA'ya en uygun kodlama yöntemi üzerinde yeniden deneyler yapılacaktır.

3.7. YSA Optimizasyonu

En İyi Kodlama Yönteminin Seçilmesi başlığı altında tanıtılan YSA mimarisi Tablo 3.9 ve Şekil 3.1 Deney için kullanılan YSA'nın genel görünümü Şekil 3.1'de tanıtılmıştır. Şekil 3.8'de ise YSA'nın en uygun kodlama yöntemine göre oluşturulmuş ağ gösterilmiştir. Bu ağda Tablo 3.15'de verilen parametreler kullanılmış olup, eğitim algoritması olarak ise yine Levenberg-Marquardt backpropagation seçilmiştir.

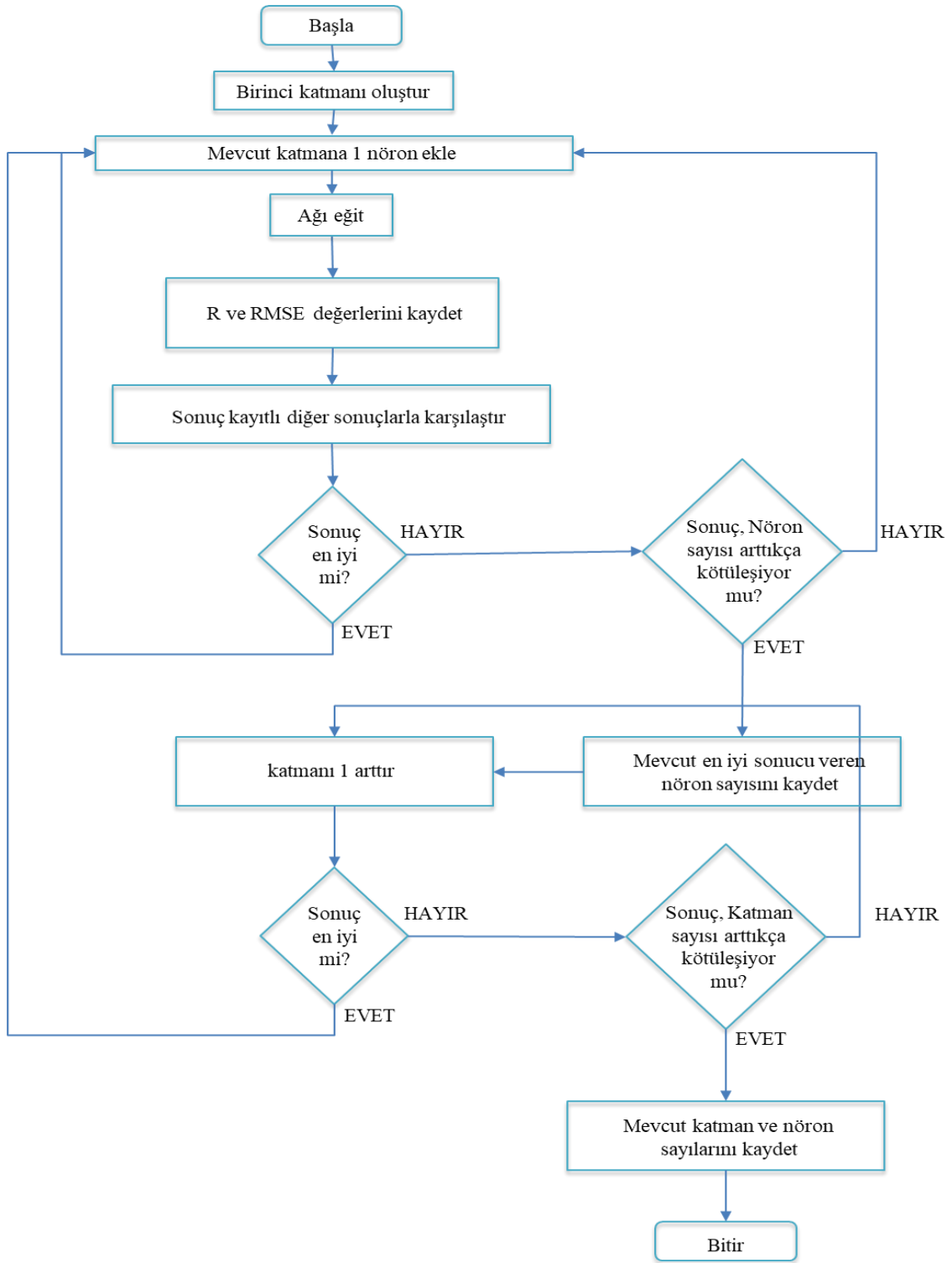


Şekil 3.8 Tasarım eforu kestirimi için kullanılan YSA görseli

Tablo 3.15 Tasarım eforu kestirimi için kullanılan YSA'nın başlangıç parametreleri

PARAMETRE TANIMI	DEĞER
Eğitim İçin Maksimum Deneme	1000
Performans Hedefi	0
İzin verilebilir en fazla doğrulama sayısı	6
Eğitimin durması için gereken en düşük performans artış oranı	1.00E-07
İlk μ değeri	0.001
μ azalma faktörü	0.1
μ arttırma faktörü	10
Kabul edilebilir en fazla μ değeri	1.00E+10

