



T.C.  
İSTANBUL ÜNİVERSİTESİ-CERRAHPAŞA  
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ



**Yüksek Lisans Tezi**

**ÇEVRE KİRLİLİĞİNİN İZLENMESİNDE MAKİNE ÖĞRENMESİ  
ALGORİTMALARI  
KULLANIMI**

**Selcen Zehra BEYZADE**

**DANIŞMAN  
Prof Dr. Sedat BALLIKAYA**

**II. DANIŞMAN  
Doç. Dr. Hasan TİRYAKİ**

**Mühendislik Bilimleri Anabilim Dalı**

**Mühendislik Bilimleri**

**İSTANBUL, 2024**

## TEZ KABUL VE ONAYI

Selcen Zehra BEYZADE tarafından, Prof. Dr. Sedat BALLIKAYA danışmanlığında hazırlanan " ÇEVRE KİRLİLİĞİNİN İZLENMESİNDE MAKİNE ÖĞRENMESİ ALGORİTMALARININ KULLANIMI" başlıklı bu çalışma, jürimiz tarafından 28/10/2024 tarihinde yapılan sınav sonucunda oy birliği ile başarılı bulunarak Yüksek Lisans Tezi olarak kabul edilmiştir.

### Tez Jürisi

	İmza	Sonuç
<b>DANIŞMAN</b>	Prof. Dr. Sedat BALLIKAYA İstanbul Üniversitesi-Cerrahpaşa Mühendislik Bilimleri Anabilim Dalı	<input type="checkbox"/> Kabul <input type="checkbox"/> Ret
<b>ÜYE</b>	Dr. Aminu YUSUF İstanbul Üniversitesi-Cerrahpaşa Mühendislik Bilimleri Anabilim Dalı	<input type="checkbox"/> Kabul <input type="checkbox"/> Ret
<b>ÜYE</b>	Prof. Dr. Ayberk YILMAZ İstanbul Üniversitesi Fizik Bölümü Yüksek Enerji ve Plazma Fiziği Anabilim Dalı	<input type="checkbox"/> Kabul <input type="checkbox"/> Ret
<b>ÜYE</b>	Doç. Dr. Cihat BOYRAZ Marmara Üniversitesi MakineMühendisliği Termodinamik Anabilim Dalı	<input type="checkbox"/> Kabul <input type="checkbox"/> Ret

---

<b>ÜYE</b>	Dr. Uğur ODABAŞI İstanbul Üniversitesi Mühendislik Bilimleri Anabilim Dalı	<input type="checkbox"/> Kabul <input type="checkbox"/> Ret
------------	---	--

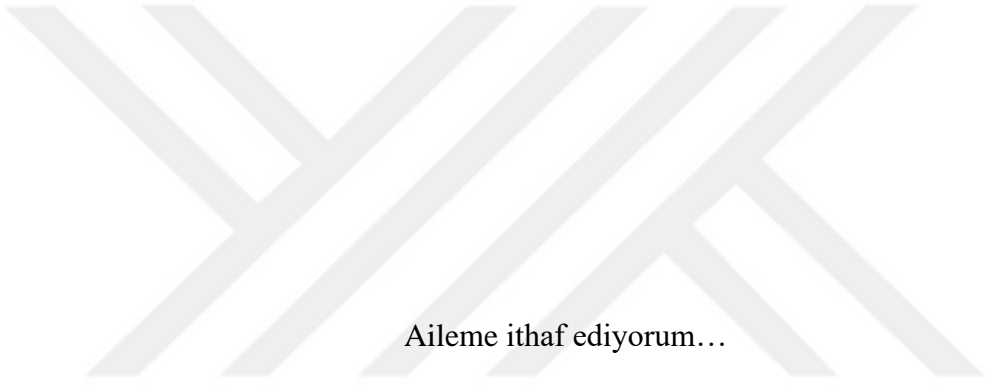
---



## BEYAN

Bu tez çalışmasının kendi çalışmam olduğunu, tezin planlanmasından yazımına kadar bütün aşamalarda etik dışı davranışımın olmadığını, bu tezdeki bütün bilgileri akademik ve bilimsel etik kuralları içinde elde ettiğimi, bu tez çalışmasıyla elde edilmeyen bütün bilgi ve yorumlara kaynak gösterdiğimi ve bu kaynakları da kaynaklar listesine aldığımı, yine bu tezin çalışılması ve yazımı sırasında patent ve telif haklarını ihlal edici bir davranışımın olmadığını ve her türlü hukuki sorumluluğu aldığımı kabul ederim.

Selcen Zehra BEYZADE



Aileme ithaf ediyorum...

## **BÜTÇE DESTEKLERİ**

### **ÇEVRE KİRLİLİĞİNİN İZLENMESİNDE MAKİNE ÖĞRENMESİ ALGORİTMALARININ KULLANIMI**

Bu tez çalışması için herhangi bir kurumdan bütçe desteği alınmamıştır.



## TEŐEKKÜR

Bu tez alıőmasında maddi ve manevi desteęini esirgemeyen deęerli danıőman hocalarım Prof. Dr Sedat BALLIKAYA ve Do. Dr. Hasan TİRYAKİ' ye, bilgi ve tecrübeleriyle tezim konusunda bana yardımcı oldukları için en içten dileklerle teőekkür ederim.

Tüm öęrenimim boyunca beni sabırla destekleyen deęerli ailemin tüm üyelerine, özellikle de oęluma ve eőime teőekkürü bor bilirim.

Ekim 2024

Selcen Zehra BEYZADE

# İÇİNDEKİLER

Sayfa no

TEZ KABUL VE ONAYI.....	ii
BEYAN .....	iii
BÜTÇE DESTEKLERİ .....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER.....	vii
ŞEKİL LİSTESİ .....	x
TABLO LİSTESİ.....	xii
SİMGE VE KISALTMA LİSTESİ .....	xiii
ÖZET .....	xiv
ABSTRACT .....	xvi
<b>1. GİRİŞ.....</b>	<b>1</b>
<b>2. KAVRAMSAL ÇERÇEVE .....</b>	<b>4</b>
2.1. Çevre Kirliliğinin İncelenmesinde Jeofizik SIP Metodu .....	4
2.2. SIP yönteminin Çevre Kirliliği Probleminde Uygulama Alanları.....	4
2.2.1. Toprak Kirliliği.....	4
2.2.2. Yeraltı Suyu Kirliliği .....	4
2.2.3. Yeraltı Gazlarının Tespiti ve Biyolojik Aktivitenin İzlenmesi .....	5
2.2.4. Atık Sahalarının İzlenmesi .....	5
2.3. Jeofizik Bilim Dalının ve SIP Yönteminin Zorlukları ve Makine Öğrenimi İlişkisi...5	
2.3.1. Ölçümde Hassasiyet ve Veri Toplama .....	6
2.3.2. Veri Yorumlamada Karşılaşılan Zorluklar .....	6
2.3.3. Çevrenin Ölçümlere Etkisi .....	6
2.3.4. Maliyet Yüksekliği ve Zaman Alıcılığı .....	7
2.3.5. Veriyi Entegre Etme ve Multidisipliner Uzman Sorunu .....	7
2.4. Makine Öğreniminin Jeofiziksel Kullanımına Dair Bazı Örnekler .....	7
<b>3. YÖNTEM.....</b>	<b>9</b>
3.1. SIP Yöntemi.....	9
3.2. Makine Öğrenimi Yöntemleri.....	15

3.2.1.	Random Committee- Rastgele Komite.....	15
3.2.2.	Random Tree- Rastgele Ağaç.....	16
3.2.3.	KStar.....	18
3.2.4.	Multi-Layer Perceptron-MLP.....	19
3.2.5.	Random Forest- Rastgele Orman .....	20
3.2.6.	Decision Table-Karar Tablosu .....	22
3.2.7.	M5P Rules-M5P Kuralları.....	23
3.2.8.	M5 Rules-M5 Kuralları .....	23
3.2.9.	Additive Regression-Eklenen Regresyon.....	24
3.2.10.	Bagging-Torbalama .....	25
3.2.11.	Linear Regression- Doğrusal Regresyon .....	27
3.2.12.	Random Sub Space- Rastgele Alt Uzay.....	27
3.2.13.	SMOreg.....	28
<b>4.</b>	<b>BULGULAR .....</b>	<b>29</b>
<b>5.</b>	<b>TARTIŞMA .....</b>	<b>39</b>
<b>6.</b>	<b>SONUÇ VE ÖNERİLER.....</b>	<b>40</b>
	<b>KAYNAKLAR.....</b>	<b>42</b>
	<b>İNTİHAL RAPORU İLK SAYFASI .....</b>	<b>47</b>
	<b>ETİK KURUL İZİN YAZISI .....</b>	<b>48</b>
	<b>KURUM İZİNİ YAZILARI.....</b>	<b>49</b>
	<b>ÖZGEÇMİŞ .....</b>	<b>50</b>

## ŞEKİL LİSTESİ

### Sayfa No

Şekil 1.1: Jeofizik Yöntemler için Genelleştirilmiş Akış Şeması.....	2
Şekil 3.1 a: IP Ölçümü Sırasında Polarizasyon Etkisinin ve Voltaj Tepkisinin Gösterimi, b-SIP Ölçümü Sırasında Polarizasyon Etkisinin ve Genlik Tepkisinin Gösterimi.....	11
Şekil 3.2: %3 Manyetit-Kum Karışımının SIP Cevabı.....	13
Şekil 3.3: Rastgele Ağaç Akış Diyagramı.....	17
Şekil 3.4: Multilayer Perceptron Giriş – Çıkışları (Işıkhani, 2020). ....	19
Şekil 3.5: Rastgele Orman Akış Diyagramı (Yılmaz, 2014). ....	21
Şekil 3.6: Torbalama Akış Diyagramı (Ay, 2019).....	26
Şekil 4.1: Faz Kayması Şeması.....	29
Şekil 4.2: Rasgele Komite Eğitim ve Test Verisi Sonuçları. ....	30
Şekil 4.3: Rastgele Ağaç Eğitim ve Test Verisi Sonuçları.....	31
Şekil 4.4: KStar Eğitim ve Test Verisi Sonuçları. ....	31
Şekil 4.5: MLP Eğitim ve Test Verisi. ....	31
Şekil 4.6: Rastgele Orman Eğitim ve Test Verisi. ....	32
Şekil 4.7: Karar Tablosu Eğitim ve Test Verisi. ....	32
Şekil 4.8: M5 Rules Eğitim ve Test Verisi.....	32
Şekil 4.9: M5P Eğitim ve Test Verisi. ....	33
Şekil 4.10: Eklenen Regresyon Eğitim ve Test Verisi.....	33
Şekil 4.11: Torbalama Eğitim ve Test Verisi.....	33

<b>Şekil 4.12:</b> Lineer Regresyon Eğitim ve Test Verisi.....	34
<b>Şekil 4.13:</b> Rastgele Alt Uzay Eğitim ve Test Verisi.....	34
<b>Şekil 4.14:</b> SMOREg Eğitim ve Test Verisi. ....	34



## TABLO LİSTESİ

	<b>Sayfa No</b>
<b>Tablo 3.1:</b> Makine Öğreniminde Kullanılan SIP Verileri .....	14
<b>Tablo 4.1:</b> Eğitim Verileriyle Kullanılan Makine Öğrenimi Yöntemleri ve Hata Oranları.....	36
<b>Tablo 4.2:</b> Test Verileriyle Kullanılan Makine Öğrenimi Yöntemleri ve Hata Oranları. ....	38



## SİMGE VE KISALTMA LİSTESİ

### Simgeler Açıklama

$M_a$	: Görünür yüklenebilirlik
$V_p$	: Verilen Akımın Herhangi Bir Anında Ölçülen Voltaj
$V_S$	: Artık Gerilim
$\Delta t$	: Artık Gerilimin Oluştığı Zaman Penceresinin Uzunluk
$t_1, t_2$	: Akım Kesildikten Sonraki Manuel Olarak Belirlenen Süreler
$ \sigma $	: İletken Büyüklüğü
$\varphi$	: Faz Farkı
$\sigma$	: Karmaşık İletkenlik
$\sigma'$	: Karmaşık İletkenlik Gerçek bileşen
$i\sigma''$	: Karmaşık İletkenlik Sanal bileşen
$i$	: $\sqrt{-1}$

### Kısaltmalar Açıklama

<b>Hz</b>	: Hertz
<b>rad</b>	: radyan
<b>IP</b>	: İndüklenmiş Polarizasyon
<b>SIP</b>	: Spektral İndüklenmiş Polarizasyon
<b>V/mV</b>	: Görünür Yüklenebilirlik Birimi, Volt Başına Milivolt

## ÖZET

### Yüksek Lisans Tezi

## ÇEVRE KİRLİLİĞİNİN İZLENMESİNDE MAKİNE ÖĞRENMESİ ALGORİTMALARININ KULLANIMI

**Selcen Zehra BEYZADE**

**İstanbul Üniversitesi-Cerrahpaşa**

**Lisansüstü Eğitim Enstitüsü**

**Mühendislik Bilimleri Anabilim Dalı**

**Mühendislik Bilimleri Tezli Yüksek Lisans Programı**

**Danışman: Prof. Dr. Sedat BALLIKAYA**

**II. Danışman: Doç. Dr. Hasan TİRYAKİ**

Artan nüfus, sanayileşme, toksik atıkların bertarafı ve sağlıksız tarımsal faaliyetler toprak kirliliğini artırarak çevresel dengeyi bozmaktadır. Bunlardan biri de hidrokarbon kirliliğidir. Hem ekosistemlerin doğal yapılarında kalabilmesi hem de insan sağlığının olumsuz etkilenmemesi için oluşan hidrokarbon kirliliğinin sınırlarının belirlenmesi, izlenmesi ve doğru şekilde müdahale edilmesi gerekmektedir.

Kimyasal deneyler yapmak bir yöntem iken bir diğer yöntem de belirli zaman aralıklarında ortamın elektriksel özelliklerini incelemektir. Kimyasal deneylere göre daha geniş alanları tarayabilen ve alınan numunelere zarar vermeden bunu yapabilen jeofizik spektral indüklenmiş polarizasyon (SIP) yöntemi son yıllarda popüler hale gelmiştir. Zamanla hidrokarbonlarla kirlenen alanlarda manyetit gibi çeşitli metalik mineraller artmaktadır. Özellikle manyetit olmak üzere farklı tipteki metalik minerallerin elektriksel özelliklerini tespit edebilen SIP yöntemi bu tip çevre kirliliğinin sınırlarını belirlemek için kullanılmaktadır.

Bu yöntemi geliřtirmek için laboratuvar deneylerine ihtiyaç duyulmaktadır. Ancak bu deneyler genellikle zaman alıcı ve maliyetli olabilmektedir.

Verilerle oluşturulan örüntüleri ve ilişkileri öğrenen makine öğrenmesi algoritmaları, bazı işlemleri daha etkili bir şekilde gerçekleştirme, zaman kaybını önleme ve maliyetleri en aza indirme potansiyeline sahiptir.

Bu tezde, laboratuvar ortamında toplanan metalik bir mineral olan manyetitın elektriksel tepkileri kullanılmıştır. SIP sonuçları farklı makine öğrenmesi algoritmalarıyla test edilmiş ve algoritmaların performansları hakkında değerlendirmeler yapılmıştır. Tezde kullanılan makine öğrenmesi algoritmalarının performansları gerçek verilerle karşılaştırılmış ve bu süreçte %95'in üzerinde başarı gösteren 13 algoritma belirlenmiştir.

Rastgele Komite ve Rastgele Ağaç algoritmalarının en düşük hata oranlarına sahip olduğu ve hızlı çalışma süreleriyle en iyi performansı gösterdiği belirlenmiştir. Ayrıca diğer algoritmaların başarı oranının %95 ve üzeri olduğu da dikkate alınmalıdır.

Bu tez, hidrokarbonlardan kaynaklanan çevre kirliliği analizlerinde kullanılan SIP yöntemi (laboratuvar koşullarında) için bir ölçek oluşturmak amacıyla makine öğrenmesi algoritmalarının nasıl kullanılabileceği konusunda önemli bilgiler sunmaktadır.

Son olarak, laboratuvar çalışmalarının zaman ve maliyetini azaltma ve elde edilen verilerin insan kaynaklı hatalardan arındırılmasına yardımcı olma potansiyeline sahip olan makine öğrenmesi algoritmalarının, çevre kirliliği sınırlarının belirlenmesi ve izlenmesinde kullanılabilecek ölçeğin oluşturulmasında kullanılması önerilmektedir.

Ağustos 2024, 66 sayfa

**Anahtar kelimeler:** SIP, saçılımlı manyetit, manyetit, makine öğrenimi, metalik mineraller

## **ABSTRACT**

**Master Thesis**

### **USE OF MACHINE LEARNING ALGORITHMS IN MONITORING ENVIRONMENTAL POLLUTION**

**Selcen Zehra BEYZADE**

**İstanbul University-Cerrahpaşa**

**Institute of Graduate Studies**

**Engineering Sciences Department**

**Engineering Sciences Thesis Master's Program**

**Supervisor : Prof. Dr. Sedat BALLIKAYA**

**Co-Supervisor: Doç. Dr. Hasan TIRYAKI**

Increasing population, industrialization, disposal of toxic wastes, and unhealthy agricultural activities increase soil pollution and disrupt the environmental balance. Hydrocarbon pollution is one of them. For both ecosystems to remain in their natural structure, and for human health not to be adversely affected, the boundaries of the hydrocarbon pollution that occurs must be determined, monitored, and intervened correctly.

