



# Fotovoltaik Sistemlerde Topluluk Öğrenmesi Temelli Hata Tespiti

*Araştırma Makalesi/Research Article*

 Ceyhun KAPUCU<sup>1</sup>,  Mete ÇUBUKÇU<sup>2</sup>

<sup>1,2</sup> Güneş Enerjisi Enstitüsü, Ege Üniversitesi, İzmir, Türkiye  
[ckapucu@gmail.com](mailto:ckapucu@gmail.com), [mete.cubukcu@ege.edu.tr](mailto:mete.cubukcu@ege.edu.tr)  
 (Geliş/Received:04.01.2019; Kabul/Accepted:05.03.2019)  
 DOI: 10.17671/gazibtd.508475

**Özet**— Fotovoltaik (FV) sistemler için doğru bir hata tespit yeteneği, işletme maliyetlerini ve bir arıza nedeniyle oluşabilecek devre dışı kalma sürelerini azaltarak FV sistemin verimliliğini artırabilir. Bu çalışmada, FV sistemler için bir hata tespit yöntemi önerilmiştir. Önerilen yöntem, topluluk öğrenmesi temelli bir modelin FV sistemlerdeki hataları sınıflandırmak amacıyla kullanılmasına dayanmaktadır. Topluluk öğrenmesi yöntemi, tek bir öğrenme algoritmasının genelleme yeteneğinin ve sağlamlığının üstüne çıkabilmek için farklı algoritmaların tahminlerini birleştirir. Bu çalışmada, sınıflandırma problemlerinde yaygın olarak kullanılan bazı öğrenme algoritmalarından bir topluluk öğrenmesi modeli oluşturulmuştur. Topluluk modeli, daha sonra parametre optimizasyonu uygulanarak geliştirilmiştir. Öğrenme algoritmalarının her biri ve bunları birleştiren topluluk modeli tahmin doğrulukları açısından karşılaştırılmıştır. Önerilen yöntem, Scikit-learn makine öğrenme kütüphanesi ile Python kullanılarak gerçekleştirilmiştir. Yöntemin deneysel geçerliliği Muğla'da (Türkiye) kurulu bir konut tipi FV sistemden elektriksel ve meteorolojik ölçüm verileri kullanılarak yapılmıştır. Sonuçlar, optimize edilmiş bir topluluk öğrenmesi modeliyle, önerilen yöntemin yalnızca sınıflandırma doğruluğunu geliştirmediğini, aynı zamanda fotovoltaik sistem hata tespiti için güçlü bir genelleme yeteneğine de sahip olduğunu göstermektedir.

**Anahtar Kelimeler**— hata tespiti, sınıflandırma, topluluk öğrenmesi

## Photovoltaic System Fault Detection Based On Ensemble Learning

**Abstract**— An accurate fault detection capability for photovoltaic (PV) systems can improve PV system productivity by reducing operational costs and possible downtimes caused by a failure. In this paper, a fault detection method for PV systems is proposed. The proposed method is based on the use of an ensemble learning based model for classifying faults in PV systems. Ensemble learning combines the predictions of different algorithms in order to improve generalizability and robustness over a single learning algorithm. In this study, an ensemble learning model is built from some learning algorithms that commonly used in the classification problems. The ensemble model is then improved via parameter optimization. Each learning algorithms and the ensemble model that combines them are compared in terms of their prediction accuracy. The proposed method was implemented using Python with Scikit-learn machine learning library. The experimental validation of the method has been performed using electrical and meteorological measurements data from a residential PV system installed in Muğla (Turkey). Results show that, with an optimized ensemble learning model, the proposed method not only improves the classification accuracy but also has a strong generalization ability for PV system fault diagnosis.

**Keywords**— fault detection, classification, ensemble learning

### 1. GİRİŞ (INTRODUCTION)

FV güç sistemlerinin izlenmesi, bu sistemlerin verimli çalışması açısından önemli bir işlemdir. Verim açısından izlenmesi gereken en önemli değişken ise gerçekleştirme

oranıdır. Son 20 yıl boyunca, FV güç sistemleri kurulumlarında gerçekleştirme oranı ortalaması %65'ten %85'e yükselmiştir. Bu yükseliş, FV güç sistemlerinin izlenmesi ve toplanan verilerin analiz edilmesiyle mümkün olmuştur. Gerçekleştirme oranını etkileyen en önemli

nedenlerin başında, güneş paneli çıkış gücünü düşüren yüksek panel sıcaklığı, panel yüzeyine gelen güneş ışığını engelleyen kir-toz ve benzeri etmenler, paneller arası uyumsuzluk ve kablolama kayıpları ile AC çıkış veren sistemlerdeki DC-AC dönüşümünü yapan evirici kayıpları gösterilebilir [1]. Bu olası kayıplarla birlikte FV güç sisteminin performansını belirlemek, hataları tanımlamak ve enerji verimliliğini takip etmek bir izleme sisteminin temel amaçlarındandır. Günümüz izleme sistemleri genellikle FV dizi seviyesinde ve evirici tarafında yapılmaktadır. Orta ve büyük ölçekli bir FV sistemde çok fazla sayıda panel olduğundan, bir hata durumunda hatalı panelleri içeren diziyi doğru ve hızlı bir şekilde tespit etmek önemli hale gelmektedir [2]. FV sistemler için hata tespiti yöntemleri genel olarak, görsel, termal ve elektriksel yöntemler şeklinde gruplandırılabilir [3]. Elektriksel yöntemi temel alan yaklaşımlar ise sırayla; güneş ışınımı, sıcaklık gibi meteorolojik verilere ihtiyaç duyulmayan yaklaşımlar, FV güç sisteminin akım, gerilim karakteristiğinin analiz edildiği yaklaşımlar, en yüksek güç noktasının takip edildiği yaklaşımlar, yapay zekâ teknikleri kullanan yaklaşımlar ve bunların dışında kalan diğer yaklaşımlar olarak ele alınabilir [4]. Son yıllarda, FV sistemlerde hata tespitinde yapay zekâ tekniklerinin kullanımı giderek artmaktadır.