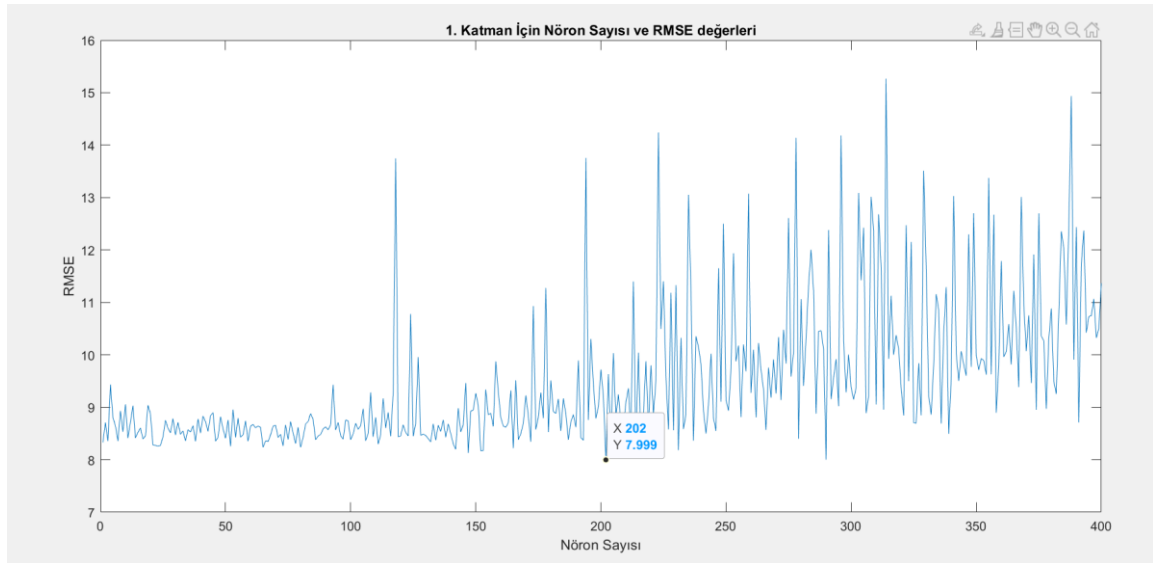
Uzman görüşü ile ayıklanmış veri üzerinde yukarıda belirtilen mimari ile yapılmış çalışmanın sonucu Tablo 3.14'de zaten hesaplanmıştır. Bu noktada RMSE değerini en aza indirmek aynı zamanda ise R değerini de yükseltmek için Şekil 3.9'de verilen algoritma kullanılmıştır.



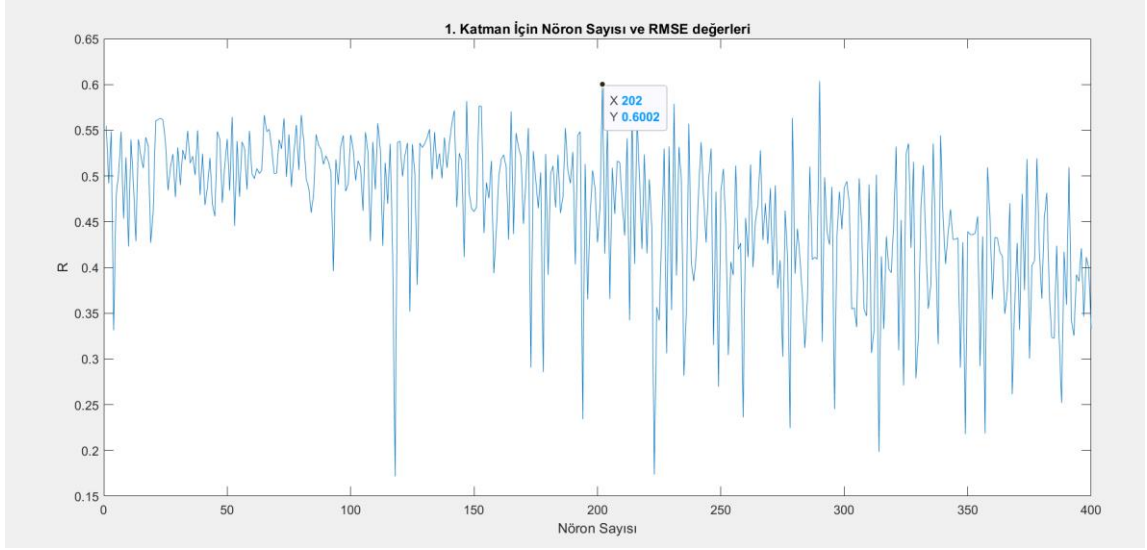
Şekil 3.9 YSA en uygun şekle sokma algoritması akışı

Bu algoritmada öncelikle mevcut katmana bir nöron ekleyerek süreç başlatılır. Tüm nöron sayıları için sistem eğitilerek en iyi RMSE ve R değeri belirlenir. Eğer mevcut katmandaki nöron en iyi değeri verdi ise sisteme bir katman daha eklenerek aynı süreç tekrarlanır. Bu noktada algoritmanın durma koşulu en iyi sonucu veren katman sayısı olacaktır. Ayrıca katman sayısı arttırıldığında kendisinden bir önceki katmanın nöron sayısı sabit tutulur.

Tanımlanan algoritmayı örneklendirebilmek için; Şekil 3.11 ve Şekil 3.10’da verilen bir katmanlı olarak eğitilmiş veride en iyi RMSE ve R değeri için ideal nöron sayısının 202 olduğu gözlemlenmiştir.

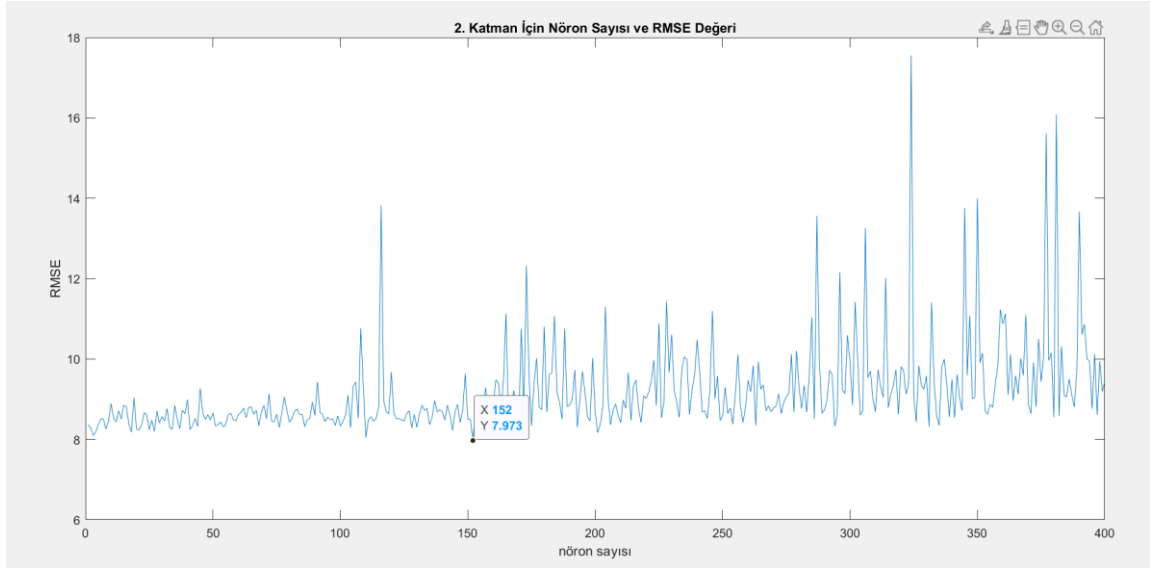


Şekil 3.10 Bir katmanlı dummy kodlama ile eğitilmiş veri setinin RMSE değerinin nöron sayısına göre değişimi

