T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

VERİ MERKEZLERİ VE IT SİSTEMLERİ İÇİN YAPAY SİNİR AĞLARI VE REGRESYON YÖNTEMLERİ İLE ENERJİ YÜK PROFİLİNİN OLUŞTURULMASI VE ANALİZİ

MURAT SALİM KARABİNAOĞLU
YÜKSEK LİSANS TEZİ
ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

GEBZE
2018

T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

VERİ MERKEZLERİ VE IT SİSTEMLERİ İÇİN YAPAY SİNİR AĞLARI VE REGRESYON YÖNTEMLERİ İLE ENERJİ YÜK PROFİLİNİN OLUŞTURULMASI VE ANALİZİ

MURAT SALİM KARABİNAOĞLU
YÜKSEK LİSANS TEZİ
ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

DANIŞMANI
DR. ÖĞR. ÜYESİ TUBA GÖZEL

GEBZE
2018

T.R.

GEBZE TECHNICAL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

LOAD FORECASTING AND ANALYZING OF DATA CENTERS AND IT SYSTEMS BY USING ARTIFICIAL NEURAL NETWORKS AND REGRESSION METHODS

MURAT SALİM KARABİNAOĞLU

A THESIS SUBMITTED FOR THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF ELECTRONIC ENGINEERING

THESIS SUPERVISOR

FACULTY MEMBER DR. TUBA GÖZEL

GEBZE

2018

GEBZE TEKNİK ÜNİVERSİTESİ	**YÜKSEK LİSANS JÜRİ ONAY FORMU**

GTÜ Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 25../..04../..2018 tarih ve 2018../..23.. sayılı kararıyla oluşturulan jüri tarafından 10/05/2018 tarihinde tez savunma sınavı yapılan Murat Salim KARABİNAOĞLU'nun tez çalışması Elektronik Mühendisliği Anabilim Dalında YÜKSEK LİSANS tezi olarak kabul edilmiştir.

JÜRİ

ÜYE (TEZ DANIŞMANI)	:Dr. Öğr. Üyesi Tuba GÖZEL	
ÜYE	:Prof. Dr. Hakan HOCAOĞLU	
ÜYE	:Dr. Öğr. Üyesi Erdem BİLGİLİ	

ONAY

Gebze Teknik Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun

...../...../..... tarih ve ...../..... sayılı kararı.

ÖZET

Veri merkezleri, sunucu, router, switch ve firewall gibi bilgi işlem cihazlarının yanı sıra destek ekipmanları olan kurtarma üniteleri (Backup Units) ve iklimlendirme cihazlarını barındıran tesislerdir. İnternet kullanımındaki artışlar, büyük veri (Big Data), nesnelerin interneti (Internet of Things), bulut hesaplamaları (Cloud Computing) gibi uygulamalarda meydana gelen gelişmeler dünya çapındaki veri merkezlerinin sayılarının ve kapasitelerinin artmasına sebep olmuştur.

Tez çalışmam, veri merkezlerinin enerji tüketim miktarlarını optimize edebilmek amacıyla otomasyon kontrol sistemleri, uygun yazılımlar ile yük tahmin yöntemlerini bir araya getirerek gerekli yazılımların geliştirilmesi sonucunda uygun yöntemler bulmayı amaçlamaktadır. Bu sebeple, yapay sinir ağları ve çeşitli regresyon yöntemleri kullanılarak veri merkezlerindeki enerji tüketimini tahmin edebilmek için gerekli verilerin toplanmasını sağlayacak Veri Toplama ve Enerji İzleme Yazılımı (DAEMS-Data Acquisition and Energy Monitoring Software) geliştirilmiştir. Bununla birlikte veri merkezlerini örnekleyebilmek amacıyla bir sunucu hazırlanarak kullanılmıştır.

Dört farklı yöntem ile yapılan enerji tüketimi tahmin sonuçlarına yer verilerek bu yöntemlerin performansları karşılaştırılmıştır.

Tahmin için kullanılan giriş parametrelerinin tüketilen güç ile olan ilişkisi 24 saatlik aralığa ait yaklaşık 86400 adet veri sayesinde korelasyon matrisi oluşturmakta kullanılmıştır. Korelasyon matrisinde tüm parametrelerin bir günlük veri üzerinden birbirleri ile olan korelasyon ilişkileri ve bağlılık seviyeleri gösterilmiştir.

Anahtar Kelimeler: Veri Merkezleri, Yük Profili Tahmini, Yapay Sinir Ağları, Regresyon Yöntemleri, Otomasyon Kontrol Sistemleri.

SUMMARY

A data center is a repository that houses computing facilities like servers, routers, switches and firewalls, as well as supporting components like backup equipment, fire suppression facilities and air conditioning. Data centers nowadays are proliferating all over the world because of increasing demand for Internet applications, Big Data, Internet of Things and Cloud Computing.

In my thesis, my goal is to find a way to optimize energy consumption of data centers by encompassing automation control system devices, appropriate software and load forecasting methods. For this reason, a software named as DAEMS (Data Acquisition and Energy Monitoring Software) was prepared for a data center to collect data for forecasting of electricity by using Artificial Neural Networks and multiple regression methods. In addition, an individual server which was configured for modelling a data center was used.

By applying four different methods there has been analyzed the performance of these methods for predicting the load profile.

The input parameters with the number of 86400 for 24 hours period which have already been described as input parameters for our prediction design were used to create a correlation matrix. It is shown that the correlation level of the each parameters with the others for one day period of data set.

Key Words: Data Center, Load Profile Forecasting, Artificial Neural Networks, Regression Methods, Automation Control Systems.

TEŞEKKÜR

Başta, yüksek lisans eğitimimde ve akademik hayatımda desteğini ve yardımlarını hiçbir zaman esirgemeyip bilgisi ile bu çalışmanın oluşmasının yolunu açan danışmanım Dr. Öğr. Üyesi Tuba GÖZEL'e, göstermiş olduğu destek ve anlayıştan dolayı sevgili eşim Tuğba ve oğlum Yağız KARABİNAOĞLU'na en içten teşekkürlerimi sunarım.

İÇİNDEKİLER

	Sayfa
ÖZET	v
SUMMARY	vi
TEŞEKKÜR	vii
İÇİNDEKİLER	viii
SİMGELER ve KISALTMALAR DİZİNİ	ix
ŞEKİLLER DİZİNİ	x
TABLOLAR DİZİNİ	xii
1. GİRİŞ	1
1.1. Tezin Amacı, Katkısı ve İçeriği	1
2. LİTERATÜR ÖZETİ	5
3. TASARIM VE YAZILIMLAR	20
3.1. Tasarım Prensipleri	20
3.2. Geliştirilen Yazılımlar	21
4. METOTLAR VE UYGULAMALAR	27
4.1. Veri Toplama	27
4.2. Yapay Sinir Ağları	36
4.3. Regresyon Yöntemleri	46
5. ANALİZLER VE KARŞILAŞTIRMALAR	52
5.1. Veri Analizleri	52
5.2. Metot Karşılaştırmaları	57
5.3. Değerlendirme	61
6. SONUÇLAR	63
KAYNAKLAR	65
ÖZGEÇMİŞ	70
EKLER	71

SİMGELER ve KISALTMALAR DİZİNİ

Simgeler ve Kısaltmalar		**Açıklamalar**
GTÜ	:	Gebze Teknik Üniversitesi
YSA	:	Yapay Sinir Ağları
MLP	:	Multi Layer Perceptron
DAEMS	:	Data Acquisation and Energy Management Software
PUE	:	Power Usage Effectiveness
RAM	:	Random Access Memory
CPU	:	Central Processing Unit
IPC	:	Industrial PC
PLC	:	Programmable Logic Controller
GtCO2	:	Gigatonnes of Carbon Dioxide
BHT	:	Bilgi ve Haberleşme Teknolojileri
IT	:	Information Technologies

ŞEKİLLER DİZİNİ

Şekil No:		Sayfa
1.1:	Veri Merkezleri Sera Gazı Emisyonları 2002-2020.	2
3.1:	Yerleşim Şeması.	20
3.2:	DAEMS Yazılımı Akış Şeması.	22
3.3:	DAEMS Kullanıcı Arayüzü.	23
3.4:	DAEMS Ayarlar Ekranı.	23
3.5:	DAEMS Kayıt İlerleme Sayfası.	24
3.6:	DAEMS Ölçüm Bilgileri Kaydet Sayfası.	24
3.7:	YSA Modeli Akış Şeması.	25
4.1:	Birinci Ethernet Portu Veri Trafiği-Sunucu Güç Tüketimi.	28
4.2:	İkinci Ethernet Portu Veri Trafiği-Sunucu Güç Tüketimi.	28
4.3:	Ram ve Cpu Kullanımı-Sunucu Güç Tüketimi.	29
4.4:	Ortam ve Cpu Sıcaklığı-Sunucu Güç Tüketimi.	29
4.5:	Ortam Nemi-Sunucu Güç Tüketimi.	30
4.6:	Giden Veri -1.	32
4.7:	Gelen Veri -1.	32
4.8:	Giden Veri -2.	32
4.9:	Gelen Veri -2.	33
4.10:	RAM Kullanımı.	33
4.11:	Cpu Kullanımı.	33
4.12:	Ortam Sıcaklığı.	34
4.13:	Cpu Sıcaklığı.	34
4.14:	Ortam Nemi.	35
4.15:	Güç Tüketimi.	35
4.16:	YSA Algoritması Akış Şeması.	37
4.17:	Sinir Ağı Modeli.	38
4.18:	Aktivasyon Fonksiyonu.	39
4.19:	Sinir Ağı Detaylı Görünüm.	41
4.20:	Hata-EPOCH Grafiği.	46
5.1:	Cpu Sıcaklığı Korelasyon Grafiği.	54

5.2: RAM Kullanımı Korelasyon Grafiği. 54
5.3: Ortam Sıcaklığı Korelasyon Grafiği. 55
5.4: Cpu Kullanımı Korelasyon Grafiği. 55
5.5: Giden Veri -1 Korelasyon Grafiği. 56
5.6: Gelen Veri -1 Korelasyon Grafiği. 56
5.7: Gerçek Güç-Doğrusal Regresyon. 57
5.8: Gerçek Güç-Doğrusal Olmayan Regresyon. 58
5.9: Gerçek Güç-Farklılaştırılmış Doğrusal Olmayan Regresyon. 58
5.10: Gerçek Güç-Yapay Sinir Ağları. 59
5.11: Doğrusal Olmayan Regresyon Hata-YSA Hata. 59

TABLOLAR DİZİNİ

Tablo No:		Sayfa
4.1:	Ethernet Portları Veri Trafiği-Güç İlişkisi Tablosu.	31
4.2:	Sunucu ve Ortam Şartları Parametreleri-Güç İlişkisi Tablosu.	31
4.3:	Doğrusal Regresyon Katsayıları.	49
4.4:	Doğrusal Olmayan Regresyon Katsayıları.	50
4.5:	Farklılaştırılmış Doğrusal Olmayan Regresyon Katsayıları.	51
4.6:	Regresyon Eğitim Başarım Oranları Karşılaştırma Tablosu.	51
5.1:	Korelasyon Matrisi Tablosu.	53
5.2:	Metot Karşılaştırma Tablosu.	61

1. GİRİŞ

Veri merkezleri, küresel anlamda dünya üzerindeki elektrik tüketiminin %3'ünü oluşturan enerji tüketim oranlarına sahiptirler. Veri merkezleri, aynı zamanda dünya sera gazı emisyonu miktarının da %2'sini oluşturan önemli etkenlerden birisi haline gelmiştir. Veri merkezleri gelecek on yıl içerisinde üç kat daha fazla enerji harcayacağı yapılan tahminlerle hesaplanabilmektedir. 2015 yılında dünya çapında 416,2 terawatt saat elektrik tüketimine sahip veri merkezleri, İngiltere'nin aynı dönemdeki 300 terawat saat'lik [1], [2] ve Türkiye'nin 265,7 terawatt saatlik elektrik tüketimlerinden oldukça fazla miktarda enerji tüketen tesisler olarak karşımıza çıkmaktadırlar [3].

Son yıllarda, nesnelerin interneti (IoT-Internet of Things), bulut temelli sistemler (Cloud-Based Systems), büyük veri (Big Data) vb. uygulamalarda yaşanan gelişmeler ile dijitalleşme, veri merkezlerinin kapasite ve sayıca artarak büyümelerine sebep olmaktadır.

Veri merkezlerinin, sera gazı emisyonunun % 2'sini oluşturması, küresel iklim değişikliği üzerindeki artan etkisine ait farkındalığı da artırmıştır. Tüm bu sebeplerden dolayı, veri merkezlerinin enerji tüketim tahmini, güç yönetimi ve optimizasyonu zaruri hale gelmiştir. Bunlar ile birlikte çevresel etki ve sorumluluklar ile enerji maliyetlerini de göz önünde bulundurduğumuzda veri merkezlerinin enerji tüketimlerini optimize edebilmek önemlidir.

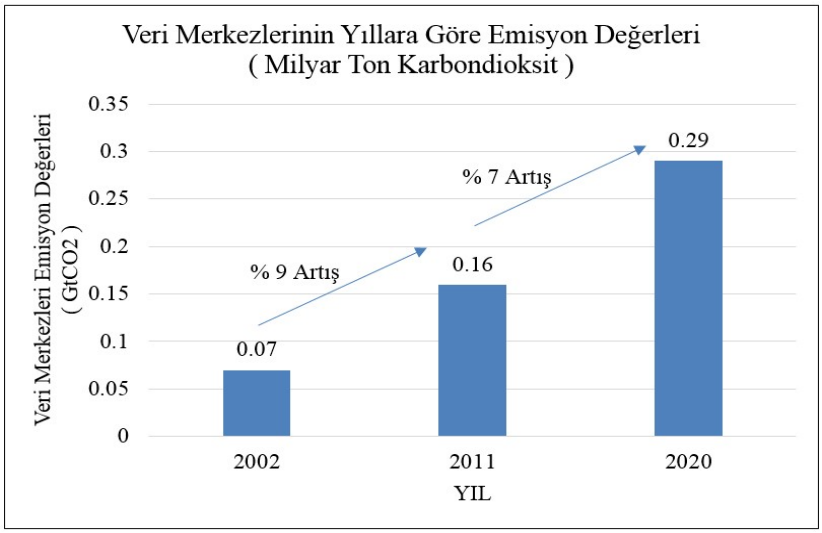
Bu merkezlerde yaşanabilecek aksaklıkları önceden anlamaya yönelik tasarım, algoritma, yazılım, donanım ve analiz çalışmalarını gerçekleştirmek ciddi kayıpların önüne geçebileceği gibi enerji tüketim miktarlarını da azaltabileceği ve önemli miktarlarda tasarruf sağlayabileceği anlaşılmaktadır.

1.1. Tezin Amacı, Katkısı ve İçeriği

Tezimin amacı, dünya enerji tüketiminde ve sera gazı emisyon oranlarında yüksek paylara sahip veri merkezlerinin enerji yük profillerinin üretilmesidir. Bu amaçla hazırlanan çalışmamda veri merkezlerininim kritik sistemler olarak düşünüldüğü için bu tür bir uygulamayı yapmanın zorlukları ile başa çıkabilmek adına bir adet sunucu vasıtasıyla veri merkezlerinin örneklenmesi hedeflenmiştir.

Bu aşamada veri merkezlerinde enerji tüketen diğer cihazlara ait parametreler ihmal edilmiştir. Örnek bir model oluşturabilmek adına bir sunucu ve ona ait çeşitli içsel ve dışsal parametreler veri merkezi gibi çalıştığı varsayımına göre kullanılmıştır. Bir sunucu üzerinde yapılan bu çalışma, erişilebilir haldeki bir veri merkezinde gerekli yazılımsal ve donanımsal düzenlemeler sayesinde rahatlıkla uygulanabilecektir. Bu çalışma ile literatüre veri merkezleri için enerji yük profili üretme konusunda bütünsel bir yaklaşım katılması da hedeflenmektedir.

Bilgi ve Haberleşme Teknolojilerine (BHT) ait endüstrilerin iklim değişikliği üzerindeki etkilerini araştıran birçok çalışma GeSI (Global e-Sustainability Initiative) [4], [5] tarafından yapılmış ancak son zamanlarda Malmodin tarafından yapılan çalışmalar ile de güçlendirilmiştir [6]. BHT endüstrilerinin üç temel sektöründe, veri merkezlerinin % 7 büyüme oranı ile 2011 yılından 2020 yılına kadar 0,29 GtCO2 oranına yükselerek en hızlı büyümeye sahip olacakları düşünülmektedir [5]. GeSI tarafından daha önceleri yapılan bu büyüme tahminleri, 2002-2020 yılları arasında küresel ayak izi ölçüsünde % 7'lik bir büyüme öngörmekteyken, 2002-2011 yılları arasında ölçülen gerçek değerler ile ulaşılan sonuç Şekil 1.1'de görülmekte olan % 9 ile hâlihazırda daha yüksek bir etki oluşturduğu ortaya çıkmıştır [4].



Şekil 1.1: Veri Merkezleri Sera Gazı Emisyonları 2002-2020.

Elektrik yük profillerinin gelecek zamanlar için tahmin edilmeye çalışılması, elektrik yüklerinin kararsızlıkları ve çeşitlilikleri sebebiyle ilgi çekici bir problem olmaya devam etmektedir [24]. Elektrik yükünün gün içerisindeki değişimi, çoğunlukla değişkenlik gösteren bir yapıya sahiptir ve düzensizdir.

Özellikle veri merkezlerindeki enerji tüketim tahmini konusunda daha önce yapılan çalışmalar [18], [25] veri merkezlerindeki her bir sunucu özelindeki içsel parametreler olan Cpu ve Ram kullanım oranı ile veri iletişim parametrelerinden, gelen ve gönderilen veri miktarlarını toplu ve eşzamanlı olarak dikkate almamıştırlar.

Veri merkezlerinin enerji tüketimleri konusunda daha önce yapılan bir diğer çalışma [25] bize göstermektedir ki veri merkezlerinde bulunan belirli cihazlara ait (Router, Switch vs.) enerji tüketim miktarlarının doğrudan ölçülemiyor oluşu kritik bir problem olarak karşımıza çıkmaktadır. Tez çalışmamda farklı bir bakış açısıyla daha doğru bir tahmin yöntemi üretebilmemiz için uygulanan tasarımın tek bir sunucu üzerinde çalıştırılması sağlanarak doğrulama işlemi yapılacaktır. Böylece, bu tasarım veri merkezlerindeki tüm cihaz ve sistemler için genişletilerek gelecekteki çalışmalarda kullanılabilecektir. Tez çalışmamda enerji analizörü kullanmamız sayesinde sunuculara ait gerçek zamanlı enerji tüketim bilgisi yüksek hassasiyetli dijital bilgiye çevrilerek kayıt altına alınabilecektir.

Bu çalışmaya ait tasarımın literatüre bir diğer katkısı ise sunucuya ve diğer tüm çevresel değişkenlere ait bilgileri eş zamanlı olarak toplayacak olan Veri Toplama ve Enerji İzleme Yazılımının (DAEMS) üretilecek olmasıdır. Bu çalışmamda, veri merkezlerinin gelecek dönemler için enerji talep miktarlarını belirleyebilmek için yük tahmini, veri merkezlerine ait içsel ve ortama ait çevresel parametreler kullanılarak yapılacaktır. Bu değişkenler fiziksel olarak bağlı cihazlara ait parametreler olan sıcaklık ve nem olabileceği gibi bunlarla birlikte sunucuya ait içsel parametreler olan Cpu ve Ram kullanım oranı, gelen ve giden veri miktarı da olabilecektir.

Otomasyon kontrol cihazları (PLC), uygun aktüatör ve sensörler, özel tasarım yazılım çözümlerinden olan Veri Toplama ve Enerji İzleme Yazılımı, Yapay Sinir Ağlar, çeşitli Regresyon Yöntemleri ve bunlara ait uygulama yazılımları gibi farklı türde cihazlar ve yazılımlar bir araya getirilerek kullanılmıştır.

Güç tüketim tahmini konusunda regresyon yöntemleri ile birlikte yapay sinir ağları gibi birçok yöntem bulunmaktadır. Daha önce yapılan araştırmalardan bir tanesinde lineer, logaritmik, çok terimli regresyon yöntemleri, yapay sinir ağları ile karşılaştırılmış ve yapay sinir ağları yönteminin gerçek tüketim verilerine en yakın değerleri en düşük hata oranı ile verdiği sonucuna ulaşılmıştır [19]. Bu araştırmaya göre YSA, tahmin etme süreçlerinde en güvenilir yöntemlerden bir tanesi olarak kabul edilmektedir. Bu ve benzeri çalışmalar ve sonuçlar incelendiğinde tez çalışmamda yük

tahmin yöntemi olarak YSA ve regresyon yöntemleri kullanılmıştır. Tez çalışmamda makine öğrenme algoritmasının tatbiki amacıyla YSA ve çeşitli regresyon yöntemleri ile birlikte farklılaştırılmış bir regresyon yöntemi de kullanılacaktır. Bu yöntemler ile veri merkezlerinin çalışmasını modellemek için kullanılan sunucunun, gelecek dönemlerdeki yük profili tahmin edilmeye çalışılacaktır.

Farklı yöntemlerle yapılan analizler sonucunda elde edilen yük eğrileri, Ortalama Karesel Hata ve Korelasyon Katsayısı parametreleri dikkate alınarak karşılaştırılacaktır. Bu karşılaştırma işlemi, tüm metotlar için aynı ham veri kullanılarak ve aynı performans ölçütleri ile yapılacağından çıkan sonuçlar yük profili üretmedeki metotların başarım ölçüsünü gösterebilecektir. Ayrıca, elde edilen sonuçlar, yük profili üretme işlemimizdeki 9 adet giriş parametresi ve bir adet tüketilen enerji miktarı parametresi dikkate alınarak toplamdaki 10 ayrı özelliğin birbirleri ile olan etkileşimine ait korelasyon matrisi tablosu oluşturmakta kullanılmıştır. Bu sayede 10 ayrı parametrenin her birinin diğerleri ile olan etkileşim seviyeleri -1 ve +1 aralığında üretilmiş, hangi parametrelerin diğerleri ile daha yüksek bir korelasyon seviyesine sahip olduğu görülmüştür.

Sürece ait detaylar aşağıdaki bölümlerde detaylı olarak açıklanacaktır ancak özetlemek gerekirse:

- Tez çalışmamın ikinci bölümünde konumuz ile ilgili daha önce yapılmış olan çalışmalar incelenmiştir.
- Üçüncü bölümde yapılan tasarım ve bu tasarıma ait detaylar ile birlikte yukarıda bahsedilen hedeflere ulaşabilmek adına geliştirilmiş gerekli yazılımlar hakkında bilgiler verilmiştir.
- Dördüncü bölümde çalışmamda kullanılan metotlardan bahsedilmiştir. Bu metotlardan YSA ve Regresyon yöntemleri için gerekli parametrelerin toplanması ve bu metotların uygulanması hakkında detaylı bilgiler paylaşılmıştır. Sonrasında, uygulanan YSA ve Regresyon metotları hakkında detaylı bilgiler açıklanacaktır.
- Beşinci bölümde veri analizi ile birlikte, uygulamaları gerçekleştirilen metotlar ile elde edilen sonuçlar ve bu sonuçların karşılaştırmalarına yer verilmiştir. Ayrıca bu çalışmanın devamı olarak gelecek dönemlerde yapılabilecek çalışmalardan bahsedilmiştir
- Sonuç bölümünde tüm analiz ve çalışma sonuçları değerlendirilmiştir.

2. LİTERATÜR ÖZETİ

Günümüzde modern veri merkezleri farklı disiplinleri bir araya getirmesi ve karmaşıklığı sebebiyle tasarımından, kurulumuna, işletmesinden sürekli devrede olması gerekliliğine kadar dikkatli ve titiz bir çalışma gerektirmektedir. Veri merkezleri için hali hazırda en geçerli ve yaygın değerlendirme kıstası Uptime Institute tarafından hazırlanan veri merkezlerine ait altyapı ve diğer özelliklere ait Tier standartlarıdır [17]. Bu standartlar veri merkezlerini topoloji açısından ve işlemsel sürdürülebilirlik açısından değerlendirebilmektedir. Topoloji standartları veri merkezlerini fonksiyonellik, kapasite ve diğer veri merkezleri ile beklenen kullanılabilirlik ya da performans açısından karşılaştırabilme temeline dayanmaktadır. Tier standartları veri merkezlerinin ölçeklerine ve önem verilen özelliklerine göre dört farklı derecelendirmeye sahiptir. Bu altyapı standartları sayesinde tüm veri merkezlerinin değerlendirme ve sertifikalandırma işlemleri dünyada en çok kabul gören standartlar ile yapılabilmektedir.

Veri merkezlerinin fonksiyon, kapasite, performans ve işletme devamlılığı açısından hedeflediği sınıflandırmayı ülkemizde Türk Standartları Enstitüsü tarafından hazırlanan Veri Merkezleri Altyapısı Standart Taslağı ile belirli bir standarda ulaştırmak istenmektedir [16].

Veri merkezleri ile ilgili yapılan yük tahmini ve optimizasyon çalışmalarının ortak özelliklerinden birisi de enerji verimliliğine odaklanmak olmuştur. Bu amaçla yapılan çalışmalardan bir tanesi de Google'ın kendine ait veri merkezlerinden bir tanesi için PUE değerinin tahmin edilmesine dayanan ve yapay sinir ağı temelli algoritma kullanılarak gerçekleştirilmiş analiz ve optimizasyon çalışmasıdır [18]. Bu çalışmada yapay sinir ağı üç katmandan oluşmaktadır. Rastgele başlatma işlemi ile aktivasyon fonksiyonuna göre ağırlık değerlerine karar verilir. İleri yayılım ve maliyet fonksiyonunun kullanılması ile birlikte geri yönlü yayılım sayesinde sinir ağının eğitim işlemi tamamlanmış olacaktır. İlgili çalışmada yapay sinir ağı, 19 adet giriş özelliği ve parametresine bağlı değerlere göre eğitilmiş olur [18]. Sonuç olarak gerçek ve güncel PUE değeri ile birlikte tahmin edilen PUE değeri hesaplanabilmektedir. Makine öğrenme yapılarını kullanarak veri merkezlerinin mevcut sensör verileri ve parametreleri kullanılarak enerji verimlilik performansını artırmak ve enerji maliyetlerinde tasarrufa gitmek mümkün olabilmektedir.