While conducting chemical experiments is one method, another method is to examine the electrical properties of the environment at certain time intervals. The geophysical spectral induction polarization (SIP) method, which can scan larger areas compared to chemical experiments and do this without damaging the samples taken, has become popular in recent years. Various metallic minerals - such as magnetite - increase in areas contaminated with

hydrocarbons over time. The SIP method, which can detect the electrical properties of different types of metallic minerals, especially magnetite, is used to determine the boundaries of this type of environmental pollution.

Laboratory experiments are needed to develop this method. However, these experiments can usually be time-consuming and costly.

Machine learning algorithms that learn patterns and relationships formed with data have the potential to perform some operations more effectively, prevent loss of time, and minimize costs.

In this thesis, electrical responses of magnetite, a metallic mineral, collected in a laboratory environment were used. SIP results were tested with different machine learning algorithms and evaluations were made about the performances of the algorithms. The performances of the machine learning algorithms used in the thesis were compared with real data, and 13 algorithms that showed over 95% success were determined in this process.

It was determined that Random Committee and Random Tree algorithms had the lowest error rates and exhibited the best performance with fast working times. In addition, it should be taken into account that the success rate of other algorithms was 95% and above.

This thesis provides important information on how machine learning algorithms can be used to create a scale for the SIP method (in laboratory conditions) used in environmental pollution analyses caused by hydrocarbons.

Finally, it is recommended that machine learning algorithms, which have the potential to reduce the time and costs of laboratory studies, and help the obtained data to be free from human-induced errors, be used to create the scale to can be used for determining and monitoring environmental pollution limits.

September 2024, 66 pages.

**Keywords:** SIP, disseminated magnetite, magnetite, machine learning, metallic minerals

## 1. GİRİŞ

Çevre kirliliği sanayi devrimiyle paralel olarak gelişti ve dünya genelinde sorunlara neden olmaya devam etmektedir. Giderek büyüyen bu problemi tanımlamak ve çözümler bulmak için bilim adamları kirlenmiş alanlar üzerinde titizlikle çalışmaktadır. Özellikle hidrokarbonla kirlenmiş alanların selameti için yapılan jeofizik çalışmalar hızla çoğalmaktadır (Kessouri ve diğ, 2020; Lund ve diğ, 2017; Rossbach ve diğ, 2016).

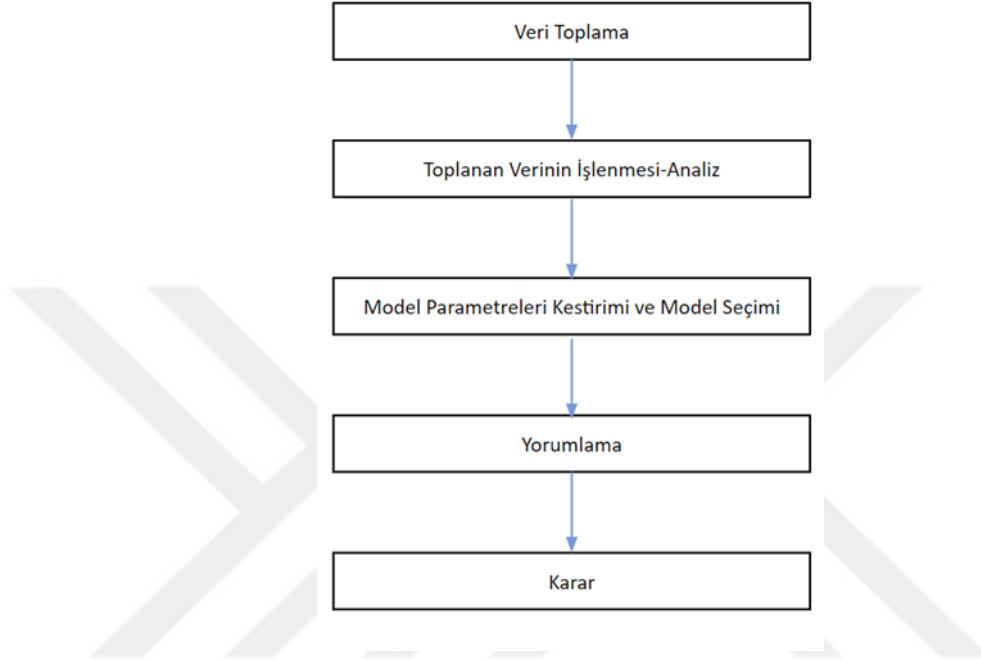
Bununla beraber bilim adamları 1980’li yıllarda doğadaki varlıkların akıllı davranışlarını taklit eden sistemler kurmayı amaçladılar (Charniak ve McDermot, 1985; Nabiyevev 2012: 25). İlerleyen çalışmalar sonucunda, insanın işlerini kolaylaştıran araçlardan da öte, tahmin eden, kümelendiren, sınıflandıran sistemler oluşturulabildi ve bu sistemlerin toplamına yapay zeka adı verildi. Bilgisayarların verilerden öğrenmesini, açıkça programlanmasalar bile tahmin etmesini ve karar vermesini sağlayan algoritmalar ve teknikler geliştirmeye odaklanan yapay zeka alanı ise makine öğrenmesi olarak adlandırıldı. Kısa zamanda makine öğrenmesi; sağlık, finans pazarlama, ulaşım, bilimsel aktiviteler ve birçok sektörlerde geniş bir uygulama yelpazesine sahip oldu.

Son yıllarda çevre kirliliği araştırmalarında da kullanılmaya başlayan jeofizik; jeolojik yapıları, yer katmanlarını, depremleri, yer altı su kaynaklarını, volkanik faaliyetleri gibi doğa olaylarını fizikteki ilke ve teknikleri kullanarak inceleyen bir bilim dalıdır. Jeofiziğin bilim dalını oluşturan iki önemli husus vardır. Bunlar; Fizik prensipleri ile ölçme ve gözlem.

Fizik Prensipleri: Jeofizik büyük ölçüde fiziğin temel prensiplerini yani mekanik, elektromanyetizma, termodinamik ve dalga yayılımı gibi ilkeleri kullanan bir bilim dalıdır. Çeşitli Dünya süreçlerini ve olaylarını anlamak için bu ilkeler teorik temeli sağlamaktadır.

Ölçme ve Gözlem: Jeofizikçiler, sismik dalgalar, ısı akışı, yerçekimi, manyetik alanlar ve elektriksel direnç dahil olmak üzere Dünyanın farklı özelliklerini gözlemlemek ve ölçmek için çeşitli alet ve tekniklerden yardım almaktadır. Elde edilen veriler jeolojik oluşumları, yeraltı yapılarını ve diğer Dünya özelliklerini yorumlamak için değerli veriler sağlamaktadır. Bu iki prensip kullanılarak jeofizik yöntemlerin kullanımının akış şeması Şekil 1.1’de gösterilmektedir. Veri toplama işlemi en önemli ve üzerinde hassas davranılması gereken bir

süreçtir. Verinin işlenmesine ve analiz yapmaya uygun olup olmadığına bakıldıktan sonra, model seçimi yapılır ve model parametreleri düzgün bir şekilde girilmektedir. Sonraki adım ise oluşan veri seti üzerinden yorumlamalar yapmaktır. En son olarak uzman personel veri hakkındaki kararı vermektedir.



**Şekil 1.1:** Jeofizik Yöntemler için Genelleştirilmiş Akış Şeması

Hangi alanlarda jeofizik yöntemlerin kullanıldığından aşağıda bahsedilmiştir

**Arama ve Kaynak Değerlendirmesi:** Doğal kaynakların araştırılmasında örneğin yeraltı suyu, petrol, gaz ve mineraller kaynakları araştırılmasında jeofizik yöntemler yaygın olarak kullanılmaktadır (Mensah ve diğ., 2024; Revuelta 2017) Jeofizikçiler, yer altı yapılarını ve özelliklerini görüntüledikten sonra potansiyel kaynak yatakları belirlenmektedir. Daha sonrasında kaynak yatakların ekonomik açıdan sürdürülebilirliğinin değerlendirilmesine jeofizik yöntemlerin kullanımı sıhhatlidir.

**Doğal Tehlike Değerlendirmesi:** Heyelanlar, depremler, tsunamiler ve volkanik patlamalar dahil olmak üzere doğal tehlikelerin değerlendirilmesinde ve azaltılmasında önemli bir rolü jeofizikçiler oynamaktadır. Bunu jeolojik verileri analiz ederek, yer deformasyonunu inceleyerek ve sismik aktiviteyi izleyerek yapmaktadırlar (Molua 2024; Florin ve diğ., 2014). Bunu sonucu olarak da Dünya sistemlerinin davranışının anlaşılmasına ve potansiyel tehlikelerin tahmin edilmesine katkıda bulunmaktadır.

Çevresel İzleme ve Yönetim: Son yıllarda yeraltı suyu kirliliği, arazi çökmesi ve toprak erozyonu gibi çevresel sorunları izlemek ve yönetmek için de jeofiziksel teknikler önerilmektedir. Bu teknikler kullanılarak jeolojik özellikleri haritalandırılması ve Dünya'nın yeraltındaki değişikliklerini izleyerek çevresel etkilerin değerlendirilmesine ve sürdürülebilir kaynak yönetimi için stratejiler geliştirilmesine olanak sağlanmaktadır (Adebayo 2023; Balwant ve diğ., 2022).

Disiplinler Arası Yaklaşım: Jeoloji, bilgisayar bilimi, matematik ve mühendislik dahil olmak üzere çeşitli bilimsel disiplinlerden gelen bilgi ve yöntemleri bir araya gelmesiyle jeofizik teknikleri ortaya çıkmaktadır. Bu disiplinler arası yaklaşım, yeni bir bakış açısı sağlamaktadır (Cao ve diğ., 2024; Zhao ve diğ., 2024). Böylelikle karmaşık jeolojik ve çevresel sorunlarının birçok perspektiften ele alınmalarına olanak tanınmaktadır.

## **2. KAVRAMSAL ÇERÇEVE**

### **2.1. Çevre Kirliliğinin İncelenmesinde Jeofizik SIP Metodu**

SIP yer kabuğu kirliliğinin ve yeraltı suyu kirliliğinin incelenmesi, izlenmesi ve değerlendirilmesinde kullanılan yöntemlerden biridir (Sanaiyan ve diğ., 2018; Atekwana ve diğ., 2010). Ayrıca kirlleticilerin mekânsal dağılımının ortaya çıkarılmasında da önemli rol oynamaktadır (Flores ve diğ., 2022). Vurgulanmak istenen şu ki; frekansa bağımlı ölçümleri sayesinde yüksek hassasiyete sahip SIP yöntemi, çevre koruma ve rehabilitasyon çalışmaları için kritik bir konumdadır.

### **2.2. SIP yönteminin Çevre Kirliliği Probleminde Uygulama Alanları**

#### **2.2.1. Toprak Kirliliği**

Tarımsal kirlilik ve endüstriyel atıkların meydana getirdiği kirliliğin tespitinde SIP kullanılabilir (Kirmizakis ve diğ.,2020; Kessouri ve diğ., 2019). Açıklamak gerekirse tarımsal alanlarda gübrelerin ve pestisitlerin kullanımı toprakta bir kirlilik oluşturmaktadır. Endüstriyel kullanım sonucu toprağa bulaşan kimyasal maddelerde aynı şekilde kirliliğe neden olmaktadır. Bu 2 tür kirliliğin ölçütü şudur ki; biriktikleri bölgelerde yük transferi ve iletkenlik sinyallerinde belirgin değişikliklerin meydana gelmesi. Bu sinyalleri kullanarak SIP, kirleticilerin toprakta birikimini ve hareketini izleyebilmektedir.

#### **2.2.2. Yeraltı Suyu Kirliliği**

Yeraltı suyunda bulunan ağır metaller ve organik kirleticiler SIP ile belirlenebilmektedir (Almpanis ve diğ., 2024; Moshe ve diğ., 2022). Bir yer altı suyu havzası ağır metal veya petrol türevleri içeriyorsa elektriksel özellikleri normalden farklı değerler üretmektedir. Bu farklılık SIP ölçümlerinde anomali olarak gözlemlenmektedir ve SIP yer altındaki bu kirliliklerin dağılımını belirleyebilmektedir.

### **2.2.3. Yeraltı Gazlarının Tespiti ve Biyolojik Aktivitenin İzlenmesi**

Metan gibi gazlar organik maddelerin bozulmasıyla oluşmaktadır. Genellikle çöplüklerde oluşan bu gazlar yeraltındaki polarizasyon sinyallerini değiştirmektedir. Bu tür gazların varlığı ve dağılımını anlamlandırmak için SIP kullanılmaktadır (Martinho ve diğ., 2023; Moser ve diğ., 2023). Bununla beraber kirlilik sonucu toprakta oluşan biyolojik aktivitenin de incelenmesi işlemlerinde SIP kullanılmaktadır (Abdel ve diğ., 2010; Ntallagiannis ve diğ., 2005).

### **2.2.4. Atık Sahalarının İzlenmesi**

Kazayla patlamış bir petrol kuyusunda saçılan petrol veya atıkların depolandığı bölgelerdeki herhangi bir sızıntı kirletici olarak tanımlanmaktadır. Toprak ve yer altı suyu için tehlike oluşturan bu kirleticilerin ne kadar yayıldığını SIP yöntemi ortaya çıkartabilmektedir (Brahmi ve diğ., 2023; Wemegah ve diğ., 2014).

## **2.3. Jeofizik Bilim Dahının ve SIP Yönteminin Zorlukları ve Makine Öğrenimi İlişkisi**

Jeofizik bilimi; yeryüzünün karmaşıklığı, ölçüm yapan araçların farklılığı ve tekniklerin çeşitliliği nedeniyle kapsamlı düşünmeyi ve birkaç yöntemin beraber kullanılmasını gerektirmektedir. Bu yöntemler aynı anda ya da sırasıyla uygulanmalıdır (Binley ve diğ., 2015; Parsekian ve diğ., 2016). Bununla beraber verisi toplanabilen büyük ölçekli doğal olaylar, herhangi bir fiziksel modelin veya matematiksel hesaplamanın ötesinde karmaşıklık sergileyebilmektedir. Yerbilimcilerin görevi olan yer yüzeyinin karmaşıklığını anlamlandırmak ve bireysel olayları incelemek için laboratuvarlar kullanılabilen veya yer yüzeyi periyodik aralıklarla izlenebilmektedir. Böylelikle basitleştirilmiş modellerden çıkarımlar yapmak için varsayımlar ve yaklaşımlar uygulanabilmektedir. Her iki durumda da incelenmek istenen büyük ölçekli sistemi tam olarak açıklamakta sorunlar çıkmaktadır. Bu durumun sonucu olarak finansal sıkıntılar ortaya çıkmakta ve jeofiziğin uygulanabileceği çeşitli araştırmalarda geniş çapta benimsenmesini engellemektedir.

Spesifik olarak SIP yöntemi de bazı sınırlamalara ve zorluklara sahiptir. Bunlar ölçümde hassasiyet ve veri toplama, veri yorumlamada karşılaşılan zorluklar, çevrenin ölçümlere etkisi,

maliyetin yüksek olması, zaman alıcılığı, veriyi entegre etme ve multidisipliner uzman sorunu olarak sıralanabilir.

### **2.3.1. Ölçümde Hassasiyet ve Veri Toplama**

Düşük frekans ölçümlerinin getirisi olarak çevrede bulunan elektriksel gürültülerden daha çabuk etkilenmesi mümkündür. Metalik objeler, kentsel alanlara yakınlık gibi etmenler gürültüyü arttırmaktadır. Bu gürültüler elde edilen verinin hassasiyetini negatif yönde etkilemektedir. Bununla birlikte alet kalibrasyonunun doğru yapılması önem arz etmektedir. Küçük değişiklikler bile verilerin yanlış yorumlanmasına neden olabilmektedir.

### **2.3.2. Veri Yorumlamada Karşılaşılan Zorluklar**

IP yönteminde kullanılan yüklenebilirlik formülasyonu (Seigel ve diğ., 1959) ve doğrusal model seçimleri (Chambers ve diğ., 2004; Li ve Oldenburg, 2000; Oldenburg ve Li, 1994) SIP yöntemi için aynı verimde çalışmamaktadır. Bu yüzden bilim adamları yeni modeller ve formülasyonlar geliştirmektedir (Blaschek ve diğ., 2008; Kemna, 2000; Shi ve diğ., 1998; Kemna ve Binley, 1996). Bu modellerin henüz gelişmekte olması ve her modelin yeryüzünün karmaşık yapısı için uygulanamaması zorluk oluşturmaktadır. Sadece model seçimi değil aynı zamanda yeraltındaki farklı malzemelerin yakın elektriksel özelliklere sahip olması verilerin yorumlanmasını zorlaştırmaktadır.