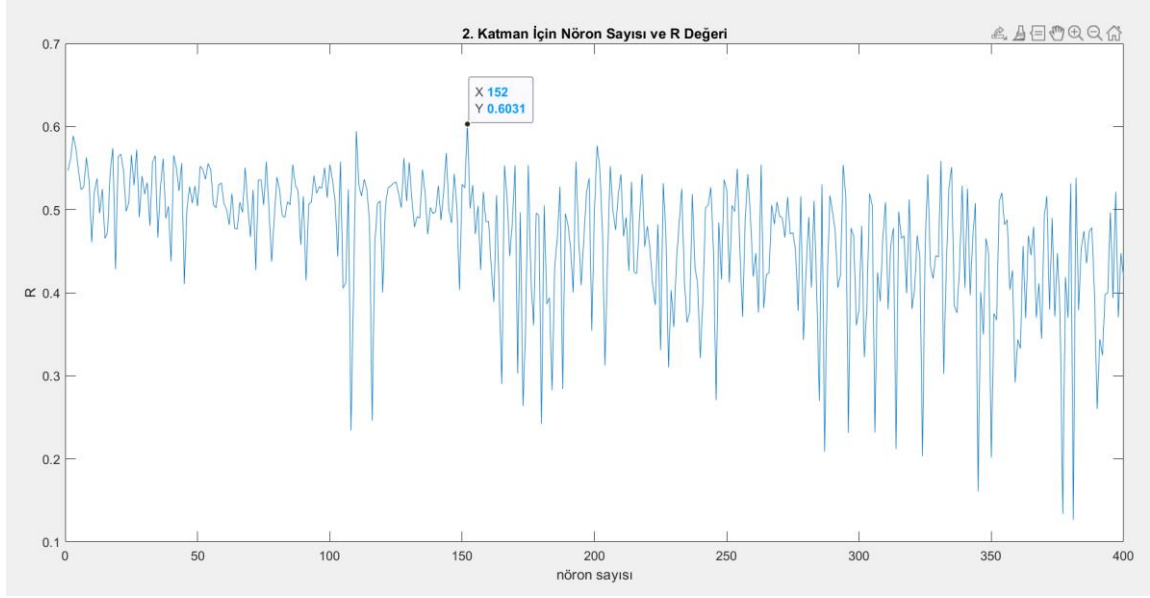


Şekil 3.11 Bir katmanlı dummy kodlama ile eğitilmiş veri setinin R değerinin nöron sayısına göre değişimi

Ardından en iyi sonucu veren 202 nörona sahip ilk katman sabit tutularak veriye bir katman eklenir ve aynı test yeniden yapılır. (Şekil 3.12 ve Şekil 3.13) Bu testin sonucunda ikinci katmanın 152 nörona sahip olması en iyi sonucu vermeye birlikte bir katmanlı ağdan da daha iyi RMSE ve R değerine sahiptir. Bu sonuçlar ışığında üçüncü bir katman eklenerek deney devam ettirilir.



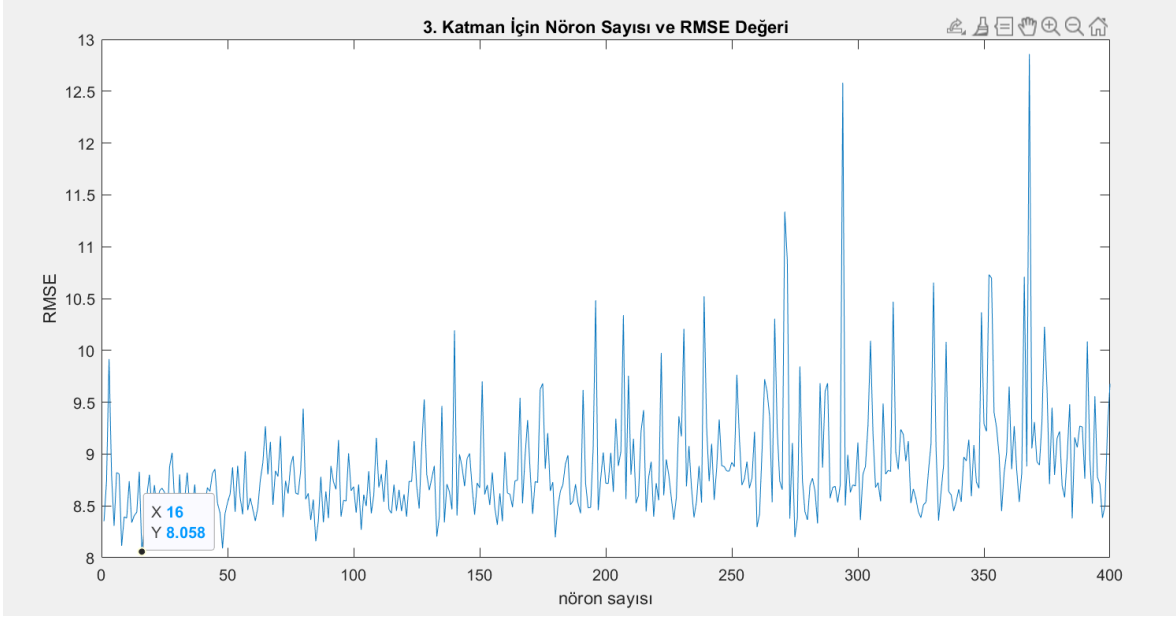
Şekil 3.12 Birinci katmanı 202 nörondan oluşan, ikinci katmanının dummy kodlama ile eğitilmiş veri setinin RMSE değerinin nöron sayısına göre değişimi