Mekki ve arkadaşları çalışmalarında, kısmi gölgeleme altındaki FV panelin üreteceği çıkış gerilimi ve akımını tahmin edebilen ve böylelikle hata durumunu yakalayabilen yapay sinir ağları (YSA) tabanlı bir yöntem geliştirmişlerdir [5]. Önerdikleri yöntem, ölçülen FV çıkış değerleri ile YSA tarafından tahmin edilen değerler arasındaki farkın belirli bir eşik değerini aşıp aşmadığının kontrolü şeklinde çalışmaktadır. Fark eşik değerini aşarsa, çalışma durumu hatalı bir durum olarak kabul edilir. Çalışmada geliştirilen YSA tabanlı modelin, FV panel çıkış değerlerini doğru olarak tahmin edebileceği ve çıkış gücünde herhangi bir düşüşü tespit edebileceği gösterilmiştir. Bununla birlikte ağırlık çıkış değerlerini yüksek doğrulukla tahmin edebilmesi için periyodik olarak eğitilmesi önerilmiştir. Önerilen yöntemde, açık devre, kısa devre gibi hata tipleri dikkate alınmamıştır. Zhao ve arkadaşları bir çalışmalarında FV sistemlerde hata tespiti yapabilmek için grafik tabanlı yarı denetimli bir öğrenme modeli önermişlerdir [6]. Önerilen yöntemle, etiketlenmiş verilerin gerçek çalışma koşullarında toplanması işleminin zorluğundan ve yüksek maliyetinden kaçınılması amaçlanmaktadır. Bununla birlikte, önerilen yöntem yalnızca açık devre ve kısa devre hatalarını tespit edebilir. Çalışmada FV sistemlerindeki kısmi gölgeleme hataları dikkate alınmamıştır. Chine ve arkadaşlarının çalışmasında FV sisteme ait akım, gerilim ve I/V eğrisindeki tepe noktalarının sayısını temel alan bir hata tespit tekniği önerilmiştir [4]. Önerilen yöntemde, FV çıkışları ve I/V eğrisindeki maksimum güç noktası sayısı, Matlab ortamında geliştirilen bir benzetim modeli kullanılarak elde edilmiştir. Elde edilen bu verilerle, algılayıcılardan okunan gerçek sisteme ait akım ve gerilim verileri arasındaki fark bir eşik değerini aşarsa, bu durumun olası bir hatanın varlığını göstereceği bildirilmiştir. Bonsignore ve arkadaşlarının çalışmasında FV sistemlerindeki düşük

ve yüksek topraklama hatalarını, kısa devre hatalarını ve kısmi gölgeleme hatalarını tanımlamak için Uyarlamalı Sinirsel-Bulanık Çıkarsama Sistemi temelli bir hata tespit yöntemi sunulmuştur [7]. Önerilen yöntem, FV panellerinin modellenmesini, bu modelden hesaplanan niteliklerin I/V karakteristik analizini ve bir Norm testinin uygulanmasını içermektedir. Daniel Riley ve Jay Johnson, bir çalışmalarında, FV sistemlerin sağlıklı çalıştığını kontrol eden ve hata durumunda uyarabilen, parçacık sürüsü optimizasyonu ile eğitilmiş YSA temelli bir yöntem önermişlerdir [8]. YSA'na en uygun çalışma durumunu öğretecek şekilde eğitmek için veriler, FV sistemin kurulmasını takiben minimum kirlilik, bozulma veya arızalarla çalıştığı varsayılan ilk iki ay boyunca toplanmıştır. Yöntem daha sonra YSA modeli tarafından tahmin edilen güç ile gerçek sistemden ölçülen gücü karşılaştırmakta ve farkın belirlenen bir eşik değerini aşması durumunda, FV sistemde bir bozulmaya veya bir arıza nedeniyle bir güç kaybına işaret etmektedir.

Literatürde sık rastlanan, ölçümlerin, beklenen değerlerle karşılaştırılmasına dayanan hata tespit yöntemleri, genel olarak olası hata durumunda uyarılmak için doğru bir eşik değerinin seçilmesini gerektirmektedir. Eşik belirleme yaklaşımı ise deneme yanılma şeklinde belirlenebilir veya FV sistem tasarımı uzmanlığını gerektirir.

Yukarıda bahsi geçen çalışmaların dışında, literatürde hatanın tespiti, sınıflandırılması ve hata yerinin tespiti konularında, yapay zekâ teknikleri kullanılan pek çok çalışma mevcuttur. Bu çalışmalardan [9] karar ağaçları temelli, [10], [11] bulanık mantık temelli, [12], [13], [14], [15], [16], [17] YSA temelli, [18] uzman sistemler temelli, [19] aykırı değer algılama kuralları temelli, [20], [21] uzatma sinir ağı temelli çalışmalardır. Bu çalışmaların dışında, matematiksel, istatistiksel, termografi, resim işleme teknikleri kullanarak hata tespiti yapan çalışmalarda vardır.

Son yıllarda literatürde pek çok alanda daha sık kullanılmaya başlanan topluluk öğrenmesi yöntemi, tek bir öğrenme algoritmasının genelleme yeteneğinin ve sağlamlığının üstüne çıkabilmek için farklı modelleri birleştirme yaklaşımına dayanmaktadır. Her ne kadar bu yaklaşım hem özellik seçimi hem de parametre optimizasyonu uzmanlığı gerektirse de, uygun özellikler seçildikten sonra, aday öğrenme algoritmalarını birleştirecek topluluk öğrenmesi modelinin sınıflandırma performansı, parametre optimizasyonu ile en üst düzeye çıkarılabilir. Topluluk öğrenmesine dayalı yöntemler, FV sistemleri dışında farklı alanlarda da aykırı değer algılama veya hata tespitinde kullanılmışlardır [22]. Ayrıca, FV enerji üretimi tahmininde [23], [24], [25] ve binalarda enerji tüketiminin kontrolünde önemli olan ısıtma ve soğutma yüklerinin tahmininde [26] topluluk öğrenmesi tabanlı yöntemlere dayanan çalışmalar vardır.

Bu çalışmada, bir FV sistemde meydana gelebilecek hata tiplerinden olan kısmi gölgeleme ve kısa devre hatalarının tespiti için topluluk öğrenmesini temel alan bir hata tespit yöntemi sunulmuştur. Önerilen yöntem, tek öğrenme

algoritmasına dayanan geleneksel hata tespiti yöntemlerinin sınırlı sınıflandırma performansının ve genelleme yeteneğinin kayda değer şekilde artırılabilmesine imkân vermektedir. Yöntem, deneysel amaçla tasarlanan ve izlenen bir FV güç sisteminin elektriksel çıktılarının yanında güneş ışınımı ve hücre sıcaklığı ölçümlerini de kullanarak hata tespiti yapabilmektedir. Çalışmanın ikinci bölümünde deneysel sistem tanıtılmaktadır. Üçüncü bölümde önerilen hata tespit yöntemi ve içerdiği topluluk öğrenmesi modelinin oluşturulması anlatılmaktadır. Çalışmanın dördüncü bölümünde modelin kullanılması ile elde edilen sonuçlar paylaşılmıştır. Son bölümde ise çalışmayla ilgili değerlendirmeler paylaşılmıştır.