Bir başka çalışmada ise güneş enerjisi üretim santrallerinin güç üretim modellerinin çıkartılması amacıyla yapay sinir ağları ve regresyon yöntemleri kullanılmış ve karşılaştırılmıştır [19]. GES'lere ait enerji üretim miktarları güneşin pozisyonuna bağlı olduğu gibi iç ve dış faktörlerle de alakalıdır. Bu faktörler, bulutluluk düzeyi, sıcaklık, rüzgâr hızı, yağmur ve nem miktarı şeklinde olabilecektir. Bu parametreler kullanılarak yük tahmin yöntemleri olan yapay sinir ağları, lineer regresyon, logaritmik regresyon ve polinom regresyon yöntemleri karşılaştırılmıştır [19]. Bir yıllık veri ile 24 saat üzerinden yapılmış olan bu çalışmada dört farklı yöntemin pratik hataları 100kWh üzerinden değerlendirilmiştir. Bu çalışma sonucunda en düşük hata oranı ve en iyi tahmin sonucuna bağlı eşleştirme analizi yapay sinir ağları ile bulunabilmiştir.

Yapay sinir ağları ve lineer olmayan regresyon yöntemi sınıflandırma problemlerinde en çok tercih edilen analiz yöntemlerindendir. Bu yöntemlerin karşılaştırılması için daha önce yapılan bir diğer çalışma ise meteoroloji için yağmur yağış oranlarının tahmini amacıyla da kullanılmıştır [20]. Bu amaçla yapılan çalışma sonucunda birinci ve ikinci derece regresyon yöntemleri ve yapay sinir ağının gizli katmanında beş ve altı nöronun bulunduğu seçenekler karşılaştırılmıştır. Bu yöntemler kullanılarak yapılan tahmin sonuçlarına göre korelasyon katsayısı ve RMS hatası dikkate alınarak altı nörona sahip yapay sinir ağı ile yapılan hesaplamaların daha iyi sonuçlar verdiği görülmüştür.

Çoklu lineer regresyon analizi yöntemi kullanılarak bağımlı değişkenler en iyi şekilde tahmin edilmeye çalışılabilir [21]. Bu yönteme ait hesaplamalar neticesinde bulunan bağımlı değişkenler ise yapay sinir ağları kullanmak suretiyle tahmin edebilirlik seviyesi iyileştirilerek daha iyi sonuçlar alınabilir [21]. Yapay sinir ağlarının lineer olmayan modelleme yeteneklerini kullanmak lineer regresyon yönteminde daha iyi sonuçlar elde etmeyi sağlayabilir. Bu yöntem ile yapılan çalışma sonucunda, lineer olmayan regresyon analizinin lineer olana göre iyileştirmeler yapılmasına rağmen hedeflenen yük tahmin sonucuna daha yakın ve uyumlu hesaplamalar yaptığı görülebilir. Hatalı sınıflandırma oranının lineer regresyon analizine göre daha düşük olduğu literatürdeki bu çalışma ile gösterilmiştir [21].

Yapay sinir ağları temelli tahmin ve varsayım uygulamalarından bir tanesi de lineer olmayan yük harmoniklerinin tahmin edilebilmesi şeklinde olmuştur [22]. Güç dağıtım sistemlerinde yükler ve kaynaklar arasındaki etkileşim karmaşık bir süreç

olmakla birlikte analitik olarak tahminde bulunmadan açıklama yaparak sonuçlara ulaşabilmek çoğunlukla mümkün olamayabilir. Bu sebeple yapay sinir ağları, öngörüde bulunabilme özelliği ile yük kaynaklı harmonikler ve güç kaynağı sebepli harmonikler arasındaki farkı bulabilmenin bir yöntemi olarak kullanılabilmektedir [22].

Bazı durumlarda ise yapay sinir ağları diğer yöntemlere göre uzun dönemli analiz sonuçları değerlendirildiğinde daha iyi sonuçlar veremeyebilir. Buna örnek olarak yapım projelerindeki maliyetlerin öngörülebilmesi için 530 adet geçmiş veri ile hazırlanan ve Çoklu Regresyon Analizi, Yapay Sinir Ağları ve Olay Temelli Muhakeme yöntemleri örnek gösterilebilir [23]. Bu yöntemlerin karşılaştırıldığı çalışma sonucunda en iyi yapay sinir ağları modelinin diğer iki yönteme göre daha doğru öngörülen sonuçlar verebildiği görülmektedir. Fakat sinir ağı modelinin eğitme ve hata düzeltme işlemleri nispeten yavaş olabilmekte ve uzun zaman alabilmektedir. Uzun dönemli değerlendirme sonuçları dikkate alındığında ise olay temelli muhakeme yönteminin daha iyi kararlılığa sahip olduğu görülebilmektedir [23]. Bu sebeple hibrit bir yöntemin daha uygun olabileceği bu çalışmada açıklanmaya çalışılmıştır.

Veri merkezlerinin yük profillerinin tahmin edilmesi ve enerji verimliliklerinin artırılması uygulamalarında farklı yöntemler kullanılabilmektedir. Bu yöntemlerden bir tanesi de Holt-Winters yöntemidir [13]. Veri merkezlerinde bulunan sistemleri için, enerji tüketimini azaltmak amacıyla birçok zamanlama ve planlama yöntemi teklif edilebilir. Bu yaklaşımlar sayesinde, veri merkezlerinin çalışma yükleri daha doğru tahmin edebilir [13]. Farklı veri setleri ile yapılan dönemsel analizler sayesinde veri merkezlerinin daha yoğun olarak çalıştığı dönemlerin daha az yoğun olunan dönemlere göre analiz yöntemlerinin farklılaşması yaklaşımları uygulanabilir [13]. Bir günlük çalışma zamanı içinde beş dakikalık örnekler alınarak günlük ve haftalık verilerin saatlik ortalamalarının alınması şeklinde yapılan literatür çalışmasında, örnek sayısı azaltıldığı durumda yük profili tahmin performansı bundan etkilenmemekle birlikte, analiz için çalıştırılan algoritmaya ayrılması gereken hafıza ve algoritma çalışma zamanı da azaltılmış olabilecektir [13]. Literatürde bu konu ile ilgili yapılan çalışmalardan bazıları ise bayezyen temelli çıkarım yöntemlerine dayanan bağımlı karışım modeli olabilmektedir [14].

Yüksek güç yoğunluğuna sahip veri merkezlerinin yük tahminlerinin yapılması ile ilgili yapılan çalışmalardan bir tanesinde bölgesel yük yoğunluğu indeksi ve

mekânsal yük tahmin yöntemlerinin kullanıldığı çalışmadır [15]. Bu çalışmada veri merkezlerine ait yük tahmini yapabilmek amacıyla kullanılan yöntemlerden Güç Kullanım Verimliliğinin (Power Usage Effectiveness, PUE) kullanılması teklif edilmiştir. Veri merkezinin soğutma kapasitesi ve güç kullanım verimliliğinin göstergeleri veri merkezlerinin yük tahminini üretmekte kullanılmıştır. Veri merkezlerinin termal yük hesaplamaları yapılmış ve PUE temelli yük tahminleri gerçekleştirilmiştir. İki farklı ölçeğe sahip ve benzer iklim şartları özelliklerindeki veri merkezi için bu yöntemler uygulanarak PUE temelli karşılaştırma yapılmıştır. Bu karşılaştırma sonucunda A isimli veri merkezinin PUE değeri daha iyi çıkmıştır. Bu sonucun sebepleri ise; A tesisinin B'den daha yeni olması ve bu sebeple kullanılan cihazların daha iyi yayılım performansına sahip olmaları, A tesisinin bina yapısının daha iyi tasarlanmış olması, A tesisindeki soğutma sisteminin daha yüksek kapasiteli olması şeklinde olduğu öngörülmektedir.

YSA ve çoklu regresyon analizinin tahmin edebilme yetenekleri sayesinde yapılan bir çalışma, Avusturalya'daki mevsimsel uzun dönemli yaz yağmurlarının tahmin edilmesi üzerine yapılmıştır [49]. İki farklı giriş veri setini, iki farklı temel iklim tarzı sayesinde kullanarak bu çalışma yapılmıştır. Bu iklim tarzları ENSO (El Nino Southern Oscillation) ve IOD (Indian Ocean Dipole) olarak isimlendirilmiştir. Bu iki temel iklim modelinin uzun dönemli yaz yağmurlarını tahmin edebilme üzerindeki etkilerini, geçmiş verileri eş zamanlı olarak kullanmak suretiyle YSA ve çoklu regresyon analizi yöntemleri ile bulabilmeyi önermektedir [49]. YSA modeli, çok katmanlı algılayıcı formunda Levenberg-Marquardt algoritması kullanılarak uygulanmıştır. YSA ve çoklu regresyon analizi yöntemleri ortalama karesel hata (MSE), ortalama mutlak hata (MAE), pearson korelasyon katsayısı (r) ve willmott anlaşma indeksi (d) parametreleri dikkate alınarak değerlendirilmiştir. Çoklu regresyon analizi modelinin test veri seti ile tahmin edebilme yetenekleri YSA ile uyumlu sonuçlar vermiştir. Bu istatistiksel analiz sonucunda YSA modelinin, potansiyeli ve sonuçları bakımından büyük ölçekli iklim tarzı kullanarak tahmin edebilme özellikleri dikkate alındığında çoklu regresyon analizine göre daha tavsiye edilebilir olduğu belirtilmiştir [49].

YSA modeli, lineer olmayan sistemlerin modellenmesi ve simülasyonunun yapılması için kullanılabilmektedir [50]. YSA modelinin en kullanışlı ve ilginç kulanım sebeplerinden birisi tahmin edebilme özelliğidir. Bu açıdan yapılan literatür

çalışmalarından birinde bir ülkenin enerji tüketim miktarı tahmin edilmek istenmiştir [50]. Bu sayede yıllar sonrası için uzun dönemli enerji tüketim tahmini yapılmıştır. Bu tahmin yapılırken YSA modelinin hesaplama hızı, kompleks ve lineer olmayan fonksiyonlarla başa çıkabilme özelliği, sağlam yapısı ve yüksek verimlilik özellikleri kullanılarak gerçekleştirilmiştir. YSA uygulaması için uzun dönemli elektrik tüketimini etkilediği düşünülen dört farklı giriş parametresi kullanılarak tüketilen gücün bulunması çalışmasına ait analizler yapılmıştır.

YSA modelinin performansı literatürdeki iki farklı metot olan lineer regresyon modeli ve destek vektör makinesi modeli hata oranları dikkate alınarak karşılaştırılmıştır. Bu karşılaştırma sonucunda YSA modelinin diğer iki metoda göre yedi farklı yıl için uzun dönemli enerji tüketim tahmini açısından doğruluk seviyesi daha yüksek sonuçlar verdiği görülmüştür. Bu çalışma ile belirtilen bir diğer konu ise YSA ve diğer metotlar için kullanılmak üzere gerçekte ölçülen verilerin metot performansları açısından önemli olduğudur [50]. Bu çalışmadaki yaklaşım sayesinde enerji tüketiminin doğru tahmin edilebilirliği kullanılarak enerji politikalarında doğru adımlar atılabilmesinin sağlanabileceği belirtilmiştir.

Türkiye'nin elektrik enerjisi tüketimini modellemek ve gelecek dönemlerdeki tüketimlerini tahmin etmek için YSA ve çoklu regresyon analizi yöntemlerini kullanarak literatürde yapılmış bir çalışma bulunmaktadır [51]. Bu amaçla altı bağımsız değişken tahmin edici giriş değişkenleri şeklinde kullanılarak elektrik tüketimini tahmin etmeye çalışmaktadır. Bu çalışmada YSA analizinde kullanılacak bağımsız değişkenlerin belirlenmesi için çoklu regresyon analizi modeli kullanılmıştır [51]. 1992 ve 2014 yılları arasındaki elektrik tüketim verileri kullanılarak oluşturulan dört farklı model, farklı bağımsız değişkenlere ait seçenekler şeklinde oluşturularak YSA modeli içerisinde uygulanmıştır.

2015-2023 yılları aralığı için Türkiye'nin elektrik tüketiminin tahmin edilmesi problemine bir çözüm olarak sunulan çalışmada, ileri yönlü çok katmanlı algılayıcı sinir ağı yapısı kullanılmıştır [51]. Türkiye'nin bu dönemdeki elektrik tüketimi, oluşturulan bu tahmin etme modeli sayesinde 337087.4 ve 385006.6 Gwh aralığında değişeceği belirlenmiştir. Bulunan bu değerler TEİAŞ (Türkiye Elektrik İletim A. Ş.) tarafından tahmin edilen enerji tüketim sonuçları ile karşılaştırıldığında daha düşük değerler elde edildiği görülmüştür.

Regresyon analizi ve YSA metotlarını kullanarak orta vadeli yük tahmini yapabilmek için Nijerya'da yapılan bir çalışmada, geçmiş elektrik tüketim verileri kullanılarak yapılan tahmin yöntemleri temelli bir uygulama yapılmıştır [52]. Üç farklı regresyon analizi ve YSA modelleri, yük verilerinin kendine özgü olmalarına rağmen orta vadeli enerji tüketiminin doğru tahmin edilebilmesi amacıyla uygulanabilirliğini göstermek için kullanılmıştır. Modellerin istatistiksel değerlendirmesi, ortalama mutlak hata yüzdesi (MAPE) ve ortalama karesel hata (RMSE) değerlerine göre yapılmıştır. Yapılan karşılaştırma çalışması sonucunda YSA modelinin yük verisini tahmin edebilme açısından MAPE değerini 0.0285 ve RMSE değerini 1.124 olarak bulan sonuçlar ürettiği görülmüştür. Bu sebeple YSA modelinin regresyon analizine göre daha iyi sonuçlar verdiği belirlenmiştir.

Yük profili tahmini, elektrik kurumlarının verimli ve ekonomik işlemler yapabilmeleri yönünden çok önemli bir rolü bulunduğu için akademisyenler ve araştırmacılar tarafından ilgi çeken bir konu olmuştur [53]. Buna benzer sebepler ile yapılan bir çalışmada, YSA ve regresyon analizi modellerine ait yöntemler kullanılarak elektrik firmalarının geçmiş verileri sayesinde günlük, haftalık ve aylık yük tahminlerinin yapılabilmesi amaçlanmıştır.

Değiştirilerek kullanılan YSA modeli, lineer regresyon yöntemi ile birleştirilerek 2020 yılına kadar olan yük tahminlerinin yapılabilmesi için kullanılmıştır. Bu teknik, geleneksel lineer regresyon modeli olarak akış eğrilerinin tahmin edilmesi ile karşılaştırılmıştır. Sonuç olarak teklif edilen düzenlenmiş modelin daha uygulanabilir ve verimli sonuçlar ürettiği görülmüştür. Bu çalışma 2020 yılına kadar yük talebinin tahmin edilebilmesi amacıyla YSA analizi uygulamasının kullanılabileceğini göstermiştir [53].

Regresyon yöntemleri, tahmin problemlerinde çokça kullanılan metotlardan bir tanesidir. Buna benzer olarak yapılan literatürdeki bir diğer çalışmada, kümelenmiş lineer regresyon yöntemi (CLR - Clusterwise Linear Regression) kullanılarak Avusturalya'nın Victoria eyaletindeki aylık yağış miktarlarının tahmini uygulaması yapılmıştır [54]. Bu yöntem, kümelenme ve regresyon yöntemlerinin bir araya getirilmesinden oluşmaktadır. Yapılan bu çalışmada, bir optimizasyon problemi olarak formüle ederek ve bunu çözebilmek için artımlı algoritma kullanarak uygulanmıştır.

Bu amaçla bahsedilen algoritma uygulanarak coğrafi olarak farklı sekiz hava istasyonundan 1889-2014 yılları arasında kaydedilen beş farklı meteorolojik değişken,

giriş verisi olarak kullanılarak aylık yağış miktarının tahmin edilebilmesi amaçlanmıştır. Bu yöntemin tahmin performansını değerlendirmek için, tahmin doğruluğu ölçümünü yapan dört farklı yöntem kullanılarak karşılaştırılmıştır.

Hesaplanan sonuçlar kullanılarak beklenti - maksimizasyon algoritması, çoklu lineer regresyon, YSA ve destek vektör makineleri kullanılarak yapılan karşılaştırma sonucunda teklif edilen algoritmanın daha iyi sonuçlar ürettiği görülmüştür [54].

Kısa dönemli yük tahmini, güç sistemleri ve akıllı şebekelerin işletmeleri için hayati öneme sahip olduğundan bu konuda yapılan bir literatür çalışmasında bir saatten bir haftaya kadar zaman dilimleri için enerji tüketim tahmini çalışması yapılmıştır [55]. Bu süreç, özellikle serbest enerji piyasası dikkate alındığında dağıtım şirketleri ve büyük kapasiteli elektrik müşterileri açısından zorunlu bir durum haline gelmiştir.

Elektrik üretimine ait planlamalar elektrik tüketim tahminlerine göre yapılmaktadır. Eğer, enerji tüketim tahmini olması gerekenden yüksek yapıldıysa, gereksiz yere fazla sayıda üretim birimi başlatılacak ve üretim maliyetinin artmasına sebep olacaktır. Tam tersi bir durum olarak enerji tüketim tahmini talepten daha düşük olursa bu kez güç sisteminde elektrik kesintileri oluşabilecektir. Bu probleme bir çözüm olarak sunulan bu çalışmada, yapay arı kolonisi ve yapay sinir ağları yöntemlerinin hibrit bir karışımı olan bir yöntem kısa dönemli elektrik tüketim tahminini yapmak üzere sunulmuştur [55].

Yapay arı kolonisi algoritması, YSA metodunun öğrenme sürecinde nöronlar arasındaki ağırlık değerlerinin optimize edilmesi amacıyla kullanılmıştır. Geçmiş yük bilgileri, sıcaklık farklılıkları ve mevsimler model için giriş parametreleri olarak tasarlanmıştır. Üç yıllık döneme ait saatlik enerji tüketim verileri ağ yapısının eğitilmesi amacıyla, bir yıllık döneme ait saatlik veri ise ağın testi amacıyla kullanılmıştır. Farklı günler ve dönemler için yapılan bu çalışma sonucunda, hibrit yöntemin ortalama mutlak yüzde hata parametresi hesaplanarak, 2.12 ortalama değere sahip sonuçlar verdiği görülmüştür [55].

Uzun dönemli elektrik talebinin doğru bir şekilde tahmin edilebilmesi, talep tarafı yönetimi (Demand Side Management), elektrik şebekesinin ağ yapısı ve işletmesi açısından çok önemli bir paya sahiptir [56]. Talebin normalden yüksek tahmini elektrik şebekesindeki varlıklara ve sistemlere gereğinden yüksek yatırım yapılması ve elektrik fiyatlarının yükselmesi anlamına gelecektir. Talebin olması

gerekenden daha düşük tahmin edilmesi ise şebekenin güvenilirliği ve güvenliği açısından olumsuz sonuçlar oluşturabilecektir.

Bu sebeplerle derin sinir ağları kullanılarak yapılan literatür çalışmasında Avusturalya'nın uzun dönemli elektrik tüketim tahmini yapılmak istenmiştir [56]. Bu çalışmada otomatik kodlayıcı ile çok katmanlı algılayıcıyı farklı bir yöntem ile bir araya getirilerek tüm ülkenin elektrik tüketiminin tahmini 1-24 aylık dönem için yapılmak istenmiştir. Bu çalışma sonucunda ulaşılan deneysel sonuçlar göstermiştir ki klasik sinir ağlarına göre bu derin sinir ağı yapısı özellikle 12-24 aylık dönem için daha yüksek performanslı sonuçlar üretmiştir [56].

Daha önceki incelenen literatür çalışmalarında da bahsedildiği gibi doğru elektrik yük tahmini büyük ekonomik ve sosyal kazançlar sağlayabilmektedir. Bir diğer literatür çalışmasında da derin sinir ağları kullanılarak elektrik tüketim tahmini yapılmak istenmiştir [57].

Bu amaçla öncelikle yük tahmini ile ilgili olan parametreler incelenerek belirlenmiştir. Derin sinir ağları yapıları incelenerek yapılan çalışmada derin sinir ağı yapısı farklı ön eğitim teknolojileri ile kullanılmıştır. Farklı sinir ağları modellerinin performansları değerlendirilerek, nispeten daha büyük veri seti kullanılarak sunulan yöntemin avantajları gösterilmiştir. Deneysel sonuçlar derin sinir ağlarının yük tahmini konusundaki avantajlarını göstermiştir [57].

Literatürde, sinir ağları temelli yapılan ve aylık, günlük ve saatlik verilerin ağ yapısı için veri kaynağı olarak değerlendirildiği bir diğer çalışmada üç boyutlu sinir ağı modeli kullanılarak bu modelin potansiyeli açıklanmıştır [58]. Üç boyutlu ağ olarak sunulan hipotezde, bu yapının seçeneklerine ve avantajlarına odaklanılmıştır.

MATLAB programı kullanılarak yapılan çalışmada bir yıllık süreye ait aylık veriler, bir aylık süreye ait günlük veriler ve on iki aylık süreye ait saatlik veriler yapay sinir ağı tasarımı için kullanılmıştır [58]. Üç boyutlu sinir ağı yapısının elektrik güç tüketimi açısından kullanılmasının en temel avantajı her saat, her gün ve her ay için güç tüketim tahmini yapılabilmesi ile elektrik ağ şebekesinin kontrol verimliliğini yükseltme ihtimalidir. Buna ek olarak bu yapı sayesinde, daha yüksek tahmin doğruluğu ile sonuçlanması amacıyla başlangıç verilerinin ortalamalarını almaya ihtiyaç olmadığı açıklanmıştır [58]. Bu çalışma göstermiştir ki çok sayıda sinir ağının kullanılması ile üç boyutlu bir yapı özelliği kazandırılmıştır. Üç boyutlu bu tasarımın,

iki boyutlu ağ yapılarına göre daha yüksek doğrulukta elektrik yük tahmini yapabileceği belirtilmiştir.

Veri merkezlerinin enerji verimliliğinin ölçümü ile ilgili oluşturulan ilk ölçüm parametresi Güç Kullanım Verimliliği (PUE) hesaplaması ve anlaması kolay bir parametre olarak 2007 yılında ortaya çıktığı belirtilen çalışmada, veri merkezlerindeki bilgi teknolojisi cihazlarının enerji tüketimlerini negatif yönde etkileyecek parametreler de belirtilmiştir [32]. Kabul gören bu parametreler, sunucuların düşük iş yükü ile kullanımı, verimsiz ağ yapısı yönetimi, sınırlı sanallaştırma uygulanması ve iş modelindeki eksiklikler şeklindedir.

Çevresel Koruma Ajansı (Environmental Protection Agency (EPA)) tarafından 2008 yılında yayınlanan bir raporda Veri Merkezleri için bir uygulama projesi hazırlanmıştır. Bu çalışma 17 vaka çalışmasını Silikon Vadisi Liderlik Grubu'na üye olan şirketler ile gerçekleştirilmiştir [33]. 2008 yılında yapılan bu çalışma, veri merkezlerindeki yeni teknolojilerin benimsenmesi konusunda yapılan öncü çalışmalardan olmuştur ve esas amacı, bugünün yeni teknolojilerini kullanarak enerji tasarrufu sağlanabildiğini göstermektir.

Bu çalışmaya göre veri merkezindeki enerji tüketimi, Bilgi Teknolojileri (BT) altyapısı ve saha altyapısı olarak sınıflandırılmıştır. BT altyapısı, sunuculardan, veri depolama ekipmanlarından ve ağ yapısından oluşmaktadır. Bu çalışma göstermiştir ki, çok çekirdekli işlemciye sahip cihazlar ve güç yönetim teknolojilerinin kullanımı yöntemleri, sunucu gibi BT altyapısının parçası olan cihazlarda daha verimli güç tüketim değerleri ile sistemlerin çalışabilmesini sağlayabilmektedir. Sunucu tabanlı sistemlerin ayrıca, konsolidasyon, sanallaştırma, verimli kaynak yönetimi ve gelişmiş kendine özgü yöntemler sayesinde daha verimli çalıştırılabildikleri ifade edilmiştir.

Aynı çalışmada [33] depolama üniteleri ve ağ yapısı da enerji tüketimi açısından analiz edilmiş ancak sunuculara göre daha az etkilere sahip olduğu anlaşılmıştır. Bu çalışmanın sonuçları enerji tasarrufu ve sera gazı etkisi açısından değerlendirilmiş ve her bir veri merkezinin enerji tasarrufu sağlayabileceğini göstermiştir.

Bulut veri merkezleri ile ilgili yapılan bir diğer çalışma [34] ise ağ yükü talebini olasılıksal ve sinir modelleri ile tahmin etmeyi amaçlamaktadır. Yük talep tahmini ve olasılıksal durum geçiş modeli yeni bir çerçeve olarak sunulmuştur. Bu model ile bulut veri merkezlerinin gerekli performans seviyeleri sağlanmak koşulu altında enerji tüketimini azaltarak en uygun seviyede bulut kaynak tahsisini sağlayabileceği ifade

edilmiştir. Bunu sağlamak için sinir ağlarından ve ardışık bağımlı lineer tahmin algoritmalarından faydalanılmıştır. Sonuç olarak, bu çalışmada sinir ağı ve lineer tahmin edici modellerin verimli bir şekilde gelecek yük tahminlerini yapabildikleri görülmüştür.

2015 yılında Almanya‘da yapılan bir çalışmada [35] veri merkezleri altyapılarının yeni kurulumu veya modernizasyonu işlemleri, bir önceki yıla göre % 10 artış göstererek yıllık 900 milyon avro değerinde yatırım yapıldığı ifade edilmiştir. 2025 yılı için veri merkezlerinin tahmin edilen yıllık enerji tüketim rakamı 16.4 Milyar kWh/Yıl şeklindedir. Ayrıca bu çalışma göstermektedir ki, veri merkezlerindeki yıllık en yüksek güç tüketimine sahip cihazların sunucular ve depolama ekipmanları olduğudur.