### **2.3.3. Çevrenin Ölçümlere Etkisi**

Toprak nemi, tuzluluk oranı ve zeminin sıcaklığı gibi çevresel etmenler polarizasyonu etkileyebilmektedir (Kessouri ve diğ., 2019; Aizebeokhai, 2014). Verilerin toplanması ve yorumlanması işlemlerinin doğru bir şekilde yapılması bu çevresel etmenlerin değişkenliği ile zorlaşabilmektedir.

Çevre kirliliği çalışmalarında SIP yaygın şekilde tercih edilse bile belirli kirleticilere karşı duyarlıdır. Daha açık ifadeyle düşük iletkenliğe sahip kirleticiler için kullanılamamaktadır. Ek olarak aşırı iletken zeminlerde ve yoğun olarak kayalardan oluşan ortamlarda doğru SIP verileri elde etmek zorlaşmaktadır. Bu durumlar SIP ölçüm alımını sınırlandırabilmektedir.

#### **2.3.4. Maliyet Yüksekliği ve Zaman Alıcılığı**

SIP yönteminde kullanılan ekipmanlar hassas ölçüm kabiliyetine sahiptir fakat pahalı ekipmanlardır. Buda büyük çevre arařtırmaları için maliyetlerde artışa neden olmaktadır (Kemna ve diğ., 2012). Aynı zamanda geniş alanlarda uzun süreli ve yüksek çözünürlüklü ölçümlerle veri toplamak zaman alıcı olmaktadır.

#### **2.3.5. Veriyi Entegre Etme ve Multidisipliner Uzman Sorunu**

Bilindiđi üzere jeofizik yöntemlerin herhangi birinin tek başına kullanılması tavsiye edilmez. Ölçüm alınan ortamda en az iki yöntemle ölçüm alınmakta ve bu yöntemler birbiriyle karşılaştırılmaktadır. Elde edilen SIP verilerinin diđer jeofizik yöntemlerle -nükleer manyetik rezonans, manyetik duyarlılık yöntemi gibi- entegrasyonu zorlu yorumlama süreçleri içermektedir (Keating ve diğ., 2015). Farklı yöntemlerden gelen çeşitli verilerin doğru bir şekilde toplanması ve yorumlanması multidisipliner bir yaklaşım gerektirmektedir. Bunun için birden fazla konuda uzman olan kişilere ihtiyacı olmaktadır. Bu da bazı durumlar için kısıtlayıcıdır.

### **2.4. Makine Öğreniminin Jeofiziksel Kullanımına Dair Bazı Örnekler**

Veri Analizi ve İşleme: Jeofizik veri bakımından zengin bir bilim dalıdır. Veriyi anlamlandırmak için analiz ve işleme kısımlarında makine öğrenimi teknikleri kullanılmaktadır (Paasche ve Tronicke, 2007; Zhang ve diğ, 2020). Ayrıca veri madenciliđi derin öğrenme ve desen tanıma gibi metotlar jeofizik verilerden bilgi çıkartabilmektedir (Reichstein ve diğ, 2019).

Jeofizik Görüntü İşleme: Jeofizik verisinin görüntülü hale getirilmesi ve yorumlanmasında da makine öğrenimi iş görmektedir. Buna örnek olarak sensörlerden alınan verinin işlenmesi ve jeolojik yapıların tanımlanması verilebilmektedir (Chen ve diğ., 2019).

Tahmin Modelleri: Yer kabuğunda meydana gelen olayları tahmin etmek için makine öğrenimi modelleri kullanılabilir. Yer altı su kaynaklarının miktar tahmini veya deprem tahmini buna örnek olarak verilebilmektedir. Elde edilen veriler analiz edilerek tahmin modelleri oluşturulabilmektedir (Li ve diğ., 2018).

Jeofiziksel Modelleme ve Simülasyon: Makine öğrenimi, yer yüzeyi ve yeraltından toplanan jeofizik verilerini kullanarak süreçlerin anlaşılmasına, modellenmesine ve simülasyonunun oluşturulmasına yardımcı olabilmektedir (Karpatne ve diğ., 2018).

Risk Değerlendirmesi ve Kaynak Keşfi: Jeofizik veriler ile yalnızca maden yatakları, petrol ve doğal gaz rezervleri gibi doğal kaynakların keşfi gerçekleşmez, aynı zamanda bu kaynakların işlenmesi için risk değerlendirmesi de yapılmaktadır ve son yıllarda makine öğrenimi bu risk değerlendirmelerinde de yerini almaktadır (Ma ve Mei, 2021; Cracknell ve diğ., 2014).

Jeofiziksel Araştırma ve Geliştirme: Makine öğrenimi uygulamalarıyla jeofiziksel araştırma ve geliştirme çalışmaları hızlanabilmekte ve yeni keşiflere olanak sağlayabilmektedir (Cao ve diğ., 2023; Yijun ve diğ., 2023; Woolhead ve Landry, 2021).

Makine öğrenimi, yukarıda bahsedildiği gibi jeofizik alanında oldukça çeşitli uygulamalara sahiptir ve bu alanların gelecekte daha da yaygınlaşması beklenmektedir. SIP yöntemi ise çevre kirliliğinin izlenmesinde başarılı sonuçlar vermektedir fakat toplanmasında ve verinin işlenmesinde zorluklarla karşılaşılabilir. Bu zorluklar makine öğrenimi algoritmaları ile çözülebilmektedir. Örneğin çeşitli mineraller hakkında elektriksel veriye laboratuvar ortamında ulaşılması süreci zaman almakta ve maliyetli olmaktadır. Veri elde etmek için harcanacak zamanı azaltmak ve maliyeti düşürmek makine öğrenmesi algoritmaları kullanımı ile mümkündür. Bu bağlamda tezin hipotezi ve temel hedefleri aşağıdaki gibidir.

Tez Hipotezi; Laboratuvar ortamında yapılan farklı yüzdeliklerde mineral-temiz kum karışımının elektriksel cevaplarını öğrenmek için makine öğrenimi algoritmaları kullanılabilir mi?

Bu tezin temel hedefleri şunlardır;

Farklı hacim konsantrasyonlarında kumla karıştırılmış manyetit, siderit ve HFO'nun SIP verisiyle makine öğrenimi algoritması sayesinde minerallerin yoğunluk, frekans ve faz değişimi ilişkisini anlamlandırmak.

Manyetitin farklı tane yarıçapı boyutunun SIP tepkisini regresyon ve sınıflama algoritmalarıyla belirlemek.

### 3. YÖNTEM

Bu bölümde tez konusu olan verinin toplanma -jeofizik SIP- yönteminden ve uygulanan makine öğrenimi algoritmalarından başarılı sonuç veren 13 algoritmadan bahsedilecektir.

#### 3.1. SIP Yöntemi

SIP yöntemi, yer yüzeyinin elektriksel özelliklerini incelemek için kullanılan jeofizik tekniklerinden bir tanesidir. Yeraltındaki sıvı içeriğinin potansiyel iletken minerallerin ve gözeneklilik değişimlerinin varlığını araştırma konusunda önemli bir yeri vardır. Çalışma prensibi; indüklenen polarizasyon sinyalinin frekansa bağlı cevabını inceleyerek geleneksel indüklenen polarizasyon (IP) yöntemini genişletmektedir. Bu yöntemin temelinde yatan mantık, indüklenmiş polarizasyon yöntemiyle karşılaştırarak anlatılırsa daha anlamlı hale gelecektir.

#### IP Temel Prensibi ve SIP ile Karşılaştırılması

IP yöntemi şu şekilde çalışmaktadır; yere çakılan elektrotlar aracılığıyla zemine doğru akım verilir kare dalga gönderilir ve bu akım kesildiğinde yeraltında elektriksel bir tepki oluşur. Oluşan tepkiye voltaj denir ve bu tepki indüklenmiş polarizasyon olarak bilinen, zaman gecikmeli bir bileşene sahiptir.

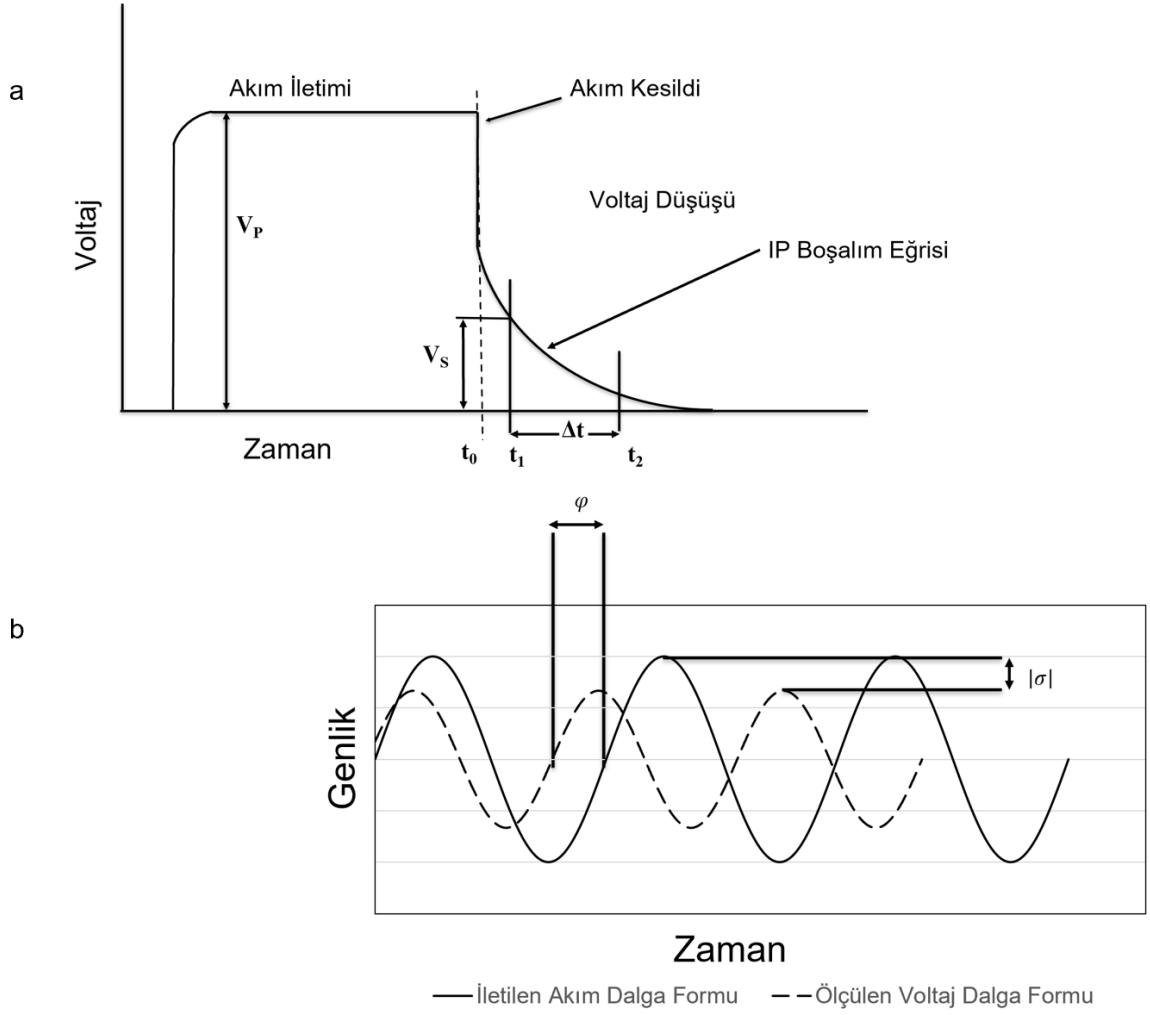
IP yönteminde görünür yüklenebilirlik ( $M_a$ ) hesaplanmaktadır (Ward, 1990). Gerilim farkı akımın kesildiği  $t_0$  anında ölçülememektedir fakat belirli bir süre geçtikten sonra IP boşalım eğrisi üzerinde  $t_1$  ve  $t_2$  zamanları arasında gerilim farkının integre edilmesiyle  $M_a$  bulunmaktadır. Bazı bilim adamlarına göre (Kemna ve diğ., 1997; Vinegar ve Waxman, 1984; Marshall ve Madden, 1959; Seigle, 1959) yüklenebilirlik ile faz açısı arasındaki doğrusal bir ilişki, (zaman alanı IP ölçümlerinde) sabit frekanslı faz açısı tepkisi varsayılarak hesaplanmaktadır. Görünür yüklenebilirlik birimi genel olarak V/mV olarak geçmektedir.

$$M_a = \frac{\int_{t_1}^{t_2} V_s dt}{V_p} \frac{1}{\Delta t}$$

(3.1 )

Bu formülde  $V_p$  verilen akımın herhangi bir anında ölçülen voltaj,  $V_s$  uygulanan akımın kesilmesinden sonra  $t_1$  ve  $t_2$  süreleri arasında tanımlanan bir zaman penceresi boyunca oluşmuş olan bir artık gerilimdir.  $t_1$  ve  $t_2$  arasındaki  $\Delta t$  artık gerilimin oluştuğu zaman penceresinin uzunluğuna eşittir.  $t_1$  ve  $t_2$  akım kesildikten sonraki manuel olarak belirlenen süre, genellikle 10-15 saniye civarındadır.





**Şekil 3.1 a:** IP Ölçümü Sırasında Polarizasyon Etkisinin ve Voltaj Tepkisinin Gösterimi, **b-** SIP Ölçümü Sırasında Polarizasyon Etkisinin ve Genlik Tepkisinin Gösterimi.

SIP yönteminde ise yer yüzeyinden verilen belirli frekanstaki sinüzoidal akıma karşılık, aynı şekilde sinüzoidal olan voltaj tepkisi gecikmeli bir şekilde ölçülmektedir (Şekil 3.1-b). Sonuçlar iletken büyüklüğü ( $|\sigma|$ ) ve faz farkı ( $\varphi$ ) olarak alınmaktadır. Bu sonuçlar eşdeğer olarak Denklem 3.2 veya Denklem 3.5' te kullanılarak karmaşık iletkenlik ( $\sigma$ ) elde edilmektedir. Karmaşık iletkenliğin gerçek ( $\sigma'$ ) ve sanal ( $\sigma''$ ) bileşenleri Denklem 3.3 ve Denklem 3.4'te gösterilmektedir:

$$\sigma = \sigma' + i\sigma'' \quad (3.2)$$

$$\sigma' = |\sigma| \cos \varphi \quad (3.3)$$

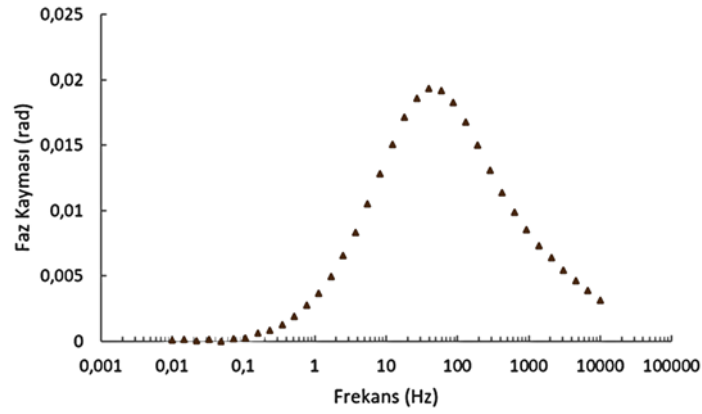
$$\sigma'' = |\sigma| \sin \varphi \quad (3.4)$$

$$\sigma = |\sigma|e^{i\varphi} \quad (3.5)$$

Denklem 3.5 teki  $i = \sqrt{-1}$  olarak ifade edilmektedir.

Karmaşık iletkenlik verisi sonucunda ölçümü alınan yüzey iletkenliği hakkında bilgi edinilmiş olmaktadır.

Tek frekans kullanan IP yönteminden farklı olarak SIP frekans değişimini kullanarak hassas ölçümler alabilmektedir.



**Şekil 3.2:** %3 Manyetit-Kum Karışımının SIP Cevabı.

Mineral ara yüzeylerinde biriken yük, elektrokimyasal reaksiyonlar, çift katmanlı kapasitans indüklenmiş polarizasyona neden olabilen etmenlerdir. Bazı durumlar -iletken mineraller veya hareketli iyonlara sahip gözenek sıvıları gibi- indüklenen polarizasyon etkisine katkıda bulunarak frekansa göre değişen karmaşık bir iletkenlik tepkisi meydana getirmektedir. Şekil 3.2 de %3 manyetit oranına sahip kum manyetit karışımının maksimum faz kaymasının gözlemlenmektedir.

Farklı derinlik ve hassasiyette ölçüm alınmak istendiğinde uygulanan akımının frekansı değiştirilmektedir. Ayrıca çeşitli yer altı malzemeleri, farklı mineraller, sıvı tuzluluğu ve sıvı fazları karakteristik frekansla cevap vermektedir. Verilen farklı frekansta tepkiler bilinen malzemelerin laboratuvar ölçümleriyle karşılaştırılmasıyla veya yer altı elektriksel özelliklerini modellemek için ters çevirme teknikleri kullanılmasıyla yorumlanmaktadır.