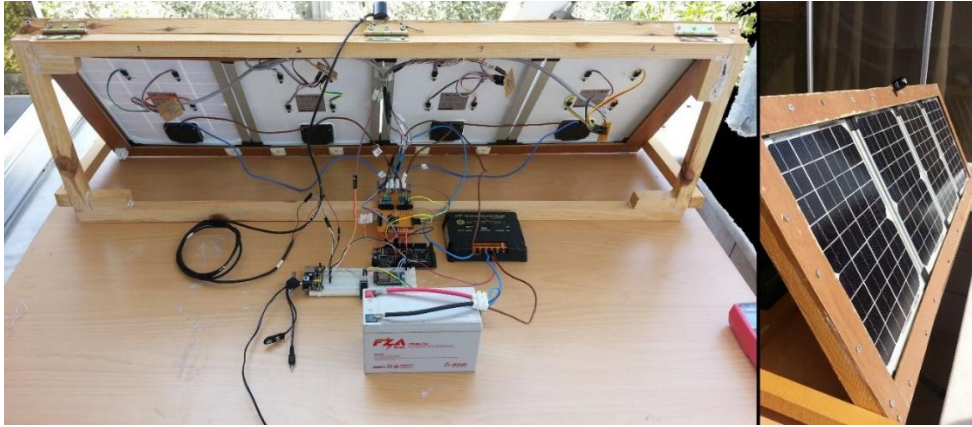
## 2. DENEYSEL SİSTEM (EXPERIMENTAL SYSTEM)

Elektriksel ve meteorolojik verilerin ölçüldüğü deneysel güneş enerjisi sistemi Şekil 1. 'de görülmektedir. Deneysel sistem Muğla ilindeki bir konut üzerinde kuruludur. Sistem, 2x2 şeklinde 4 adet güneş paneli, en yüksek güç noktası takibi (MPPT) yapabilen bir şarj kontrol cihazı, batarya, doğru akımla çalışan elektriksel bir yük ve temel parametreleri anlık olarak ölçen algılayıcılarla birlikte bunları kontrol eden ATmega2560 mikro-denetleyici ve ESP8266 WIFI yongası içeren Arduino tabanlı geliştirme kartı içermektedir. Kullanılan algılayıcılar, FV panellerin üzerine düşen güneş ışınımını ölçen bir piranometre (Apogee, PYR-S), panellerin arka yüzey sıcaklıklarını ölçen sıcaklık algılayıcıları (Maxim, DS18B20) ve her FV

dizininin güç çıkışını ölçen algılayıcılarıdır (Texas Instruments, INA226).

FV sistemlerdeki panellerin gölgelenmesi durumları ve kısa devre hataları Şekil 2 ve Şekil 3'te görüldüğü gibi gerçekleştirilerek hata senaryoları oluşturulmuş ve bu durumlarda ölçümlere devam edilerek ilgili hata durumlarında FV sistemin elektriksel çıktıları kaydedilmiştir. Tek ve çift panel üzerindeki gölgelenme senaryoları için Şekil 2'te görülen yarı saydam paneller kullanılarak FV panel üzerine düşen güneş ışınımı miktarının belirli bir seviyede azalması sağlanmıştır. FV paneller için kısa devre hata senaryosu ise FV panel geçiş (by-pass) diyotunun bir iletken kablo aracılığıyla kısa devre edilmesi şeklinde gerçekleştirilmiştir. Şekil 4'de bu çalışmada tespit edilen hata tipleri için hata senaryolarının nasıl gerçekleştirildiği şematik olarak gösterilmiştir. Şekil 4.(a) hatasız çalışma senaryosunun, Şekil 4.(b) tek paneldeki gölgelenme senaryosunun, Şekil 4.(c) iki paneldeki gölgelenme senaryosunun, Şekil 4.(d) ise tek paneldeki kısa devre senaryosunun nasıl gerçekleştirildiğini göstermektedir.

Algılayıcılar tarafından 30 saniyelik aralıklarla ölçülen veriler, hafıza kartına depolanmış ve internet erişiminin mümkün olduğu durumlarda da bulut bilişim katmanındaki veri tabanına kaydedilmiştir. Bulut bilişim katmanına gönderilen bu veriler hata tespiti için geliştirilen topluluk öğrenmesi modelinin eğitilmesinde ve test edilmesinde kullanılmıştır.



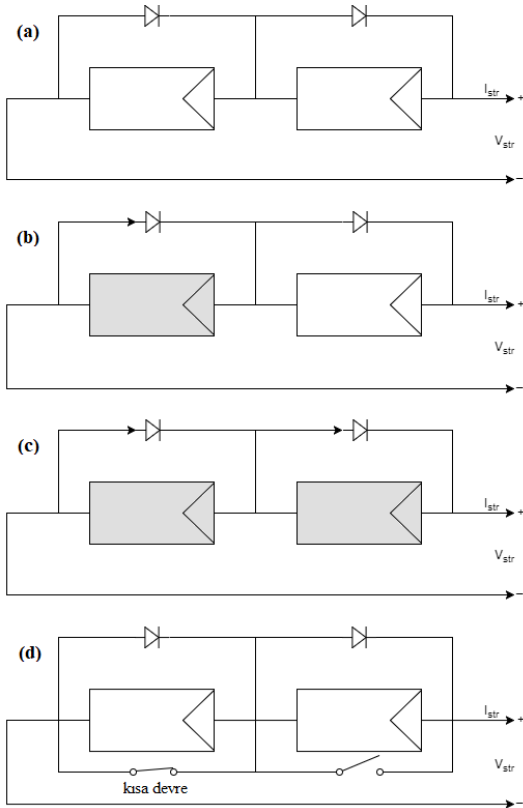
Şekil 1. Deneysel sistem (Experimental system)



Şekil 2. Bir paneldeki gölgelenme senaryosu  
(Shading scenario in one module)



Şekil 3. Bir paneldeki kısa devre senaryosu  
(Short-circuit scenario in one module)



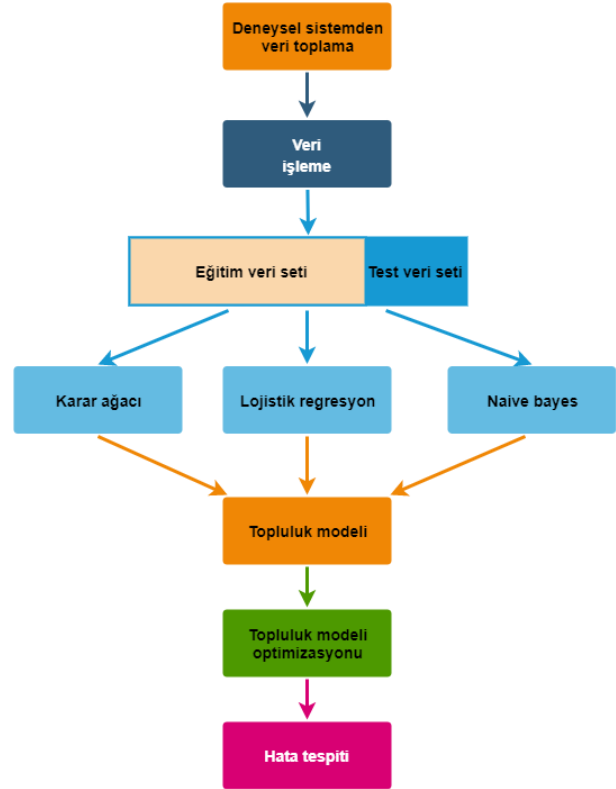
Şekil 4. Hata senaryoları gerçekleştirme şeması  
(Scheme of performing fault scenarios)