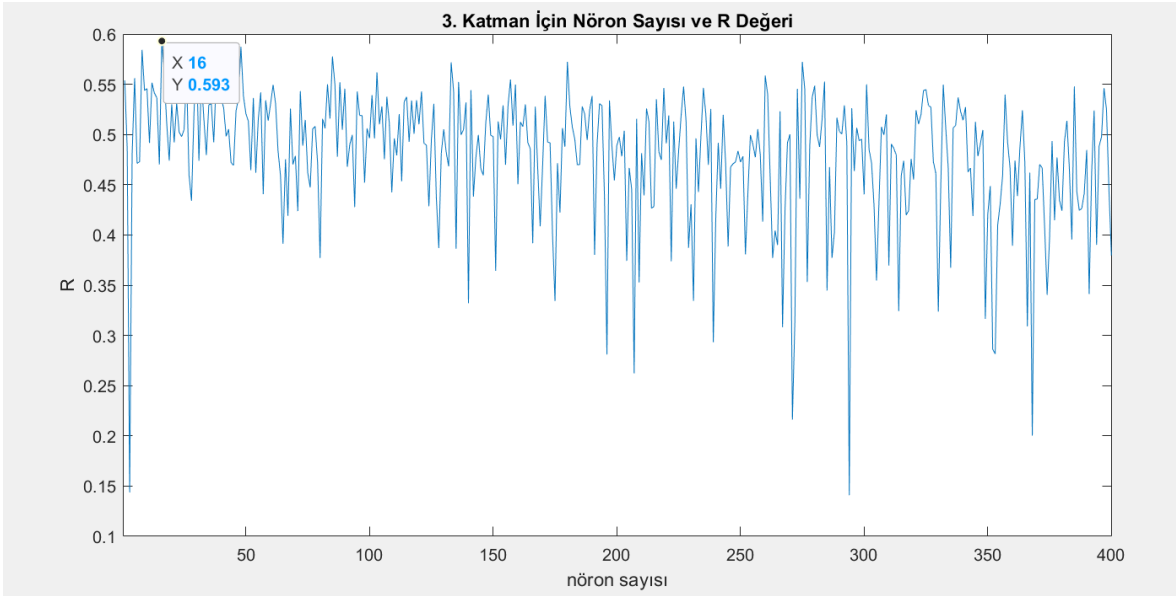
Şekil 3.13 Birinci katmanı 202 nörondan oluşan, ikinci katmanının dummy kodlama ile eğitilmiş veri setinin R değerinin nöron sayısına göre değişimi

İlk iki katmanda iyileşme sağlandığı için üçüncü katmana geçilerek deney devam ettirilir. Üçüncü katmandaki nöronların artması ile yapılan deney ise YSA'nın performansının düştüğü gözlemlendiğinden deney durdurulur. (Şekil 3.14 ve Şekil 3.15) Bu noktada elde edilen ağ Şekil 3.16'de gösterilmiştir.

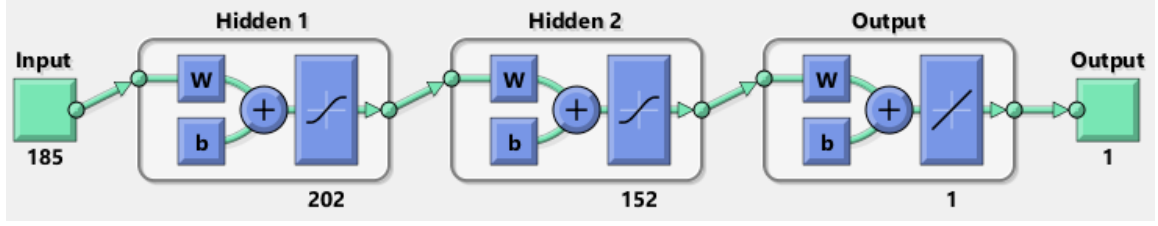
Takım tasarım eforu tahmini için en iyi olarak hesaplanan dummy kodlama ve YSA makine öğrenmesi yöntemi Şekil 3.16'de en iyi hale getirilerek 7.9727 RMSE hata değerinde ve 0.6031 R değeri ile orta düzeyde bir korelasyon ilişkisi kurulmuştur. Bu noktadan sonra makine öğrenme mimarisini en iyi hale getirebilmek için, tüm girdi verisi kodlama yöntemleri YSA için yeniden gözden geçirilerek daha iyi bir sonuç elde edebilmek adına deneyler yapılacaktır.



Şekil 3.14 Birinci katmanı 202 nörondan oluşan ve ikinci katmanı 152 nörondan oluşan, üçüncü katmanının dummy kodlama ile eğitilmiş veri setinin RMSE değerinin nöron sayısına göre değişimi



Şekil 3.15 Birinci katmanı 202 nörondan oluşan ve ikinci katmanı 152 nörondan oluşan, üçüncü katmanının dummy kodlama ile eğitilmiş veri setinin R değerinin nöron sayısına göre değişimi



Şekil 3.16 Deneyler sonucunda en iyi hale getirilmiş YSA

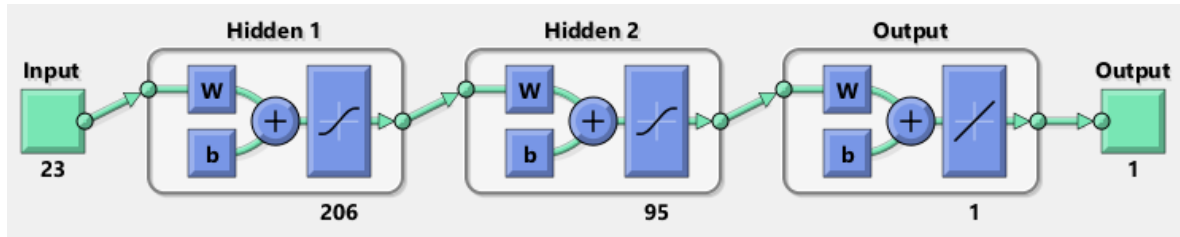
3.8. En Verimli Makine Öğrenme Yöntemi İçin Kodlama Deneyleri ve Sonuçları

Bu bölümde, bir önceki başlıkta hesaplanan en iyi mimari üzerinde kodlama yöntemlerinin yeniden denenerek daha iyi bir sonuca ulaşmanın mümkün olabilirdiği araştırılmıştır. Şekil 3.9’de gösterilen algoritma takip edilerek yapılan çalışmada Tablo 3.16’de verilen sonuçlara ulaşılmıştır.