2013 yılında yayınlanan bir çalışmada veri merkezlerinin enerji kullanımlarını tahmin edebilen bir yazılımın sunumu yapılmıştır [36]. Bu uygulama 15500 metrekare alana kurulu, 332 kilowatt kurulu güce sahip bilgi teknolojileri cihazları içeren konvansiyonel bir veri merkezi için gerçekleştirilmiştir. Yazılım, veri merkezinin soğutma sisteminin, elektrik güç dağıtım sisteminin, elektrik enerjisi kullanımının ve sahadan tahliye edilen sıcaklık miktarının tahminlerini yapabilmektedir. Bilgisayar ekipmanları tarafından gerçekte tüketilen güç değerleri ise, her bir ekipmanda kurulu bulunan Güç Dağıtım Üniteleri üzerinden ölçülmüştür. Yazılımın tahmin yeteneği ön analizler yapıldıktan sonra daha doğru bilgilerin girişi yapılarak iyileştirilmiş ve varyans değeri % 1.2 değerine kadar indirilmiştir. Veri merkezindeki mevcut yük ve fan sistemleri açısından farklı senaryolar için denemeler gerçekleştirilmiştir. Açıklanan yöntemler ile yapılan bu uygulama ve yazılım sayesinde veri merkezinin enerji tüketimi simüle edilerek, enerji verimliliğini artırabilecek modeller üretebileceği ifade edilmiştir [36].

Veri merkezlerinin enerji tüketimlerinin % 50’ye yakın miktarını soğutma sistemleri oluşturmaktadır [37]. Bu yüzden termal yönetim, modern veri merkezleri için önemli bir alan oluşturmaktadır. Yeni bir yaklaşım sunan bir diğer çalışmada ise veri merkezi tasarımındaki dinamikler ve çalışma yükünden kaynaklı aşırı ısınma durumlarını tahmin edebilecek önleyici bir model düşünülmüştür [37].

Bu termal tahmin yöntemi modelinin temelinde veri merkezindeki sunucuların etrafındaki sıcaklık değerlerinin tahmin edilmesi bulunmaktadır [37]. Bu amacı başarabilmek için, siber fiziksel etkileşimleri kaydedebilme ve bu verilerden otomatik

olarak öğrenebilme, veri merkezleri büyüklüklerine göre hesaplama ve fiziksel açıdan ölçeklenebilir olma ve gerçek zamanlı sensör ölçümleri ile sürekli tahmin yapabilme özellikleri bulunmaktadır. Bu çalışmanın literatüre katkıları ise sistematik bir yaklaşım sunarak, fiziksel yasaları ve sensör ile ölçüm tekniklerini kullanmak, veri merkezinin siber fiziksel yapısına ait parametrelerin öğrenilmesini sensör verilerini kullanan bir algoritma sayesinde uygulamaya çalışmak şeklinde olmuştur [37].

Sahadaki ölçüm verilerinden elde edilen öğrenme yapısı sayesinde bu analiz ile sunucuların termal durumlarına ait tahminler yapılmıştır [37]. Bu algoritma ile sunucuların termal durumlarını önceden tahmin edebilen bir yapı kurulmaya çalışılmıştır. Bu modelde kullanılan öğrenme algoritması katsayıların öğrenmesi amacıyla lineer en küçük kareler yöntemini kullanmıştır [37].

Veri merkezlerinin sıcaklık dağılımları karmaşık olsa da bu çalışmada, yoğun bir ensturmantasyon ve tanımı yapılan gri-kutu termodinamik modeli ile sunucuların termal durumlarının gerçek zamanlı olarak izlenebildiği görülmüştür [37]. Makine öğrenme süreci sonunda 4.2 dakika önceden termal alarmların başarılı bir şekilde tahmin edilebildiği görülmüştür.

Veri merkezlerindeki enerji tüketimlerini tahmin edebilmek için kullanılan yöntemlerden bir tanesi bu yapının en önemli parçalarından olan sunucuların enerji tüketimlerine odaklanan çalışmalar olmuştur [38]. Bu çalışmalardan bir tanesinde güç tüketimin tahmin etmek için kullanılan modelde incelenen parametreler, sunucunu işlemcisinin, rastgele erişimli belleğinin (RAM) ve ağ arabirim denetleyicisinin (Network Interface Controller) tükettiği güç olmuştur. Modelde, Lasso lineer regresyon yöntemi kullanılmıştır [38].

Bu çalışmada üç adet donanım yapısını işleme katarak giriş parametresi olarak kullanan, negatif olmayan katsayılar ile tasarlanmış Lasso modeli temelli lineer regresyon yöntemi kullanılmıştır [38]. Lineer bir güç tahmin yöntemi kullanılması ile 30 adet regresyon parametresi başlangıç değerleri verilerek işlem yapılmıştır. 2730 adet toplam veri ile analiz gerçekleştirilmiştir. Bu çalışma ile sunucuların en yüksek aktivite anlarında daha fazla ya da daha az enerji tüketmelerine bağlı olarak yapılacak yönlendirmeler ile enerji tasarrufu sağlanabileceği ifade edilmektedir [38].

Veri merkezlerinin enerji tüketimlerini ve dolayısı ile enerji maliyetlerini düşürebilmek için kullanılan yöntemlerden bir tanesi de sunucu konsolidasyon teknikleri olmuştur. Konsolidasyon kelimesi, veri merkezlerindeki benzer yapıdaki

cihazların birleştirilmesi manasında kullanılan bir terimdir. Veri merkezlerinde birden fazla tek başına çalışan cihazların bir ağ sunucusu üzerinde birleştirilmesi konsolidasyon işlemi için örnek olarak gösterilebilir.

Bu konuda yapılan çalışmalardan bir tanesinde konsolide edilebilme karakteristiğini bulabilme açısından kurumsal bir sunucunun iş yükü analizlerine yer verilmiştir [39]. Bu çalışma ile anlaşılmıştır ki iş yükü ve uygulama talebinin en az olduğu durumlarda konsolidasyon işlemi önemli bir güç tasarrufu sağlayabilme imkanı sunmaktadır. Ancak, uygulamalar arasındaki korelasyon konsolidasyon işleminde hesaba katılmaz ise önemli kapasite aşımlarına sebebiyet verebileceği görülmüştür. Bu çalışma sonucunda PCP(Peak Clustering based Placement) olarak isimlendirilen konsolidasyon metodu, üstün güç tasarrufu sağlayabilmiş, düşük kapasite aşımları oluşmuş ve sunucular arasında başarılı bir yük dengesi kurduğu görülmüştür [39].

Enerji verimli veri merkezleri oluşturabilmek için yapılan konsolidayson yöntemi içeren çalışmalarından bir diğerinde sahada bulunan cihazların açılıp kapatılması, güce duyarlı konsolidasyon algoritmalarının çalıştırılması ve makine öğrenme teknikleri kullanılarak "akıllı" olarak isimlendirilen bir konsolide etme tekniği sunulmuştur [40].

Makine öğrenme yaklaşımı kullanılarak geçmiş sistem davranışlarından öğrenebilen tasarımlar ile güç tüketim seviyeleri, işlemci yük oranları tahminleri yapılabilmekte ve hizmet seviyesi anlaşmalarına (Service Level Agreement) ait zaman planlaması oluşturulabileceği belirtilmiştir. Bu çalışmada güç tüketimini lineer olmayan bir ilişki ile etkileyen bir parametre olan Cpu kullanım oranı bulunmaktadır. Güç tüketimini tahmin etmek basit lineer regresyon ile analiz etmeye göre daha karmaşık bir işlem olduğu için M5P (M5 Model Trees) makine öğrenme algoritması kullanılmıştır [40].

Bu çalışmadaki yaklaşım, watt olarak yapılan tüketim değerinden iş yükü özelliklerine, çapraz disiplinleri ve farklı teknikleri bir araya getirdiği için dikey bir yapıda oluşturulmuştur [40]. Gerçek bir senaryonun tüm kontrol döngülerini içeren, heterojen iş yüklerini temsili olarak barındıran bir çalışma olarak geleneksel uygulamalara alternatif olarak yapılmıştır.

Sonuç olarak zamanlama fonksiyonları dahilinde operasyonların yapılabilmesi, hizmet seviyesi anlaşmalarının zamanlamaları ve konsolidasyon hareketleri üzerinde makine öğrenme modellerinin oldukça faydalı olduğu anlaşılmıştır. Ayrıca

konsolidasyon kullanmayan yöntemlerin daha düşük enerji verimliliklerine sahip olduğu görülmüştür.

Veri merkezlerinin enerjiye duyarlı yönlendirme (routing) bakış açısıyla güç tüketimlerinde tasarruf yapılıp yapılamayacağını inceleyen çalışmalar bulunmaktadır [41]. Veri merkezleri çoğunlukla tam kapasiteden uzak bir iş yükü ile çalışmaktadırlar. Birçok veri merkezi yöneticisini rahatsız eden ise bu durumda boşta bekleyen ağ cihazlarının tükettikleri enerji olmaktadır. Enerjiye duyarlı yönlendirme olarak isimlendirilen bu yöntemde ise en kritik nokta ağ performansında bir olumsuzluk oluşturmayacak şekilde mümkün olan en az sayıda ağ cihazı ile çalışabilmek olmaktadır.

Bunun için boşta bekleyen cihazların enerji tasarrufu amacıyla kapatılabileceği ya da uyku moduna alınabileceği belirtilmiştir. Bu çalışma dâhilinde yapılan simülasyonlar, enerji duyarlı yönlendirme işlemlerinin ağdaki iş yükü yüksek seviyelerde değilken enerji tasarrufuna fayda sağladığını göstermiştir.

Bir diğer çalışmada veri merkezindeki trafik yükünü değiştirerek güç tasarrufu yapabilecek bir yöntemden bahsedilmiştir [42]. “Elastik Ağaç” olarak isimlendirilen bu yöntem dinamik olarak aktif ağ elemanlarını düzenlemekte ve veri trafiğini yönlendirmektedir. “Elastik Ağaç” yapısı aynı zamanda tüm ağ üzerinde çalışabilecek bir güç yöneticisi olarak da isimlendirilebilir. Bu çalışmada enerji verimliliği, performans ve sağlamlık açısından gerçek bir e-ticaret sitesi üzerinden incelenmiştir. Bahsedilen bu yönlendirme işlemleri için minimum güç tüketimine sahip alt ağ yapıları için farklı senaryolar test edilmiştir [42]. Bu yöntem ile veri merkezlerinde trafik dalgalanmaları ile başa çıkılarak devamlılık sağlanmakta ve % 50’ye yakın bir enerji tasarrufu oluşturulabileceği belirtilmektedir [42].

Veri merkezindeki enerji tüketimini azaltmak için veri trafiğini birleştirme yöntemleri bazı çalışmalar ile sunulan tekniklerden bir tanesi olmuştur. Benzer bir çalışmada veri merkezlerinin çoğunlukla % 5 ile % 25 aralığında iş yükü oranları ile çalıştıkları, ancak tüketilen enerji miktarının maksimuma yakın olduğu belirtilmiştir [43]. Buna çözüm olarak iş yükü ile orantılı enerji çeken “enerji orantılı” veri merkezleri fikri benimsenmiştir. Trafik toplanması (traffic aggregation) olarak da isimlendirilebilecek bu yöntem ile üzerinde düşük ağ trafiği bulunan cihazlardaki yük, daha yüksek ağ trafiği olan switch gibi cihazlarda birleştirilebilmektedir.

Bu çalışmada ağ topolojisi olarak yüksek performanslı veri merkezi ağlarında FBLY (the flattened butterfly) olarak isimledirilen topolojinin doğasından kaynaklı sebeplerle daha iyi güç verimliliğine sahip olduğu belirtilmiştir [43]. Veri merkezlerinde ağ trafiğini daha az yoğun hatlardan daha yoğun hatlara yönlendirerek bu fikir uygulanmaya çalışılmıştır. Daha önceleri geliştirilen trafik modelleri ve simülasyonlar kullanılarak % 5 iş yükünde % 49 enerji tasarrufu sağlandığı görülmüştür. Bu literatür çalışmasının bir diğer sonucu ise, ağların birleştirilmesi yöntemi ile veri merkezlerinin çok yalın bir hal alabileceğinin anlaşılmış olmasıdır [43].

Veri merkezlerinin, karlılığı artırabilmek için güç tüketimini ve barındırılan uygulamalardaki hizmet seviyesi anlaşmalarındaki ihlalleri en aza indirebilmesi, günümüzdeki en önemli amaçlarından olmaktadır. Veri merkezlerinin güç yönetimi güç tasarrufunun önemli bir parçası olmaktadır. Bu amaçla yapılan çalışmalardan bir tanesi servis kalitesine duyarlı güç yönetiminin sanallaştırılmış veri merkezlerinde kullanılması uygulamasıdır [44].

Modern veri merkezleri günümüzde rutin olarak sanallaştırma işlemlerini, uygulama taşınabilirliği, uygulama izolasyonu sağlarken kaynak çoklayıcı amaçlarıyla kullanmaktadırlar. Sanallaştırma, veri merkezlerindeki birden çok sanal makinanın tek bir fiziksel makine üzerinde çalıştırılarak fiziksel kaynakların dinamik paylaşımlarına imkân vermektedir. Harcanan güç değerini en aza indirmeye çalışırken önemli bir gereklilik ise güvenilir servis kalitesinin müşterilere verilmesidir.

Entegre bir yönetim çözümünün sanal makine ölçeklendirme ve sunucu konsolide etmenin avantajlarını kullanarak sanallaştırılmış veri merkezlerindeki enerji verimliliğini ve servis kalitesini artırmayı başardığı görülmüştür [44]. Daha az sayıda fiziksel sunucu üzerinde konsolide edilen çok sayıda sanal makine güç tüketimi açısından verimli sonuçlar vermiştir. Bu çalışmanın daha öncekilere göre farklılıkları ise lineer programlama, karınca kolonisi optimizasyonu ve kontrol teorisi tekniklerini kullanmasıdır [44].

Son yıllarda veri merkezlerinin enerji tüketimlerinin artışı ile birlikte enerji verimli tasarımlar ve soğutma sistemlerinin kontrolü ve takibi alanında birçok çalışma yapılmıştır. Enerji sistem simülasyonları bu iş için kullanılabilecek verimli bir araçtır. Fakat veri merkezlerindeki mevcut bilgi teknolojileri sistemlerine ait modeller, veri

merkezinin soğutma ihtiyacı ya da talebi üzerinde BT cihazlarının tasarım ve çalışma durumlarının etkilerini iyi bir şekilde temsil edememektedir.

Bu probleme çözüm bulabilmek için yapılan bir çalışmada veri merkezi enerji sistem simülasyonlarında doğrudan kullanılabilecek BT cihazları için güç tüketim modeli sunulmuştur [45]. Bu tasarım, sunucuların modelleri, ağ cihazlarının modelleri ve güç dağıtım ekipmanlarına ait modellerden oluşmaktadır. Bu enerji modeli, veri merkezindeki tipik bir IT cihazının enerji performansını gerçek zamanlı dinamik işlem şartlarında simüle etmek için kullanılabilmektedir.

Veri merkezleri ve onların enerji yük profilleri ile ilgili çalışmalar özellikle dünya enerji tüketimindeki önemi sebebiyle yakın zamanda daha çok ilgi çeken bir konu halini almıştır. Artan dünya nüfusu ve gelişen internet tabanlı teknolojiler sebebiyle veri merkezlerinin önemi ve sayısı gittikçe artmaktadır. Bu sebeple veri merkezlerinin tükettiği enerji miktarları da sürekli artarak dünya enerji tüketiminden daha yüksek paylar almaktadırlar.

Tez çalışmamda, bu konuda daha önce yapılan literatür çalışmaları incelenmiş, konumuza en yakın ve uygun olanları seçilerek yukarıdaki bölümlerde bahsedilmiştir. Veri merkezlerinin enerji tüketimleri üzerine yapılan çalışmaların geçmişi çok eski tarihlere dayanmadığı için bu konuda çok yüksek adette ve çeşitte çalışma bulmak mümkün olamamıştır. Ancak araştırılan birçok çalışmada veri merkezleri ve onlara ait enerji yük profillerinin, genel parametreler ile enerji verimliliği üzerindeki etkilerinin incelendiği görülmüştür.

Bu çalışmamda ise daha önce yapılan çalışmalardan farklı bir bakış açısı ile veri merkezleri ve IT sistemleri için cihaz bazında enerji tüketimi incelenerek bir yük profili oluşturulmaya çalışılmıştır. Bu model daha sonra veri merkezleri için rahatlıkla kullanılabilecektir. Yük profili belirlenecek cihazlara ait içsel ve dışsal parametrelerin eş zamanlı olarak kaydedilmesi sayesinde daha doğru yük profilleri üretebilmek hedeflenmektedir. Literatür çalışmalarında, veri merkezleri ve IT sistemlerinde çalışan cihazların içsel parametreleri olan Cpu Kullanımı, Ram Kullanımı veya Cpu Sıcaklığı gibi parametrelerin bir arada incelenmediği görülmüştür. Bu parametreleri dikkate alarak veri merkezleri ve IT sistemlerine ait cihazlar için yük profili üretme çalışmaları anlamında eksiklikler olduğu düşünülmüş ve bu alanda bir çalışma yapılarak literatüre katkı sağlanması hedeflenmiştir.

3. TASARIM ve YAZILIMLAR

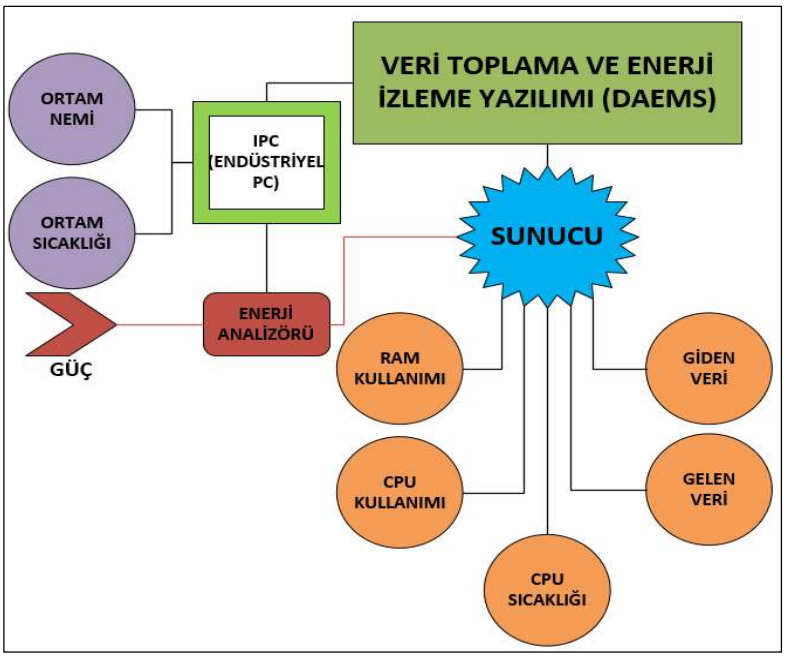
3.1. Tasarım Prensipleri

Tez çalışmamda, karmaşık veri merkezlerinin yerine bir sunucu kullanarak yapay sinir ağları ve regresyon yöntemleri sayesinde yük tahminleri yapılacaktır. Öncelikle tahmin yöntemlerinin giriş parametrelerini oluşturacak olan özel giriş değişkenleri belirlenecektir. Bu değişkenler aşağıdaki şekilde

- PC tarafından tüketilen enerji ile ilgili: Voltaj, Akım, Güç,
- PC'nin Ethernet bağlantıları ile ilgili: Gelen Veri Miktarı, Giden Veri Miktarı,
- Ortam şartları ile ilgili: Ortam Sıcaklığı, Ortam Nem Oranı,
- PC'nin işlemcisi ile ilgili: Cpu Kullanım Oranı, Ram Kullanımı, Cpu Sıcaklığı

olarak listelenmiştir.

Veriler, endüstriyel PC (PLC) vasıtasıyla Şekil 3.1'de gösterilen yerleşim şemasına göre eş zamanlı bir şekilde toplanacaktır. Bu tasarım, DAEMS yazılımı, endüstriyel PC, sunucu, enerji analizörü, sıcaklık ve nem sensörleri vs.den oluşmaktadır.



Şekil 3.1: Yerleşim Şeması.

DAEMS yazılımının üzerinde çalıştığı sunucu, iki adet Ethernet portuna sahiptir. Bu sayede hem günlük internet ağ trafiğinin çalışmasına izin vermekte hem de endüstriyel PC ile olması gereken ağ haberleşme bağlantısı görevini üstlenmektedir.

Endüstriyel PC, uygun giriş/çıkış modülleri sayesinde sahada bulunan sıcaklık ve nem sensörleri ve enerji analizörü ile bağlantılıdır, EtherCAT protokolü sayesinde anlık ve eşzamanlı olarak haberleşebilmektedir. Endüstriyel PC, sunucunun bulunduğu ortama ait sıcaklık ve nem bilgilerinin yanı sıra sunucun enerji beslemesine uygun bağlantılar yapılmak suretiyle elde edilen enerji analizörü sayesinde sunucuya ait güç tüketim bilgilerini de anlık olarak toplamaktadır.

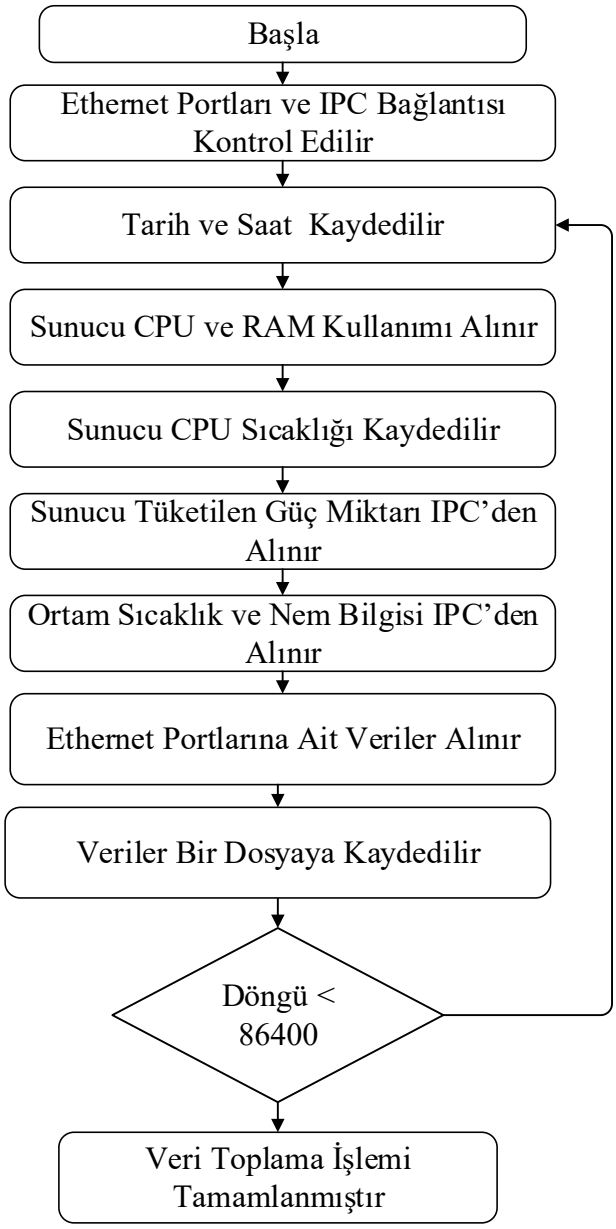
Enerji analizörü sunucuya ait sadece güç bilgisini değil aynı zamanda akım ve gerilim bilgilerini de anlık olarak endüstriyel PC'ye göndermektedir. Enerji analizörü, endüstriyel PC'ye EtherCAT protokolü ve uygun giriş/çıkış modülü sayesinde bağlanabilir. Şekil 3.1'de yerleşim şemasında görülebileceği üzere, sunucu üzerinden alınmak istenen parametrelerden Ram ve Cpu kullanımı, Cpu sıcaklığı ile Gelen ve Giden Veri parametreleri yine sunucu üzerinde çalışmakta olan DAEMS yazılımı sayesinde gerekli analiz ve karşılaştırma işlemlerimiz için kayıt altına alınacaktır.

3.2. Geliştirilen Yazılımlar

Lineer olmayan bağımlılıklar, enerji verimliliğini anlayabilmek ve optimize edebilmek amacıyla elektrik yükünü tahmin etmeyi zorlaştırmaktadırlar. Bu sebeple, özel tasarlanmış DAEMS yazılımı, ilgili parametreleri belirleyerek bunları kayıt altına alabilmek için geliştirilmiştir.

DAEMS yazılımı sunucu üzerinde çalışır ve endüstriyel PC ile EtherCAT protokolü üzerinden bağlantılıdır. DAEMS yazılımının sunucu üzerinde çalışma sebebi Şekil 3.1'deki yerleşim şemasında görülebileceği üzere bir adet sunucu ile bu çalışmayı yapacağımız için ilave bir bilgisayara ihtiyaç duymayışımızdır. Aynı şekilde bu çalışma, birden fazla sunucu veya IT cihazı için yapılmak istenirse bu yazılımı sunuculardan bir tanesinin üzerinde çalıştırarak aynı süreç daha çok cihaz için gerçekleştirilebilecektir.