### 3. ÖNERİLEN HATA TESPİT YÖNTEMİ (THE PROPOSED FAULT DETECTION METHOD)

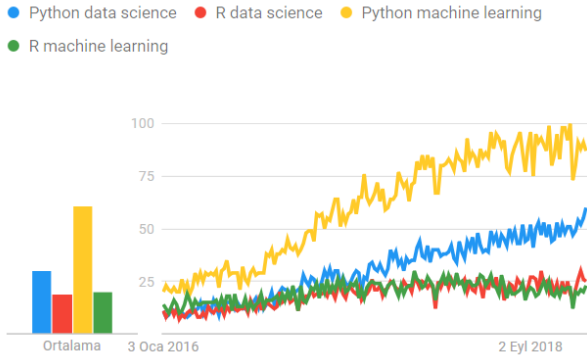
Bu çalışmada önerilen topluluk öğrenmesi temelli hata tespit sisteminin oluşturulma aşamaları Şekil 5'te akış şeması halinde gösterilmiştir.

Deneyisel sistemden toplanan ölçüm verileri ilk önce veri bütünlüğü açısından kontrol edilmelidir. Ölçüm sırasında bir veya daha fazla algılayıcıdan veri okunmadığı veya okunup kaydedilemediği durumlardaki ölçüm örnekleri eksik veri içereceğinden bu örnekler veri işleme adımında veri setinden temizlenirler. Çalışmanın sonraki bölümlerinde bahsedilen özellik seçimi ve hedef sınıfların oluşturulması işlemleri yine bu veri işleme adımında gerçekleştirilen işlemlerdir. Veri işleme adımından sonra gelen diğer adımlar, çalışmanın sonraki bölümlerinde detaylı olarak ele alınmıştır.

Yapılan çalışmada ölçüm verilerinin işlenmesi, eğitim ve test veri setlerinin oluşturulması, öğrenme algoritmaları ve topluluk modelinin oluşturulması, tüm modellerin eğitimi, test edilmesi ve topluluk modelinin optimizasyonu adımları Scikit-learn makine öğrenme kütüphanesi [27] ile Python kullanılarak gerçekleştirilmiştir. Şekil 6'da görüldüğü gibi veri biliminde ve makine öğrenmesi temelli yöntemlerde Python kullanımı son yıllarda giderek artmaktadır [28]. Python ve R eko-sistemi açık kaynak kodlu yapıları ve çok sayıda yardımcı kütüphaneleri ile en çok tercih edilen iki eko-sistem olarak karşımıza çıkmaktadır.



Şekil 5. Hata tespit sisteminin oluşturulması  
(Constructing fault detection system)



Şekil 6. Veri bilimi ve makine öğrenmesinde Python ve R kullanımını [28]

(Python and R usage in data science and machine learning [28])

### 3.1. Topluluk Öğrenmesi (Ensemble Learning)

Geleneksel olarak, makine öğrenme modellerine dayanan FV hata tespit yöntemleri, model geliştirme süreci boyunca bir öğrenme algoritması kullanılarak geliştirilirler. Tek öğrenme algoritmasına dayanan bu geleneksel yaklaşım, kullanılan modelin kendi sınırlamalarından olumsuz etkilenmektedir. Örneğin, bazı öğrenme algoritmalarının çıktısı, giriş verilerinde yapılan küçük değişikliklerin ardından önemli ölçüde değişebilir. Bu kararsızlık problemi, FV sistemler için hata tespiti gibi birçok değişkenin etken olduğu saha uygulamaları için yanlış tahminlerle sonuçlanabilir. Sınıflandırma, kümeleme ve regresyon problemlerinde, tek bir öğrenme algoritmasının sahip olduğu kararlılığı ve tahmin doğruluğunu arttırmak için, topluluk öğrenmesi kavramı ortaya atılmıştır [29]. Topluluk öğrenmesi temelli bir model, bireysel olarak eğitilmiş modelleri bir araya getirir ve tek bir nihai tahmin üretir. Topluluk öğrenmesi yöntemi, heterojen topluluk ve homojen topluluk olmak üzere iki kategoride sınıflandırılabilir. Heterojen topluluk modelinde, aynı eğitim verisi farklı öğrenme algoritmalarına veya aynı algoritmalara farklı parametre ayarları ile uygulanır. Homojen topluluk modelinde ise, orijinal eğitim verilerinden örneklenen farklı veri kümeleri, aynı parametre ayarları ile aynı öğrenme algoritmasına uygulanır [30]. Böylece, heterojen topluluk modeli farklı öğrenme algoritmalarının birbirlerini tamamlamasını sağlayarak, homojen topluluk modeli ise temel olarak seçilen bir öğrenme algoritmasını, eğitim verilerinin farklı dağılımlarıyla eğiterek tahmin doğruluklarını arttırlar. Literatürde sadece karar ağaçlarından oluşan bir topluluk öğrenmesi modeli olan rastgele orman modelinin kullanıldığı çalışma [31] homojen topluluk öğrenmesine dayalı hata tespitinin FV sistemlerde kullanılmasına örnek olarak gösterilebilir.