Bulunan en ideal yöntem Şekil 3.16’da verilen ağ mimarisi üzerinde ikili kodlama ile oluşturulmuştur. Bu yöntem 7.184 RMSE hata değerinde ve 0.7075 R değeri ile yüksek düzeyde bir korelasyon ilişkisi kurmayı başarmıştır.

Tablo 3.16 YSA mimarisi ve kodlama yöntemi en iyi hale getirme deneyleri

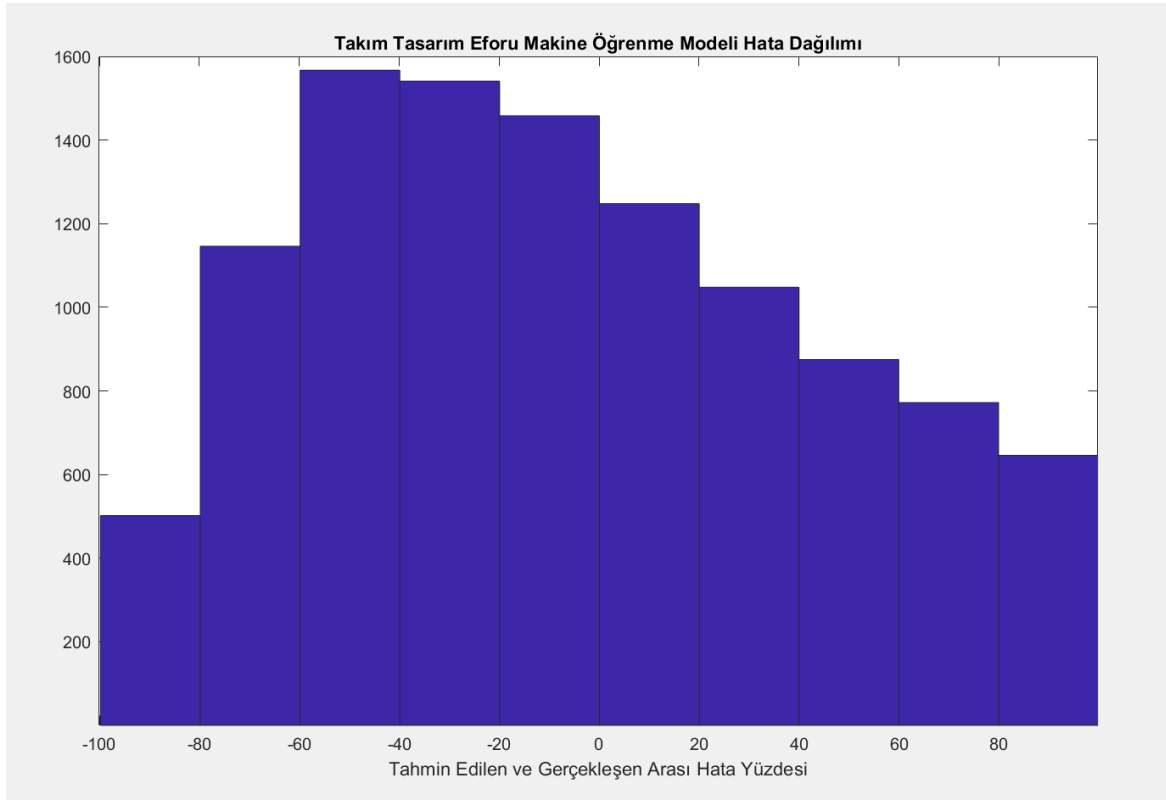
Kodlama Yöntemleri	Birinci Katman			İkinci Katman		
	Nöron Sayısı	RMSE	R	Nöron Sayısı	RMSE	R
Sıralı	337	8.6720	0.5089	35	8.5070	0.5299
İkili	206	7.5229	0.6803	95	7.1840	0.7075
One Hot	4	8.0520	0.5925	11	8.2262	0.5681
Dummy	202	7.9985	0.6002	152	7.9728	0.6031
Efekt / Deviasyon	210	8.0138	0.6004	68	8.1059	0.5850
Frekans	168	8.5431	0.5218	126	8.4545	0.5430
Ortalama	119	8.1697	0.6053	323	7.9741	0.6355
Kodlama Yöntemleri	Üçüncü Katman			Dördüncü Katman		
	Nöron Sayısı	RMSE	R	Nöron Sayısı	RMSE	R
Sıralı	17	8.4930	0.5274	60	8.5343	0.5215
İkili	87	7.3071	0.6863	-	-	-
One Hot	-	-	-	-	-	-
Dummy	16	8.0583	0.5930	-	-	-
Efekt / Deviasyon	-	-	-	-	-	-
Frekans	54	8.3674	0.5477	33	8.4286	0.5377
Ortalama	330	8.2390	0.5982	-	-	-
Deney durdurulmuştur						
En iyi sonucu veren katman ve nöron sayısı						



Şekil 3.17 İkili kodlama için hesaplanan en iyi YSA mimarisi

3.9 Hesaplama Sonuçlarının Değerlendirilmesi

Bir önceki bölümde en iyi kodlama yöntemi ve YSA mimarisinden oluşan model oluşturulmuştur. Bu model uç değerlerden arındırılmış gerçekleşen tasarım eforu değerleri ile karşılaştırıldığında RMSE değeri 10.3404 ve R değeri olarak 0.5933 olarak hesaplanmıştır. Şekil 3.18'de hata dağılımı verilen modelin tahmin gücünün, gerçekleşen değer altında değer hesaplama eğiliminde olduğu gözlemlenmiştir.