SIP yöntemi genel olarak maden yataklarını tanımlama, akifer özelliklerini belirleme, yeraltı suyu akış yollarını haritalandırma ve yeraltındaki kirletici madde hareketlerini değerlendirme amacıyla kullanılmaktadır. Kimyevi yöntemler gibi yeraltına müdahale ederek yapısını bozmadığı ve daha geniş alanları tarayabildiği için daha çok tercih edilmektedir. Tablo 3.1'de bu tez için kullanılmış, laboratuvar ortamında ölçülmüş SIP verileri gösterilmektedir.

**Tablo 3.1:** Makine Öğreniminde Kullanılan SIP Verileri

	<b>Frekans (Hz)</b>	<b>Hacimsel Yüzelik</b>	<b>Tane Çapı (mm)</b>	<b>Özgül Ağırlığı (gr/cm<sup>3</sup>)</b>	<b>Faz Kayması (rad)</b>
1.	1389,495	0,09	45	5,2	0,000277
2.	3059,95	0,19	45	5,2	0,001505
3.	3059,95	0,39	45	5,2	0,003404
4.	3981,072	0,79	45	5,2	0,013534
5.	6738,627	1,56	45	5,2	0,024692
6.	10000	3,1	45	5,2	0,074387
7.	10000	5	45	5,2	0,111511
8.	936,3292	0,09	106	5,2	0,000538
9.	1389,495	0,19	106	5,2	0,001275
10.	1389,495	0,39	106	5,2	0,003905
11.	3059,95	0,76	106	5,2	0,012995
12.	3059,95	1,56	106	5,2	0,021919
13.	4540,91	3,1	106	5,2	0,05259
14.	4540,91	5	106	5,2	0,088753
15.	3059,95	0,9	250	5,2	0,001753
16.	1389,495	0,19	250	5,2	0,002573
17.	1389,495	0,39	250	5,2	0,004757
18.	1389,495	0,78	250	5,2	0,008572
19.	2061,986	1,56	250	5,2	0,013635
20.	3981,072	3,1	250	5,2	0,032413
21.	1000	5	250	5,2	0,064088
22.	193,0698	0,39	1000	5,2	0,00311
23.	193,0698	0,78	1000	5,2	0,005992
24.	193,0698	1,56	1000	5,2	0,013052
25.	286,512	3,1	1000	5,2	0,024895
26.	425,1786	5	1000	5,2	0,049858
27.	425,1786	6,2	1000	5,2	0,063209
28.	130,1025	0,09	2000	5,2	0,000808
29.	130,1025	0,19	2000	5,2	0,001001
30.	59,07838	0,39	2000	5,2	0,003685
31.	59,07838	0,78	2000	5,2	0,005793
32.	39,81072	1,56	2000	5,2	0,014429
33.	39,81072	3,1	2000	5,2	0,019175
34.	39,81072	5	2000	5,2	0,031616
35.	10000	0,78	1000	3,96	0,004049
36.	10000	1,56	1000	3,96	0,001744
37.	10000	3,1	1000	3,96	0,002414
38.	10000	5	1000	3,96	0,004733
39.	10000	0,78	1000	2,65	0,002356
40.	10000	1,56	1000	2,65	0,002153

41.	10000	3,1	1000	2,65	0,00245
42.	10000	5	1000	2,65	0,00283
43.	10000	6,2	1000	2,65	0,002958
44.	10000	12,5	1000	2,65	0,004529

### 3.2. Makine Öğrenimi Yöntemleri

Bu bölümde kullanılacak sınıflandırma ve regresyon terimleri su şekilde tanımlanmaktadır.

Sınıflandırma: Makine öğrenmesinde etiketlenmiş veri kullanılarak modelin eğitilmesine ve kümelere veya sınıflara ayrılmasına denilmektedir.

Regresyon: Makine öğrenmesinde bağımlı ve bağımsız değişken arasındaki ilişkinin matematiksel modellemesidir.

SIP laboratuvar ortamında toplanan verilerin üzerinde başarılı olan 13 program hakkında özet bilgi sunulmuştur.

#### 3.2.1. Random Committee- Rastgele Komite

Sınıflandırma ve regresyon işlemleri için kullanılan bir makine öğrenmesi algoritması olan Random Committee, birden fazla zayıf öğrenciden (çoğunlukla karar ağaçları, K-en yakın komşu veya benzeri basit modeller) oluşan ve genel performansı arttırmak için tahminlerini birleştiren bir topluluk öğrenme yöntemidir. Regresyon kararlarında zayıf öğrencilerden gelen tahminlerin ortalaması kullanılırken sınıflandırma kararlarında oy çokluğu ile belirlenmektedir.

Birden fazla modelin tahminlerini birleştirdiği için tahmin performansını arttırabilmektedir ve aşırı uyum riskinin azaltılmasına yardımcı olmaktadır.

Çalışma prensibi üç aşamada incelenebilmektedir:

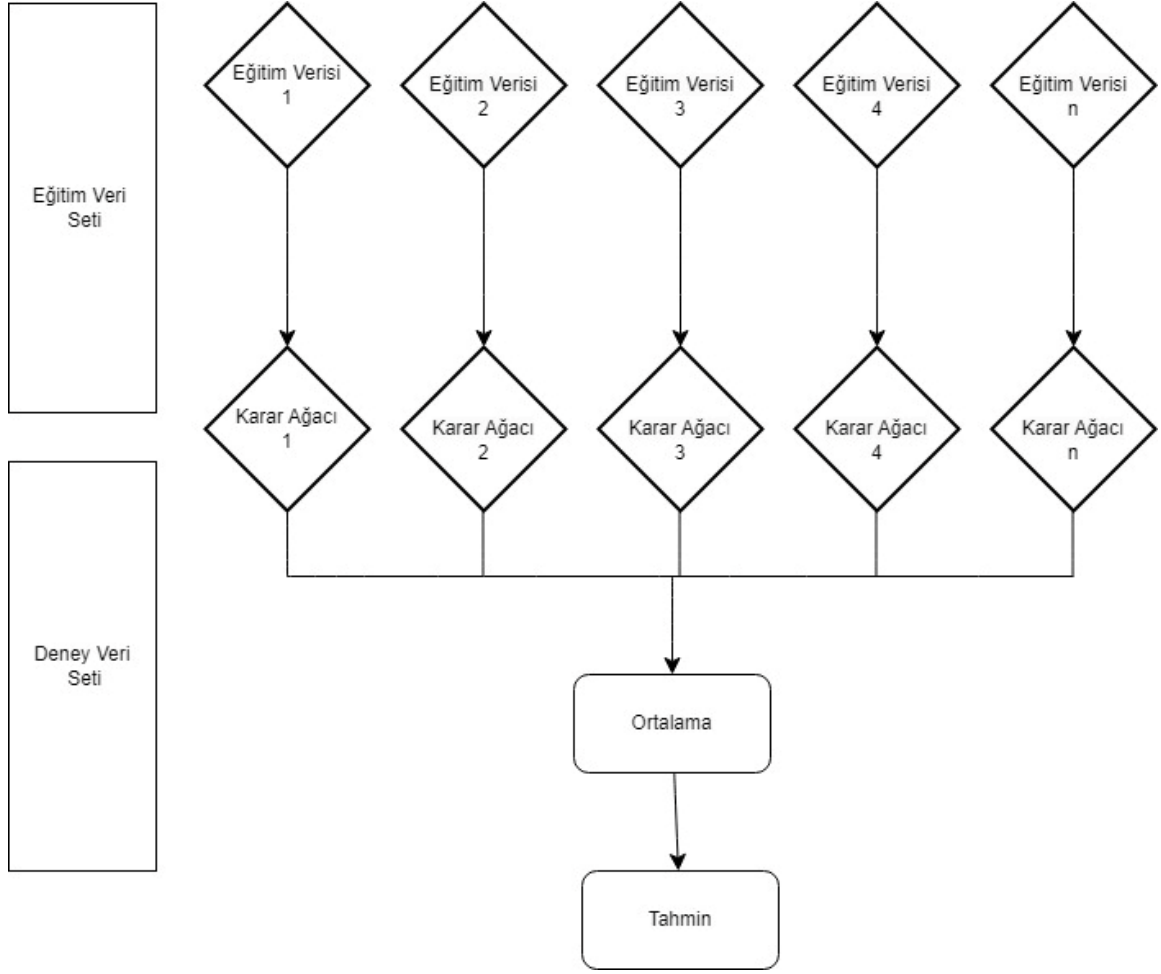
- Zayıf Öğrencilerin Eğitimi; Eğitim veri seti bootstrap yöntemi (Efron, 1992, syf.569) veya rastgele alt örnekleme (Kish. 1965, syf.564) yöntemlerinden biri kullanılarak farklı alt kümelere bölünür. Bu alt kümelerin her biri üzerinde zayıf öğrencilerin birden fazla örneği -aynı veri kümesinde olmak kaidesiyle – algoritma tarafından eğitilir. Eğitim sonucu farklı tahminlerin çıkma olasılığı vardır ve bunun nedeni zayıf öğrencilerin veri setinin farklı alt kümesi üzerinde eğitilmesidir.

- b) Tahminleri Birleřtirme: Eđitimden sonra, eđitilen her zayıf öđrencinin tahmini ile yeni bir veri noktaları bulunur ve bu tahminler birleřtirilir. Regresyon problemlerinde bütün tahminlerin ortalaması ile iřlem yapılırken sınıflandırma problemlerinde çođunluk oyu nihai sınıf etiketini ortaya çıkarmaktadır.
- c) Doğruluđu Arttırma: Bir modelde doğru ve sağlam tahmin sayısının varyantı yüksekken birden fazla modelin birleřmesi varyantsın azaltılmasının yolunu açabilmektedir.

### 3.2.2. Random Tree- Rastgele Ađaç

Hem sınıflandırma hem regresyon görevlerinde geçerli cevaplar sunan bir karar ađacı türüdür. Bu görevleri ađaç oluřturma sürecine rastgelelik ekleyerek yerine getirmektedir. Karar ađaçlarının sađlamlıđını ve çeřitliliđini arttırmak için üretilmiř bir yöntemdir. Bir topluluk yöntemi olan Random Forest yöntemi için temel öđrenenler olarak Rastgele Ađaçlar kullanılabilir.

“Bishop 2006 makalesinde bahsedildiđi üzere, yöntemin modellerde çeřitliliđe neden olması avantajlarındandır.” Geleneksel karar ađaçlarında karřılařılan aşırım uyumu rastgelelik uygulandıđı için azaltılabilmektedir. Verimliliđi ise diđer ađaç tabanlı yöntemlere kıyasla daha yüksektir. Çalışma prensibi ařađıda açıklanan Rastgele Ađaç’ın akıř diyagramı Őekil 3.3’ de gösterilmiřtir (javapoit).



**Şekil 3.3:** Rastgele Ağaç Akış Diyagramı.

Çalışma prensibi aşağıdaki gibi incelenebilmektedir:

- Ağaç Oluşturma;** Sınıflandırma görevinde hedef etiketleri en iyi şekilde ayıran, regresyon görevinde ise verileri- hedef değerleri tahmin eden özellik değerlerine göre alt kümelere yinelemeli biçimde bölerek meydana getirmektedir. Standart bir karar ağacından farkı olarak bu ayırma işlemini rastgeleliğin 2 çeşidini (rastgele özellik seçimi ve rastgele bölmeler) kullanarak yapmaktadır. Rastgele özellik seçimi, tüm özellikleri dikkate almadan sadece özelliklerin rastgele bir alt kümesini dikkate almaktadır. Rastgele bölmeler ise, rastgele seçilen özellik içinde bölme noktası rastgele seçilebilmektedir.
- Bölme Seçim Metotları;** Varyant azaltma, bilgi kazanımı ve Gini safsızlığı gibi bölme kriterleri rastgele seçilen özellik alt kümesine uygulanmaktadır. Böylelikle aşırı uyum olasılığı daha düşük farklı ağaçlar oluşturmaya olanak sağlanmaktadır.

- c) Ağaç Büyümesi; Ağacın büyümesi, bir düğümde çok az örnek olması, maksimum derinliğe ulaşma veya yapraklarda belirli bir saflığa ulaşma ve benzeri durdurma koşuluyla karşılaşınca kadar büyümeye devam etmektedir (Barros ve diğ., 2015).

### 3.2.3. KStar

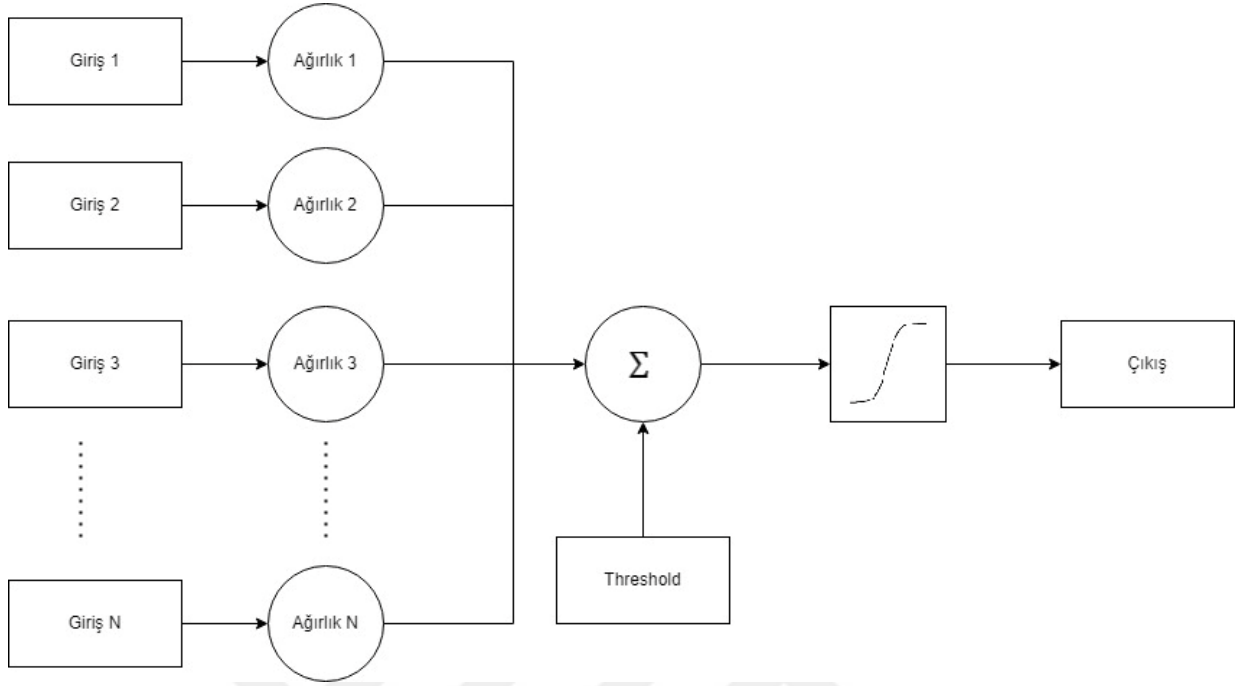
Sınıflandırma görevleri için kullanılmaktadır. Literatürde en yaygın kullanımı olan örnek tabanlı öğrenme algoritmasının (Okamoto ve Yugami, 2003: 207) konseptine dayanmaktadır. KStar entropi tabanlı bir mesafe fonksiyonu kullanır ve en yüksek olasılığı seçer (Painuli ve diğ., 2014: 2641).

Entropi tabanlı mesafe fonksiyonu sayesinde KStar daha etkili bir şekilde çeşitli veri türlerini işlemektedir. Bunun yanında katı geometrik mesafe fonksiyonları yerine olasılıksal dönüşümleri dikkate aldığı için gürültülü verilere göre daha dayanıklı olabilmektedir. En büyük avantajı ise uygulaması ve anlaşılması diğer yöntemlere göre basit ve kolaydır.

Çalışma durumu aşağıdaki şekilde gruplanabilmektedir:

- a) Örnek Tabanlı Öğrenme: Eğitim sırasında genel bir model oluşturmadan eğitim örnekleri toplanıp depolanmaktadır. Daha sonrasında bu örnekler doğrudan tahminlerde bulunmak için kullanılmaktadır.
- b) Entropi Tabanlı Mesafe Fonksiyonu: Dönüşüm karmaşıklığına göre örnekler arasındaki mesafeyi hesaplatan benzersiz bir fonksiyondur. Mesafeler arası ölçü ile veri dönüşümlerinin olasılıksal davranışını hesaba kattığı için geleneksel mesafe ölçümlerinden daha esnek bir yaklaşım sağlamaktadır.
- c) Olasılıksal Benzerlik Ölçüsü: Sınıflandırılan yeni örneğe olan benzerliklerine göre eğitim örneklerinin ağırlıklarını yeniden belirlemektedir. Benzerlik entropi tabanlı mesafe fonksiyonuna göre hesaplandıktan sonra algoritma sınıf etiketini en yakın komşuların ağırlıklı çoğul oyuna göre tahmin etmektedir.
- d) Parametre Ayarlaması: Dönüşüm sırasındaki rastgeleliği kontrol eden bir parametre KStar tarafından kullanılır. Bu parametrenin değişimi, modeldeki önvargı-varyant (bias-variance) dengesini etkileyebilir.