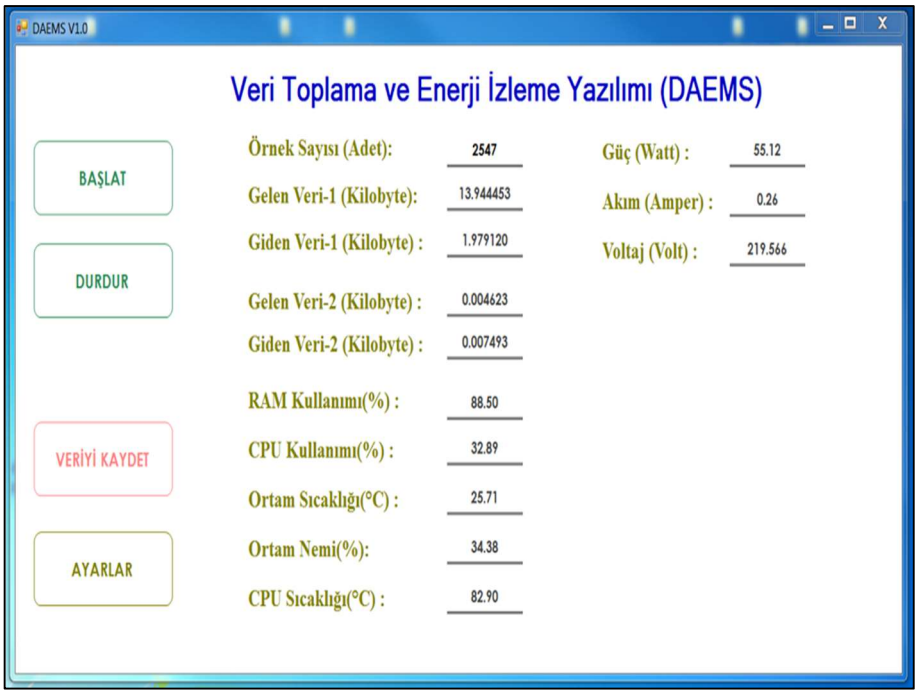
DAEMS, sunucuya ait içsel parameterlere erişebilme yetisine sahip olup bütün bağlı cihazlar ile eşzamanlı çalışarak veri okuyabilme özelliğine sahiptir. Şekil 3.2'de DAEMS yazılımına ait akış şeması görülmektedir.



Şekil 3.2: DAEMS Yazılımı Akış Şeması.

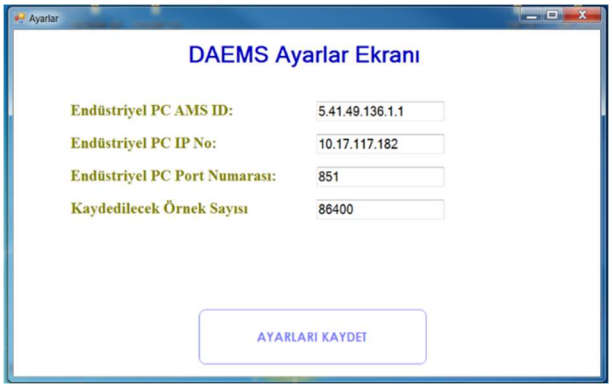
DAEMS yazılımı temelde sunucu-istemci tabanlı bir program olup, bağlantılı birçok endüstriyel PC ve sunucu ile aynı anda çalışabilme özelliğine sahiptir. DAEMS yazılımının kullanıcı arayüzü ana sayfasına ait ekran görüntüsü Şekil 3.3'te görülmektedir.

BAŞLAT butonuna basıldığında yazılım gerekli tüm bilgileri çevre cihazlardan ve üzerinde çalıştığı sunucudan alarak uygun formatta kaydedilecek hale getirmektedir.



Şekil 3.3: DAEMS Kullanıcı Arayüzü.

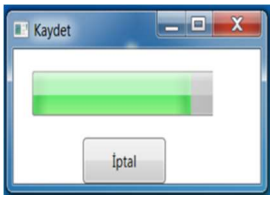
DAEMS yazılımının istenildiği şekilde çalışabilmesi için gerekli parametreler Şekil 3.4'te görülebilen Ayarlar sayfasından girişleri yapılarak yazılıma kaydedilmekte ve uygulanmaktadır.



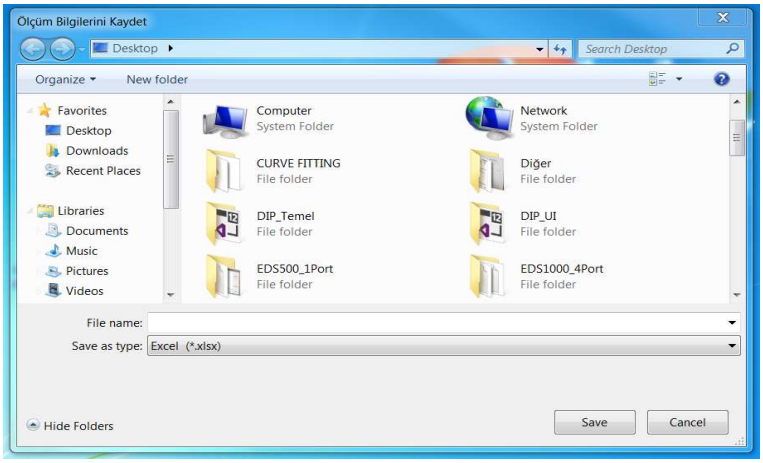
Şekil 3.4: DAEMS Ayarlar Ekranı.

Yapılan ölçüm işlemleri, DURDUR butonuna basıldığında ya da Ayarlar sayfasındaki kaydedilecek örnek sayısına ulaşıldığında tamamlanacaktır. Veri toplama işlemi tamamlandığında, VERİYİ KAYDET butonuna basılması durumunda Kaydet ekranı açılacaktır. Yazılım, ölçüm yapılan veri sayısına göre Şekil 3.5'te görülmekte olan Kaydet sayfasındaki ilerleme çubuğundan görülebileceği üzere veriyi kaydetmek

için hazırlayacaktır. Veri, kaydedilmek üzere hazır olduğunda Şekil 3.6'da görülen Ölçüm Bilgilerini Kaydet sayfası açılmakta ve veri dosyası bilgisayarda istenilen konuma kaydedilmektedir.



Şekil 3.5: DAEMS Kayıt İlerleme Sayfası.



Şekil 3.6: DAEMS Ölçüm Bilgileri Kaydet Sayfası.

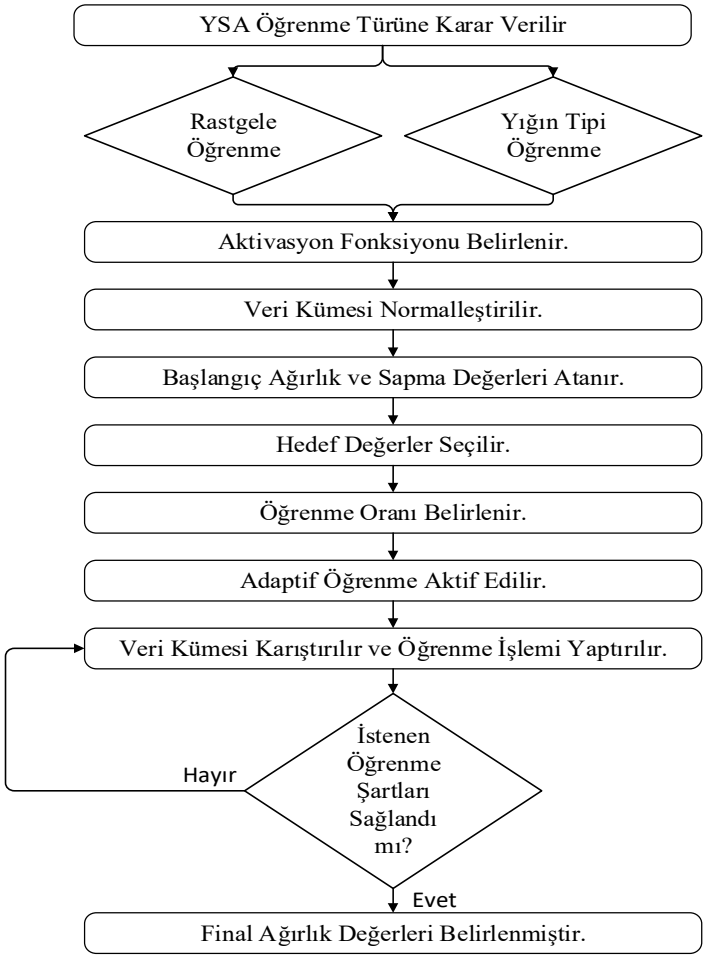
DAEMS, endüstriyel PC'den sıcaklık, nem ve güç tüketimi ile ilgili bilgileri alırken aynı zamanda sunucuya ait içsel parametreleri de toplayarak kaydeder. DAEMS, sunucuya ait şu içsel parametreler olan; Ram ve Cpu kullanım oranı, Cpu sıcaklığı, iki adet Ethernet portuna ait Giden ve Gelen veri miktarı bilgilerini toplayabilmektedir. DAEMS yazılımı endüstriyel PC'ye EtherCAT protokolü ile bağlı olduğu için sıcaklık, nem ve güç tüketimi hakkındaki bilgileri eş zamanlı olarak toplayabilmektedir.

Yazılım, istenen döngü sayısına ulaştığında ya da DURDUR butonuna basıldığında durdurulacaktır. Bu işlemler sonucunda elde edilen kayıtlı dosyadaki ham veriler, gerekli işlemlerden geçirildikten sonra yapay sinir ağına ve regresyon yöntemlerine ait denklemleri eğitmek, performanslarını test etmek ve karşılaştırmak için kullanılacaktır.

Endüstriyel PC üzerinde çalışan bir başka yazılım da “Structure Text” programlama dili ile hazırlanan PLC yazılımıdır. “Structure Text” bir PLC yazılımı programlama dili olup IEC61131-3 Uluslararası standartları ile tanımlanmıştır. Bu yazılım, ortam sıcaklık ve nem bilgileri ile güç tüketimini kaydetmek için kullanılır.

Endüstriyel PC, dijital ve analog giriş-çıkış modülleri sayesinde sensörlerden ve enerji analizöründen veri okumak için kullanılır. Enerji analizörü, 15 bit çözünürlük ile sunucunun tükettiği voltaj, akım ve güç bilgilerini ölçebilme yeteneğine sahiptir.

Sunucunun bulunduğu ortama ait sıcaklık bilgisi, ortamda bulunan PT100 (Sınıf B) tipi $0 \cdots 50$ °C ölçüm aralığına sahip sensörden gelen bilginin endüstriyel PC’de bulunan PT100 okuma modülü sayesinde okunması ile alınmıştır. Ortamın nem bilgisi ise, $5 \cdots 95$ % r.H. ölçüm aralığına sahip 4-20mA çıkış verebilen nem sensöründen gelen verinin yine 4-20mA analog giriş modülü sayesinde okunarak kayıt altına alınabilmiştir.



Şekil 3.7: YSA Modeli Akış Şeması.

Çalışmamda yapay sinir ağlarının uygulanabilmesi için C++ programlama dili ile tez çalışmama özel tasarlanarak hazırlanan YSA uygulama yazılımı kullanılmıştır. Bu yazılıma ait akış şeması Şekil 3.7'de görülmektedir.

Bu şemadan da anlaşılabileceği üzere YSA uygulaması için öncelikle öğrenme türüne karar verilecektir. Sonrasında, yazılımdaki hesaplamalarda kullanılacak aktivasyon fonksiyonuna karar verilerek veri kümesi normalleştirilecektir. Ağ yapısında kullanılacak nöron ve sapma değeri sayılarına göre, başlangıç ağırlıkları ve sapma değerleri atanacaktır.

Veri kümesi için oluşturulan tasarımda hedef değerler belirlenerek ağ yapısının başlangıç öğrenme oranı belirlenecektir. Ağın yaptığı hata miktarına bağlı olarak değişecek adaptif öğrenme özelliği aktif edilecektir. Veri kümesi karıştırılarak öğrenme süreci başlatılacaktır. Yazılımdan istenen şartların üretildiği sonuçlar alınana kadar öğrenme süreci sürekli devam edecektir. İstenen şartlar sağlandığında ise ağ yapısına ait final ağırlık değerleri belirlenmiş olacaktır.

Tez çalışmamda kullanılan diğer analiz metotlarından olan tüm regresyon yöntemleri için Curve Expert isimli analiz programından faydalanılmıştır. Giriş ve çıkış değişkenlerine bağlı olarak oluşturulan fonksiyonlar bu yazılım sayesinde analiz edilerek gerçek ve tahmin edilen yük profilleri arasındaki en yakın ve uygun eğri üretilmeye çalışılmıştır.

Bu istatistiksel test yöntemi değişkenlere ait değerlerin, bağımlı değişkene ait hesaplanan hata oranına bağlı olarak yeniden hesaplanması ve daha doğru değerlere yakınlaştırılması şeklinde uygulanması ile elde edilmiştir.

4. METOTLAR VE UYGULAMALAR

Bu bölümde öncelikle veri toplama metodumuza ait bilgiler verilecektir. Daha sonra, tez çalışmamda uyguladığım üç tanesi regresyon yöntemlerinden oluşan metotlar olarak, yapay sinir ağları da dâhil olmak üzere 4 farklı modele ait detaylı bilgiler verilecektir. Bu modeller

- yapay sinir ağları modeli
- doğrusal regresyon modeli,
- doğrusal olmayan regresyon modeli,
- farklılaştırılmış doğrusal olmayan regresyon modeli,

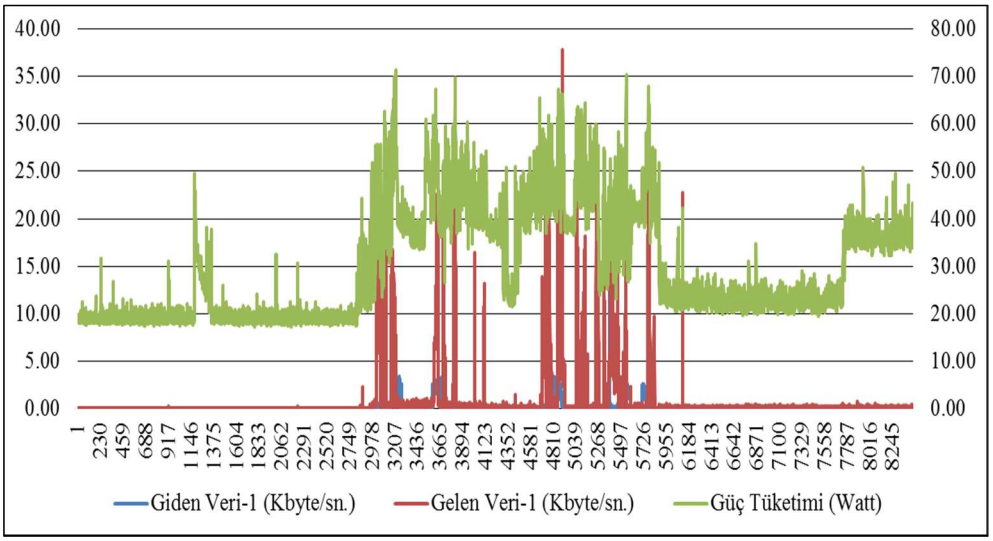
şeklindedir.

Bu tasarıma ait tüm bu test metotları, aynı veri topluluğu kullanılarak, bu verilerin gerektiğinde ham halleriyle gerektiğinde ise normalleştirme ve bir dizi işlemlerden geçirildikten sonra kullanılmıştır. Bu sayede test metotlarına ait gerekli eğitme ve parametre bulma işlemleri sağlıklı bir şekilde uygulanabilmiştir. Bu parametrelere göre metotların analizleri yapılmış, performans testleri gerçekleştirilmiştir.

4.1. Veri Toplama

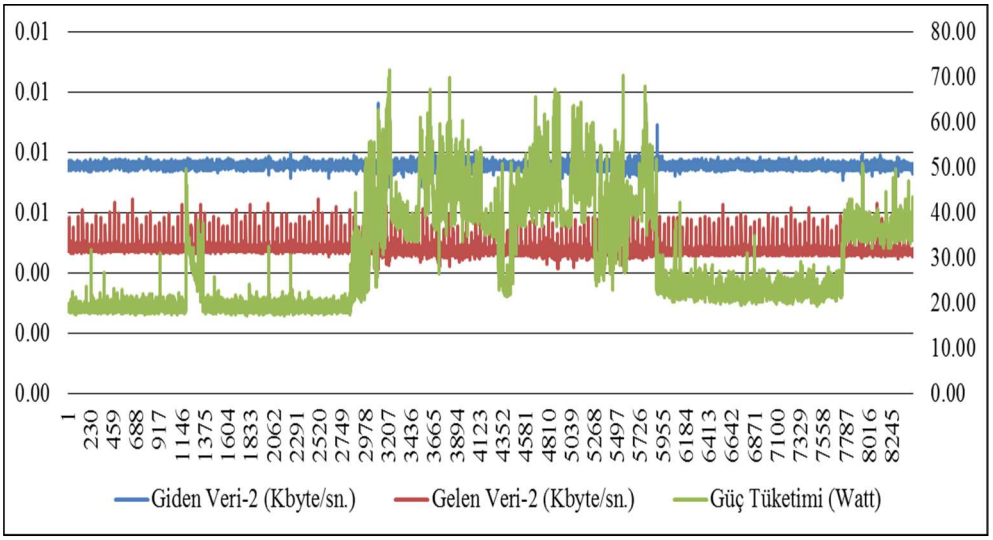
Tez çalışmamdaki veri toplama işi için geliştirilen DAEMS yazılımı sayesinde dokuz farklı parametre için veriler toplanmıştır. Bu parametreler 24 saatlik zaman diliminde 86400 adet verinin bir sn. periyotlarla toplanması ile oluşturulmuştur. Veri topluluğu, yaklaşık 10 saniyelik değerlerin ortalamaları alınarak 8451 adet olarak düzenlenmiştir. Dakikada 60 adet veri alma uygulamasının amacı daha hızlı periyotlarla verileri alarak daha yüksek hassasiyete sahip giriş parametrelerini dört farklı metodumuz için uygulayabilmektir. 24 saat süre için elde edilecek veri sayısı oldukça yüksek olacağından ortalama alınarak bu parametrelerin daha uygulanabilir hale getirilmesi amaçlanmıştır.

Şekil 4.1 grafiği sunucu üzerindeki birinci Ethernet portuna ait gelen ve giden veri trafiğinin, harcanan güç ile olan ilişkisini göstermektedir.



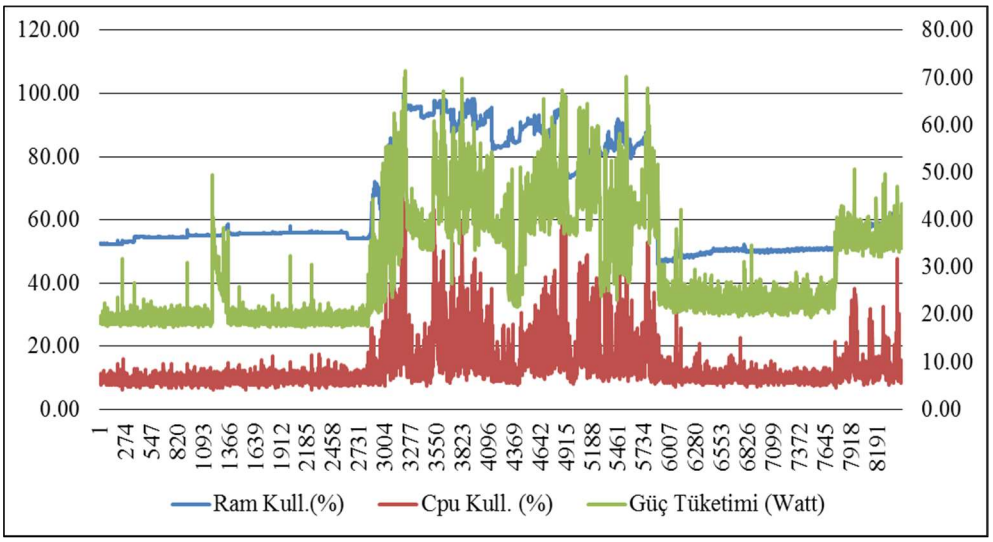
Şekil 4.1: Birinci Ethernet Portu Veri Trafiği-Sunucu Güç Tüketimi.

Şekil 4.2 grafiği, sunucu üzerindeki ikinci Ethernet portuna ait gelen ve giden veri trafiğinin, harcanan güç ile olan ilişkisini göstermektedir. Şekil 4.1 ve Şekil 4.2 karşılaştırıldığında bir numaralı Ethernet portuna ait veri trafiği oranının, iki numaralı Ethernet portundan daha yüksek kapasiteli veri trafiği oranlarına sahip olduğu görülmektedir. Şekil 4.1 ve Şekil 4.2 grafiklerinde birincil y ekseni veri trafiğini Kbyte/sn. olarak, ikincil y ekseni ise harcanan güç değerini Watt birimi cinsinden göstermektedir.



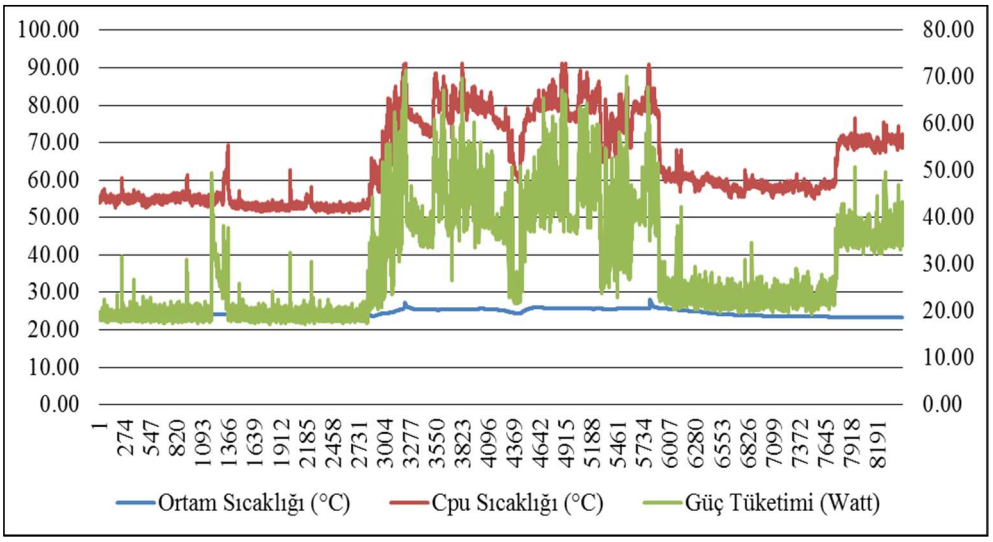
Şekil 4.2: İkinci Ethernet Portu Veri Trafiği-Sunucu Güç Tüketimi.

Şekil 4.3 grafiği, sunucunun anlık Ram ve Cpu kullanım yüzdelerinin sunucunun güç tüketimi ile olan değişimlerini göstermektedir.



Şekil 4.3: Ram ve Cpu Kullanımı-Sunucu Güç Tüketimi.

Şekil 4.3 grafiğinden görülebileceği üzere birincil y ekseni yüzde olarak sunucunun Ram ve Cpu kullanım oranlarını, ikincil y ekseni ise sunucunun güç tüketimini Watt birimi cinsinden göstermektedir.

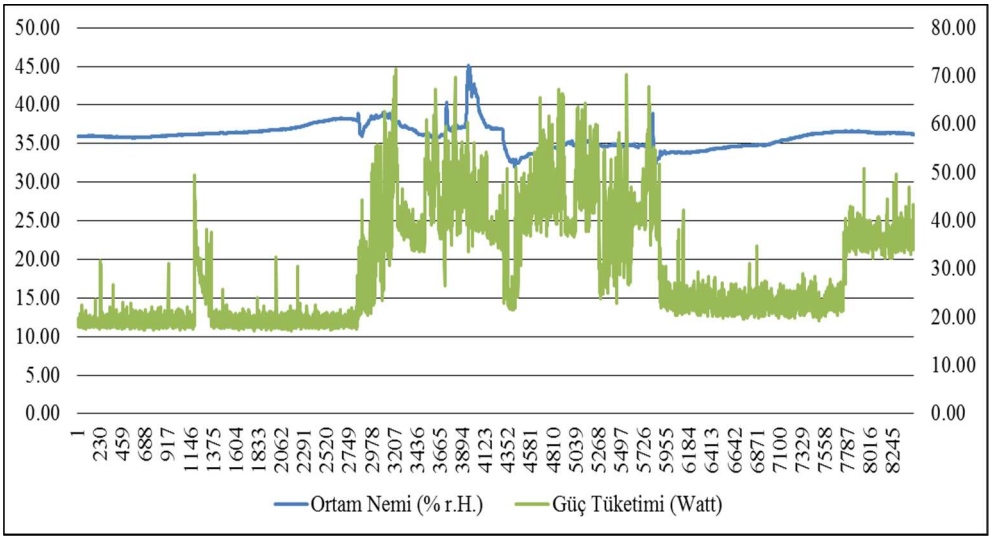


Şekil 4.4: Ortam ve Cpu Sıcaklığı-Sunucu Güç Tüketimi.

Şekil 4.4 grafiği ortam sıcaklığı ve sunucunun Cpu sıcaklığı değerlerinin güç tüketimi ile olan ilişkisini göstermektedir. Bu grafikteki birincil y ekseni derece

santigrat birimi türünden sıcaklıkları göstermekte, ikincil y ekseni ise sunucu güç tüketimini Watt birimi cinsinden göstermektedir.

Şekil 4.4'te, sunucunun işlemcisine ait sıcaklık değişiminin, sunucunun güç tüketimi ile oldukça benzer bir eğriye sahip olduğu görülebilmektedir.



Şekil 4.5: Ortam Nemi-Sunucu Güç Tüketimi.

Şekil 4.5 grafiği, sunucunun bulunduğu ortamdaki nem değerinin, sunucunun güç tüketimi ile olan ilişkisini göstermektedir. Bu grafikteki birincil y ekseni bağıl nem oranını ve ikincil y ekseni ise Watt biriminden sunucunun güç tüketimini göstermektedir.

Şekil 4.1'den Şekil 4.5'e kadar ki grafiklerden de görülebileceği üzere, bazı giriş parametreleri ile çıkış parametresi olan tüketilen güç arasında Cpu sıcaklığı ve Ram kullanım verileri dikkate alındığında bir ilişki olabileceği, ancak ikinci Ethernet portuna ait giden ve gelen veri değişimlerinin tüketilen güce göre çok sınırlı kaldığı görülmektedir. Bu ve benzeri etkileşimler göstermektedir ki giriş parametreleri ile çıkış parametresi olan güç tüketimi arasında benzer bir grafik eğrisi oluşabileceği gibi farklı profilde bir eğri de olabilmektedir.

Tablo 4.1'de sunucunu Ethernet portlarındaki veri trafiğinin güç ile değişimine ait veri setlerinden 10 adet örnek gösterilmektedir.