### 3.2. Topluluk Öğrenmesi Modelinin Oluşturulması (Constructing The Ensemble Learning Model)

Deneysel sistemden elde edilen ölçüm veri setinde 9 özellik bulunmaktadır. Bu özellikler FV diziden ölçülen voltaj ( $V_{str}$ ), akım ( $I_{str}$ ) ve hesaplanan FV dizi gücü ( $P_{str}$ ),

FV panellerin üzerine düşen güneş ışınımı ( $G$ ), her panelin arkasındaki dört adet DS18B20 sıcaklık algılayıcısından okunan FV panel arka yüzey sıcaklıkları ( $T_0, T_1, T_2, T_3$ ) ve bunların ortalamasıdır ( $T_m$ ). Panel arka yüzey sıcaklıklarının ortalama değeri  $T_m$ , ayrı ayrı ölçülen arka yüzey sıcaklıkları  $T_0, T_1, T_2$  ve  $T_3$  değerlerinin yerine temsilen kullanılabilir. Panel arka yüzey sıcaklığı ve hücre sıcaklığı belirgin bir şekilde farklı olabileceğinden, hücre sıcaklığı, denklem 1'de verilen basit ilişki ile tahmin edilebilir [32].

$$T_c = T_m + \frac{G}{G_{ref}} \Delta T \quad (1)$$

Bu denklemde,  $T_c$  hücre sıcaklığı ( $^{\circ}\text{C}$ ),  $T_m$  panel arka yüzey ortalama sıcaklığı ( $^{\circ}\text{C}$ ),  $G$  panel yüzeyindeki ışınım miktarı ( $\text{W}/\text{m}^2$ ),  $G_{ref}$  referans ışınım miktarı ( $1000\text{W}/\text{m}^2$ ),  $\Delta T$  ise malzemeye göre değişen sıcaklık farkı katsayısıdır (cam-hücre-cam şeklinde lamine edilmiş panel için  $2^{\circ}\text{C}$ , cam-hücre-polimer tabaka için ise  $3^{\circ}\text{C}$ ). Hücre sıcaklığının elde edilmesinin ardından kullanılacak tüm özellikler sırasıyla  $V_{str}, I_{str}, P_{str}, G, T_c$  olarak belirlenmiştir.

Hata tespiti bir sınıflandırma problemi olduğundan ve sınıflandırma, denetimli öğrenmenin alt kategorisi olduğundan, eğitim veri setindeki her örnek, izlenen FV dizinin normal ve hatalı koşullarda çalıştığını ifade eden "Case" hedef sınıfı altında sınıflandırılmıştır. Sınıflandırılan tüm tipler Tablo 1'de listelenmiştir. Seçilen özelliklerden oluşan veri seti daha sonra iki alt gruba ayrılır. Örneklerin %70'i eğitim için, %30'u ise eğitilmiş öğrenme algoritmalarının daha önce hiç karşılaşmadıkları örneklerle karşı genelleme yapma yeteneğini test etmek amacıyla kullanılmıştır.

Tablo 1. Hata tiplerinin sınıflandırılması (Classification of fault types)

Case #	Hata koşulu
Case 0	Normal çalışma.
Case 1	Dizideki bir panelin kısmi gölgelenmesi.
Case 2	Dizideki iki panelin kısmi gölgelenmesi.
Case 3	Dizideki bir panelde kısa devre olması.

Nihai topluluk modelini oluşturmak için, Scikit-learn kütüphanesindeki "VotingClassifier" sınıflandırıcısı kullanılmıştır. Bu sınıflandırıcının amacı, kavramsal olarak farklı öğrenme algoritmalarını birleştirmek ve hedef sınıfı tahmin etmek için oy çokluğu (hard-voting) veya tahmin olasılıklarının ortalaması (soft-voting) kullanmaktır. Örneğin, oy çokluğu tipindeki bir topluluk modeli üç farklı öğrenme algoritması içeriyorsa ve algoritmaların üçte ikisi test örneğini "Sınıf-1" olarak, üçte biri ise "Sınıf-2" olarak sınıflandırırsa, test örneği oy çokluğuyla "Sınıf-1" olarak sınıflandırılır. Bu çalışmada kullanılan nihai topluluk modelinin oluşturulmasında karar ağacı, lojistik regresyon

ve naive bayes öğrenme algoritmaları kullanılmıştır ve oluşturulan model heterojen bir topluluk modelidir.

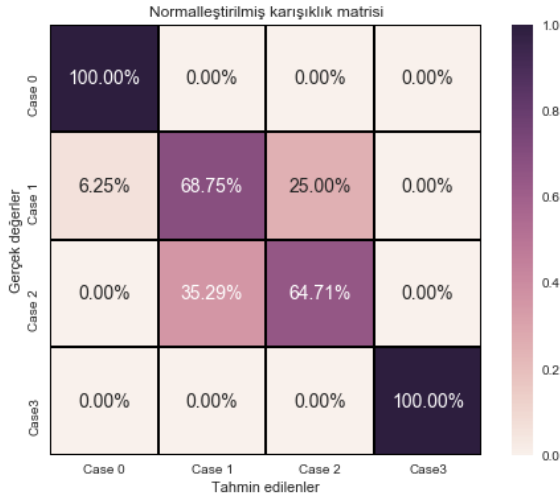
#### 4. SONUÇLAR ve TARTIŞMA (RESULTS AND DISCUSSION)

Bu bölümde topluluk modeli ve onu oluşturan algoritmaların tahmin doğrulukları karşılaştırılmıştır. Tüm modellerin eğitim ve test veri setleri için 5 katlı çapraz doğrulama yöntemiyle elde edilmiş tahmin doğrulukları ortalamaları Tablo 2.'de listelenmiştir.

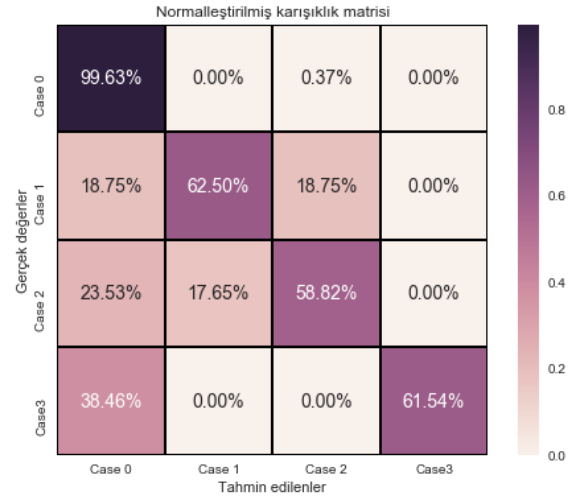
Tablo 2. Tüm modeller için tahmin doğrulukları  
(Prediction accuracy for all models)

Model	Eğitim veri seti	Test veri seti
Karar ağacı	%96.71	%96.17
Lojistik regresyon	%95.10	%96.17
Naive bayes	%94.34	%94.93
Topluluk modeli	%97.02	%96.52

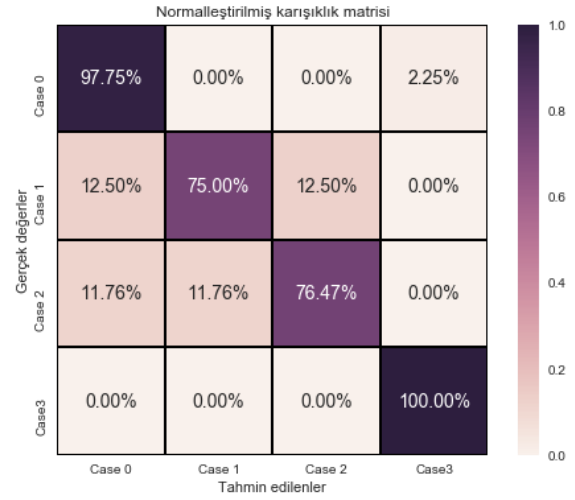
Şekil 7, Şekil 8, Şekil 9 ve Şekil 10'da sırasıyla karar ağacı, lojistik regresyon, naive bayes ve henüz optimize edilmemiş topluluk modeli için test veri setinde elde edilen sınıflandırma performanslarının görsel ifadeleri olan karışıklık matrisleri görülmektedir.