### 3.2.4. Multi-Layer Perceptron-MLP



Şekil 3.4: Multilayer Perceptron Giriş – Çıkışları (Işıksan, 2020).

Çok katmalı algılayıcı olarak da bilinen MLP hem sınıflandırma hem de regresyon görevlerinde kullanılan bir yapay sinir ağıdır. Bulunan her katman bir sonrakine tamamen bağlı olan çok sayıda düğüm katmanlarından meydana gelmektedir. Çeşitli uygulamalar için güçlü bir araç haline gelmesinin ana nedeni ise geri yayılım yoluyla verilerden karmaşık desenleri ve gösterimleri öğrenme becerisidir. Yapısı giriş katmanı, ağırlık belirleme, aktivasyon fonksiyonu ve çıktı katmanından oluşmaktadır. Şekil 3.4’de multilayer perceptron giriş ve çıkışları şematik olarak gösterilmiştir.

MLP karmaşık doğrusal olmayan ilişkileri yakalayabilmektedir. Ayrıca sınıflandırma ve regresyon görevleri dışında zaman serisi tahmini dahil olmak üzere çeşitli görevler içinde uygulanabilmektedir.

Çalışma prensibi aşağıdaki gibi açıklanabilmektedir:

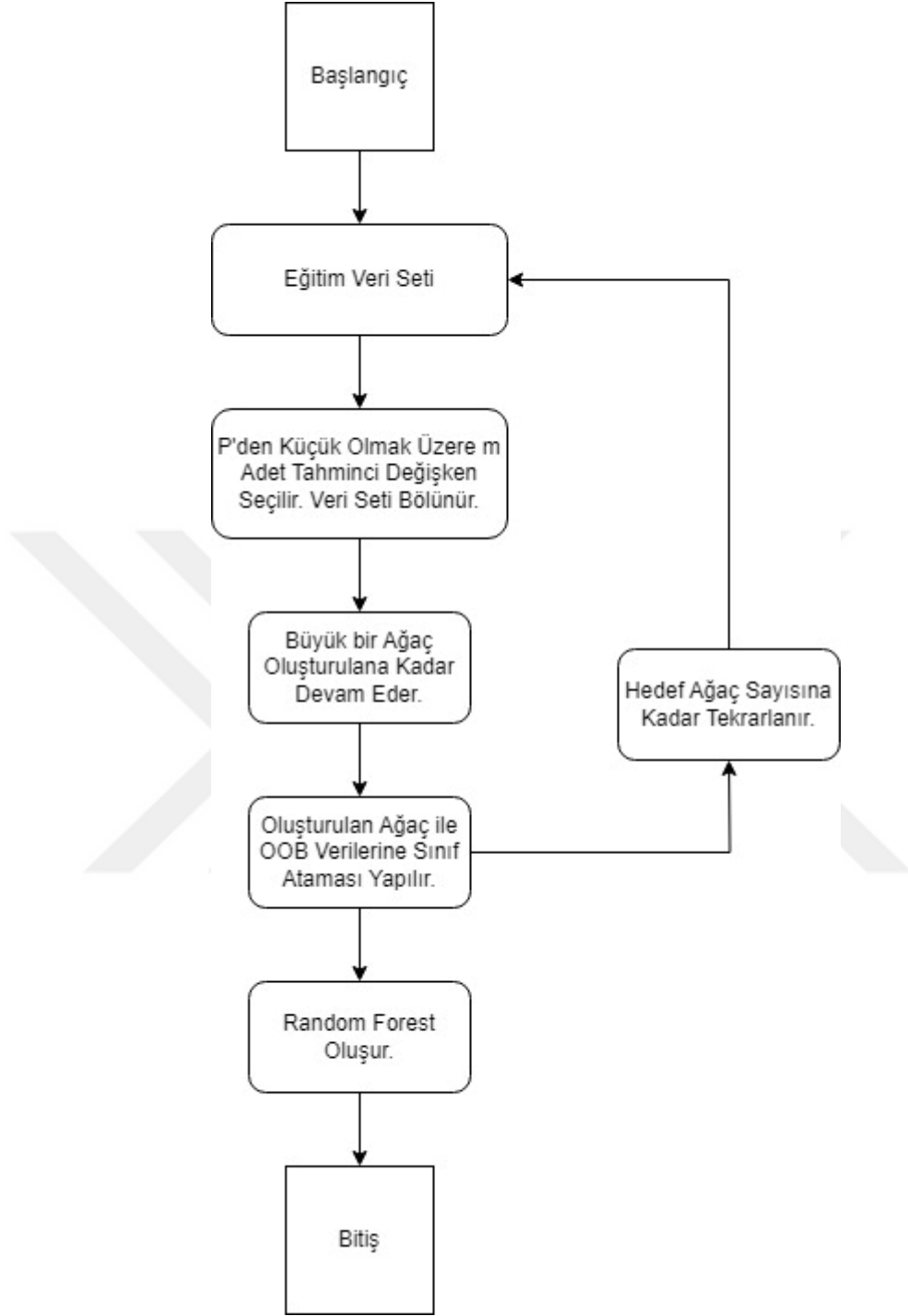
- İleri Yayılım: Girilen verileri ağdan katman katman geçmektedir. Bir katmandaki her nöron, girdilerinin ağırlıklı toplamını hesapladıktan sonra bir aktivasyon fonksiyonu uygular ve çıktıyı bir sonraki katmana iletilmektedir.

- b) Aktivasyon Fonksiyonları: Ağa lineer olmayanlık dahil ederek karmaşık desenlerin öğrenmesini sağlayan fonksiyonlardır.
- c) Kayıp Fonksiyonu: Gerçek hedef değerlerinin ağı tahminlerinden olan uzaklığını ölçer. Regresyon için ortalama karesel hata, sınıflandırma içinse çapraz entropi yaygın kayıp fonksiyonları arasında yer almaktadır.
- d) Geri Yayılım ve Optimizasyon: Geri yayılım ağdaki her bir ağırlığa göre kayıp fonksiyonunun eğiminin hesaplama işlemidir. Bu işlem, kayıp fonksiyonunu yinelemeli olarak en aza indirerek ağı performansını iyileştirmektedir.

### **3.2.5. Random Forest- Rastgele Orman**

Bir topluluk öğrenme yöntemi olan Random Forest hem sınıflandırma hem de regresyon görevleri için uygun görülmektedir. Çok sayıda karar ağacı eğitim sırasında oluşturulmaktadır ve bu karar ağaçları arasında sınıf modu sınıflandırma için veya bireysel ağaçların ortalama tahmini regresyon için bir çıktı üretmektedir.

Bireysel karar ağaçlarına kıyasla daha yüksek doğruluk ve genelleme potansiyeline sahiptir. Verilerdeki gürültüye karşı dayanıklıdır. Bununla birlikte özellik önemi tahminleriyle en etkili özelliği belirlemeye yardımcı olabilmektedir.



**Şekil 3.5:** Rastgele Orman Akış Diyagramı (Yılmaz, 2014).

Rastgele Orman akış diyagramı Şekil 3.5’da gösterilmiştir ve çalışma yöntemi şu şekilde açıklanabilmektedir:

- a) Önyükleme Örnekleme: Rastgele örnekleme verilerinin yerlerini değiştirerek eğitim verilerinin birden fazla alt kümesini oluşturmaktadır. Bu alt kümelerin her biri farklı

birer karar ağacını eğitmek için kullanılmaktadır. Bu işleme torbalama (bagging) denir ve aşırı uyumu önlemek ve varyantı azaltmak için önemlidir.

- b) Rastgele Özellik Seçimi: Karar ağaçlarının aksine her karar ağacının inşası sırasında düğümleri bölmemektedir yani tüm özellikleri dikkate almak yerine rastgele bir özellik kümesi seçmekte ve ağaçlar arasında çeşitlilik meydana getirmektedir. Bu da aralarındaki korelasyonu azaltmaya yardımcı olmaktadır.
- c) Karar Ağaçları Oluşturma: Önyükleme ve rastgele özellik seçiminden sonra ağaçlar budamadan maksimum derinliklerine büyütülür. Bu durum ağaçların ilgili örneklerinde doğru sonuçlar elde etmesini sağlamaktadır fakat aşırı uyum sorununa neden olabilmektedir.
- d) Tahminleri Birleştirme: Tüm ağaçların tahminleri arasında en çok oyu alan sınıfa atamasıyla sınıflandırma görevini, tüm ağaçlardan gelen tahminlerin ortalamasıyla da regresyon görevini gerçekleştirmektedir.

### 3.2.6. Decision Table-Karar Tablosu

Karar analizinde ve makine öğrenmesinde kullanılan basit fakat güçlü bir yöntemdir. Kısacası girdi özelliklerini çıktı veya kararlara eşleyen kurallar tablosu olarak düşünülebilmektedir.

Karar alma süreçlerini net ve öz bir yöntemle temsil etmektedir. Tüm koşulları ve eylemleri listelemesi tutarlılığı sağlamaya yardımcı olmakla kalmayıp aynı zamanda hiçbir senaryonun gözden kaçmasına izin vermemektedir. Koşul ve eylemlerin sık sık değişikliğe uğramadığı durumlar için iyi sonuçlar vermektedir.

Karar tablosu birleşenleri şu şekilde sıralanabilir:

- a) Koşullar: Girdi özellikleridir. Genel olarak karar tablosunun sütunlarında yer almaktadır.
- b) Eylemler: Belirlenmiş koşul kombinasyonlarından ortaya çıkan sonuç ve kararlardır. Koşullara karşı gelen ayrı sütunlarda yer almaktadır.
- c) Kurallar: Koşulların bir kombinasyonunu ve ortaya çıkan eylemi ayrıntılı olarak anlatan belirlenmiş bir kuralı temsil eden satırdır.
- d) Koşul Girişleri: Oluşturulmuş bir kuraldaki her koşul için değerleri veya aralıkları temsil etmektedir.
- e) Eylem Girişleri: Kural içerisindeki koşul gerçekleştiğinde gerçekleştirilecek eylemi belirtmektedir.

### 3.2.7. M5P Rules-M5P Kuralları

Karar ağacı algoritması olan M5P özellikle regresyon görevlerinde kullanılmaktadır. Klasik karar ağacı ve doğrusal regresyon temel yapı taşlarıdır. Karar ağaçlarının sezgisel bölme yeteneğiyle doğrusal modellerin doğruluğunu birleştiren metot, karmaşık ve büyük veri kümelerinde etkili ve hassas sonuçlar ortaya koymaktadır. Verinin basit ya da karmaşık olması algoritmanın performansını etkilememektedir.

Algoritmanın çalışma şekli şöyledir ki:

- a) Veri Girişini Hazırlama: Çok boyutlu veri girişiyle başlar yani veri kümesi bağımlı ve bağımsız değişkenler içermektedir.
- b) Ağacın Oluşturulması: Çok boyutlu veriyi parçalara ayırarak bir ağaç oluşturmaktadır. Bu bölme işlemi veri kümesi farklı eşik değerlerinde test edilerek yapılmaktadır. Bağımlı değişken üzerindeki en büyük varyant azalmasının olduğu değer veya özellik baz alınarak bölmeler yapılmaktadır. Ağacın her bir dalındaki veri belli bir değere göre ayrılmaktadır ve bu döngü ağacın yaprak düğümleri elde edilene kadar sürdürülmektedir.
- c) Doğrusal Regresyon Modellerinin Oluşturulması: Yaprak düğümlerindeki verilerin en iyi şekilde temsili için doğrusal model oluşturulmaktadır. Bu işlem her yaprak düğüm için tekrarlanmaktadır.
- d) Budama: Lüzumsuz veya çok küçük veri kümeleri üzerinde inşa edilen dalların budanması gerekmektedir. Böylelikle veri sadeleşme gerçekleşmektedir. Ayrıca yaprak düğümünde meydana gelen doğrusal model, alt dallardaki modelden daha başarılıysa bu durumda alt dallar budanmaktadır. Eğitim sonrası algoritma hazır durumundadır.
- e) Tahmin: Yeni veri örnekleri karar ağacı üzerindeki uygun dalı izler ve son yaprak düğümdeki doğrusal model kullanılarak tahminde bulunmaktadır.

### 3.2.8. M5 Rules-M5 Kuralları

Genel olarak M5P olarak bilinen bir algoritmadan türetilen M5 Rules regresyon görevleri için kullanılır. Karar ağaçları ve kural tabanlı öğrenme algoritmaları birleşiminden oluşmaktadır. Bu birleşim sonucu oluşan kuralları tahmin yapmak için kullanılmaktadır. Böylelikle sayısal büyük verilerin tahmin sorunlarına doğrulanabilirlik ve yorumlanabilirlik çözümlerini sunmaktadır. Kural tabanlı yapısı modelin daha kolay anlaşılmasına ve yorumlanmasına yardımcı olmaktadır. Ancak model ağacının oluşturulmasına ve budama eşiğine özen gösterilmezse karmaşıklığa neden olabilmektedir.

M5 Rules çalışma prensipleri su şekilde sıralanabilmektedir:

- a) Model Ağacı Oluşturma: Karar ağacı modeline benzemektedir; eğitim verilerinden bir karar ağacı oluşturularak başlanmaktadır. Fakat karar ağacının yapraklarında bulunan sabit değerler yerine doğrusal regresyon modelleri bulunan model ağacı oluşturulmaktadır. Yapraklarda oluşturulan bu modeller sayesinde her yaprağın girdi özelliklerinin doğrusal bir kombinasyonuna bağlı devamlı bir değeri tahmin etmesini sağlanmaktadır.
- b) Budama: Bu işlem genelleme yeteneğini geliştirmek ve aşırı uyumu önlemek için yapılmaktadır. Tahmin doğruluğuna büyük ölçüde katkısı bulunmayan dallar budanmaktadır.
- c) Kural Çıkarımı: Budama işlemi sonucunda başlangıçtaki tam ağaçtan daha az dallı ve bunun sonucu olarak daha anlaşılabilir olan modelden kural çıkarımı yapılmaktadır. Şöyle ki; ağacın başlangıç noktasından yaprak düğüme kadar izlenen belirli bir yol vardır ve bu yol üzerindeki her bir dal bir koşul olarak kural oluşturulmasına yardım etmektedir. Böylelikle her yaprak düğüm için bu şekilde kurallar belirlenmektedir.
- d) Kural Basitleştirme: Benzer kurallar bir araya getirilmekte ve gereksiz koşullar kaldırılmaktadır. Bu işlem kuralları daha yorumlanabilir hale getirmektedir.
- e) Tahmin: Kurallar belirlendikten sonra M5 modeli yeni bir veri geldiğinde bu verinin hangi yaprak düğümünde olduğu tahmin edilmektedir. Modelin performansına göre oldukça hassas ve doğrusal olabilmektedir.