Tablo 4.2'de ise sunucu ve ortam şartlarına ait parametrelerin güç ile olan ilişkisi yine 10 adetlik örnek veri seti üzerinden gösterilmiştir.

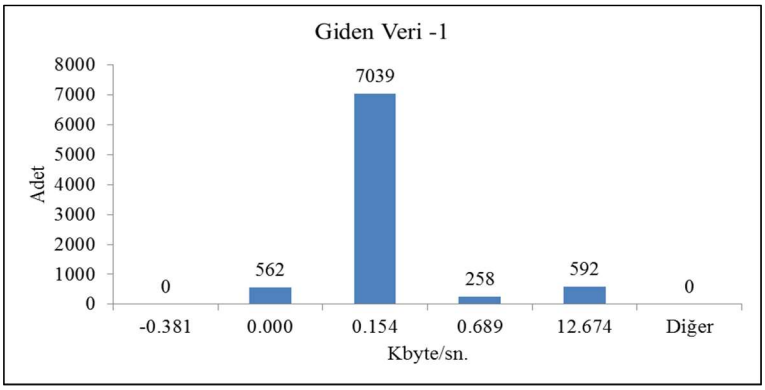
Tablo 4.1: Ethernet Portları Veri Trafiği-Güç İlişkisi Tablosu.

Sayı	Giden Veri-1 (Kbyte/sn.)	Gelen Veri-1 (Kbyte/sn.)	Giden Veri-2 (Kbyte/sn.)	Gelen Veri-2 (Kbyte/sn.)	Güç Tüketimi (Watt)
1	1.719570	26.815711	0.007126	0.004528	64.63
2	1.616441	21.966175	0.007158	0.004501	63.48
3	2.831501	9.148106	0.007274	0.004617	56.47
4	3.693689	8.897463	0.007545	0.004864	55.19
5	2.852139	10.500571	0.007578	0.004770	67.29
6	2.759543	9.131500	0.007269	0.004581	59.09
7	1.791705	4.411211	0.007438	0.004961	54.04
8	1.842717	6.959824	0.007828	0.004842	55.87
9	2.103602	4.805787	0.007192	0.004432	55.20
10	2.373731	6.795384	0.007552	0.004730	51.14

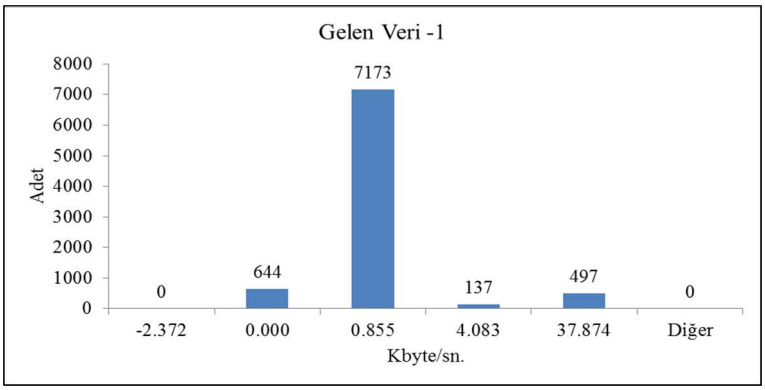
Tablo 4.2: Sunucu ve Ortam Şartları Parametreleri-Güç İlişkisi Tablosu.

Sayı	Ram Kull. (%)	Cpu Kull. (%)	Ortam Sıcaklığı (°C)	Ortam Nemi (% r.H.)	Cpu Sıcaklığı (°C)	Güç Tüketimi (Watt)
1	96.16	37.66	25.66	34.59	90.70	64.63
2	95.45	31.18	25.66	34.59	91.10	63.48
3	93.78	38.26	25.67	34.58	89.90	56.47
4	93.02	33.84	25.66	34.58	89.90	55.19
5	93.20	64.24	25.67	34.59	90.30	67.29
6	94.10	32.39	25.67	34.61	90.20	59.09
7	92.46	26.59	25.66	34.60	88.20	54.04
8	92.18	21.69	25.66	34.60	87.40	55.87
9	92.33	35.32	25.67	34.63	86.20	55.20
10	92.43	26.56	25.66	34.64	85.30	51.14

Toplanan bu parametre kayıtlarından birinci Ethernet portuna ait giden ve gelen veriler, Kbyte/sn. biriminden nasıl bir değişim gösterdikleri Şekil 4.6 ve Şekil 4.7‘de bulunan veri yoğunluk dağılımı grafiklerinde görülebilmektedir.

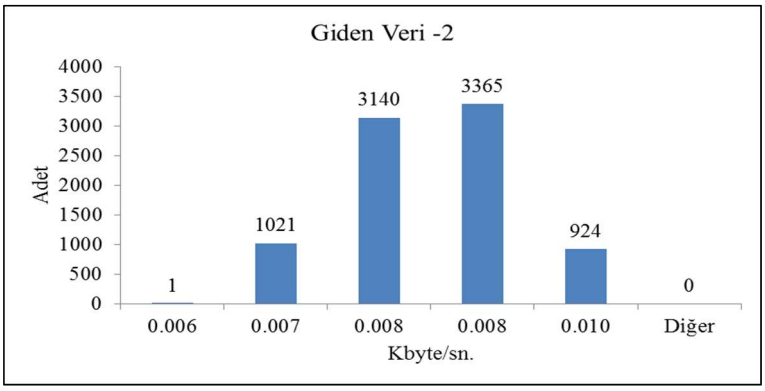


Şekil 4.6: Giden Veri -1.

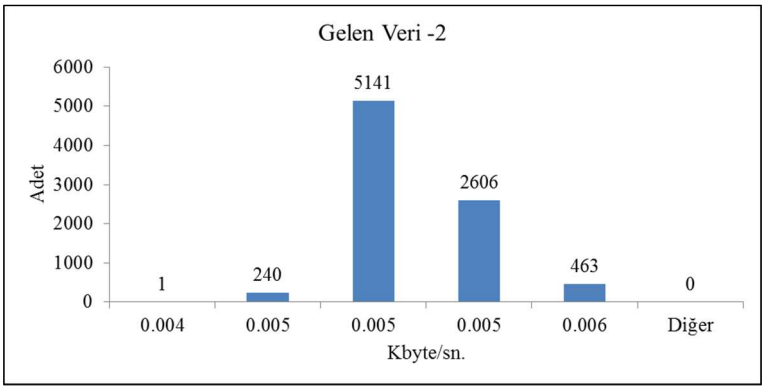


Şekil 4.7: Gelen Veri -1.

Birinci Ethernet bağlantısına ait giden ve gelen veri miktarlarının, ikinci Ethernet portuna göre daha yüksek olmalarının sebebi ise sunucunun internet bağlantısının bu ağ adaptörü üzerinden yapılıyor olmasıdır. Şekil 4.6 ve Şekil 4.7 grafiklerinde, birinci Ethernet bağlantısına ait gelen veri miktarının, giden veri miktarından yaklaşık 5.5 kat fazla olduğu görülebilmektedir.

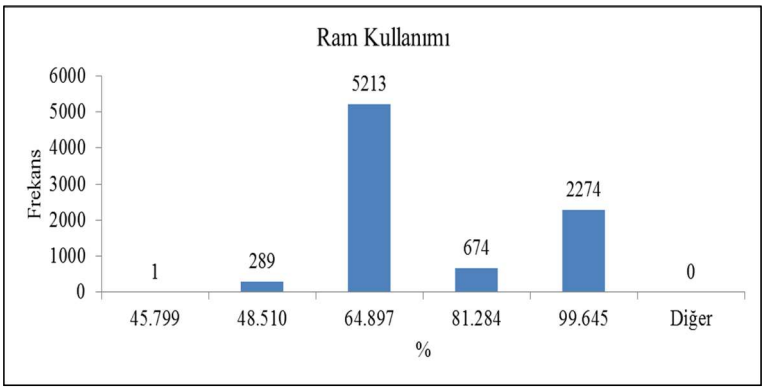


Şekil 4.8: Giden Veri -2.

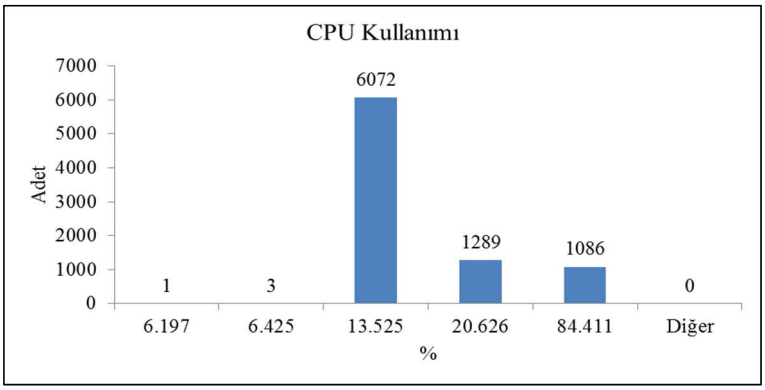


Şekil 4.9: Gelen Veri -2.

Şekil 4.8 ve Şekil 4.9 veri yoğunluk dağılımı grafikleri, ikinci Ethernet portundaki veri yoğunluklarını göstermektedir. Üzerinde çalışma yapılan sunucunun ikinci Ethernet bağlantısı, IPC cihazı ile olan haberleşme için kullanıldığından yüksek bir ağ trafiğine sahip olmadığı görülmüştür.



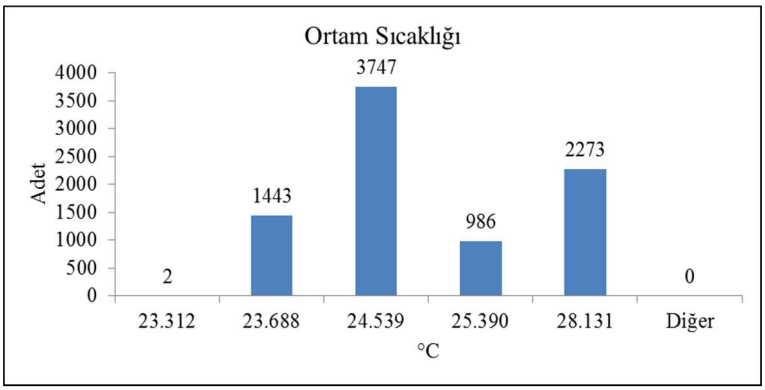
Şekil 4.10: Ram Kullanımı.



Şekil 4.11: Cpu Kullanımı.

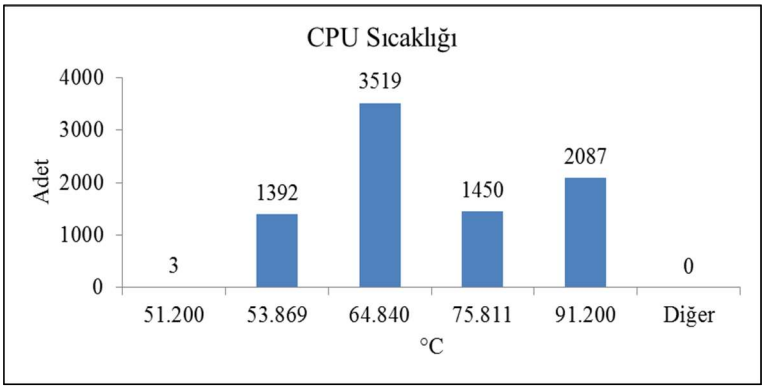
Şekil 4.10 ve Şekil 4.11 grafikleri, sunucunun Ram ve Cpu kullanım oranlarına ait veri yoğunluk dağılımlarını gösteren grafiklerdir. Ölçüm boyunca sunucunun Ram kullanım oranına bakıldığında, verilerin yaklaşık % 62'sinde Ram'ın % 65'inin, % 27'sinde ise Ram'ın yaklaşık tamamının kullanılmış olduğu görülmektedir.

Şekil 4.11 grafiğinde görülebilen Cpu kullanım oranlarına göre, ölçümün yaklaşık % 72'sinde Cpu'nun % 13.5 oranında kullanıldığı, % 13'ünde ise kullanımın % 84'lere çıktığı görülmektedir.



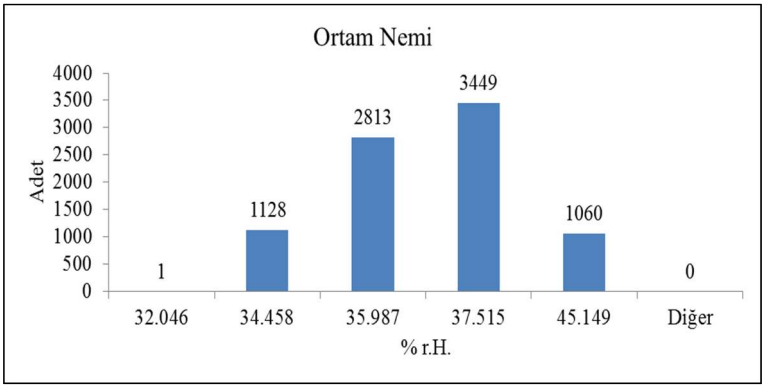
Şekil 4.12: Ortam Sıcaklığı.

Ham verilerin toplanması aşamasında IPC cihazı, ethercat protokolü üzerinden gönderdiği DAEMS yazılımına gönderdiği ortam sıcaklığına ait verileri $0 \cdots 50$ °C aralığına göre yaklaşık $0 \cdots 6000$ birim aralığında ölçeklendirerek göndermektedir. Bu sebeple verilerin kaydedildiği aralık olan $0 \cdots 6000$ değerinden $0 \cdots 50$ °C aralığına çevirme işlemi yapılmıştır.



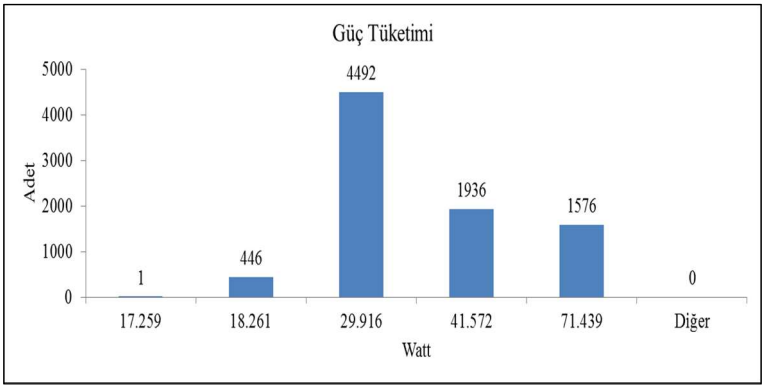
Şekil 4.13: Cpu Sıcaklığı.

Şekil 4.12 ve Şekil 4.13 grafikleri, daha önce Şekil 4.4'te tüm veri kümesi için güç ile değişimleri gösterilen sunucunun bulunduğu ortam sıcaklığı ve Cpu sıcaklığına ait veri yoğunluk dağılımlarını gösteren grafiklerdir. Sunucunun bulunduğu ortamın sıcaklığı Şekil 4.12'de görülebildiği üzere yaklaşık 23 ile 28 °C arasında değişim göstermektedir. Ölçümün % 41'lerinde Cpu sıcaklığı 24.5 °C iken, % 28'inde ise 28 °C civarında olduğu görülmektedir. Sunucunu işlemcisine ait sıcaklık parametresinin Şekil 4.13'te görülebileceği üzere $50 \cdots 95$ °C aralığında değiştiği gözlemlenmiştir.



Şekil 4.14: Ortam Nemi.

Şekil 4.14 grafiği, daha önce Şekil 4.5'te tüm veri kümesi için güç ile değişimleri gösterilen sunucunun bulunduğu ortam nemine ait veri yoğunluk dağılımını gösteren grafiktir. IPC cihazında bulunan 4-20 mA analog okuma modülü, sensörden okuduğu ortam nem bilgisini yaklaşık $0 \cdots 4096$ aralığında ölçekleyerek DAEMS yazılımına kayıt amacıyla gönderdiği için bu değeri $0 \cdots 95$ % r.H. aralığına taşıma işlemi bütün veri setinin normalizasyon aşamasından önce yapılmıştır.



Şekil 4.15: Güç Tüketimi.

9 farklı parametrenin etkilediği düşünülen sunucuya ait güç tüketimini gösteren veri yoğunluk dağılımı grafiği ise, Şekil 4.15'te gösterilmiştir. Bu grafikten anlaşılacağı üzere, sunucuya ait güç tüketim oranlarının 24 saatlik çalışma zamanı için $15 \cdots 75$ Watt aralığında değişim göstermektedir.

Sunucu, yapılan ölçümün yaklaşık % 53'ünde 30 watt enerji tüketirken, yaklaşık % 19'unda 71 watt enerji tüketmiştir. Tüm ölçümün yaklaşık % 5'inde ise 18 watt tükettiği görülmektedir.

Bu veriler, yapay sinir ağının ve regresyon formüllerinin eğitilmesi ve test edilmesi için kullanılacaktır.

4.2. Yapay Sinir Ağları

Sinir ağları, yapay nöronlar arasındaki etkileşim vasıtasıyla kavramsal davranışları taklit eden bir makine öğrenme algoritması sınıfıdır [24]. Sinir ağları karmaşık sistemleri modellemede avantajlara sahiptir.

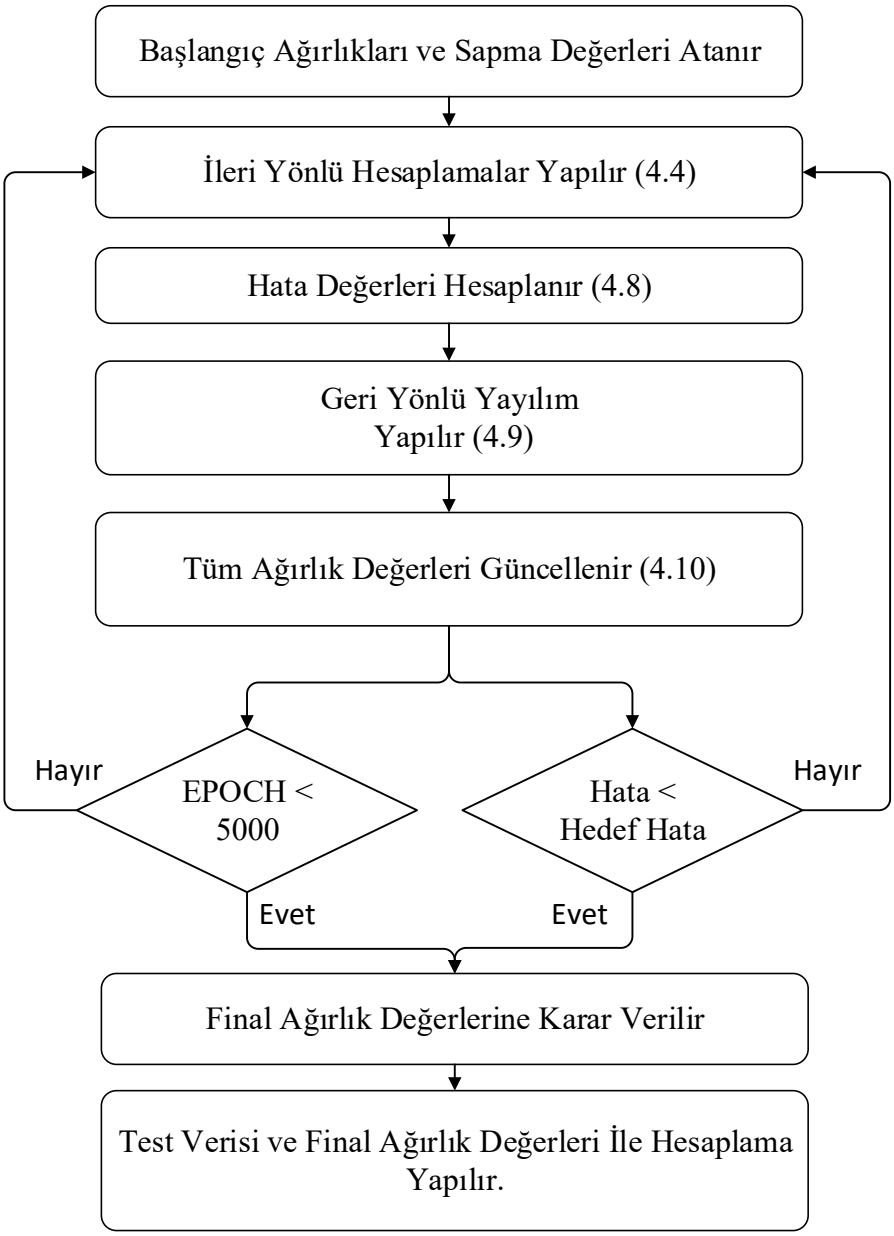
Sinir ağları kullanıcının, veri içerisindeki ilişkiyi varsayan modelde bulunan parametre etkileşimlerini önceden tanımlamasına ihtiyaç duymaz. Bunun yerine sinir ağları, en iyi modeli otomatik olarak üretebilmek için parametreler arasındaki ilişkiyi ve örüntüyü bulmaya çalışır.

Makine öğrenmesinin bu alandaki en yaygın uygulamaları; konuşma tanıma (speech recognition), resim işleme (image processing) ve otonom temsilci (autonomous agents) olarak gösterilebilir.

YSA genel olarak insan beyninin ya da merkezi sinir sisteminin çalışma prensiplerinin taklit eden bilgi işleme sistemleridir. YSA ile basit biyolojik sinir sisteminin çalışma şekli taklit edilir. Taklit edilen sinir hücreleri nöronlar içerirler ve bu nöronlar çeşitli şekillerde birbirlerine bağlanarak ağı oluştururlar.

Şekil 4.16, bu çalışmada yapılan YSA uygulamasının akış şeması olarak gösterilmiştir. Bu akış şeması YSA algoritmasına ait adımları göstermekle birlikte ilerleyen bölümlerde detayları verilecek olan formüllere de atıfta bulunmaktadır.

Şekil 4.16'da sinir ağı yapısında kullanılan algoritmanın işleyiş adımlarını içeren akış şeması bulunmaktadır.

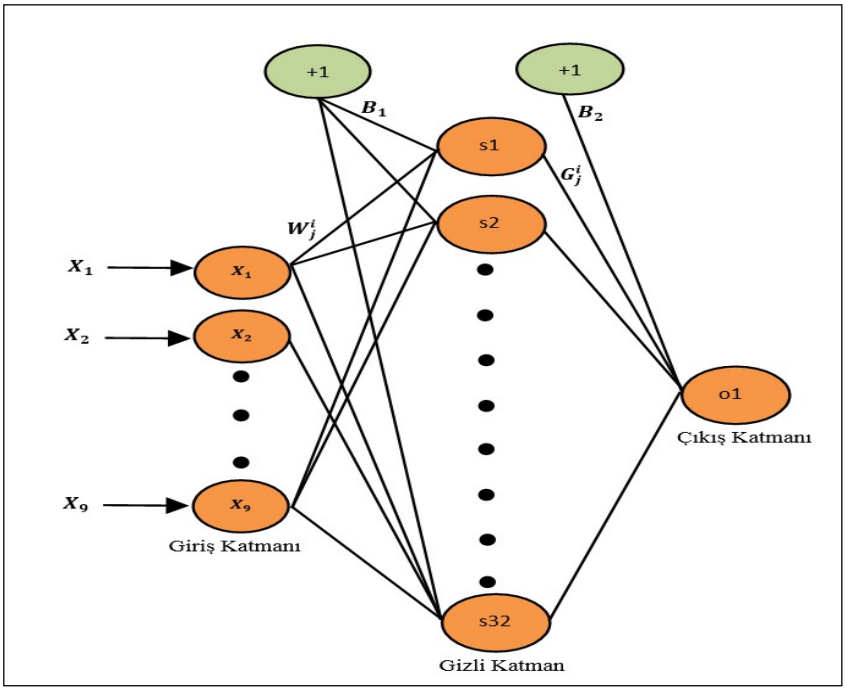


Şekil 4.16: YSA Algoritması Akış Şeması.

Şekil 4.16'da görülebileceği üzere yapılan YSA uygulaması için öncelikle sinir ağının katman ve nöron sayılarına göre başlangıç ağırlıkları ve sapma değerleri atanacaktır. İleri yönlü hesaplamalar gerçekleştirilir. İleri yönlü hesaplama sonucunda oluşan hata değeri hesaplanacaktır. Bu hata değerine bağlı olarak geri yayılım metodu ile tüm ağırlıklar güncellenecektir.

Yazılım ile ulaşılması hedeflenen Epoch ya da hata değerine ulaşıldığında, final ağırlık değerlerine karar verilmiş olacaktır. Yazılım, test verisi ve final ağırlık değerlerini kullanarak hesaplama yapar. Böylece program döngüsü tamamlanmış olmaktadır.

Tez çalışmamda aşağıdaki gibi gösterilen yapay sinir ağı modeli kullanılmıştır;



Şekil 4.17: Sinir Ağı Modeli.

Bu çalışmada kurulması planlanan yapı için birçok girişe ihtiyaç duyulacağından bir adet nöron yeterli görülmeyecektir. Paralel işlem yapan birden fazla nörona ihtiyaç duyulduğunda katman kavramı devreye girmektedir.

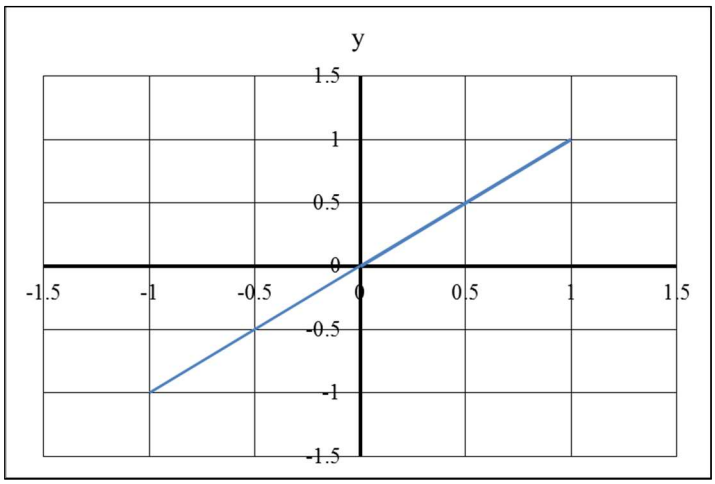
Şekil 4.17‘de görülebileceği üzere YSA modelimiz çok katmanlı Perseptron (MLP) modeli için, dokuz adet nörona sahip Giriş katmanı, 32 adet nörona sahip Gizli katman ve bir adet nörona sahip Çıkış katmanından oluşmaktadır. Bu tasarım sayesinde üç katman arasında toplam 320 adet ağırlık (weight) değeri oluşmaktadır. Gizli katmanda bulunan 32 adet nöron sayısı tahmini olarak seçilmiş olup ağın performansına göre gerekirse artırılıp azaltılabilecektir.