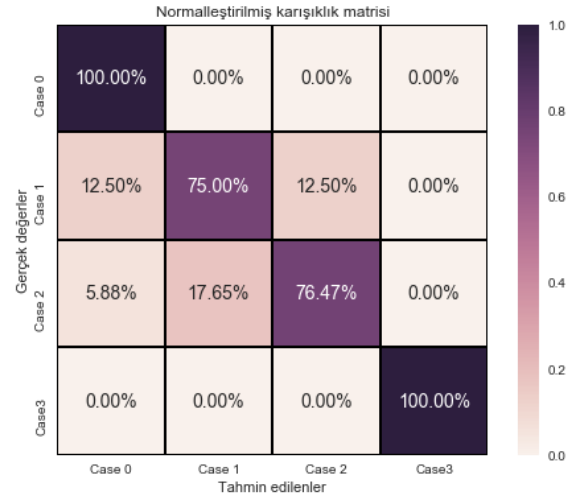
Şekil 7. Karar ağacı karışıklık matrisi  
(Decision tree confusion matrix)



Şekil 8. Lojistik regresyon karışıklık matrisi  
(Logistic regression confusion matrix)



Şekil 9. Naive bayes karışıklık matrisi  
(Naive bayes confusion matrix)



Şekil 10. Topluluk modeli karışıklık matrisi  
(Ensemble model confusion matrix)

Matrisler hem doğru hem de yanlış sınıflandırmaların oranını gösterir. Her bir hata tipi için veri setlerindeki etiketlenmiş veri sayısı farklı olduğundan karışıklık matrisleri normalleştirilmiştir. İlk topluluk modeli oluşturulurken varsayılan ayarlar kullanılmıştır. Bu ayarlar; topluluk modelinin bir örneği sınıflandırırken, kendisini oluşturan algoritmaların tahminlerinden oy çokluğuna göre seçim yapması ve her algoritmanın eşit ağırlık katsayısına sahip olmasıdır.

Tüm algoritmalar test veri setinde Case 0 için yüksek sınıflandırma performansı gösterirken, Case 1 ve Case 2 için düşük sınıflandırma performansı ortaya koymaktadırlar. Karar ağacı ve naive bayes algoritmaları Case 3 için sınıflandırma hatası yapmazken, lojistik regresyon Case 3 için düşük sınıflandırma performansı göstermiştir. Topluluk modeli karışıklık matrisinde görüldüğü gibi, test veri setinde Case 0 ve Case 3 için yanlış sınıflandırılma olmadığı anlaşılmaktadır. Aynı veri setinde Case 1 ile ilişkili örneklerin % 12,50'si Case 0 olarak, %12,50'si ise Case 2 olarak sınıflandırılırken, Case 2 ile ilişkili örneklerin % 5,88'i Case 0 olarak, %17,65'i ise Case 1 olarak sınıflandırılmıştır.

Topluluk modeli, parametre optimizasyonu ile daha yüksek sınıflandırma performansı elde edebilir. Bunun için Scikit-learn şebeke-arama fonksiyonu kullanılmıştır. Şebeke-arama, topluluk öğrenmesi modeli için en iyi parametreler kombinasyonunu bulmak amacıyla belirtilen parametre değerleri üzerinde kapsamlı bir arama yapar. Arama işlemi, sonucun daha güvenilir ve tarafsız olması için 5 katlı çapraz doğrulama ile yapılmıştır. Oluşturulan topluluk modeli için oylama tipi ve ağırlıklar parametreleri üzerinde şebeke-araması yapılmıştır.

Şekil 11'de görüldüğü gibi oylama tipi için oy çokluğu (hard) ve tahmin olasılıklarının ortalaması (soft) seçenekleri üzerinde arama yapılırken, ağırlıklar için ise [0,0,1], [0,1,0], [1,0,0], [0,1,1], [1,1,0], [1,0,1], [1,1,1], [1,1,2], [1,2,1], [2,1,1], [1,2,2], [2,2,1], [2,1,2], [2,2,2], [0,0,2], [0,2,0], [2,0,0], [0,2,2], [2,2,0], [2,0,2] seçenekleri üzerinde arama yapılmıştır.

```
params = {'weights':[[0,0,1],[0,1,0],[1,0,0],[0,1,1],[1,1,0],
                    [1,0,1],[1,1,1],[1,1,2],[1,2,1],[2,1,1],
                    [1,2,2],[2,2,1],[2,1,2],[2,2,2],[0,0,2],
                    [0,2,0],[2,0,0],[0,2,2],[2,2,0],[2,0,2]],
          'voting':['soft', 'hard']}
```

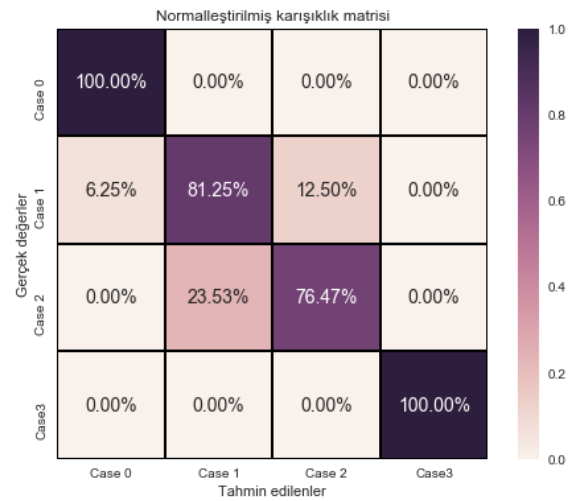
Şekil 11. Topluluk modeli için şebeke-arama ayarları  
(Grid-search parameters for ensemble model)