Şekil 3.18 Takım Tasarım Eforu Makine Öğrenme Modeli Hata Dağılımı

TUSAŞ Üretim Planlama ve Kontrol (ÜPK) iş zekâsı grubu proje planlarının yürütülmesinden sorumlu olan bölümdür. Bu bölüm tüm proje girdilerini değerlendirerek genel bir iş planı yapar. Bu plan günlük olarak takip edilir ve gün içerisinde ihtiyaçları belirleyip ilgili bölümlere bilgilendirme yapar. Benzer şekilde Takım Mühendislik Bölümü de günlük proje eforu tahminlerini gün içerisinde ÜPK ekiplerine bildirir.

Günlük olarak değerlendirilen bu bilgiler toplanarak üst yönetime sunulur ve ilgili aksiyon planları ilgili bölümler ile değerlendirilir. Bu noktada genel olarak şirketin efor tahmininin beklenen den kısa çıkması aksiyon alınması gereken bir sorun olarak kabul edilir. Belirtilen bu “aksiyon” genellikle plan dışı fazla çalışma olduğundan dolayı şirkete maliyet olarak döner.

Uzun ölçekte sürece bakıldığında ise takım tasarım proje planlamasında ise efor hesaplamaları yapılırken proje takvimleri yıllar ile ifade edildiğinden dolayı hesaplamalar yapılırken “saat” birimi yerine “gün”, bazı durumlarda ise “ay” birimleri kullanmak genelde tercih edilmektedir. Ayrıca tahminler verilirken takım tasarım mühendisliğine ayrılan süre projenin toplam süresinin %20 sini geçmediği sürece tahminler doğru kabul edilmektedir. Örneklendirmek gerekir ise; 5 yıl sürecek bir projede takım tasarımı için ayrılan süre yaklaşık 1 yıl olacaktır. O nedenle tasarım tahmininin gerçekleşen değerden yüksek çıkması, tahminin doğru olarak kabul edilmesinde yeterli olacaktır. Ayrıca en iyi performansı veren modelde tahmin 42.16 saat olmakta ve bu değer yaklaşık 5 güne⁶ tekabül etmektedir. Bu durumda gerçekleşmesi “0” çıkan bir projede en kötü durumda maksimum 5 gün hata olacaktır.

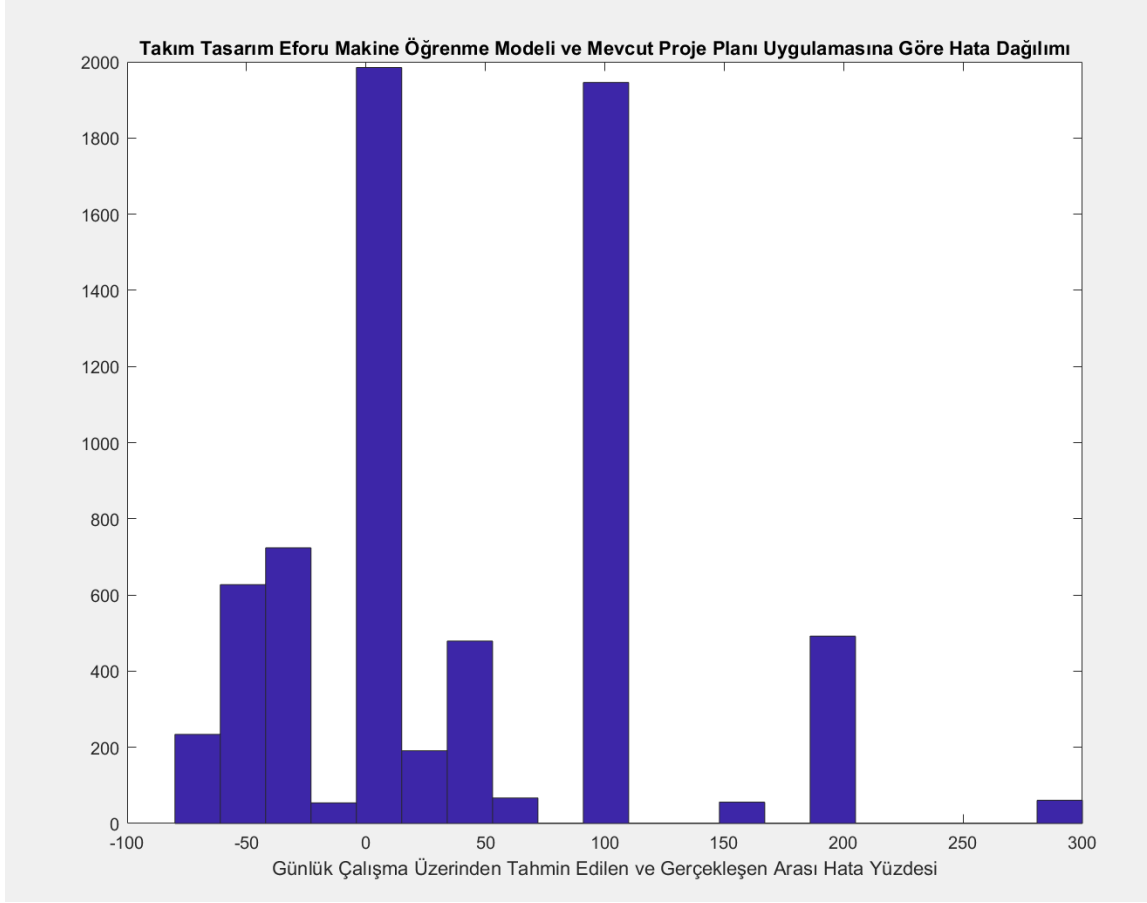
Yukarıda tanımlanan durum ile yapılan çalışma beraber değerlendirildiğinde aşağıdaki şekilde çıkarımlar yapılabilir:

- Proje tahminleri minimum 1 gün olarak tanımlanır.
- Proje efor tahminlerinde en fazla gerçekleşmenin %20 altında tahmine izin verilir.
- Yüksek tahminler her zaman doğru kabul edilir;

$$\text{eğer (eğitilmiş model tahmini} > \text{gerçekleşen proje saati)} \rightarrow \text{tahmin doğrudur} \quad (3.3)$$

Oluşturulan modeli yukarıda belirtilen şekilde değerlendirebilmek için gerçekleşen değerler ve tahmin edilen değerler tam sayı olarak güne çevrildiğinde ise Şekil 3.19'daki durum oluşmaktadır. Ayrıca 3.3 numaralı sözde kod ile tanımlandığı üzere “0” üzeri değerler doğru olarak değerlendirilir. Bu şekilde yapılan doğru tahminler toplan verinin %84.1373 ünü içermektedir.