### **3.2.9. Additive Regression-Eklenen Regresyon**

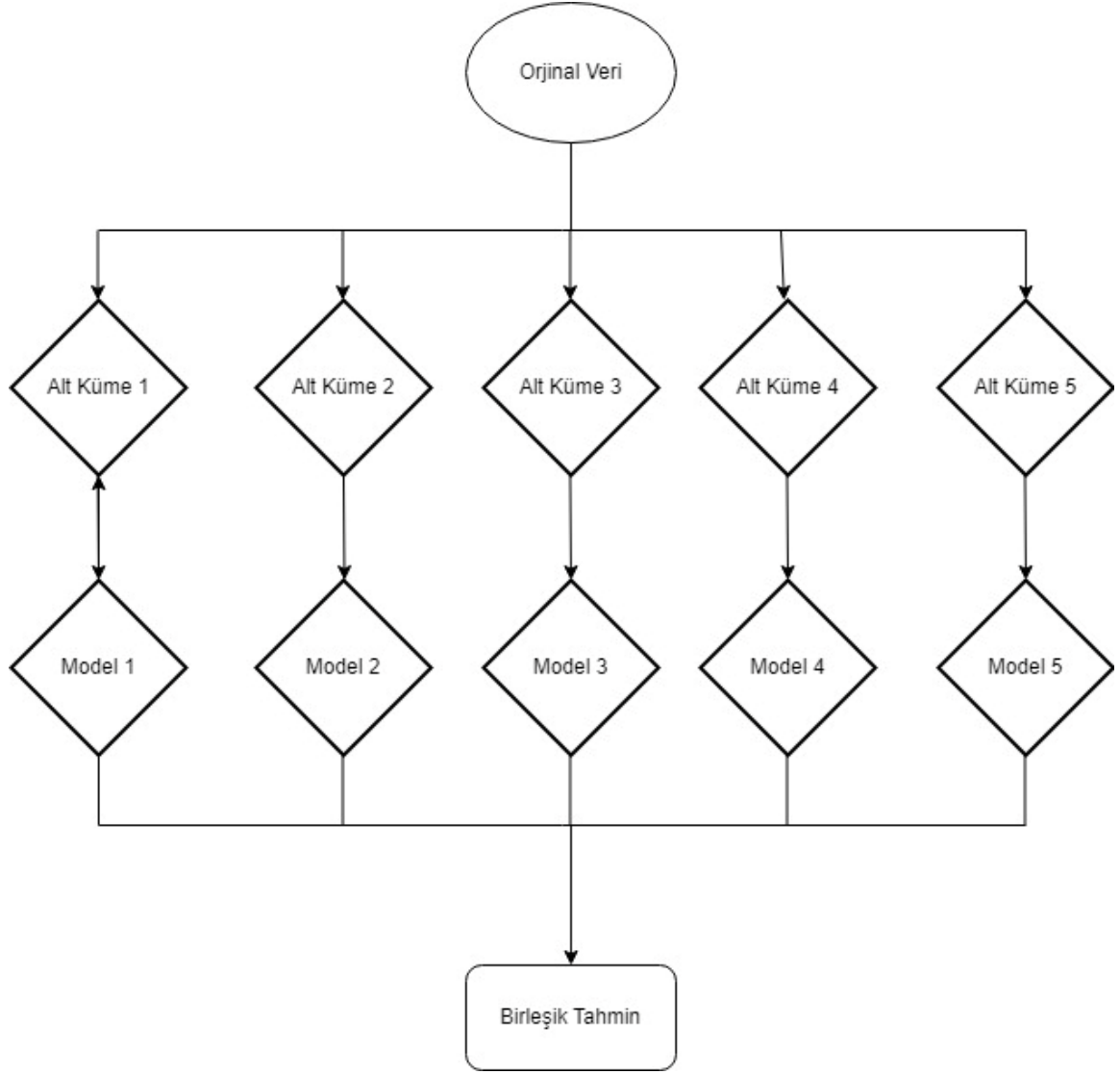
Regresyon görevleri için kullanılan bir makine öğrenme tekniğidir. Zayıf öğrenenlerden oluşan bir grubu güçlü öğrenene dönüştürmeyi hedeflemektedir. Bunun için basit ve zayıf öğrenciler bir araya getirilerek oluşturulan birden fazla modelin tahminlerini birleştirmektedir. Bir sonraki model tüm önceki modellerin hatalarını düzeltmek üzere eğitildiği, sıralı bir şekilde daha iyiye giden bir model topluluğu meydana getirmektedir. Yöntem çeşitli veri ve sorun türlerine uygulama esnekliğine sahiptir. Zaman alıcı bir yöntem olsa bile çok sayıda yinelemeli veya karmaşık temel modellere eklenen yeni model ve nihai tahmine bu yeni modelin etkisi gözlemlenebilmektedir. Bu da yorumlanabilirliği arttırmaktadır.

Çalışma prensibi şu şekildedir:

- a) Başlatma: Eğitim verileri üzerinde tahminlerde bulunmak için hedef değerlerin ortalaması gibi basit bir model kullanarak bir başlangıç modeli oluşturulmaktadır.
- b) Ardışık Eğitim: Yeni modeller topluluğa eklenmektedir. Yineleme sonucu tüm eski modellerin birleşik topluluğu hataları için yeni bir model eğitilmektedir. Bu hatalar gerçek hedef değerler ile mevcut topluluk tarafından yapılan tahminler arasındaki farka bakılarak belirlenmektedir.
- c) Model Ekleme: Uygun bir ağırlık belirlenerek yeni model topluluğa eklenmektedir. Belirlenen ağırlık modelin son tahmini hangi derecede etkilediğini belirlemektedir. Bu işlem sonucunda yeni model birleştirilmiş modele eklenip güncellenmektedir.
- d) Tahmin: Yeni veri için tahmin birleştirilmiş modeldeki tüm modellerin ağırlıklı tahminlerinin toplanmasıyla ortaya çıkmaktadır.

### **3.2.10. Bagging-Torbalama**

Sınıflandırma ve regresyon görevleri için idealdir. Bunun yanında herhangi bir algoritma da toplama için de kullanılmaktadır. Birden fazla modelin tahminlerini birleştirmesi aşırı uyumu önlemeye yardımcı olmakla beraber daha doğru ve kararlı modeller oluşturmaktadır. Özellikle yüksek varyansa sahip modellerde bu yöntemin kullanımı önerilmektedir (Kuhn ve Johnson, 2013). Model eğitimleri paralel olarak gerçekleştirildiğinden dolayı zaman tasarrufu sağlamaktadır. Özetle Bagging akış diyagramı Şekil 3.6' de gösterilmiştir



**Şekil 3.6:** Torbalama Akış Diyagramı (Ay, 2019).

Çalışma prensibi şu şekildedir:

**Bootstrap Örnekleme:** Art arda değiştirilen orijinal eğitim verilerinden rastgele alt veri kümeleri oluşturma prensibine dayanan önyükleme (bootstrap) işlemi, bagging işleminin ilk aşamasıdır. Alt kümelerin her biri orijinal veri setiyle aynı boyuttadır. Bu adımda bazı örnekler hiç kullanılmazken bazıları çok defa kullanılabilir.

**Birden Fazla Modelin Eğitilmesi:** Her model önyükleme sonucu oluşturulan yeni örnekleri üzerinden eğitime tabii tutulmaktadır. Eğitilen bu modeller benzer türde olsa bile farklı örnekler üzerinde eğitildiklerinden dolayı farklı tahminler doğurabilmektedir.

Tahminlerin Birleştirilmesi: Her modelin tahmini hesaplandıktan sonra regresyon görevleri için tahminlerin ortalaması nihai tahmin, sınıflandırma görevleri için ise en sık tahmin edilen sınıf çoğunluk oylaması ile belirlenmektedir.

### **3.2.11. Linear Regression- Doğrusal Regresyon**

Bağımlı ve bağımsız değişkenler arasında doğrusal bir ilişki olması ilkesi üzerine kurulmuş basit ve güçlü bir algoritmadır. Birçok karmaşık algoritma için temel olarak kabul edilmektedir. Veri analizinden tahmini modellemeye kadar yaygın bir kullanıma sahiptir. Basit doğrusal regresyon ve çoklu doğrusal regresyon iki türüdür. Bu algoritmanın performansı değişkenler arasındaki ilişkinin gerçekten doğrusal olup olmadığına bağlıdır. Aykırı değerler modelin doğruluğu üzerinde ciddi bir etkiye sahipse de kolayca yorumlanabilirliği, hızlı ve basit oluşu bu türü avantajlı kılmaktadır.

Uygulama süreci şu şekildedir:

Veri Toplama ve Ön İşleme: Bağımlı ve bağımsız değişkenler ve bu değişkenlere en uygun katsayılar belirlenmektedir. Ölçekleme, eksik değerler sorunu, aykırı değerleri bulma gibi ön işlemlerden geçirilmektedir.

Modelin Kurulması: Model hazırlanmış olan veri içerisinde olan eğitim verisi kullanılarak eğitilmektedir.

Model Test Etme: Eğitilen model test verisi kullanılarak test edilip performansı ölçülmektedir. Gerekli görüldüğü halde hiper parametre ayarları veya farklı özelliklerin eklenmesi yollarıyla model iyileştirilmesi yapılabilmektedir.

Tahmin Yapma: Model tahmin yapma işi için yeni verilere uygulanmaktadır.

### **3.2.12. Random Sub Space- Rastgele Alt Uzay**

Sınıflandırma ve regresyon görevlerinde kullanılmaktadır. Bir topluluk oluşturma yöntemi olan rastgele alt uzay yönteminde veri seti içerisinde rastgele seçimde bulunularak modelleri eğitmektedir. Yüksek boyutlu veri setleri için bu yöntemle çalışmak modeli daha güçlü ve kararlı kılmaktadır. Model ve özellik alt küme sayısı büyük olduğunda maliyetli olabilmekte fakat aynı durum sayesinde topluluk yeni görülmemiş verileri daha iyi genelleme yaparak modellerin performansını ve sağlamlığını arttırmaktadır.

Çalışma yöntemi şu şekildedir:

Özellik Alt Kümesinin Oluşturma: Toplanan veri seti birçok özellik içermektedir. Bu özelliklerin içinden bir tanesi rastgele olarak seçilerek bir özellik alt kümesi oluşturulmaktadır.

Seçilen her özellik alt kümesi veri setindeki toplam özelliklerin rastgele bir örneklemesini oluşturmaktadır.

**Model Eğitimi:** Oluşturulan özellik alt kümeleri ayrı ayrı modelleri (karar ağacı, destek vektör makinesi veya başka bir sınıflandırıcı/regresör) tekrar tekrar eğitmektedir. Farklı bir özellik kümesinden öğrenen modeller kendi aralarında çeşitlilik sağlamış olmaktadır.

**Tahminlerin Birleştirilmesi:** Eğitilmiş olan çeşit çeşit modellerden gelen tahminler bir araya getirilmektedir. Sınıflandırma görevleri için çoğunluk oylaması, regresyon görevleri için ortalama tahminler değerlendirilmektedir.

**Nihai Model Oluşturulması:** Değerlendirilen bütün alt küme tahminleri bir araya getirilerek nihai bir karar verilmektedir. Bu şekilde model çeşitliliğinden yararlanarak genelleme hatalarının azaltıldığı bir nihai bir model oluşturulmuş olmaktadır.

### **3.2.13. SMOreg**

Sıralı minimal optimizasyon anlamına gelen SMOreg, destek vektör makinelerini regresyon görevleri için hazırlamaktadır. Bu yöntem büyük eğitim verileri için zaman kaybına ve yüksek maliyete neden olabilmektedir. Buna karşın genelleme performansı yüksektir ve doğrusal olmayan veri ilişkilerini iyi derecede modelleyebilmektedir.

Aşağıdaki prensipler üzerine bina edilmiştir:

**Kernel Fonksiyonları:** Doğrusal olmayan veri ilişkilerini modellemek için Kernel fonksiyonlarını kullanmaktadır. Fonksiyonun amacı orijinal veri uzayını daha yüksek boyutlu bir uzaya uyarlayarak doğrusal hale getirmeye çalışmaktadır. Lineer Kernel, Polinom Kernel ve RBF- Radial Basis Function- Kernel yaygın kullanılan fonksiyon türleridir.

**Destek Vektörleri:** Destek vektör makineleri üyesi olan çeşitli destek vektörleri karar sınırlarını belirlemektedir.

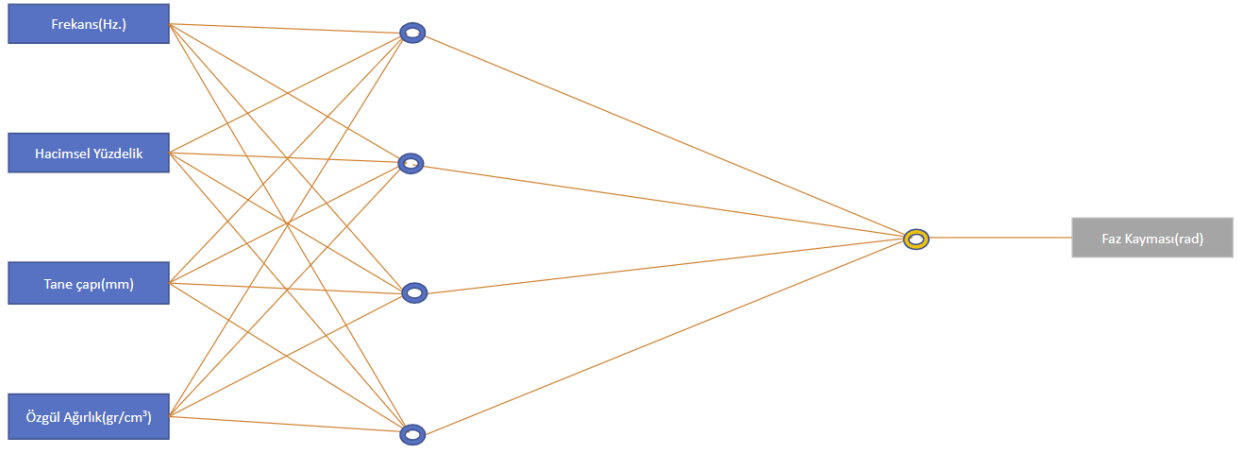
**Hata Marjı:** Model tahminleri ile gerçek değerler arasındaki kabul edilebilir hata aralığını tanımlamaktadır.

**Optimizasyon:** Problemleri hızlı bir şekilde çözmek için SMO algoritmasını kullanmaktadır. Bu algoritmanın amacı model parametrelerini optimize etmektir.

**Regülerizasyon:** Aşırı uyumun önüne geçmek için regülerizasyon parametresi kullanılmaktadır.

#### 4. BULGULAR

Şekil 4.1’de görüldüğü üzere giriş verisi olarak tane çapı, frekans değeri, hacimsel yüzdelik ve mineral cinsi, çıkış olarak ise faz kayması verileri kullanılarak temel model yapısı oluşturulmuştur.



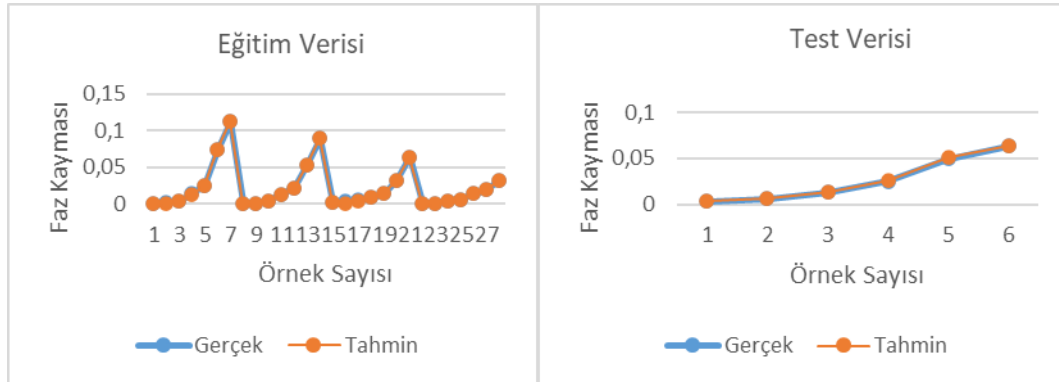
Şekil 4.1: Faz Kayması Şeması.

Elde bulunan 44 adet farklı örneğin ölçümünden oluşan SIP verileri, toplamda 1683 tanedir (0.01-10000 Hz. aralığında farklı frekanslar) ve bu verinin 1647 tanesi eğitimde, 36 tanesi ise testte kullanılmıştır. Test verileri eğitim aşamasında kullanılmaksızın başarılı sonuçlar elde edilmesi amaçlanmaktadır.

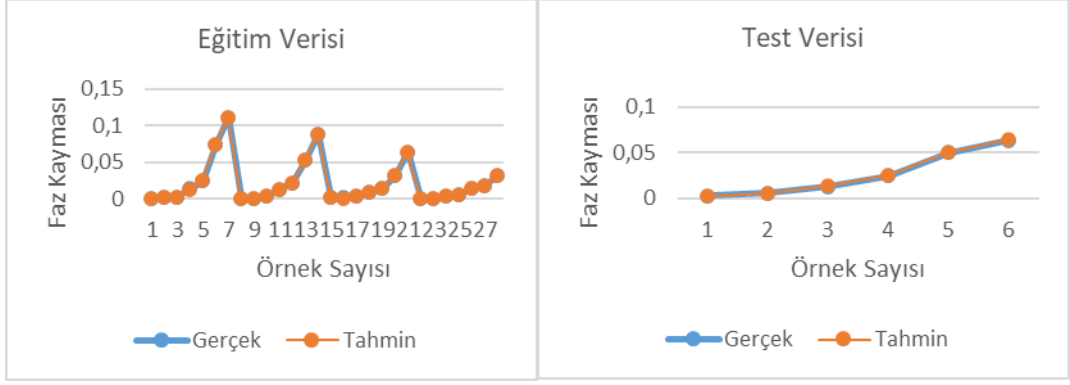
İlk denemede öncelikle 1683 adet veri normalize (min-max normalizasyonu) edilmiştir. Ardından 1647 tanesi (43 örneğin verisi) eğitimde, geriye kalan 36 tanesi (1 örneğin verisi) ise testte kullanılmıştır. Ancak giriş frekansı 630 Hertz’in üzerine çıktığında algoritmalar açısından tutarlı olmayan sonuçlar elde edilmiştir. Bunun üzerine kör nokta olarak tabir ettiğimiz aralıkları detaylandırabilmek için Origin programında 44 adet örneğin verilerinin her birini 1000 adet veriye evrilerek veri sayısı 44.000’e çıkarılmıştır. Böylelikle ikinci deneme için bu 47.000 veri önce normalize (min-max normalizasyonu) edilmiş ve ardından 43.000 tanesi (43 örneğin verisi) eğitimde, geriye kalan 1000 tanesi (1 örneğin verisi) ise testte kullanılmıştır. Ancak yine de algoritmalarla tutarlı olmayan sonuçlara ulaşıldığı görülmüştür.

Veriler herhangi bir nedenle normalizasyona uygun olmayabilir düşüncesiyle bu kez normalize etmeden arttırılmış veriler kullanılarak yeni bir deneme yapılmıştır. Ancak yine makine öğrenimi algoritmaları için tutarlı olmayan sonuçlara ulaşılmıştır. Özetle ilk amaca ulaşmak bu veri setiyle mümkün olamamaktadır.