Burada her 3'lü gruptaki rakamlar sırasıyla karar ağacı, lojistik regresyon ve naive bayes algoritmalarının ağırlık katsayılarıdır. Örneğin oylama tipinin "soft" ve ağırlık katsayılarının [1,1,1] şeklinde ayarlandığı bir topluluk modeli ile sınıflandırmada her bir algoritmanın bir örnek için yaptığı tahmin olasılığı 1 ile çarpıldıktan sonra tüm olasılıkların ortalamaları hesaplanır ve topluluk modelinin tahmin sonucu olarak belirlenir. Eğer katsayılar [2,1,1] şeklinde ayarlanmışsa, karar ağacının tahmin olasılığı 2 ile çarpılırken, lojistik regresyon ve naive bayes

algoritmalarının tahmin olasılıkları 1 ile çarpılır ve elde edilen olasılıkların ortalaması alınır. Şebeke-araması, topluluk modeli için en iyi tahmin doğruluğunun ağırlık katsayılarının [1,2,1] ve oylama tipinin ise "soft" şeklinde olduğu ayarlarla elde edildiği sonucunu vermiştir. En iyi parametreler, Scikit-learn şebeke-arama fonksiyonunun "best\_params\_" yöntemi ile raporlanabilir. Belirtilen şebeke-arama fonksiyonunun çalışması, 8 çekirdekli Intel Core i7-4790 @ 3.60GHz işlemci ve 16 GB RAM donanımlı masaüstü bilgisayarında, tüm işlemci gücünün bu iş için koşurulması durumunda yaklaşık 40 saniye sürmektedir.

Tablo 3. Optimize topluluk modeli için tahmin doğrulukları  
(Prediction accuracy for optimized ensemble model)

Model	Eğitim veri seti	Test veri seti
Topluluk modeli	%97.64	%97.49



Şekil 12. Optimize topluluk modeli karışıklık matrisi  
(Optimized ensemble model confusion matrix)

Elde edilen parametrelerle yapılan optimizasyon sonrası topluluk modeli karışıklık matrisinde görüldüğü gibi, test veri setinde Case 1 için doğru sınıflandırma oranı %75'ten, %81,25'e çıkmıştır. Diğer hata tipleri olan Case 0, Case 2 ve Case 3 için sınıflandırma performanslarında herhangi bir değişiklik olmamıştır. Optimizasyon öncesinde topluluk modeli Case 2 ile ilişkili örneklerin % 5,88'i Case 0 olarak, %17,65'i ise Case 1 olarak sınıflandırırken, optimizasyonun ardından yanlış sınıflandırmalar sadece Case 1 yanlış tahmini altında toplanmıştır.

Karışıklık matrislerinde görüldüğü gibi, topluluk modeli kullanımı ile, optimizasyon olmadan bile, her bir hata tipinde en iyi sınıflandırma yapan algoritmanın performansı yakalanabilmiş ve tüm algoritmaların tahmin doğrulukları ortalamalarından daha iyi sonuç elde edilmiştir. Optimizasyon sonrası ise Şekil 12'de görüldüğü gibi topluluk modelinin düşük başarı elde ettiği Case 1 sınıflandırmasında performans artışı gerçekleşmiş ve

Tablo 3’te görüldüğü gibi test veri setindeki örneklerin ait olduğu hata tipi sınıflarının tahmin edilmesinde, ortalama tahmin doğruluğu %96.52’den %97.49’a yükselmiştir.

## 5. DEĞERLENDİRME (CONCLUSION)

Bu çalışmada FV sistemlerde hata tespiti için ilk kez karar ağacı, lojistik regresyon ve naive bayes gibi farklı öğrenme algoritmaları birleştirilerek heterojen bir topluluk öğrenmesi modeli oluşturulmuş ve bu model hata sınıflandırılmasında kullanılmıştır. Bu amaçla, deneysel bir güç sistemi kurulmuş ve bu sistem üzerinde farklı güneş ışınımı seviyelerinde, kısmi gölgelenme ve kısa devre hata senaryoları gerçekleştirilmiş ve toplanan ölçüm verileri kaydedilmiştir. Kaydedilen ölçüm verileri %70/%30 oranında ayrılarak, %70’lik kısmı sırasıyla karar ağacı, lojistik regresyon ve naive bayes algoritmalarının ve bu algoritmalarla oluşturulan bir topluluk modelinin eğitiminde kullanılmıştır. Ardından tüm modeller, genelleme yapma yeteneklerini ölçmek amacıyla verilerin daha önce karşılaşmadıkları %30’luk kısmıyla test edilmişlerdir. Oluşturulan topluluk modelinin sınıflandırma performansı, kendisini oluşturan algoritmalar ile karşılaştırılmış ve topluluk modelinin her iki veri setinde de daha iyi tahmin doğruluğuna sahip olduğu görülmüştür. Bununla birlikte topluluk modeli, şebeke-arama algoritması ile optimize edildiğinde sınıflandırma performansında artış gözlenmiştir. Sonuç olarak, optimize edilmiş bir topluluk öğrenmesi modeli ile önerilen hata tespit yöntemi, yüksek sınıflandırma performansı ve güçlü genelleme yapma yeteneği ile FV sistemler için hata tespitinde yüksek başarıyla uygulanabilecek bir yöntem olduğunu göstermektedir. Gelecek çalışmalar kapsamında topluluk modelinin farklı öğrenme algoritmaları kullanılarak denenmesi, veri setinin var olan özelliklerini kullanarak yeni özellik çıkarma ile zenginleştirilmesi ve bu yeni özellikler ile test edilmesi düşünülmektedir. Ek olarak, geliştirilen hata tespit modelinin bulut bilişim katmanında çalışır hale getirilerek FV sistemin çalışması esnasında elde edilecek verilerin buluta gönderilip değerlendirilmesi ve varsa hata tespitinin yapılması planlanmaktadır.