⁶ Günlük çalışma süresi 9 saattir.



Şekil 3.19 Günlük çalışma üzerinden tahmin edilen ve gerçekleşen takım tasarım eforunun hata dağılımı

4. SONUÇ VE ÖNERİLER

Takım tasarım eforu kestirimi, havacılık endüstrisindeki öngörülemeyen süreçlerden biridir. Çünkü dikkate alınması gereken çok fazla faktör bulunmaktadır. Bu araştırmada, takım tasarımı eforunun tahmin edilmesinde makine öğrenimini kullanabilmek için bir yaklaşım geliştirilmiştir. Ayrıca, gereksinimler iyi tanımlanmışsa ve önceki efor verileri mevcutsa, bu yaklaşım diğer birçok tasarım sürecine uyarlanabilir.

Diğer yandan, bu çalışmada kullanılan imalat mühendisliği takım gereksinimi dokümanı verileri içerisinde hatalı ve tutarsız veriler daha az olabilir ve makine öğrenmesi için girdi olarak kullanılan girdi parametreleri çeşitliliği artırılabilir ise modelin doğru tahmin yeteneğinin de artacağı öngörülmektedir.

İkili kodlama ve en iyi hale getirilmiş Yapay Sinir Ağının kullanıldığı modelde ön çalışmalardan elde edilen bilgiler ışığında karar ağacı yönteminin de oldukça iyi sonuçlar verdiği görülmüştür. Bu nedenle ikincil olarak karar ağacı yöntemi üzerinde de kodlama yöntemi ve makine öğrenme yöntemi optimizasyonu yapılarak sonuçlar mevcut model ile karşılaştırılabilir.

Girdi parametrelerinin kodlanması sürecinde kullanılan yaklaşımdan farklı olarak Principal Component Analysis (PCA) tabanlı boyut indirgeme ve Autoencoder yöntemlerinin denenmesinin de sonuca etkili olabileceği düşünülmektedir. Ek olarak girdi parametrelerindeki mükerrer verileri düzenlerken ortalama çıktı değerinin alınması yerine, farklı istatistiksel metriklerin kullanılması mevcut mimari üzerinde başarı oranına etkiyebileceğinden dolayı gelecek çalışmalarda göz önünde bulundurulması faydalı olacaktır. Örneğin, mükerrer verinin çıktısı en büyük ya da en küçük olan değer alınarak mevcut makine öğrenme yöntemleri yeniden değerlendirilebilir.

Bu çalışmada makine öğrenme mimarileri içerisinde seçilen yöntemlere ek olarak Random Forest, Naive Bayes ve derin öğrenme mimarileri de kullanılarak bu tezde önerilen mimarinin performansı artırılabilir.

Yukarıda belirtilen öneriye ek olarak, bu çalışmada takım gereksinimlerinde Şekil 2.2’de soldaki resimde gösterilen “MANUFACTURING INSTRUCTIONS” alanı hesaba katılmamıştır. İlgili alanda metin verilerinin yorumlanması ve kodlanması için bir algoritma tasarlanabilir. Örneğin, kodlama için Word2Vec⁷ kullanılabilir.

⁷ Geniş metin havuzundan kelime ilişkilerini öğrenmek için kullanılan bir sinir ağı modeli.

KAYNAKLAR

- [1] C. Poli, Design for Manufacturing, A structured Approach, Melbourne New Delhi: Boston Oxford Auckland Johannesburg, 2001.
- [2] A. Dentsoras, «Information generation during design: Information importance and design effort,» *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, pp. 19-32, 03 01 2004;.
- [3] K. I. Mark V. Martin, «Design For Variety: A Methodology for Understanding The Costs of Product Proliferation,» %1 içinde *The 1996 ASME Design Engineering Technical Conferences and Computers in Engineering Conference*, Irvine, California, 1996.
- [4] S. D. E. D. E. W. Maria Carrascosa, «Using design structure matrix to estimate development time,» %1 içinde *ASME Design Engineering Technical Conference*, Georgia, USA, 1998.
- [5] J. P. & T. W. Simpson, «Development of a production cost estimation framework to support product family design,» *International Journal of Production Research Volume 43, 2005 - Issue 4*, 22 02 2007.
- [6] S. C. T. G. Rajkumar Roy, «Estimating the cost of a new technology intensive automotive product: A case study approach,» *International Journal of Production Economics 97*, p. (2005) 210–226, 2005.
- [7] Y. S. B. N. S. S. Sumeet Kaur Sehraa, «Research patterns and trends in software effort estimation,» *Information and Software Technology, Volume 91*, pp. 1-21, 11 2017.
- [8] C. B. F. J. M.-M. J. Renau, « μ Complexity: Estimating Processor Design Effort,» %1 içinde *Proceedings of the 38th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO'05)*, Barselona, İspanya, 2005.
- [9] B. W. Boehm, «Software Engineering Economics,» %1 içinde *IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, VOL. SE-10,*, Redondo Beach, Prentice-Hall, 1984.
- [10] A. Albrecht, «Measuring Application Development Productivity,» %1 içinde *SHARE/GUIDE IBM Applications Development Symposium*, Monterey, 1979.