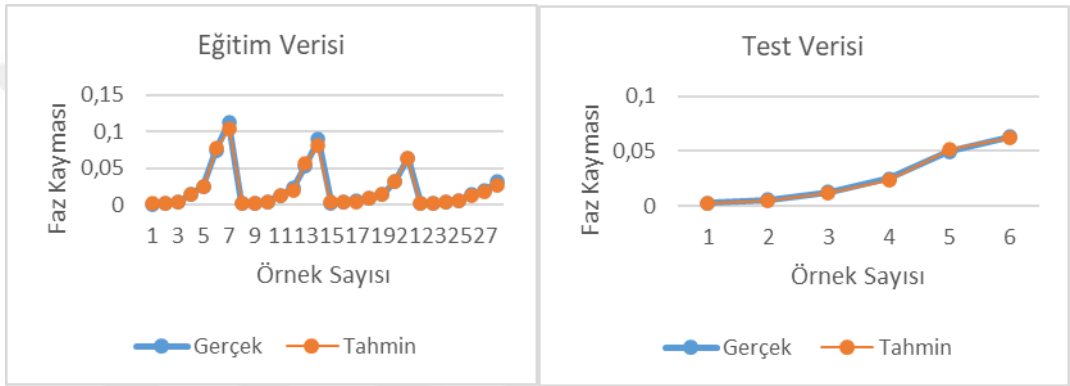
Bu sebeple ham datayı sınıflandırma yaparak en uygun makine öğrenimi yönteminin belirlenmesi şeklinde tez çalışmasını nihayete erdirmekten başka seçenek kalmadığı düşünülmüştür. Ancak jeofizik bölümünde uzmanlaşmış kişilerle yapılan istişareler sonucunda verinin tamamını işleme almak yerine faz farkının en yüksek olduğu noktaları işleme almak fikri doğrultusunda hareket etmenin daha anlamlı olacağı varsayılmıştır. Elde bulunan 44 örneğin 34 adeti metalik mineral olan manyetitten oluşmaktadır ve tepe noktaları belirgin şekilde gözlemlenebilmektedir. Bu sebeple çalışmaya bu veri seti ile devam edilmesine karar verilmiştir. Amaç, maksimum noktalara ait verileri kullanarak makine öğrenimi algoritmaları kısmında toplam 34 adet veriden 28 tanesini eğitimde, geriye kalan 6 tanesini de testte kullanarak test verilerini eğitimde hiç vermeden başarılı sonuçlar elde etmek olarak belirlenmiştir. İlk başta kullanıldığı gibi çıkış olarak faz kayması verileri (1000 ile çarpılarak), giriş verisi olarak tane çapı, frekans değeri ve mineral cinsi verileri kullanılarak temel model yapısı oluşturulmuştur.



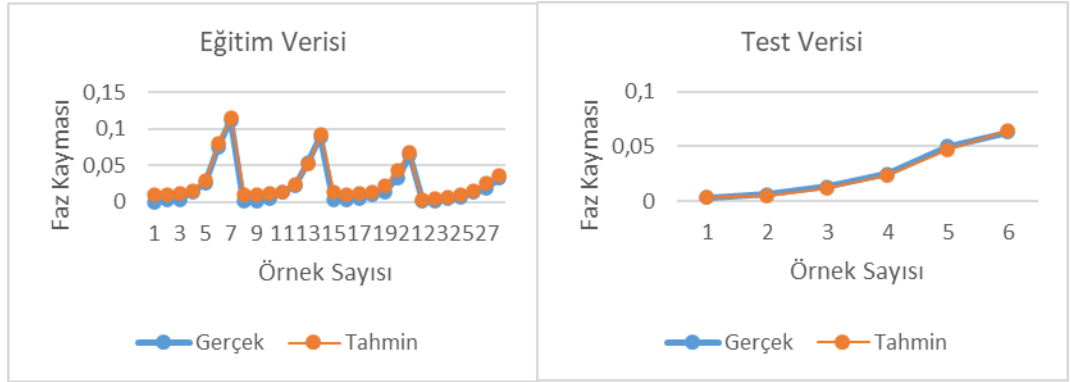
Şekil 4.2: Rasgele Komite Eğitim ve Test Verisi Sonuçları.



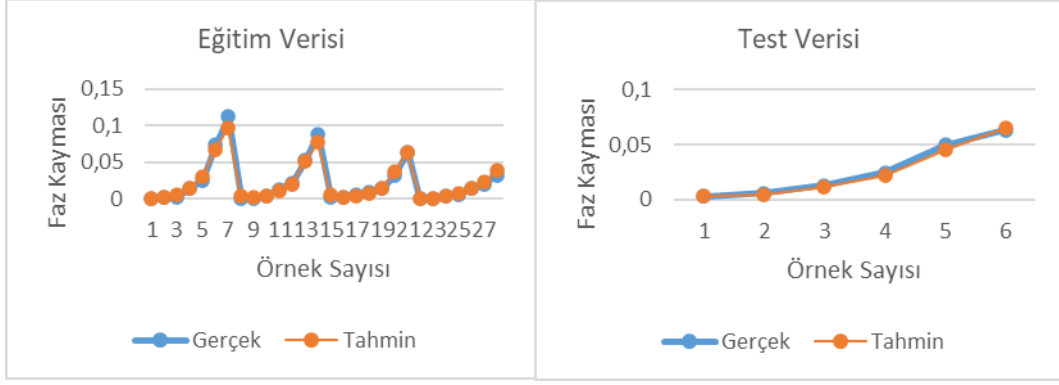
Şekil 4.3: Rastgele Ağaç Eğitim ve Test Verisi Sonuçları.



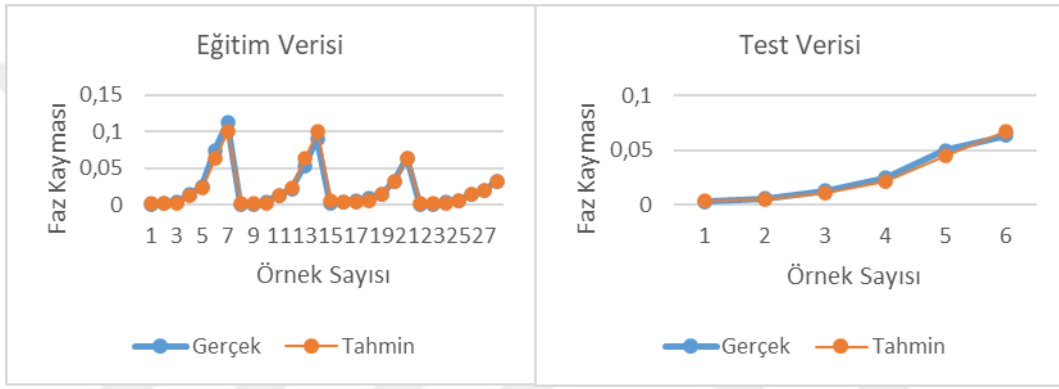
Şekil 4.4: KStar Eğitim ve Test Verisi Sonuçları.



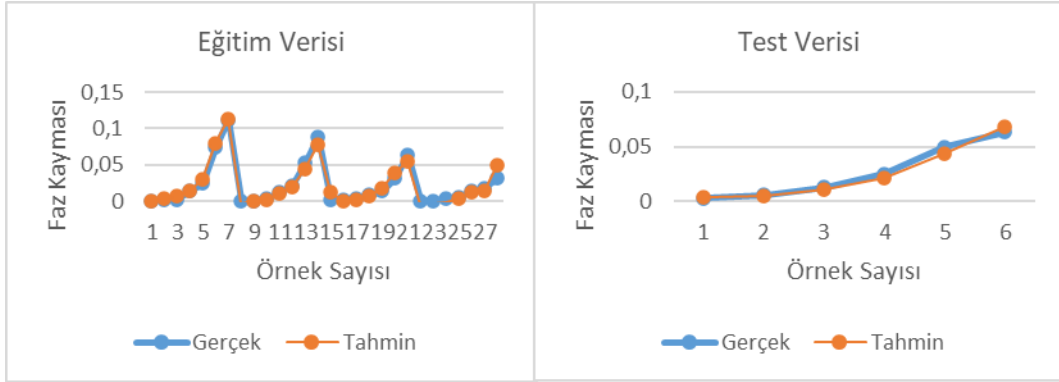
Şekil 4.5: MLP Eğitim ve Test Verisi.



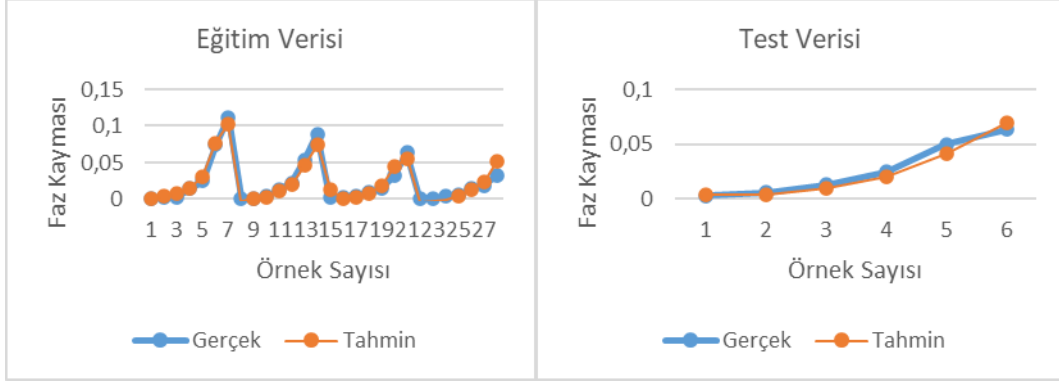
Şekil 4.6: Rastgele Orman Eğitim ve Test Verisi.



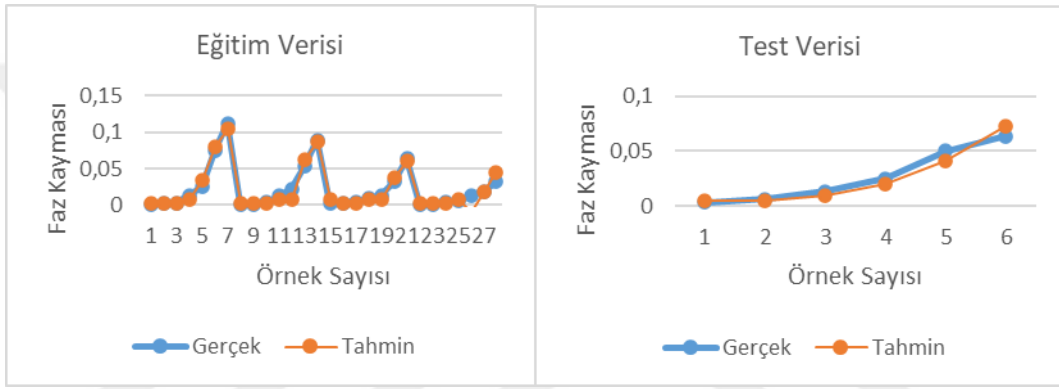
Şekil 4.7: Karar Tablosu Eğitim ve Test Verisi.



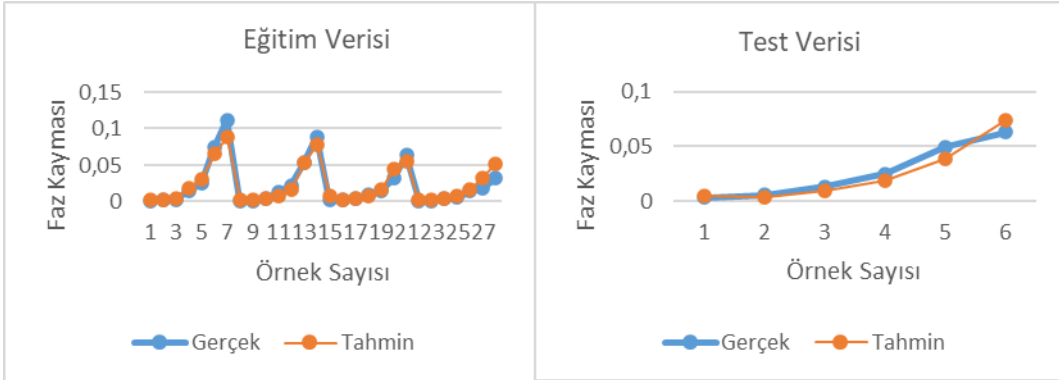
Şekil 4.8: M5 Rules Eğitim ve Test Verisi.



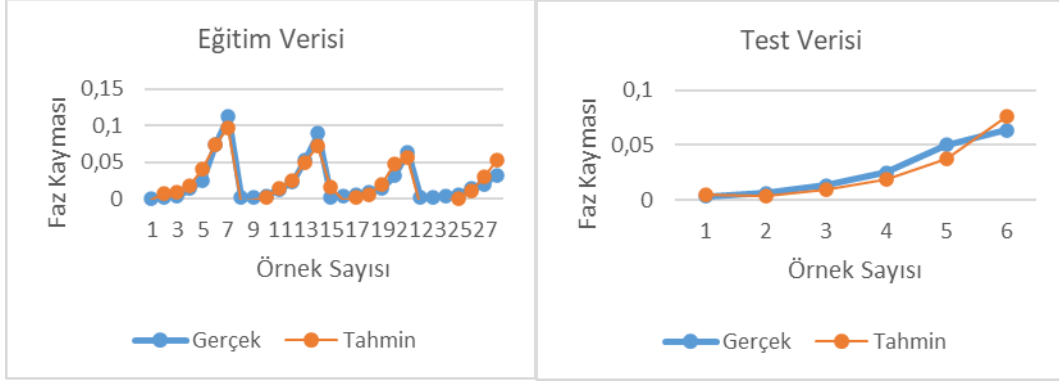
Şekil 4.9: MSP Eğitim ve Test Verisi.



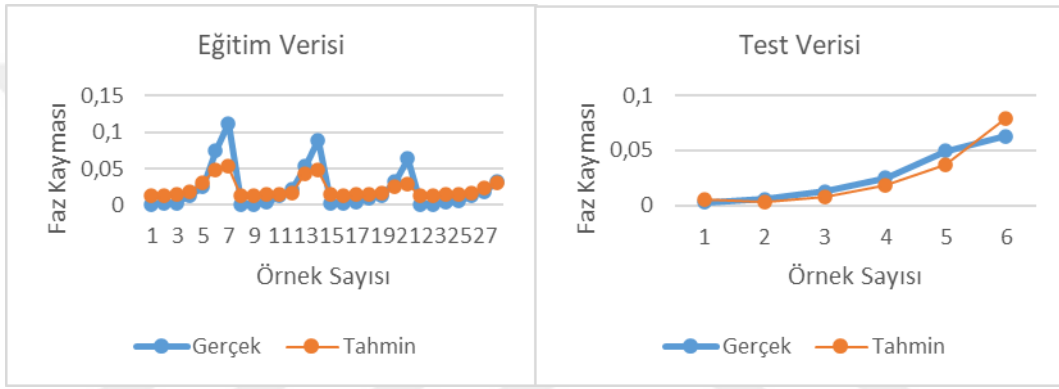
Şekil 4.10: Eklenen Regresyon Eğitim ve Test Verisi.



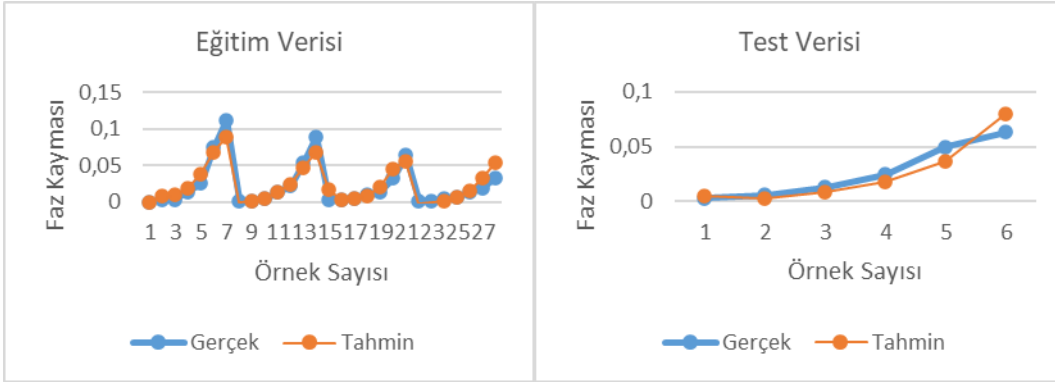
Şekil 4.11: Torbalama Eğitim ve Test Verisi.



**Şekil 4.12:** Linear Regresyon Eğitim ve Test Verisi.



**Şekil 4.13:** Rastgele Alt Uzay Eğitim ve Test Verisi.



**Şekil 4.14:** SMOreg Eğitim ve Test Verisi.

Toplam 34 adet veriden 28 tanesi ( $7 \times 4 = 28$ , tane çapı 45 mm, 106 mm, 250 mm ve 2000 mm olan manyetit minerali-kum karışımı deney sonuçları) eğitimde, geriye kalan 6 tanesi (tane çapı 1000 mm olan manyetit minerali-kum karışımı deney sonuçları) ise testte kullanılmasına karar verilmiştir.