Ayrıca iki adet Bias girişleri Gizli ve Çıkış katmanında ilave ağırlık ve nöron bağlantısı şeklinde gösterilmiştir. Bu değerler parametrik olarak B_1 ve B_2 olarak tanımlanmıştır. Bu girişlerin kullanılma sebebi sinir ağının daha iyi öğrenmesini sağlamaya çalışmaktır.

Tüm hesaplamalar için ağda kullanılacak aktivasyon fonksiyonu olarak doğrusal fonksiyon belirlenmiştir. Sinir ağında kullanılan doğrusal aktivasyon fonksiyonu aşağıdaki gibidir;

$$f(x) = x \tag{4.1}$$

Şekil 4.18'de doğrusal fonksiyona ait grafik görülebilmektedir.



Şekil 4.18: Aktivasyon Fonksiyonu.

DAEMS yazılımı tarafından toplanan veriler, yapay sinir ağı için giriş parametreleri ile çıkış parametresini oluşturacaktır. Tüm veri, YSA uygulamamız amacıyla tez çalışmamda kullanılan doğrusal aktivasyon fonksiyonuna göre normalleştirme işlemi yapılarak kullanılmıştır.

Toplanan verinin $-1 \cdots +1$ aralığına, 4.2 ve 4.3 formüllerine göre normalize edilmesi ile birlikte giriş-çıkış parametreleri, sinir ağında çok katmanlı ve ileri yönlü hesaplamalar için kullanılmıştır.

$$X_{ort} = \frac{X_{max} - X_{min}}{2} \tag{4.2}$$

$$X_{norm} = \frac{X - X_{ort}}{X_{ort}} \tag{4.3}$$

Bu hesaplamalar giriş katmanından adım adım çıkış katmanına doğru yapılmıştır. Sinir ağının bu hesaplamalar sonucunda nasıl öğrenme yapacağını belirlemek için verinin ağa veriliş şekli belirlenmelidir.

Bunun için Rastgele veya Yığın tipi öğrenme yöntemlerini kullanabileceğiz. Ağın eğitim senaryosundaki her bir iterasyonda veriler yığın halinde veya rastgele olarak verilebilir. Rastgele öğrenme yöntemi genel olarak aşağıdaki sebeplerden dolayı tercih edilebilmektedir;

- Rastgele öğrenme Yığın tipi öğrenmeden daha hızlıdır.
- Rastgele öğrenme çoğunlukla daha iyi sonuçlar üretir.
- Rastgele öğrenme, değişiklikleri takip edebilmemize imkân vermektedir.

Rastgele öğrenme çoğunlukla güncelleme işlemlerinde oluşan hata değişimleri sayesinde daha iyi çözümlerle sonuçlanmaktadır. Çünkü bu değişimler sayesinde nöronlar arasındaki ağırlık değerleri daha geniş bir değişim aralığına sahip olabilecektir [11].

Yaygın olarak kullanılan endüstriyel uygulamalarda veri dağılımı eğer zamanla değişiyorsa Rastgele öğrenme iyi bir yöntem olabilecektir. Eğer öğrenme işini yapan yazılım değişiklikleri anlayamaz ve takip edemez ise veriyi düzgün bir şekilde öğrenemeyecek ve hatalar oluşmasına sebep olacaktır [11].

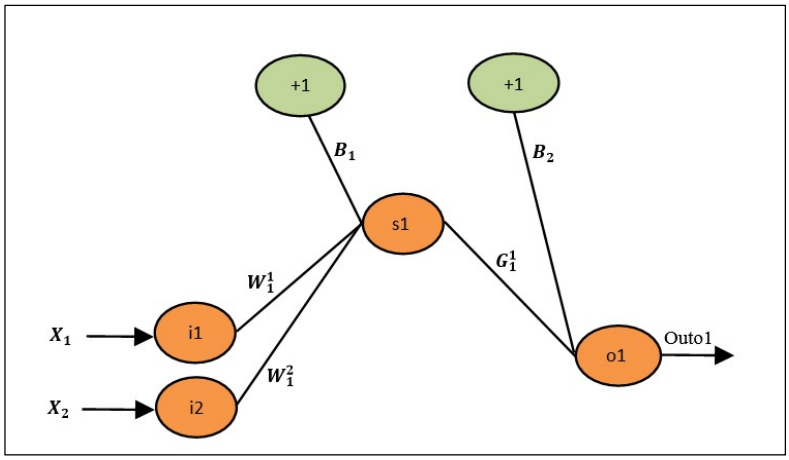
Rastgele öğrenmenin bu avantajlarına rağmen yığın tipi öğrenmenin kullanılması için de bazı avantajlı özellikler bulunmaktadır. Bunlar;

- Yakınsama durumunun iyi bir şekilde sonuçlanabilmesi,
- Bazı hızlandırma tekniklerinin bu yöntemde uygulanabilir olması,
- Ağırlık değişimlerinin ve yakınsama oranlarının teorik analizlerinin daha basit

olmasıdır.

Yığın tipi öğrenmenin bu avantajlarına rağmen Rastgele öğrenme büyük veri setleri ile uğraşılan durumlarda basit bir yöntem olarak daha hızlı olduğu için en çok tercih edilen yöntemlerden olmaktadır [11].

Bu sebeplerle sinir ağının eğitilmesi amacıyla rastgele öğrenme yöntemi tercih edilmiştir. Ağ yapısının küçük bir kesiti Şekil 4.19'da gösterilmektedir.



Şekil 4.19: Sinir Ağı Detaylı Görünüm.

Giriş değişkenleri ile birlikte ileri yönlü hesaplamalar gizli katmandaki s1 nöronu için net değer aşağıdaki şekilde yapılmıştır;

$$Nets1 = X_1 * \mathrm{W}_1^1 + X_2 * \mathrm{W}_1^2 + \cdots + X_9 * \mathrm{W}_1^9 + 1 * B_1 \tag{4.4}$$

Şekil 4.18'de görülebilen ve bu çalışmada kullanılan doğrusal aktivasyon fonksiyonuna göre üretilen s1 nöronunun çıkış değeri;

$$Outs1 = Nets1 \tag{4.5}$$

şeklinde bulunmuştur. Farklı aktivasyon fonksiyonları kullanılmak istenirse 4.5 formülünün o fonksiyona göre belirlenmesi gerekmektedir. Aynı hesaplamalar, çıkış katmanında bulunan o1 nöronu için net ve çıkış değerleri olarak şu şekilde yapılmıştır;

$$Neto1 = Outs1 * \mathrm{G}_1^1 + Outs2 * \mathrm{G}_1^2 + \cdots + Outs32 * \mathrm{G}_1^{32} + 1 * B_2 \tag{4.6}$$

$$Outo1 = Neto1 \tag{4.7}$$

Tüm sinir ağı için bu hesaplamalar aynı şekilde gerçekleştirilmiştir. Sonuç olarak sinir ağının eğitilmesi için kullanılan veri örneklerindeki giriş parametrelerine karşılık gelen çıkış değeri $Outo1$ bulunacaktır. Ağın eğitilmesi amacıyla çıkış nöronundaki hata değerinin bulunması gerekmektedir.

Toplam hata değeri;

$$E_{toplam} = \sum \frac{1}{2}(Hedef - Outo1)^2 \tag{4.8}$$

olarak hesaplanacaktır.

Sinir ağının eğitilmesi amacıyla Geri Yayılım (Back propagation) algoritması kullanılmıştır. Geri yayılım, sinir ağlarının eğitilmesi amacıyla yaygın olarak kullanılan algoritmalardan bir tanesidir. Geri yayılım algoritmasının geçmişi 1970'lere kadar gitmektedir. Ancak 1986 yılında yapılan bir çalışma [30] ile bilinirliği ve uygulama sayıları artış göstermiştir. Bu çalışmada bahsedilen geri yayılım yaklaşımın daha önceki yaklaşımlarda çözülemez görülen problemlere bir çözüm sunduğu görülmüştür. Geri yayılım algoritması kavramsal olarak basit, hesaplama olarak verimli sonuçlar veren bir yöntemdir.

Geri yayılım algoritmasında, yapay nöronlar katmanların içinde tasarlanmıştır ve bu nöronlar sinyalleri ileri yönde iletirken hataları da geri yönde yayarlar. Yapay sinir ağını eğitme işleminde geri yayılım algoritmasının avantajları ve sonuca ulaşmadaki etkinliğinden yararlanılarak kullanılmıştır.

Algoritma gerçeklemesinin yapılması ve analizlerin daha kolay anlaşılır olabilmesi amacıyla bazı parametreler için keyfi seçimler yapılması gerekmektedir. Başlangıç ağırlıkları (initial weights), sapma değerleri (the biases), eğitim işlemi için gerekli giriş ve çıkış değerleri (training inputs / outputs), öğrenme oranı (initial learning rate), hedef hata oranı (tartget error rate) ve hedef algoritma deneme sayısı (EPOCH) için başlangıç değerleri vererek algoritma gerçeklemesi yapılmıştır.

Yukarıda yapılan hata hesaplama işlemi sonucunda, öğrenme algoritmasına bağlı olarak bu hata değerine göre ağda bulunan ağırlık değerleri güncellenecektir.

Örneğin G_1^1 ağırlık değerinin güncellenmesi işlemi;

$$\frac{\partial E_{toplam}}{\partial G_1^1} = \frac{\partial E_{toplam}}{\partial Outo1} * \frac{\partial Outo1}{\partial Neto1} * \frac{\partial Neto1}{\partial G_1^1} \tag{4.9}$$

formülüne ek olarak

$$G_1^{1^+} = G_1^1 - \mu * \frac{\partial E_{toplam}}{\partial G_1^1} \tag{4.10}$$

formülü ile yeni $G_1^{1^+}$ değerinin bulunması ile gerçekleştirilmiştir. Burada kullanılan μ parametresi sinir ağının öğrenme oranıdır. Geri yayılım algoritması sayesinde çıkıştaki hata oranının bu ağırlık değeri tarafından ne kadar etkilendiğine bağlı olarak G_1^1 değeri güncellenmektedir.

Benzer şekilde W_1^1 ağırlık değeri;

$$\frac{\partial E_{toplam}}{\partial W_1^1} = \frac{\partial E_{toplam}}{\partial Outs1} * \frac{\partial Outs1}{\partial Nets1} * \frac{\partial Nets1}{\partial W_1^1} \tag{4.11}$$

$$W_1^{1^+} = W_1^1 - \mu * \frac{\partial E_{toplam}}{\partial W_1^1} \tag{4.12}$$

formülü ile hesaplanmıştır. Öğrenme algoritmasının verimini artırabilmek için bazı özel metotlardan olan

- stokastik öğrenme,
- örneklerin karıştırılması,
- girişlerin normalleştirilmesi,
- doğrusal aktivasyon fonksiyonunun kullanılması,
- hedef değerlerin uygun seçimi,
- ağırlıkların başlangıç değerleri,
- öğrenme oranının seçimi,
- adaptif öğrenme oranı

yöntemleri uygulanmıştır [11].

Önceki sayfalarda açıklandığı üzere bu yöntemlerden stokastik öğrenme işlemi, YSA yapısının öğrenme sürecinde veri seti içerisinden rastgele öğrenme yapılarak ağırlık değerlerinin belirlenmesi şeklinde uygulanmıştır.

Eğitim veri setindeki tüm örnekler ağın eğitilmesi amacıyla C++ programında eğitim başlangıcında rastgele bir veri seti sıra numarası üretilerek, o veri setinin

kullanılması ile yapılmaktadır. Bu şekilde karıştırılan sırada örnek eğitim seti verileri YSA yapısına verilmektedir.

YSA modeline ait uygulamada ham verinin normalizasyon işlemi yapılmıştır. Normalizasyon işleminin ham verilerimize uygulanmasının sebebi, böylece YSA tasarımının daha verimli ve performanslı sonuçlar vereceğinin literatürde gösterilmiş olmasıdır [48].

Girişlerin normalleştirilmesi veya standartlaştırılması öğrenme sürecini kolaylaştırmaktadır. Normalizasyon amacıyla $f(x) = x$ doğrusal fonksiyonuna göre kullanılacak veri, -1 ve +1 aralığında normalize edilmiştir. YSA tasarımındaki adımları hızlandırma yönünde faydalı olacağı literatürde yapılan çalışmalar ile desteklendiği için bu normalizasyon yöntemi seçilmiştir [46].

Eğer her bir giriş değişkeninin ortalaması sıfıra yakın ise öğrenme süreci daha hızlı olabilmektedir [11]. Giriş parametrelerinin aynı değer aralığı için normalleştirilmesinin istisnası bazı giriş parametrelerinin diğerlerine göre daha düşük öneme sahip olabilmesidir. Böyle bir durumda daha düşük önemdeki giriş değerlini daha düşük değerlere göre normalleştirmek onların öğrenme sürecinde daha az görünür olmalarını sağlayacaktır.

Sinir ağında kullanılan giriş parametrelerinin, seçtiğimiz doğrusal aktivasyon fonksiyonuna göre -1⋯ +1 aralığına göre normalleştirilmesinin bir diğer sebebi ise tasarımdaki tüm giriş parametrelerinin aynı önem derecesine sahip olduğunun varsayılması olmuştur. Literatürde çok katmanlı perceptron ağları ile yapılan tasarımlarda sinir ağları için farklı aktivasyon fonksiyonlarının test edildiği çalışmalar yapılmıştır. Literatürdeki bu tür bir çalışmada, aktivasyon fonksiyonları çeşitlerinden olan iki kutuplu sigmoid, tek kutuplu sigmoid, hiperbolik tanjant, konik kesit ve radyal temelli fonksiyon yöntemleri karşılaştırılmıştır [47]. Bu çalışmanın sonucu olarak, ağ performansını veriyi tanıma doğruluğu üzerinden değerlendirildiğinde, hiperbolik tanjant fonksiyonunun ağın tüm bölgelerinde diğer aktivasyon fonksiyonlarına göre daha iyi sonuçlar verdiği görülmüştür.

Ancak veri merkezleri için bir sunucu vasıtasıyla yapılan bu çalışmada, sunucunun tükettiği güç üzerinde etkili olan giriş parametrelerinden doğrusal bir etkiye sahip parametre sayısı daha fazla olduğu düşünüldüğü için YSA tasarımında doğrusal aktivasyon fonksiyonu kullanılmıştır. Bu fonksiyon ile daha iyi sonuçlar elde edeceğimiz düşünüldüğü için bu fonksiyon seçilmiştir.

Ağ yapısındaki ağırlıkların başlangıç değerleri parametrik olarak atanmıştır. Atanan bu değerlerin eğitim süreci üzerinde çok belirgin bir etkisi vardır. Bu sebeple giriş verilerinin normalleştirildiği varsayılarak ağırlık değerlerinin ortalamalarının doğrusal fonksiyona bağlı olarak 0 seçilmesi öğrenme sürecine katkı sağlayan işlemlerden bir tanesi olarak uygulanmıştır.

Sinir ağının öğrenme oranını, ağın ileri yönlü yaptığı hata oranına bağlı olarak azaltarak ya da artırarak daha hızlı bir öğrenme süreci oluşturulmaya çalışılmıştır. YSA yapısında ileri yönlü yapılan işlemler sonucunda hesaplanan hata değerine bağlı olarak ağırlık değerlerinin 4.10 ve 4.12 formüllerine göre güncellenmesi işleminde μ öğrenme oranı adaptif bir değişiklik göstererek uygulanmıştır. Öğrenme oranının başlangıç değeri olarak $\mu = 0.0001$ alınmıştır.

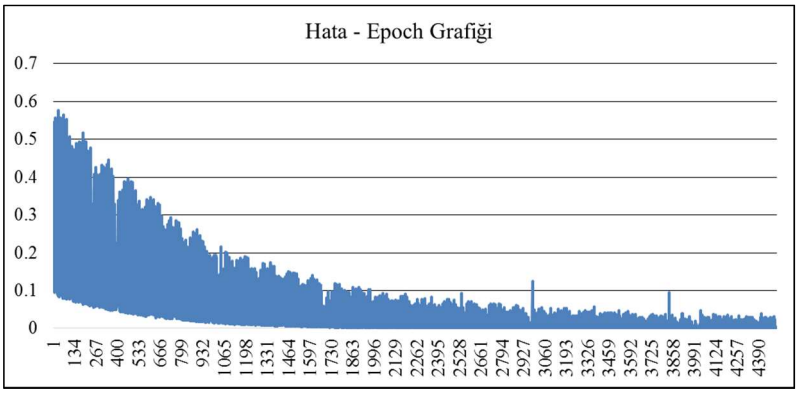
Bu varsayılan öğrenme oranı değeri, yapılan denemeler sonucunda daha iyi sonuçlar elde etmemizi sağlayacağı düşünüldüğü için seçilmiştir. Ağın yaptığı hata oranına bağlı olarak uygulama yazılımı içerisinde dinamik olarak belirlenmiş katsayılar oranında bu değer değiştirilmektedir.

Sinir ağı uygulaması olarak yapılan bu çalışmada

- giriş değişkenleri ve başlangıç değişkenleri kullanılarak ağın ileri yöndeki gizli katman ve çıkış katmanı üzerinden çıkış değerinin hesaplanması,
- hesaplanana çıkış değerine göre toplam hatanın bulunması,
- bulunan hata değeri kullanılarak geri yayılım yapılması ve tüm nöronlar arasındaki ağırlık değerlerinin güncelleştirilmesi işlemleri

sırasıyla uygulanmıştır. Uygun miktarda tekrarlanan hesaplama sayısından sonra istenen sonuca ulaşılmıştır. Veri setinin yaklaşık % 75'i ağın eğitimi için kullanılmış olup kalan % 25'i ise performans testine ayrılmıştır. Tüm ağa ait 320 adet başlangıç ağırlık değeri seçilerek eğitim işlemi yapılmıştır. Sinir ağı, kabul edilebilir hata değerine ulaşana veya hedef EPOCH (Deneme) sayısına erişene kadar eğitim işlemi tekrar eder.

Bu çalışmada 5000 EPOCH değeri, YSA modeline ait öğrenme işleminin bitirilmesi için gereken şart olarak tanımlanmış ve uygulanmıştır. Sinir ağı eğitildikten sonra Hata - Epoch grafiği Şekil 4.20'de gösterilmektedir;



Şekil 4.20: Hata-EPOCH Grafiği.

Ağın eğitim süreci, uygun ağırlık değerlerinin belirlenmesi ve Hata-Epoch grafiğinin üretilmesi ile son bulur. Bu eğitim işlemi sonucunda oluşan standart hata değeri,

$$Standart\ Hata = 0.25 \tag{4.13}$$

olarak bulunmuştur.

Ardından test veri seti, gerçek ve tahmin edilen güç sonuçları ile sinir ağının performansını doğrulamak için kullanılır. Test veri seti sinir ağının giriş katmanına verilir. Her bir nöron arasında daha önce belirlenen ağırlık değerleri kullanılarak çıkış güç değeri hesaplanır.

4.3. Regresyon Yöntemleri

İstatistik çözümlemelerinde iki adet değişken arasında bir ilişki olup olmadığını tespit etmek amacıyla yaygın olarak kullanılan yöntemlerden bir tanesi de regresyon analizidir. Regresyon analizi, verilerin dağılımına göre;

- Doğrusal regresyon analizi
- Doğrusal olmayan regresyon analizi

Değişken sayısına göre;

- Basit regresyon analizi (Tek Değişken)
- Çoklu regresyon analizi (Çoklu Değişken)

Verilerin kaynağına göre;

- Ana kütle verileriyle regresyon analizi
- Örnek verileri ile regresyon analizi
- Zaman serilerinde regresyon analizi

şeklinde gruplandırılabilir [29].

Eğri uydurma, bir grup veri noktasına en iyi yaklaşma özelliğine sahip sürecin veya matematiksel fonksiyonun üretilmesi anlamına gelmektedir. Tez çalışmamda uygulanan regresyon analizleri, bir adet bağımlı değişkenin en az bir ya da daha fazla bağımsız değişkene bağlı analiz edilmesi amacıyla kullanılacaktır. Bağımlı değişken olan Y ile bağımsız değişkenler olan X'ler arasında bir ilişki olup olmadığını ve varsa bu ilişkinin ölçülmesini ve analiz edilmesi amaçlanmaktadır.

Basit lineer regresyon iki değişkenli (x, y) olmakta ve düz bir çizgi formunda gösterilebilmektedir. Lineer regresyonda değişkenler arasındaki korelasyon lineer ve belirgin olmaktadır. İki değişken arasındaki doğrusal ilişkiyi veren lineer eşitlik;

$$y = \beta_1 + \beta_2 \mathrm{x} + u \tag{4.14}$$

şeklindedir. Burada y bağımlı ve x bağımsız değişkenlerdir. β_1 sabit bir sayıdır. β_2, regresyon katsayısıdır. u terimi hata miktarı olarak ifade edilir.

Tez çalışmamda, veri modelleme eşitliğindeki parametreleri bulabilmek için Levenberg - Marquardt (LM) yöntemi kullanan "Curve Expert (Trial)" yazılımının hesaplama kabiliyetlerinden faydalanılmıştır [31]. Bu yazılım, Dik İniş Yöntemi (Steepest-Descent) ve Taylor Serisi temelli metotları bir araya getirerek lineer olmayan optimizasyon problemleri için hızlı ve güvenilir bir teknik olmayı hedeflediği için seçilmiştir. Bu iki metottan hiçbiri her zaman ideal çözüm olmadığı için LM algoritması bu iki metot arasında iterasyon sayıları arttıkça yumuşak geçişlere izin verecek şekilde tasarlanmıştır.

$$X^2(\vec{a}) = \sum_{i=1}^{n} \left(\frac{y_i - y(\vec{x_i}; \vec{a})}{\sigma_i} \right)^2 \tag{4.15}$$

Buradan N veri noktalarının sayısını, $\vec{x_i}$, x veri noktalarını ve $\vec{y_i}$, y veri noktalarını, σ_i değeri ise i noktasındaki standart sapmayı ve $y(\vec{x_i}; \vec{a})$ i. veri noktası için uygulanan lineer olmayan modeli ifade etmektedir. Bu fonksiyon temel olarak veri noktaları ile parametrik model arasındaki ilişkiyi ölçmek için kullanılmıştır. $X^2(\vec{a})$ fonkisyonu ne kadar küçük ise o kadar gerçek değerlere yaklaştığı anlamına gelecektir. Bu fonksiyon aynı zamanda ki-kare testi olarak da ifade edilir.

Tez çalışmam ile yük profili tahmini yapacağım tasarım için dokuz adet bağımsız değişken ve bir adet bağımlı değişken ile uygulaması yapılacak çoklu doğrusal regresyon eşitliği;

$$y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \beta_7 X_{i7} + \beta_8 X_{i8} + \beta_9 X_{i9} + \beta_{10} X_{i10} + u_i \tag{4.16}$$

olarak ifade edilir.$i = 1,2,\cdots,$ n'ye kadar eğitim setindeki tüm verilerin sayısı kadar olacaktır.

Burada 10 bilinmeyenli n denklemden oluşan eşitliğimizi aşağıdaki gibi yazabiliriz:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{12} & \cdots & X_{110} \\ 1 & X_{22} & \cdots & X_{210} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n2} & \cdots & X_{n10} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{10} \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \tag{4.17}$$

Bu modeli matris cebiri olarak ifade etmek için eşitlikteki büyüklükler:

$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, $n \times 1$ boyutlu bağımlı değişken gözlemleri vektörünü,

$X = \begin{bmatrix} 1 & X_{12} & \cdots & X_{110} \\ 1 & X_{22} & \cdots & X_{210} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n2} & \cdots & X_{n10} \end{bmatrix}$, $n \times 10$ boyutlu açıklayıcı değişken verileri matrisini,

$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{10} \end{bmatrix}$, 10×1 boyutlu katsayılar vektörünü,

$u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$, 10×1 boyutlu hata terimleri vektörünü göstermektedir.

Regresyon katsayılarnı bu formül ile elde ederiz;

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{4.18}$$

4.16 eşitliğindeki çoklu doğrusal regresyon formülü ile 6540 adet veri seti için yapılan çoklu doğrusal regresyon analizi sonucunda elde edilen katsayılar Tablo 4.3'te gösterildiği gibi bulunmuştur.

Tablo 4.3: Doğrusal Regresyon Katsayıları.

$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$	$\boldsymbol{\beta_3}$	$\boldsymbol{\beta_4}$	$\boldsymbol{\beta_5}$
3.2939	0.0059	0.0961	−975.303	278.8747
$\boldsymbol{\beta_6}$	$\boldsymbol{\beta_7}$	$\boldsymbol{\beta_8}$	$\boldsymbol{\beta_9}$	$\boldsymbol{\beta_{10}}$
0.0001	0.3390	−0.0140	0.0002	0.9093

Çalışmamdaki doğrusal olmayan regresyon modeli ikinci dereceden eşitlik kullanılarak yapılmıştır. Veri setimizdeki tüm örneklerin yaklaşık % 75'i regresyon eşitliğimizdeki değişkenlerin bulunması için kullanılmıştır. YSA ve Çoklu Doğrusal Olmayan Regresyon analizi sonrasında giriş ve çıkış değişkenleri arasındaki ilişkiyi bulabilmek adına Çoklu Doğrusal Olmayan Regresyon analizi şu matematiksel formül ile yapılmıştır;

$$y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \beta_7 X_{i7} + \beta_8 X_{i8} + \beta_9 X_{i9} + \beta_{10} X_{i10} + (\alpha_1 X_{i2} + \alpha_2 X_{i3} + \alpha_3 X_{i4} + \alpha_4 X_{i5} + \alpha_5 X_{i6} + \alpha_6 X_{i7} + \alpha_7 X_{i8} + \alpha_8 X_{i9} + \alpha_9 X_{i1})^2 + u_i \quad (4.19)$$

Önceki bölümlerde detaylı çalışma mantığı açıklanan Curve Expert yazılımının 4.19 formülüne göre koşturulması sonucunda elde edilen 19 farklı değişken Tablo 4.4'te gösterildiği gibi bulunmuştur;

Tablo 4.4: Doğrusal Olmayan Regresyon Katsayıları.