## KAYNAKLAR (REFERENCES)

- [1] A. Woyte, M. Richter, D. Moser, S. Mau, N. H. Reich, U. Jahn, “Monitoring of Photovoltaic Systems: Good Practices and Systematic Analyses”, **28th European PV Solar Energy Conference and Exhibition**, Paris, France, 2013.
- [2] P. Guerriero, V. D’Alessandro, L. Petrazzuoli, G. Vallone, and S. Daliento, “Effective real-time performance monitoring and diagnostics of individual panels in PV plants”, **4th International Conference on Clean Electrical Power: Renewable Energy Resources Impact, ICCEP 2013**, Hamburg, Germany, 14–19, 2013.
- [3] B. Ando, S. Baglio, A. Pistorio, G. M. Tina, and C. Ventura, “Sentinella: Smart Monitoring of Photovoltaic Systems at Panel Level”, *IEEE Transactions on Instrumentation and Measurement*, 64(8), 2188–2199, 2015.
- [4] W. Chine, a. Mellit, V. Lugh, a. Malek, G. Sulligoi, and a. Massi Pavan, “A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks”, *Renewable Energy*, 90, 501–512, 2016.
- [5] H. Mekki, A. Mellit, and H. Salhi, “Artificial neural network-based modelling and fault detection of partial shaded photovoltaic modules”, *Simulation Modelling Practice and Theory*, 67, 1–13, 2016.
- [6] Y. Zhao, R. Ball, J. Mosesian, J.-F. de Palma, and B. Lehman, “Graph-Based Semi-supervised Learning for Fault Detection and Classification in Solar Photovoltaic Arrays”, *IEEE Transactions on Power Electronics*, 30(5), 2848–2858, 2015.
- [7] L. Bonsignore, M. Davarifar, A. Rabhi, G. M. Tina, and A. Elhajjaji, “Neuro-Fuzzy Fault Detection Method for Photovoltaic Systems”, *Energy Procedia*, 62, 431–441, 2014.
- [8] D. Riley and J. Johnson, “Photovoltaic prognostics and health management using learning algorithms”, **Conference Record of the IEEE Photovoltaic Specialists Conference**, Austin, TX, USA, 1535–1539, 2012.
- [9] Y. Zhao, L. Yang, B. Lehman, J.-F. de Palma, J. Mosesian, and R. Lyons, “Decision tree-based fault detection and classification in solar photovoltaic arrays”, **2012 Twenty-Seventh Annual IEEE Applied Power Electronics Conference and Exposition (APEC)**, Orlando, FL, USA, 93–99, 2012.
- [10] S. Spataru, D. Sera, T. Kerekes, and R. Teodorescu, “Diagnostic method for photovoltaic systems based on light I-V measurements”, *Solar Energy*, 119, 29–44, 2015.
- [11] P. Ducange, M. Fazzolari, B. Lazzarini, and F. Marcelloni, “An intelligent system for detecting faults in photovoltaic fields”, **International Conference on Intelligent Systems Design and Applications, ISDA**, Cordoba, Spain, 1341–1346, 2011.
- [12] L. L. Jiang and D. L. Maskell, “Automatic fault detection and diagnosis for photovoltaic systems using combined artificial neural network and analytical based methods”, **2015 International Joint Conference on Neural Networks (IJCNN)**, Killarney, Ireland, 1–8, 2015.
- [13] C. B. Jones, J. S. Stein, S. Gonzalez, and B. H. King, “Photovoltaic system fault detection and diagnostics using Laterally Primed Adaptive Resonance Theory neural network”, **2015 IEEE 42nd Photovoltaic Specialist Conference (PVSC)**, New Orleans, Louisiana, USA, 1–6, 2015.
- [14] M. N. Akram and S. Lotfifard, “Modeling and Health Monitoring of DC Side of Photovoltaic Array”, *IEEE Transactions on Sustainable Energy*, 6(4), 1245–1253, 2015.
- [15] K. Chao, P. Chen, M. Wang, and C. Chen, “An Intelligent Fault Detection Method of a Photovoltaic Module Array Using Wireless Sensor Networks”, *International Journal of Distributed Sensor Networks*, 10(5), 12, 2014.
- [16] E. Karatepe and T. Hiyama, “Controlling of artificial neural network for fault diagnosis of photovoltaic array”, **2011 16th International Conference on Intelligent System Applications to Power Systems**, Crete, Greece, 1–6, 2011.
- [17] K. H. Chao, C. T. Chen, M. H. Wang, C. F. Wu, “A novel fault diagnosis method based-on modified neural networks for photovoltaic systems”, **Lecture Notes in Computer Science, vol 6146**. Springer, Berlin, Heidelberg, 531–539, 2010.



- [18] Y. Yagi *et al.*, “Diagnostic technology and an expert system for photovoltaic systems using the learning method”, *Solar Energy Materials and Solar Cells*, 75(3–4), 655–663, 2003.
- [19] Y. Zhao, F. Balboni, T. Arnaud, J. Mosesian, R. Ball, and B. Lehman, “Fault experiments in a commercial-scale PV laboratory and fault detection using local outlier factor”, **2014 IEEE 40th Photovoltaic Specialist Conference, PVSC 2014**, Denver, CO, USA, 3398–3403, 2014.
- [20] C.-T. Hsieh, H.-T. Yau, and J. Shiu, “Chaos Synchronization Based Novel Real-Time Intelligent Fault Diagnosis for Photovoltaic Systems”, *International Journal of Photoenergy*, 2014, 1–9, 2014.
- [21] K. H. Chao, S. H. Ho, and M. H. Wang, “Modeling and fault diagnosis of a photovoltaic system”, *Electric Power Systems Research*, 78(1), 97–105, 2008.
- [22] Z. Y. Wang, C. Lu, and B. Zhou, “Fault diagnosis for rotary machinery with selective ensemble neural networks”, *Mechanical Systems and Signal Processing*, 113, 112–130, 2018.
- [23] M. W. Ahmad, M. Mourshed, and Y. Rezgui, “Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression”, *Energy*, 164, 465–474, 2018.
- [24] M. Pierro *et al.*, “Multi-Model Ensemble for day ahead prediction of photovoltaic power generation”, *Solar Energy*, 134, 132–146, 2016.
- [25] M. Rana, I. Koprinska, and V. G. Agelidis, “Forecasting solar power generated by grid connected PV systems using ensembles of neural networks”, **Proceedings of the International Joint Conference on Neural Networks**, Killarney, Ireland, 1–8, 2015.
- [26] M. Peker, O. Özkaraca, and B. Kesimal, “Enerji Tasarruflu Bina Tasarımı İçin Isıtma ve Soğutma Yüklerini Regresyon Tabanlı Makine Öğrenmesi Algoritmaları İle Modelleme”, *Bilişim Teknolojileri Dergisi*, 443–449, 2017.
- [27] Scikit-learn: Machine Learning in Python, <http://scikit-learn.org>, 11-07-2017.
- [28] Python data science, R data science, Python machine learning, R machine learning - Google Trends, [https://trends.google.com/trends/explore?date=2016-01-01 2019-02-01&q=Python data science,R data science,Python machine learning,R machine learning](https://trends.google.com/trends/explore?date=2016-01-01%2019-02-01&q=Python%20data%20science,R%20data%20science,Python%20machine%20learning,R%20machine%20learning), 01-02-2019.
- [29] T. G. Dietterich, “Ensemble Methods in Machine Learning”, in *Oncogene*, 12(2), 2000, 1–15.
- [30] Z. Wang, Y. Wang, and R. S. Srinivasan, “A novel ensemble learning approach to support building energy use prediction”, *Energy and Buildings*, 159, 109–122, 2018.
- [31] Z. Chen *et al.*, “Random forest based intelligent fault diagnosis for PV arrays using array voltage and string currents”, *Energy Conversion and Management*, 178(August), 250–264, 2018.
- [32] D. L. King, W. E. Boyson, and J. A. Kratochvil, **Photovoltaic array performance model**, Sandia Report No. 2004-3535, 1–19, 2004.