- [11] H. B. a. M. M. e. S. Bernardes, «Proposed framework for estimating effort in design projects,» %1 içinde *International Journal of Managing Projects in Business*, Hindistan, 2018.
- [12] H. A. B. V. Thomson, «Estimating Design Complexity,» %1 içinde *Journal of Engineering Design*, 1999.
- [13] V. T. Hamdi A. Bashir, «Estimating design effort for GE hydro projects,» %1 içinde *Computers & Industrial Engineering* 46 (2004) 195–204, Canada, 2004.
- [14] H. A. Bashir, «Modeling of development time for hydroelectric generators using factor and multiple regression analyses,» *International Journal of Project Management*, pp. 457-464, 02 08 2007.
- [15] C. O.-Y. a. T. S. Lin, «Developing an Integrated Framework for Feature-Based Early Manufacturing Cost Estimation,» %1 içinde *The International Journal of Advanced Manufacturing Technology*, Londra, 1997.
- [16] N. B. Adil Salam, «Estimating design effort using parametric models: A case study at Pratt & Whitney Canada,» %1 içinde *Concurrent Engineering: Research and Applications*, Montreal, QC, Kanada, 2016.
- [17] T. S. P. C. D. P. Kedar Potdar, «A comparative study of categorical variable encoding techniques for neural network classifiers,» *International Journal of Computer Applications* (0975 – 8887) , p. Volume 175 – No.4, 2017.
- [18] G. V. Patricio Cerda, «Encoding high-cardinality string categorical variables,» *IEEE Transactions on Knowledge and Data Engineering, Institute of Electrical and Electronics Engineers*, pp. 1-1, 2020.
- [19] «Siemens Digital Industries Software,» Siemens, 14 10 2020. [Çevrimiçi]. Available: <https://www.plm.automation.siemens.com/global/en/products/teamcenter/>. [Erişildi: 14 10 2020].
- [20] «Collaborative Innovation,» Dassault Systems, 11 11 2020. [Çevrimiçi]. Available: <https://www.3ds.com/products-services/enovia/>. [Erişildi: 14 11 2020].
- [21] «About SAP SE 2020,» SAP, 14 11 2020. [Çevrimiçi]. Available: <https://www.sap.com/corporate/en.html>. [Erişildi: 14 11 2020].

- [22] «Agile Product Lifecycle Management | ORACLE 2020,» ORACLE, 14 11 2020. [Çevrimiçi]. Available: <https://www.oracle.com/applications/agile-product-lifecycle-management/>. [Erişildi: 14 11 2020].
- [23] «NX 2020,» Siemens, 14 11 2020. [Çevrimiçi]. Available: <https://www.plm.automation.siemens.com/global/en/products/nx/>. [Erişildi: 14 11 2020].
- [24] «Design Engineering | CATIA Dassault Systems 2020,» Dassault, 14 11 2020. [Çevrimiçi]. Available: <https://www.3ds.com/products-services/catia/>. [Erişildi: 14 11 2020].
- [25] H. Alkharusi, «Categorical Variables in Regression Analysis: A Comparison of Dummy and Effect Coding,» *International Journal of Education*, pp. 202-210, 04 2012.
- [26] «Mathworks Help Center,» Mathworks, 05 12 2020. [Çevrimiçi]. Available: <https://www.mathworks.com/help/stats/what-is-linear-regression.html>. [Erişildi: 05 12 2020].
- [27] L. J. H. F. R. A. O. a. C. J. S. reiman, «Classification and Regression Trees,» %1 içinde *Boca Raton, FL: Chapman & Hall*, 1984.
- [28] R.-E. P.-H. C. a. C.-J. L. Fan, «Working set selection using second order information for training support vector machines,» *Journal of Machine Learning Research Vol 6*, p. 1889–1918, 2005.
- [29] D. F. Specht, «A general regression neural network,» *IEEE Transactions on Neural Networks*, p. 568–576, 06 08 2002.
- [30] K. Levenberg, «A Method for the Solution of Certain Non-Linear Problems in Least Squares,» *Quarterly of Applied Mathematics*, p. 164–168, 1944.

EKLER

EK 1: YSA OPTİMİZASYONU İÇİN MATLAB KODU

EK 1: YSA OPTİMİZASYONU İÇİN MATLAB KODU

```
1 %YSA OPTİMİZASYONU İÇİN MATLAB KODU
2
3 close all;
4 clear all;
5 DATA=readtable ('<DRIVE>\Path...\Vencoded_data.xlsx');
6
7 input=DATA(:,1:##);
8 target=DATA(:,##+1);
9
10 inputs=transpose (table2array (input));
11 targets=transpose (table2array (target));
12 ers=double.empty (100,0);
13 Rvalue=double.empty (100,0);
14 neurons=double.empty (100,0);
15
16 for s = 1:400
17
18     net=fitnet ([katman1_noron katman2_noron s]);
19
20 net=configure (net,inputs,targets);
21 view (net);
22 net.trainFcn='trainlm';
23 [net,tr]=train (net,inputs,targets,'usegpu','yes');
24
25 outputs = net (inputs);
26 errors = gsubtract (outputs,targets);
27 performance = perform (net,targets,outputs);
28
29 Req = fitim (outputs,targets);
30
31 close all
32
33 Rvalue (s) =sqrt (Rsq.Required,Ordinary);
34 ers (s) =sqrt (performance);
35 neurons (s) =s;
36 end
37
38 figure, scatter3 (neurons,ers,Rvalue)
39
40 MSE=transpose (ers);
41 R=transpose (Rvalue);
42
```

```
1
2
3 ENCODE EDİLMİŞ VERİ OKUNUR
4
5 GİRDİ PARAMETRELERİ İÇİN EXCEL KOLONLARI ##'e KADAR OKUNUR
6
7 GİRDİ VE ÇIKTI MAREİSİ TRANSPOSE EDİLİR
8
9 ŞERH BİR DENEYDEKİ RMSE DEĞERİ
10
11 ŞERH BİR DENEYDEKİ R DEĞERİ
12
13 BELİRTİLEN KATMAN İÇİN "s" NORUN İLE DENEY YAPILIR
14
15 KATMANLAR VE NORUN SAYILARI TANIMLANIR
16
17 YSA OLUŞTURULUR
18
19 YSA GORSELLİ OLUŞTURULUR
20
21 EGİTİM ALGORİTMASI TANIMLANIR
22
23 YSA EGİTİLİR
24
25 EGİTİLMİŞ VERİ İLE HER NORUN İÇİN SONUÇ VERİSİ KAYDEDİLİR
26
27 ŞATA HESAPLANIR
28
29 "RMSE" DEĞERİ HESAPLANIR
30
31 GERÇEKLEŞMELER VE SONUÇLAR ARASINDA REGRESYON MODELİ OLUŞTURULUR
32
33 "R" DEĞERİ HESAPLANIR
34
35 SONUÇ İÇİN 3D DAĞILIM GRAFIGİ OLUŞTURULUR
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
```