$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$	$\boldsymbol{\beta_3}$	$\boldsymbol{\beta_4}$	$\boldsymbol{\beta_5}$
17.355	0.896	0.862	−1232.472	961.669
$\boldsymbol{\beta_6}$	$\boldsymbol{\beta_7}$	$\boldsymbol{\beta_8}$	$\boldsymbol{\beta_9}$	$\boldsymbol{\beta_{10}}$
−0.001	0.158	−0.023	0.003	0.560
$\boldsymbol{\alpha_1}$	$\boldsymbol{\alpha_2}$	$\boldsymbol{\alpha_3}$	$\boldsymbol{\alpha_4}$	$\boldsymbol{\alpha_5}$
−0.149	−0.114	83.155	−157.321	0.0001
$\boldsymbol{\alpha_6}$	$\boldsymbol{\alpha_7}$	$\boldsymbol{\alpha_8}$	$\boldsymbol{\alpha_9}$	
0.021	0.001	0.0004	0.066	

Regresyon modelini doğrusal olmaktan çıkararak doğrusal olmayan bir model haline getirebilmek için 4.19 formülünde de görülebileceği üzere parametreler ilave edilmiştir.

Doğrusal olmayan çok değişkenli regresyon analizi ile yapılan bu değişken bulma işlemi sonrasında gerçek ve tahmin edilen veri arasındaki bağlantıyı bulabilmek için yeni bir matematiksel denklem aşağıdaki 4.20 eşitliğinde gösterildiği gibi oluşturulmuştur.

$$y_i = \beta_1 + \beta_{11}(\beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \beta_7 X_{i7} + \beta_8 X_{i8} + \beta_9 X_{i9} + \beta_{10} X_{i10}) + \beta_{12}(\beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \beta_7 X_{i7} + \beta_8 X_{i8} + \beta_9 X_{i9} + \beta_{10} X_{i10})^2 + u_i \quad (4.20)$$

Farklılaştırılmış doğrusal olmayan regresyon olarak ifade edilen eşitlikte bulunan β_1, β_{11} ve β_{12} değişkenleri isteğe bağlı sabit terimler olup β_2 den β_{10} a kadar olan katsayılar ise x'in sıralı artan katsayılarına karşılık gelen değişkenlerdir.

Regresyon analizlerinde veri merkezlerinin yük profilini oluşturmak için kullanılan veri topluluğunun standartlaştırılmasına gerek yoktur. Bu sebeple veri hiçbir işleme tabi tutulmadan ham halde lineer olmayan regresyon formülümüzdeki değişkenleri bulabilmek için kullanılmıştır. İlgili yazılımın 4.20 formülü için koşturulması sonucunda 12 farklı değişken Tablo 4.5'te görüldüğü şekilde hesaplanmıştır.

Tablo 4.5: Farklılaştırılmış Doğrusal Olmayan Regresyon Katsayıları.

$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$	$\boldsymbol{\beta_3}$	$\boldsymbol{\beta_4}$	$\boldsymbol{\beta_5}$
11.158	−56.900	6.380	−156374.99	44269.476
$\boldsymbol{\beta_6}$	$\boldsymbol{\beta_7}$	$\boldsymbol{\beta_8}$	$\boldsymbol{\beta_9}$	$\boldsymbol{\beta_{10}}$
0.026	71.346	−4.572	−0.019	268.542
$\boldsymbol{\beta_{11}}$	$\boldsymbol{\beta_{12}}$			
0.002	0.00000009			

Veri setinin, 4.16, 4.19 ve 4.20 numaralı formüllerde gösterilen tüm regresyon modellerine göre analiz edilmesi amacıyla kullanılan "Curve Expert" programı ile yapılan işlemler sonucunda, elde edilen başarım oranları aşağıdaki Tablo 4.6'da görüldüğü şekliyle sonuçlanmıştır.

Tablo 4.6: Regresyon Eğitim Başarım Oranları Karşılaştırma Tablosu.

Metot İsmi	**Standart Hata**	**Korelasyon Katsayısı**
Doğrusal Regresyon	2.933	0.968
Doğrusal Olmayan Regresyon	2.837	0.970
Farklılaştırılmış Doğrusal Olmayan Regresyon	2.909	0.969

5. ANALİZLER ve KARŞILAŞTIRMALAR

Yapay sinir ağları ve regresyon yöntemlerinin uygulanabilirliği ve karşılaştırılması geniş bir çalışma alanı oluşturmaktadır [19], [20], [6]. Güç üretim tahmini üzerine yapılan çalışmalardan bir tanesi de lineer, logaritmik, polinominal regresyon ve yapay sinir ağlar yöntemlerinin karşılaştırılması şeklinde olmuştur [19]. Öngörme/Tahmin etme problemlerine karşı yapay sinir ağları uygulamaları son zamanlarda üzerinde çokça çalışmalar yapılan bir konudur. Başarılı olan deneysel ve pratik testler ile ilgili fazlaca bilimsel makale ve rapor yayınlanmaktadır.

Tez çalışmam ile amacım, veri merkezleri ve IT sistemlerinin enerji tüketim miktarlarını uygun yazılım ve cihazlar kullanarak önceden nasıl tahmin edebileceğimizi açıklamaya çalışmaktır. Büyük kapasiteli bu tesislerin enerji tüketimlerini azaltmak ve optimize etmek için birçok ekonomik ve çevresel sebepler bulunmaktadır. Yapılan tasarım, büyük kapasiteli veri merkezlerine ulaşabilme ve uygulayabilme problemlerinden dolayı bir sunucu vasıtasıyla gerçekleştirilmiştir.

Bu çalışmamda başarıya ulaşabilmek için birçok değişik türde ve özellikte cihazlar bir araya getirilerek kullanılmıştır. Bunun yanı sıra bu tez çalışması için, regresyon yöntemlerinin test işlemleri amacıyla hazır bir yazılım kullanılmıştır. Ayrıca yapay sinir ağları uygulama yazılımları dâhil olmak üzere iki farklı yazılım geliştirilmiştir. Yapay sinir ağının verimli ve başarılı bir şekilde eğitilebilmesi için bazı özel ve pratik metotlar ile çok katmanlı algılayıcı (MLP) modeli kullanılmıştır.

Giriş parametreleri ile çıkıştaki gerçek güç tüketim arasındaki bağımlılığı bulabilmek için tasarlanan, programlanan ve düzenlenen her şey bir araya getirilerek bu uygulama yapılmıştır.

5.1. Veri Analizleri

Korelasyon katsayısı ise iki farklı değişkenin birbirleri ile ne kadar güçlü bir ilişkiye sahip olduğunu gösteren parametrelerden bir tanesidir. Bu katsayı, gerçek güç ve tahmin edilen güç değerleri arasındaki ilişkiye bağlı olarak -1 ve +1 arasında değişebilmektedir. Korelasyonun en düşük olduğu -1 değeri ideal negatif noktasını gösterir. Sıfır değeri korelasyon olmadığını ve en yüksek nokta olan bir değeri ise ideal pozitif noktayı göstermektedir [8].

Bu çalışmamdaki hedeflerden bir tanesi de veri merkezleri ve bilgi teknoloji sistemlerinde tüketilen güce en çok etki eden giriş parametresinin ne olduğunu bulmaktır. Bu sebeple, uygulanan tasarım için tüketilen güce etki ettiği düşünülen dokuz farklı giriş parametresi kullanılmıştır.

Giriş parametreleri ve tüketilen güç değeri arasındaki ilişki Tablo 5.1'de gösterilmiştir. Tüm bu giriş ve çıkış verileri arasındaki korelasyon ilişkisine ait parametreler 5.2 formülü ile gösterilen korelasyon hesaplaması ile yapılmıştır. Bu tablo korelasyon matrisi olarak isimlendirilmekte ve giriş parametrelerinin hem kendi içerisinde hem de tüketilen güç değeri ile arasındaki korelasyonu göstermektedir.

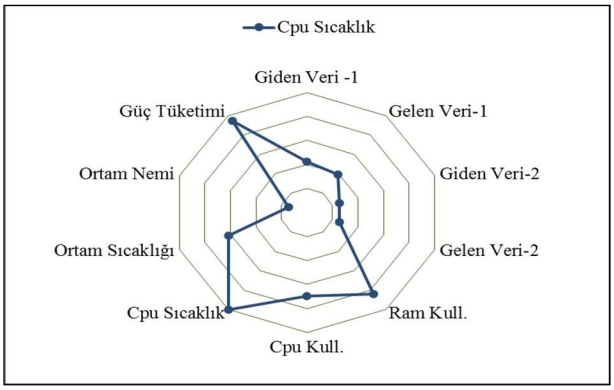
Tablo 5.1'de görülebileceği üzere sunucunun tüketilen güç değerine 0.947 korelasyon değeri ile en yüksek bağımlılığı olan parametrenin Cpu Sıcaklık parametresi olduğu görülmektedir. Sırasıyla Ram kullanımı ve Cpu kullanım değerleri de en yüksek etki oranına sahip parametreler olduğu görülmektedir.

Tablo 5.1: Korelasyon Matrisi Tablosu.

	Giden Veri-1	Gelen Veri-1	Giden Veri-2	Gelen Veri-2	RAM Kullanım	CPU Kullanım	CPU Sıcaklık	Ortam Sıcaklık	Ortam Nem	Güç Tüketimi
Giden Veri-1	1.000	0.328	0.096	0.096	0.426	0.397	0.426	0.303	0.010	0.422
Gelen Veri-1	0.328	1.000	0.025	0.025	0.307	0.519	0.388	0.181	0.007	0.452
Giden Veri-2	0.096	0.025	1.000	1.000	0.174	0.089	0.254	0.797	0.728	0.115
Gelen Veri-2	0.096	0.025	1.000	1.000	0.174	0.089	0.254	0.797	0.729	0.115
RAM Kullanım	0.426	0.307	0.174	0.174	1.000	0.604	0.840	0.582	0.170	0.793
CPU Kullanım	0.397	0.519	0.089	0.089	0.604	1.000	0.696	0.366	0.068	0.782
CPU Sıcaklık	0.426	0.388	0.254	0.254	0.840	0.696	1.000	0.618	0.144	0.947
Ortam Sıcaklık	0.303	0.181	0.797	0.797	0.582	0.366	0.618	1.000	0.490	0.469
Ortam Nem	0.010	0.007	0.728	0.729	0.170	0.068	0.144	0.490	1.000	0.072
Güç Tüketimi	0.422	0.452	0.115	0.115	0.793	0.782	0.947	0.469	0.072	1.000

Bunun yanında Ram ve Cpu kullanım oranı ile Ortam Sıcaklığı değerleri tüketilen güç ile yüksek korelasyon katsayısına sahip olduğu görülmektedir.

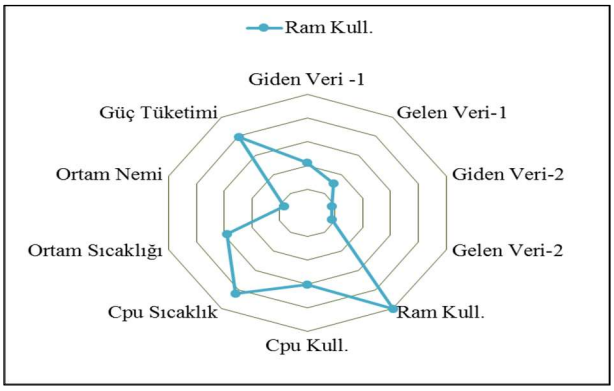
Cpu sıcaklık değerinin diğer parametreler ile ilişkisi Şekil 5.1'de gösterilmiştir.



Şekil 5.1: Cpu Sıcaklığı Korelasyon Grafiği.

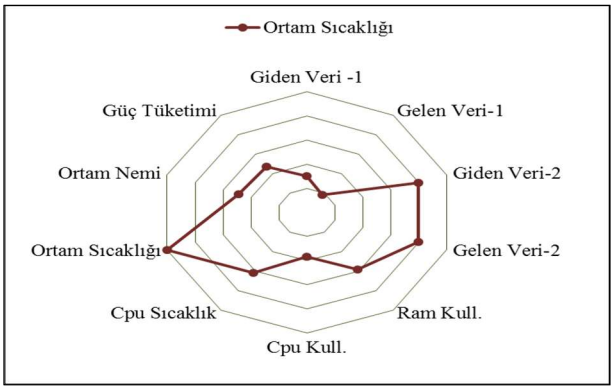
Gerçek güç tüketim oranı değeri üzerinde en büyük etkiye sahip Cpu sıcaklık değeri ile Ram kullanım oranı arasında 0.840 korelasyon değeri ile güçlü bir bağlantı olduğu görülmektedir.

Ram Kullanımına ait korelasyon grafiği Şekil 5.2'de gösterilmiştir. Ram kullanımını en çok etkileyen parametreler arasında Cpu sıcaklığı, Güç Tüketimi ve Cpu kullanım oranı bulunmaktadır.



Şekil 5.2: Ram Kullanımı Korelasyon Grafiği.

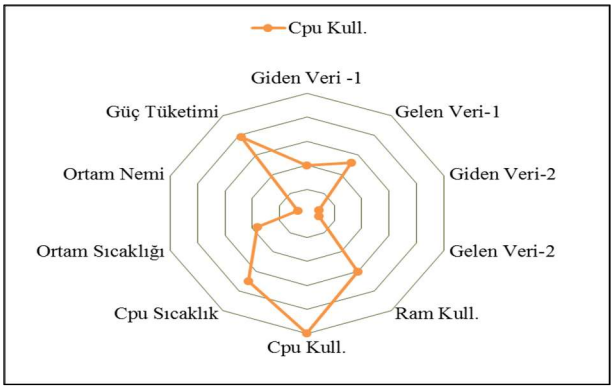
Cpu sıcaklığı parametresi üzerinde etkili diğer parametreler ise Ortam Sıcaklığı ve Cpu Kullanım Oranıdır. Cpu sıcaklığı parametresi, ortam sıcaklığı ve Cpu kullanım değerleri ile yüksek etkileşime sahiptir.



Şekil 5.3: Ortam Sıcaklığı Korelasyon Grafiği.

Ortam sıcaklığı parametresinin diğer parametreler ile olan korelasyon ilişkisi Şekil 5.3'te gösterilmiştir.

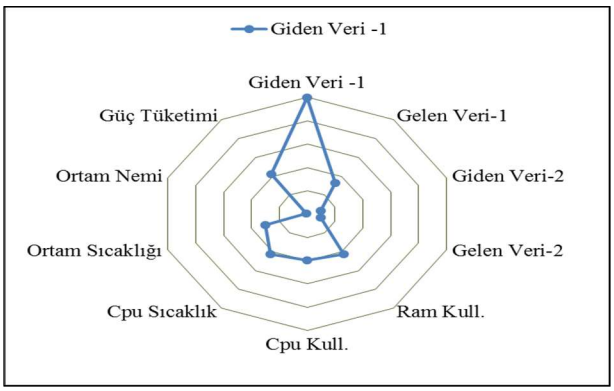
Gerçek Güç tüketimi üzerinde etkili bir diğer parametre de 0.782 korelasyon katsayısı ile Cpu Kullanım oranıdır. Bu değerin diğer parametreler ile olan etkileşimi Şekil 5.4'te gösterilmektedir.



Şekil 5.4: Cpu Kullanımı Korelasyon Grafiği.

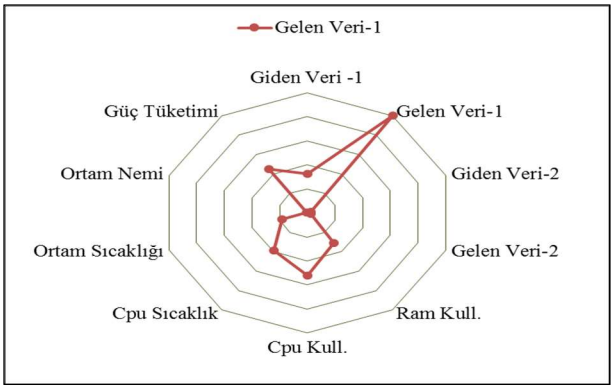
Bu çalışma ile literatüre katılmak istenen bir diğer şey ise veri merkezlerinin güç tüketimleri üzerinde sunucuların internet trafiğinin ne kadar etkili olduğunu anlayabilmektir. Çalışmamda kullanılan sunucunun iki adet ağ bağlantısı vardır. Bunun sebepleri ise; gerçek bir veri merkezinde sayıları oldukça fazla olan ağ bağlantılarının modellenebilmesini sağlayabilmek ve otomasyon sistemi ile eş zamanlı veri alışverişi yapabilmeyi olanaklı hale getirmektir. bir numaralı ağ bağlantısı üzerinden gelen ve giden veri miktarının yoğunluğu iki numaralı ağ bağlantısından daha yüksek olduğu görülebilir.

Bu sebeple tüketilen güç üzerinde bir numaralı ağ bağlantısının giden ve gelen veri parametreleri dikkate alındığında korelasyon matrisi tablosunda da görülebileceği üzere daha etkili olduğu anlaşılmaktadır. Bu parametrelerden Giden Veri -1'in diğer özellikler ile olan etkileşimi Şekil 5.5'te gösterilmiştir.



Şekil 5.5: Giden Veri -1 Korelasyon Grafiği.

Giden Veri -1'in en yüksek etkileşim oranına sahip olduğu parametre 0.426 korelasyon oranları ile Ram Kullanım oranı ve Cpu Sıcaklığıdır. Aynı Ethernet portuna ait Gelen Veri -1 parametresinin korelasyon etkileşim grafiği Şekil 5.6'da gösterilmiştir.



Şekil 5.6: Gelen Veri -1 Korelasyon Grafiği.

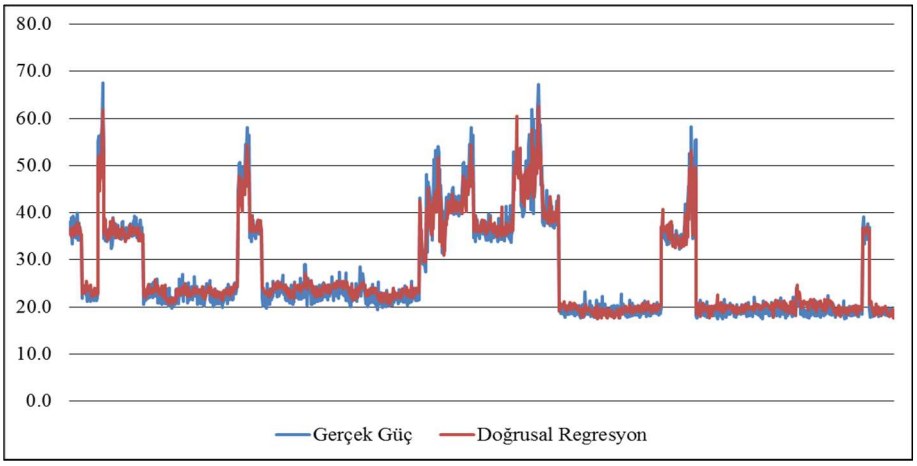
Gelen Veri-1 parametresi en yüksek korelasyon oranına 0.519 değeri ile Cpu Kullanım oranı parametresi ile etkileşimde olduğu görülmüştür.

İkinci ağ bağlantısının giden ve gelen veri parametreleri ile ortam nem değerinin tüketilen güç üzerindeki etkisinin çok sınırlı olduğu görülebilir. Korelasyon matrisi

tablosunda en düşük korelasyon katsayı değerlerinin bu parametreler için oluştuğu anlaşılmıştır.

5.2. Metot Karşılaştırmaları

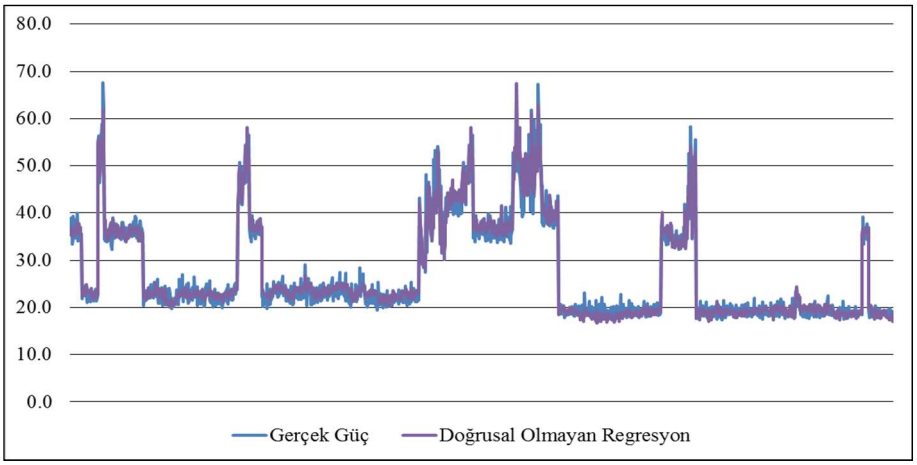
Sunucuya ait gerçek güç tüketim miktarı ile doğrusal, doğrusal olmayan ve farklılaştırılmış doğrusal olmayan regresyon yöntemleri ile yapılan yük tahmininin karşılaştırılması sonucunda elde edilen sonuçlar Şekil 5.7, Şekil 5.8 ve Şekil 5.9'daki grafiklerde gösterilmiştir;



Şekil 5.7: Gerçek Güç-Doğrusal Regresyon.

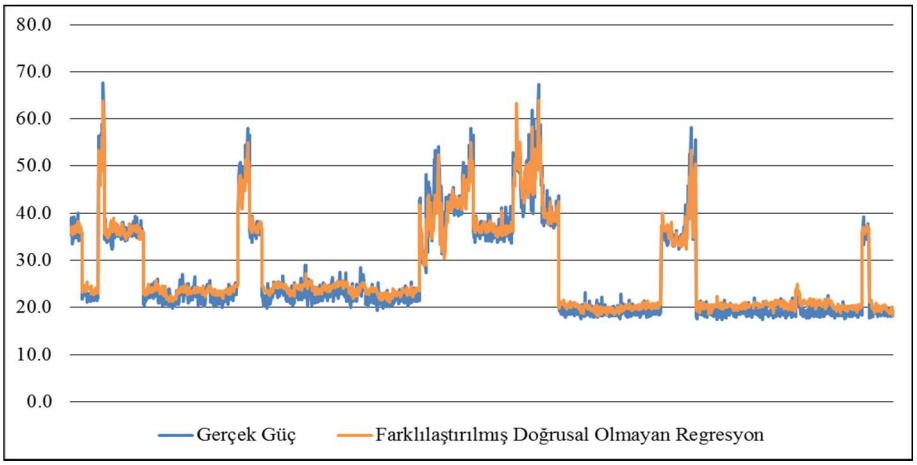
Şekil 5.7'de, sunucuya ait Gerçek Güç Tüketimi ve Doğrusal Regresyon yöntemi ile üretilen 1774 adet verinin tahmini yük profiline ait güç eğrileri gösterilmiştir. Geçmiş yük eğrisi verilerinden alınan ve tüm metotlar için kullanılan eğitim verisi sayesinde bu değişkenler "Curve Expert" programı ile hesaplanır.

Elde edilen değişkenler 1774 adet veri için 4.16 formülünde yerine yazılmak suretiyle işleme alınır. Elde edilen sonuçlar ile gerçek ölçüm değerleri bu şekilde karşılaştırılmıştır. Böylece gelecek dönemlere ait yük profilinin doğrusal regresyon yöntemi ile tahmin edilebilmesi amacıyla, test verisinin gerçek enerji tüketim değerleri ve doğrusal regresyon modelinin oluşturduğu tüketim değerleri ilişkisine ait grafik Şekil 5.7'de gösterilmiştir.



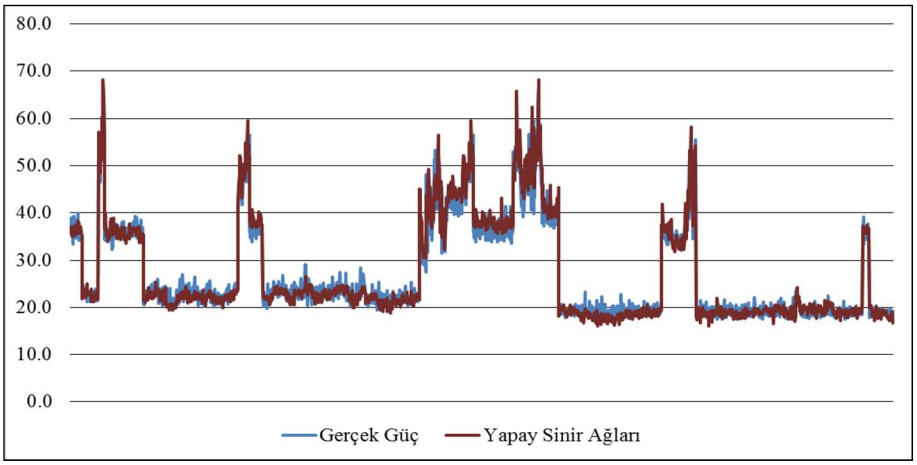
Şekil 5.8: Gerçek Güç-Doğrusal Olmayan Regresyon.

Şekil 5.8'de gösterilen 1774 adet veri için Gerçek Güç tüketimi ve Doğrusal Olmayan Regresyon yöntemi ile üretilen yük profili grafiği, doğrusal regresyon yönteminde olduğu gibi aynı eğitim ve test verileri kullanılarak oluşturulmuştur. Bu modelin doğrusal regresyon modeli ile farklılıkları ise, hesaplama yapılan ve daha önce 4.19 eşitliğinde gösterilmiş olan doğrusal olmayan regresyon formülünün kullanılması ve buna bağlı olarak değişken miktarındaki artış olmuştur.



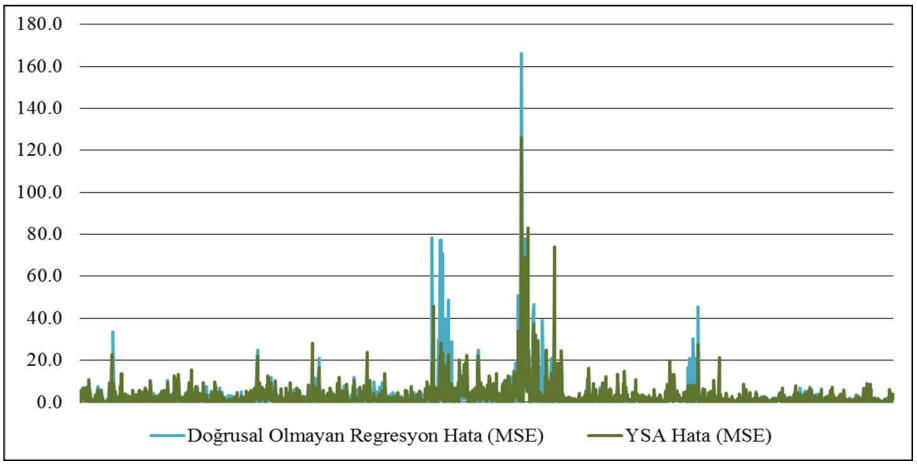
Şekil 5.9: Gerçek Güç-Farklılaştırılmış Doğrusal Olmayan Regresyon.

Şekil 5.9'da gösterilen Gerçek Güç Tüketimi ve Farklılaştırılmış Doğrusal Olmayan Regresyon yöntemi ile üretilmiş yük profili aynı eğitim ve test süreçlerinden geçilerek bulunmuştur. 4.19 formülündeki Doğrusal Olmayan Regresyon Yöntemine göre, 4.20 formülünde ek olarak β_{11} ve β_{12} parametreleri kullanılmıştır. Ayrıca 4.20 formülünde, 4.19 formülünde kullanılan α parametreleri kullanılmamıştır.



Şekil 5.10: Gerçek Güç-Yapay Sinir Ağları.

Bir diğer yük profili tahmin yöntemimiz olan YSA ile yapılan tahmin sonuçlarının gerçek güç değeri ile olan ilişkisi ise Şekil 5.10'daki grafikte görülebilmektedir.



Şekil 5.11: Farklılaştırılmış Doğrusal Olmayan Regresyon Hata-YSA Hata.

Dört farklı tahmin yöntem içerisinde gerçek güç tüketim eğrisine en yakın sonuçlar veren Doğrusal Olmayan Regresyon ve YSA yönteminin yük profili üretirken yaptıkları hata miktarları dikkate alındığında bu iki yöntemin karşılaştırma grafiği Şekil 5.11'de gösterilmiştir. YSA ile üretilen yük profili sonucunda elde edilen enerji tüketim değerlerinin ortalama hatası 2.87, Doğrusal Olmayan Regresyon yöntemi ile üretilen enerji tüketim değerlerine ait ortalama hata değeri ise 2.92 olmuştur. Bu sonuçlara göre YSA modelinin doğrusal olmayan regresyon modeline göre daha düşük oranda hata yaptığı anlaşılmıştır.

Bu tasarımın sonucu olarak elde edilen tahmin değerlerinin gerçek güç tüketim verilerine ne kadar iyi yaklaştığının, verimli ve doğru sonuçlar verdiğinin ölçülebilmesi için Ortalama Karesel Hata (RMSE-Root Mean Squar Error) ve Korelasyon Katsayısı parametreleri hesaplanacaktır.

Kare ortalamalarının hata oranları yöntemini kullanarak eğri uydurma grafiği üzerindeki gerçek değer ile tahmin edilen arasındaki hata miktarı hesaplanmıştır. Sonrasında bu değerlerin kareleri alınıp eğri üzerindeki tüm noktalar için bu değerler toplandıktan sonra örnek sayısına bölünerek çıkan sonucun karesi alınarak ortalama karesel hata değeri hesaplanmıştır. Bu hata ölçüm metodunu kullanmamızın sebebi, çok farklı alanlarda [27], [28] en yaygın kullanılan yöntemlerden bir tanesi olduğu içindir.

Hesaplanan bu değer ne kadar küçük olursa gerçek güç tüketim verisine o kadar yakın tahminlerde bulunmuş olduğumuz anlamına gelmektedir. İstatistikte bu yöntem regresyonun standart hatası ya da uygunlaştırma standart hatası olarak isimlendirilebilir [7].

Ortalama Karesel Hata değeri, 5.1 formülünde de görülebileceği üzere hata büyüklüğünü ölçmek için kullanılan ve gerçek değer ile tahmin edilen arasındaki farkın karesinin ortalamalarının karekökü olan değerdir.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{5.1}$$

X ve Y gibi iki parametre arasındaki korelasyon analizi, bu iki parametrenin arasındaki benzerliğin veya ilişkinin yönü ve büyüklüğü hakkında bilgi edinebileceğimiz 5.2 formülü ile hesaplanan bir diğer değerdir.

Korelasyon ifadesi;

$$\mathrm{Korelasyon}(x, y) = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}} \tag{5.2}$$

Bu parametreler aynı zamanda tahmin edilen ve gerçek güç tüketim değerlerinin birbirlerine ne kadar yakın olduğu sonuçlarını Tablo 5.2'de göstermektedir. Bu parametreler, dört farklı yük tahmin yöntemi için de aynı örnek veri toplulukları için aynı yöntemle hesaplanmıştır;

Tablo 5.2: Metot Karşılaştırma Tablosu.

Metot İsmi	Örnek Sayısı	Ortalama Karesel Hata (RMSE)	Korelasyon Katsayısı
Doğrusal Regresyon	1774	1.842	0.987
Doğrusal Olmayan Regresyon	1774	1.709	0.986
Farklılaştırılmış Doğrusal Olmayan Regresyon	1774	1.925	0.987
Yapay Sinir Ağı	1774	1.693	0.988

5.3. Değerlendirme

1774 adet test verisi için yapılan analiz sonucunda tahmin yöntemlerinin performansları Ortalama Karesel Hata ve Korelasyon parametreleri dikkate alınarak Tablo 5.2'de gösterildiği üzere hesaplanmıştır. Bu sonuçlardan, YSA modelinin, tüm Regreyon yöntemleri ile yapılan tahmin sonuçlarına göre daha düşük Ortalama Karesel Hata ve daha yüksek Korelasyon Katsayısı parametresine sahip olduğu görülmüştür.

Analiz sonuçlarında, YSA yöntemi ile yapılan çalışmanın diğer yöntemlere göre daha iyi sonuçlar ürettiği görülmüştür. Ayrıca doğrusal olmayan regresyon yöntemi ile elde edilen sonuçların, YSA modeli ile elde edilen sonuçlara oldukça yakın olduğu da görülmüştür. Bu sebeple farklı ham veriler kullanılarak analizler yeniden yapılabilecek ve sonuçlar değerlendirilebilecektir. Çalışmamda kullanılan bu uygulama modeli, veri merkezleri ve IT sistemleri için aşağıdaki sonuçlar olan

- mevcut veya yeni veri merkezleri ve IT sistemlerine ait yük profillerinin üretilebilmesini bağımsız değişkenlere bağlı olarak tahmin edebilmek,
- veri merkezlerinin enerji tüketimlerini etkileyen parametrelerin etki oranlarını bulabilmek ve tüm giriş ve çıkış parametreleri arasındaki etkileşim ile ilgili analizler yaparak, tasarımlar geliştirmek,

- doğrusal olmayan regresyon ve YSA yöntemlerinin enerji tüketim tahmini dikkate alınarak elde edilen sonuçlarda daha doğru ve tutarlı sonuçlar üretebildiği görülebilmiştir.

6. SONUÇLAR

Regresyon metotları ve YSA yöntemi ile tahmin problemlerinin çözümün ilişkin oldukça fazla sayıda uygulama ve analiz çalışmaları geçmiş literatür örneklerinde gerçekleştirilmiştir. Bu konuda oldukça fazla sayıda makale yayınlanmış ve sonuçlar paylaşılmıştır.

Bu tez çalışmamdaki esas amacım, sadece veri merkezleri ve bilgi teknolojileri sistemlerinin enerji tüketim miktarlarını nasıl tahmin edildiğini uygun yazılım ve cihazlar kullanarak açıklanması değil aynı zamanda dört farklı tahmin yönteminin karşılaştırılması, analiz edilmesi ve parametre ilişkilerinin değerlendirmesi şeklindedir.

Dünya enerji tüketiminde bu kadar yüksek paya sahip tesislerin enerji tüketimlerini azaltmak ve optimize edebilmek için birçok sebep bulunmaktadır. Büyük kapasiteli veri merkezlerine olan ulaşılabilirlik problemi ve yapılacak çalışmanın uygulanabilirliği dikkate alındığında çalışmam bir adet sunucu üzerinde gerçekleştirilmiştir.

Bir adet sunucu ile bu çalışmayı yapmamızın sağladığı avantaj, daha uygulanabilir olmasının yanında farklı giriş parametrelerinin az sayıda sensör ve giriş -çıkış modülü kullanılarak eş zamanlı olarak alınabilmesi olmuştur. Fazla sayıda sunucunun ve diğer IT cihazlarının olduğu bir ortamda farklı noktalardan uygun ölçümlerin yapılması gerekliliği olduğu kaçınılmazıdır. Bu uygulamanın dezavantajı ise veri merkezlerindeki farklı cihazların uygulanan modellerde ve analizlerde işleme dahil edilemeyişi olmaktadır.

Tez çalışmamın amacına ulaşabilmesi için birçok farklı türde cihazlar (Endüstriyel Bilgisayar, Sunucu, Enerji Analizörü, Sıcaklık ve Nem Sensörleri) bir araya getirilmiştir. Ayrıca iki farklı yazılım bu çalışmam için veri toplama işlemi ve YSA yöntemlerinin uygulanması amacıyla geliştirilmiştir. Çok Katmanlı Algılayıcı (MLP) modeli ve Geri Yayılım Algoritması (Backpropagation) YSA modelini en verimli şekilde eğitmek için kullanılmıştır. Regresyon analizlerini uygulayabilmek amacıyla “Curve Expert” programı düzenlenerek kullanılmıştır.

Bütün bu yazılım ve donanım cihazları giriş parametreleri ile gerçek güç tüketim değeri olan bağımsız değişkenler arasındaki ilişkiyi bulabilmek amacıyla bir araya getirilmiş, tasarlanmış, programlanmış ve düzenlenmiştir.

Giriş parametreleri ve çıkış parametresinin her birinin diğerleri ile olan etkileşimine ait Korelasyon Matrisi Tablosu oluşturulmuştur. Bu tablo sayesinde hangi parametrenin diğer parametreleri ne kadar etkilediği görülmüştür.

Dört farklı metot ve dokuz farklı giriş parametresi ile yapılan analizler sonucunda elde edilen yük profilleri gösterilmiş, bu profillerin başarı oranlarının karşılaştırılması amacıyla Ortalama Karesel Hata ve Korelasyon Katsayısı parametreleri bulunmuştur.

Böylece, dünya enerji tüketiminin en büyük müşterilerinden olan veri merkezleri ve bilgi teknolojileri sistemlerinin enerji tüketimlerini bir adet sunucu üzerinde uygulanan metotlar ile tahmin edebilmek bu çalışmanın esas amaçlarından birisi olmuştur. Sonuç olarak, YSA metodunun diğer Regresyon yöntemlerine göre veri merkezlerinin enerji tüketimlerini tahmin etmede daha başarılı olduğu ve daha iyi sonuçlar ürettiği görülmüştür.

Tez çalışmam, literatürdeki benzer bir çalışma ile birçok ortak özellik göstermesine rağmen, ortam sıcaklığı ve nemi parametreleri, Cpu sıcaklığı vs. gibi farklı parametreler kullanması ve ayrıca farklı güç tahmin yöntemlerini de kullanarak sonuçlar vermesi açısından farklılıklar göstermektedir [38]. Bir diğer çalışmada bahsedilen yönlendirme tekniği ise tez çalışmamda gerçekleştirilen öğrenme algoritmaları ile uygulanarak aktif veri merkezleri için bir çalışma yapılabilecektir [42]. Simülasyon temelli bir yöntem ile yapılan literatür çalışmasında, tez çalışmamdaki gibi gerçek bir veri merkezi ya da sunucuya ait yük profili üretme yazılımları ile birlikte çalıştırılarak optimizasyon seçenekleri değerlendirilebilecektir [43]. Bu çalışmam, farklı metotlar ve algoritmalar için ilave giriş parametreleri de dikkate alınmak suretiyle geliştirilebilecektir. Bir adet sunucu üzerinde yapılan veri merkezlerinin modellenmesi çalışmam, büyük kapasiteli veri merkezleri için ilave özelliklerden olan soğutma sistemi parametreleri, dış hava sıcaklıkları vs. parametreleri de dikkate alınarak yeni çalışmalar yapılabilir. Ayrıca gelecek dönemlerdeki hedef çalışmalar, enerji verimlilik performansının ölçülmesi ve enerji tüketimine ait optimizasyon seçeneklerinin belirlenmesi şeklinde geliştirilebilecektir.

KAYNAKLAR

[1] Web 1, (2016), http://www.independent.co.uk/environment/global-warming-data-centres-to-consume-three-times-as-much-energy-in-next-decade-experts-warn-a6830086.html, (Erişim Tarihi: 23/01/2017).

[2] Lee H., Pitelis C., Kim K., Choi Y., Dinev B., Yonkov Y., Hocaoglu H., Gozel T., Yenice S., Aktas M., (2017), "A training network for green data centres operations management through staff exchange", 9th International Conference on Education and New Learning Technologies, Barcelona, Spain, 3-5 July.

[3] Web 2, (2017), http://www.enerji.gov.tr/tr-TR/Sayfalar/Elektrik, (Erişim Tarihi: 08/05/2017).

[4] Whitehead B., Andrews D., Shah A., and Maidment G., (2014), "Assessing the environmental impact of data centres Part 1: Background, energy use and metrics", Building Environ., 82, 151-159.

[5] Web 3, (2017), https://gesi.org/report/detail/gesi-smarter2020-the-role-of-ict-in-driving-a-sustainable-future, (Erişim Tarihi: 02/10/2017).

[6] Malmodin J., Bergmark P., Lunden D., (2013), "The future carbon footprint of the ICT and E&M sectors", Proceedings of the first international conference on information and communication technologies for sustainability, 12-20, Zurich, Switzerland, 14-16 February.

[7] Shcherbakov M. V., Brebels A., Shcherbakova N. L., Tyukov A. P., Janovsky T. A., Kamaev V. A., (2013), "A Survey of Forecast Erro Measures," World Applied Sciences Journal 24 (Information Technologies in Modern Industry, Education & Society), 171-176.

[8] Kelly H. Z., Tuncali K., Silverman S. G., (2003), "Correlation and Simple Linear Regression", Radiology, 227 (3), 617-628.

[9] Haykin S., (2008), "Neural Networks and Learning Machines", 3rd Edition, Pearson Prentice Hall.

[10] Web 4, (2017), https://www.otago.ac.nz/ripe/otago104254.pdf , (Erişim Tarihi: 12/08/2017).

[11] LeCun Y., Bottou L., Orr G.B., Müller K.R., (1998), "Efficient BackProp" Neural Networks: Tricks of the Trade, 7700, 2012, 9-48.

[12] Karabinaoglu M. S., Gözel T., (2017), "Load forecasting Modelling of Data Centers and IT systems by Using Artificial Neural Networks," ELECO 10th International Conference on Electrical and Electronics Engineering, 178, Bursa-Turkey, 30 November-2 December.

[13] Rossi M., Brunelli D., (2015), ”Forecasting data centers power consumption with the Holt-Winters method”, Environmental, Energy and Structural Monitoring Systems (EESMS), 210-214, Trento, Italy, 9-10 July.

[14] Khan A. R., (2016), “Data Center Load Forecast Using Dependent Mixture Model,” Open Public Research Access Instutional Repository and Information Exchange, Theses and Dissertations, South Dakota University.

[15] Zhou C. C., Zhang F., Yuan Z., Zhou L. M., Wang F. M., Li W. and Yang J. H., (2016), ”Study on load forecasting to data centers of high power density based on power usage effectiveness,” 2016 International Conference on New Energy and Future Energy System (NEFES 2016), 40 (1), Bejing, China, 19-22 August.

[16] Web 5, (2013), https://statik.tse.org.tr/ upload/tr/dosya/icerikyonetimi/2222/17032015164319-3.pdf, (Erişim Tarihi: 21/03/2018).

[17] Web 6, (2018), https://uptimeinstitute.com/publications/asset/tier-standard-topology, (Erişim Tarihi: 21/03/2018).

[18] Gao J., (2014), “Machine learning applications for data center optimization”, Google White Paper, 27 October.

[19] Verma T., Tiwana A. P. S., Reddy C. C., Arora V., Devanand P., (2016), ”Data Analysis to Generate Models Based on Neural Network and Regression for Solar Power Generation Forecasting”, 7th International Conference on Intelligent Systems, Modelling and Simulation, 97-100, Bangkok, Thailand, 25-27 January.

[20] Srivastava S., Tripathi K.C., (2012), ”Artificial Neural Network and Non-Linear Regression:A Comperative Study”, International Journal of Scientific and Research Publications, 2, 12.

[21] Landi A., Piaggi P., Laurino M., Menicucci D., (2010), ”Artificial Neural Networks for Nonlinear Regression and Classification”, 2010 10th International Conference on Intelligent Systems Design and Applications, 210, Cairo, Egypt, 29 November-1 December.

[22] Mazumder J., Harley R. G., Lembert F. C., Venayagamoorthy G. K., (2007), ”Neural Network Based Method for Prediction Nonlinear Load Harmonics”, IEEE Transactions On Power Electronics, 22, 3.

[23] Kim G. H., An S. H., Kang K. I., (2004), ”Comparsion of Construction Cost Estimating Models Based On Regression Analysis, Neural Networks, and Case-Based Reasoning”, Building and Environment 39, 1235-1242.

[24] Vinutha H. D., Gouda K. C., Chandan K. N., (2014), “Electric Load Forecasting using a Neural Network Approach,“ International Journal of Computer Trends and Technology (IJCTT), 11, 6.

[25] Dayarathna M., Wen Y., Fan R., (2016), “Data center energy consumption modeling: A survey”, IEEE Communications Surveys & Tutorials, 18 (1), 732-794.

[26] Rawlings J. O., Pantula S. G., Dickey D. A., (1998), ”Applied regression analysis: a research tool, 2nd edition”, Springer Texts in Statistics.

[27] Gooijer, J. G. D., Hyndman R. J., (2006), “25 years of time series forecasting”, International Journal of Forecasting, 22 (3), 443-473.

[28] Hyndman R. J., Koehler A. B., (2006), “Another look at measures of forecast accuracy”, International Journal of Forecasting, 22 (4), 679-688.

[29] Karaca C., Karacan H., (2016), “Çoklu Regresyon Metoduyla Elektrik Tüketim Talebini Etkileyen Faktörlerin İncelenmesi”, Sujest, 4 (3).

[30] Rumelhart D. E., Hinton G. E., Williams R. J., (1986), “Learning Internal Representations By Error Propagation”, Parallel distributed processing: explorations in the microstructure of cognition, 1, 318-362

[31] Hyams D. G., (2018), “CurveExpert Professional Documentation”, Hyams Development, 2.6.5, 15 March.

[32] NRDC, (2014), Data Center Efficiency Assessment, The Natural Resources Defense Council.

[33] SVLG, (2008), Data Center Energy Forecast Report, Silicon Valley Leadership Group.

[34] Prevost J. J., Nagothu K., Kelley B., Jamshidi M., (2011), “Prediction of Cloud Data Center Networks Loads Using Stochastic and Neural Models”, 2011 6th International Conference on System of Systems Engineering (SoSE), 276-281, Albuquerque, New Mexico, USA, 27-30 June.

[35] Hintemann R., (2015), “Energy consumption of data centers continues to increase -2015 update”, Borderstep Institute, Germany.

[36] Coles H., Greenberg S., Tschudi W., (2013), “Demonstration of Data Center Energy Use Prediction Software”, Ernest Orlando Lawrence Berkeley National Laboratory, September.

[37] Li L., Liang C. M., Liu J., Nath S., Terzis A., Faloutsos C., “ThermoCast: A Cyber-Physical Forecasting Model For Data Centers”, (2011), 17th ACM SIGKDD international conference on Knowledge discovery and data mining, 1370-1378, San Diego, California, USA, 21-24 August.

[38] Makris T., (2017), “Measuring and Analyzing Energy Consumption of the Data Center”, The Degree of Master of Science in Technology, Aalto University School of Electrical Engineering.

[39] Verma A., Dasgupta G., Nayak T.K., De P., Kothari R., (2009), "Server Workload Analysis for Power Minimization using Consolidation", USENIX'09 Proceedings of the 2009 conference on USENIX Annual technical conference, 28-28, San Diego, California, 14-19 June.

[40] Berral J. L., Goiri I., Nou R., Julia F., Guitart J., Gavalda R., Torres J., (2010), "Towards energy-aware scheduling in data centers using machine learning", 1st International Conference on Energy - Efficient Computing and Networking, 215-224, Passau, Germany, 13-15 April.

[41] Shang Y., Li D., Xu M., (2010), "Energy-aware Routing in Data Center Network", Proceeding of the first ACM SIGCOMM workshop on Green networking, 1-8, New Delhi, India, 30 August.

[42] Heller B., Seetharman S., Mahadevan P., Yiakoumis Y., Sharma P., Banerjee S., McKeown N., (2010), "ElasticTree: Saving Energy in Data Center Networks", NSDI'10 Proceedings of the 7th USENIX conference on Networked systems design and implementation", 17-17, San Jose, California, 28-30 April.

[43] Carrega A., Singh S., Bruschi R., Bolla R., (2012), "Traffic Merging for Energy-Efficient Datacenter Networks", 2012 International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS), 1-5, Genoa, Italy, 8-11 July.

[44] Gao Y., Guan H., Qi Z., Wang B., Liu L., (2013), "Quality of service aware power management for virtualized data centers", Journal of Systems Architecture, 59, 245-259.

[45] Cheung H., Wang S., Zhuang C., Gu J., (2018), "A simplified power consumption model of information technology(IT) equipment in data centers for energy system real-time dynamic simulation", Applied Energy, 222, 329-342.

[46] Jayalakshmi T., Santhakumaran A., (2011), "Statistical Normalization and Back Propagation for Classification", International Journal of Computer Theory and Engineering, 3 (1), 1793-8201.

[47] Karlik B., Olgaç A. V., (2010), "Performance Analysis of Various Activation Functions in Generalized MLP Architectures of Neural Networks", International Journal of Artificial Intelligence And Expert Systems (IJAE), 1 (4).

[48] Masters T., (1993), "Practical Neural Network Recipes in C++", 1st Edition, Academic Press.

[49] Mekanik F., Imteaz M.A., Gato-Trinidad S., Elmahdi A., (2013), "Multiple regression and Artificial Neural Network for long-term rainfall forecasting using scale climate modes", Journal of Hydrology, 11-21.

[50] Tamizharasi G., Kathiresan S., Sreenivasan K.S., (2014), "Energy Forecasting using Artificial Neural Networks", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 3 (3).

[51] Birim S., Tümtürk A., (2016), "Modeling and Forecasting Turkey's Electricity Consumption by using Artificial Neural Network", American Scientific Research Journal for Engineering Technology and Sciences, 192-208.

[52] Samuel I. A., Emmanuel A., Odigwe I. A., Felly-Njoku C., (2017), "A Comparative Study of Regression Analysis and Artificial Neural Network Methods for Medium-Term Load Forecasting", Indian Journal of Science and Technology, 10 (10).

[53] Badran S. M., (2012), "Neural Network Integrated With Regression Methods To Forecast Electrical Load", 11th IET International Conference on Developments in Power Systems Protection, 1-6, Birmingham, UK, 23-26 April.

[54] Bagirov A. M., Mahmood A., Barton A., (2017), "Prediction of monthly rainfall in Victoria, Australia: Clusterwise linear regression approach", Atmospheric Research, 188, 20-29.

[55] Çevik H. H., Harmancı H., Çunkaş M., (2016), "Short-term Load Forecasting based on ABC and ANN for Smart Grids", International Journal of Intelligent Systems and Applications in Engineering, 4, 38-43.

[56] Hamedmoghadam H., Joorabloo N., Jalili M., (2018), "Australia's long-term electricity demand forecasting using deep neural networks", CoRR, 1801, 02148.

[57] He W., (2014), "Deep neural network based load forecast", Computer Modelling & New Technologies, 18 (3), 258-262.

[58] Dzhaparova D. A., Kalinin V. F., Zyablov N. M., Kobelev A. V., Kochergin S. V., Korolyova L. Y., (2017), "Electric Load Forecasting in Electrical Power Supply Systems", International Journal of Applied Engineering Research, 12 (24), 15278-15286.

ÖZGEÇMİŞ

Murat Salim KARABİNAOĞLU 1988 yılında Trabzon'da doğdu. 2004 yılında başladığı Erciyes Üniversitesi Mühendislik Fakültesi Elektrik-Elektronik Mühendisliği Bölümünü 2009 yılında başarıyla tamamladı. 2016 yılında Gebze Teknik Üniversitesi Fen Bilimleri Enstitüsü Elektronik Mühendisliği Anabilim Dalında başladı. 2009 -2015 yılları arasında özel sektörde çalıştıktan sonra 2015 yılından bu yana TÜBİTAK UME 'de Ar-Ge Destek ve Bilgi Sistemleri Laboratuvarında Uzman Araştırmacı olarak çalışmaktadır.

EKLER

Ek A: Tez Çalışması Kapsamında Yapılan Yayınlar

Karabinaoglu M. S., Gözel T., (2017), "Veri Merkezleri & IT Sistemleri için YSA ile Enerji Yük Profillerinin Oluşturulması," GTÜ Lisansüstü Araştırmalar Sempozyumu, 113, Kocaeli, Türkiye, 17-18 Mayıs.

Karabinaoglu M. S., Gözel T., (2017), "Load forecasting Modelling of Data Centers and IT systems by Using Artificial Neural Networks," ELECO 10th International Conference on Electrical and Electronics Engineering, 178, Bursa, Turkey, 30 November-2 December.