Şekil 4.2 ile Şekil 4.14 arasındaki eğitim verisi şu şekilde açıklanabilir; 1 ile 6. örnek sayısı için tane çapı aynı olan fakat yüzdelik manyetit miktarı örnek sayısı ile doğru orantılı olarak arttırılan karışımı ifade etmektedir. Devam eden örnek sayıları da aynı şekilde 7-14, 15-21 ve 22-27 arası örneklerin tane çapı aynı ve manyetit miktarı örnek sayısı ile doğru orantılı olarak arttırılarak faz kayması karşılıkları gösterilmektedir. Eğitim verisinde de 1-6 arası örnekler tane çapı aynı olan fakat yüzdelik manyetit miktarı arttırılmış karışımı ifade etmektedir. Tablo 4.1’de gösterildiği üzere verideki yüzey miktarı büyüdükçe faz kayması büyüklüğü azalmaktadır. Bu Şekil 4.2 ile Şekil 4.14 arasındaki eğitim ve test veri setindeki maksimum noktaların örnek sayısı arttıkça tepe noktalarının faz kayma değerlerinin daha düşük olmasından da anlaşılmaktadır.

Şekil 4.2 ile Şekil 4.14 arasındaki test ve eğitim verilerinin grafikleri algoritmaların gerçek ve tahmin değerlerinin ne derece uyumlu olduğunu göstermiştir. Test aşamasında kullanılan her bir yöntem için tutarlı olarak hacimsel yüzde arttıkça hatanın da azaldığı, bir başka deyişle kurulan makine öğrenimi algoritmalarının daha da iyi tahminler üretebildiği gözlemlenmiştir.

**Tablo 4.1:** Eğitim Verileriyle Kullanılan Makine Öğrenimi Yöntemleri ve Hata Oranları.

Metot	Yöntem süresi (saniye)	Korelasyon Katsayısı (R)	Ort. Mutlak Hata (MAE)	Ort. Karesel Hatanın Kökü (RMSE)	Bağıl Mutlak Hatanın Kökü (RAE - %)	Bağıl Karesel Hatanın Kökü (RRSE - %)
Random Committee	0,03	0,9999	0,0002	0,0003	0,9002	1,1741
Random Tree	0,02	0,9999	0,0002	0,0004	0,8553	1,3129
KStar	0,05	0,9966	0,0015	0,0027	6,8655	9,2733
Multilayer Perceptron (ANN)	32,58	0,9942	0,0054	0,0062	28,8574	21,3556
Random Forest	0,04	0,9924	0,0028	0,0045	12,9667	15,3996
Decision Table	0,02	0,9884	0,0024	0,0044	11,1658	15,1550
M5 Rules	0,04	0,9796	0,0044	0,0059	20,0813	20,1660
M5P	0,03	0,9745	0,0047	0,0066	21,7449	22,5769
Additive Regression	0,02	0,9723	0,0047	0,0069	21,7846	23,8510
Bagging	0,03	0,9665	0,0053	0,0080	24,5426	27,5682
Linear Regression	0,02	0,9544	0,0069	0,0087	31,8140	29,8532
Random Sub Space	0,03	0,9528	0,0131	0,0180	60,3215	62,0586
SMOreg (SVM)	0,03	0,9525	0,0063	0,0093	29,1080	32,1823

Tablo 4.1 de gösterildiği gibi, farklı makine öğrenimi yöntemlerinin performans metrikleri karşılaştırılmıştır. Korelasyon Katsayısı (R) incelendiğinde kullanılan bütün algoritmaların oldukça yüksek bir korelasyon katsayısına sahip olduğu görülmektedir. Bu da tahmin edilen değerlerin gerçek değerlere çok yakın olduğunun göstergesidir ve en yüksek korelasyon katsayısı, Random Committee ve Random Tree yöntemlerinde gözlemlenmiştir.

Random Tree, Random Committee, Decision Table, MP5, Additive Regression, Bagging, Linear Regression ve Random Sub Space algoritmaları daha hızlı sonuç veren (0,02 veya 0,03 saniye) yöntemlerken Multilayer Perceptron 32,58 saniye ile en uzun süre işlem yapan bir algoritmadır.

Ortalama Mutlak Hata (MAE) veri setinde bulunan gerçek değerler ile tahmin değerleri arasındaki farkın mutlak değerini alarak toplanmasından sonra örnek miktarına bölünmesi ile bulunmaktadır. Ortalama Karesel Hata Karekökü (RMSE) ise bu farkların karelerinin toplamının örnek sayısına bölünmüş halinin karekökünü alarak elde edilmektedir. Random Committee ve Random Tree, en düşük MAE ve RMSE değerlerine sahiptir. Buna karşılık,

Multilayer Perceptron (ANN) ve Random Sub Space algoritmalarının en yüksek hata değerlerine sahip olduğu görülmektedir.

Bağıl Mutlak Hata (RAE %) modelin oluşturduğu toplam mutlak hatanın basit bir referans modelin yaptığı hataya bölünmesiyle hesaplanmaktadır. Modelin performansını ölçmek için kullanılan bir parametredir. Random Tree, en düşük RAE değerine sahip (en yüksek doğruluk), buna karşılık Multilayer Perceptron ve Random Sub Space en yüksek RAE değerine sahiptir.

Bağıl Karesel Hatanın Kökü (RRSE %) hata karelerinin kökünü alarak büyük hataları daha fazla ön plana çıkarmaktadır. Random Committee ve Random Tree en düşük RRSE değerlerine sahiptir. Buna karşılık, KStar ve Multilayer Perceptron, daha yüksek RRSE değerleriyle göze çarpmaktadır.

**Tablo 4.2:** Test Verileriyle Kullanılan Makine Öğrenimi Yöntemleri ve Hata Oranları.

Metot	Süre (saniye)	Maks. % Hata	Min % Hata	Ortala % Hata
Random Committee	0,02	6,7524	1,0932	2,9557
Random Tree	0,01	9,9678	1,2514	4,2856
KStar	0,01	13,1833	1,9127	6,2087
Multilayer Perceptron (ANN)	2,30	18,9711	2,8335	9,2540
Random Forest	0,03	22,1865	4,4155	11,5342
Decision Table	0,01	28,6174	5,9976	14,6666
M5 Rules	0,02	31,8328	7,5796	16,6311
M5P	0,02	35,0482	10,7437	21,3129
Additive Regression	0,01	44,6945	13,9078	25,0753
Bagging	0,01	57,5563	17,0719	29,4840
Linear Regression	0,01	63,9871	20,2360	35,4945
Random Sub Space	0,03	73,6334	24,9822	39,7508
SMOreg (SVM)	0,02	76,8489	26,5643	41,4519

Tablo 4.2’de makine öğrenim algoritmalarının Maksimum % Hata, Minimum % Hata ve Ortalama % Hata değerleri ve çalışma süreleri birbiriyle karşılaştırılmıştır. Random Committee, Random Tree, KStar, Decision Table ve Additive Regression yöntemleri en hızlı çalışan yöntemlerdir (0,01 veya 0,02 saniye). Multilayer Perceptron (ANN) ise en yavaş çalışan yöntemdir (2,30 saniye). En düşük Maksimum % Hata’ya Random Committee sahiptir (6,7524). SMOreg ve Random Sub Space ise en yüksek Maksimum % Hata oranlarına sahiptir (sırasıyla 76,8489 ve 73,6334).

Random Committee ve Random Tree en düşük Minimum % Hata değerleriyle (sırasıyla 1,0932 ve 1,2514) gözlemlenmektedir. En yüksek Minimum % Hata ise %26,5643 ve %24,9822 değerleriyle SMOreg ve Random Sub Space algoritmalarıdır.

En düşük Ortalama % Hata payı Random Committee ‘ye aittir (%2,9557). Sırasıyla %41,4519 ve %39,7508 Ortalama % Hata oranlarıyla en yüksek değerlere SMOreg ve Random Sub space algoritmaları sahiptir.

## 5. TARTIŞMA

Yapılan analizde, eğitim ve test verileri ışığında en iyi performansı sergileyen algoritmaların Random Committee ve Random Tree olduğu görülmektedir. Bu algoritmaların hem hız açısından hem de minimum, maksimum ve ortalama hata değerlerinin düşük olmasıyla ön plana çıktığı tespit edilmiştir. Özellikle, bu iki algoritmanın doğruluk payı ve hız gibi kritik performans göstergelerinde diğer algoritmalara kıyasla üstünlük sağladığı gözlemlenmiştir. Random Committee'nin çeşitli alt modelleri bir araya getirerek hata oranlarını minimize etmesi, bu başarının temelini oluşturduğu düşünülmektedir. Aynı şekilde, Random Tree algoritmasının doğrudan karar ağaçlarını kullanarak hem hızlı hesaplama yeteneği hem de veriyi etkili bir şekilde modelleme kapasitesi, özellikle sınırlı zaman dilimlerinde ve hızlı sonuç gerektiren projelerde büyük avantaj sağlamaktadır.

Öte yandan, MLP yani yapay sinir ağları gibi algoritmaların daha yüksek hata oranlarına sahip olduğu görülmektedir. Ancak bu durum, MLP'nin zayıf olduğu anlamına gelmemektedir. MLP, derin öğrenme tabanlı yöntemlerin karmaşıklığını ve gücünü yansıtan bir algoritma olup, yer yüzeyi gibi karmaşık ortamların çözümü için potansiyel olarak güçlü bir adaydır. Bu tür karmaşık yapıları modellemek için yapay sinir ağları geniş parametreler sunar, ancak aynı zamanda büyük veri setleri üzerinde hesaplama süresi açısından dezavantajlı olabilirler. Veri setinin büyüdüğü durumlarda, MLP'nin işlem süresi önemli ölçüde artış gösterebilir. Bu nedenle, bu algoritmanın kullanılacağı projelerde veri setinin boyutu ve modelin optimizasyonu dikkatlice ele alınmalıdır.

Bir başka önemli bulgu, en yüksek hata oranlarına sahip algoritmaların SMOReg ve Random Sub Space gibi modeller olduğudur. Bu algoritmalar, belirli veri setlerinde düşük performans gösterebilmekte veya daha hassas ayarlamalar gerektirebilmektedir. SMOReg, genellikle yüksek boyutlu verilerde güçlü performans gösterir, ancak küçük ve spesifik veri setlerinde yetersiz kalabilir. Özellikle parametrik ayarlamaların (hiperparametre optimizasyonu gibi) yapılmaması durumunda, bu algoritmaların hata oranlarının yüksek olma eğilimi gösterdiği gözlemlenmiştir. Bu da algoritmaların her veri setine ve probleme uygun olmadığını, özellikle model optimizasyonunun önemini vurgulamaktadır.

Araştırmanın önemli bir diğer noktası, metalik minerallerin temizlenmiş kum taneleriyle homojen bir şekilde karıştırılması ve deneyde başka bir metalik mineral türünün kullanılmaması durumunun, elde edilen yüksek başarı oranlarına önemli bir katkı sağladığı

yönündeki bulgudur. Bu tür ideal koşullar altında algoritmaların yüksek performans göstermesi beklenmektedir. Ancak, makine öğrenimi algoritmalarının tane boyutu, faz farkı ve mineral çeşitliliği gibi daha karmaşık değişkenlerle karşılaştığında nasıl bir performans göstereceği konusunda daha fazla araştırmaya ihtiyaç duyulmaktadır. Bu çalışma, algoritmaların bu tür değişkenlerle olan ilişkisini incelemek için bir başlangıç noktası olarak kabul edilebilir. Gelecekte, metalik minerallerin farklı türlerinin ve tane boyutlarının değiştirilerek yapılacak çalışmaların eklenmesiyle, algoritmaların performansının daha kapsamlı bir şekilde test edileceği öngörülmektedir.

Mevcut literatür incelendiğinde, metalik minerallerin tane boyutu, yüklenebilirlik ve yüzdelik ilişkilerinin çevre kirliliği izleme çalışmalarında da kritik bir rol oynadığı anlaşılmaktadır (Yang ve diğ., 2023; Gurin ve diğ., 2015; Placencia ve diğ., 2014). Özellikle çevre kirliliğinin yeraltı sularına etkisi ve metalik minerallerin yüzey ve yeraltı arasındaki hareketleri konusunda yapılacak izleme çalışmalarında makine öğrenimi algoritmalarının kullanımı karmaşık çevresel süreçleri modelleme potansiyelini iyileştirebilir.

Elde edilen sonuçlar Yokus (2022) araştırmacısının elde ettiği verileri güçlendirebilir, zaman ve maliyet bakımından çalışmasının daha üst düzeye taşınmasına katkı sağlayabilir. Bu çalışma, metalik minerallerin tane boyutu, yüklenebilirlik ve yüzdelik ilişkilerini ortaya çıkartmakta ve makine öğrenimi algoritmalarının çevresel süreçleri anlamak için nasıl kullanılabileceğini test etmede önemli bir adım olarak kabul edilebilir. Algoritmaların mevcut performansları ve potansiyel kullanım alanları, jeofiziksel çalışmalar ve çevresel izleme uygulamaları için umut verici sonuçlar sunmaktadır. İlerleyen dönemlerde yapılacak daha geniş kapsamlı çalışmalar ve farklı veri setleri üzerinde yapılacak optimizasyonlar, algoritmaların çevresel kirlenmenin tespit edilmesinde ve yer kabuğu analizlerindeki katkılarını daha net ortaya koyacaktır.

## **6. SONUÇ VE ÖNERİLER**

Çalışma bulgularının ışığında varılan olumlu sonuçlar makine öğrenimi algoritmalarının karışımdaki manyetit tane boyutu, yüzdelik dilim-faz kayması ilişkisini başarılı bir şekilde öngörebilmiştir.

Başarı oranı en az %95 olan 13 algoritma yukarıda yorumlanmıştır. Yapılan ilk çalışmaların olumsuz sonuçlanması eldeki verinin iyi anlaşılması ve analiz edilmesinin makine öğrenimi algoritmalarının kullanımı için ne kadar önemli olduğunu göstermiştir.

Daha sonrasında eğitim ve test verileri ile ortaya çıkan sonuçlara göre düşük hata oranları ve hızlı çalıştırma süreleriyle, Random Committee ve Random Tree en iyi seçenekler olarak öne çıkmaktadır. Bununla beraber tüm yöntemler oldukça yüksek bir korelasyon katsayısına sahiptir, bu da tahmin edilen değerlerin gerçek değerlere çok yakın olduğunu göstermektedir. Veri kümeleri daha karmaşık tane büyüklükleri ve farklı metalik mineral türlerinden oluştuğu diğer yöntemler de göz önünde bulundurmaya daha tutarlı sonuçlar alınmasına neden olacaktır.

Makine öğrenimi algoritmaları kullanarak veri elde edilmesiyle, SIP Yönteminin Zorlukları kısmında da bahsedilen deney sırasında insan kaynaklı olabilecek hataların da önüne geçebilmek söz konusudur

Yapılan farklı türlerdeki mineral kum karışımı verilerinin işlenmesiyle algoritmalar daha da güçlenebileceği için veri eklenerek öğrenme sürecine katkı sağlanması tavsiye edilmektedir. Elde edilecek daha fazla veri ile bir metalik mineral-faz kayması cevabı tablosu inşa etme durumu SIP yöntemine katkı sağlamak durumundadır. Bu tablonun makine öğrenimi ortamında kurulması ve geliştirilmesi, zamanı ve maliyeti düşürme, insan hatasını en aza indirme gibi kazanımlar sağlayabilmektedir.

## KAYNAKLAR

- Abdel Aal, G.Z., Atekwana, E. and Atekwana, E., 2010. Effect of bioclogging in porous media on complex conductivity signatures. *Journal of geophysical research*, 115, pp.1–11.
- Adebayo, M.B., 2023. Hydrogeophysical monitoring and modeling of subsurface hydrologic response to ecosystem-scale coastal flooding. The University of Toledo.
- Aizebeokhai, A.P., 2014. Assessment of soil salinity using electrical resistivity imaging and induced polarization methods. *African journal of agricultural research*, 9(45), pp.3369-3378.
- Almpanis, A., Slater, L. and Power, C., 2024. Monitoring carbon-based remediation of DNAPL-contaminated groundwater via spectral induced polarization. *journal of environmental management*, 368, 122111.
- Atekwana, E.A. and Atekwana, E.A., 2010. Geophysical signatures of microbial activity at hydrocarbon contaminated sites: A review. *Surveys in geophysics*, 31, pp.247-283.
- Ay, Ş., 2019. Ensemble Learning — Bagging and boosting. Available at: <<https://medium.com/deep-learning-turkiye/ensemble-learning-bagging-ve-boosting-50643428b22b>> [Accessed 14 Jun. 2022].
- Balwant, P.L., Pujari, P., Dhyani, S., Bramhanwade, K. and Veligeti, J., 2022. Geophysical methods for assessing microbial processes in soil: A critical review. *Indian journal of pure and applied physics*, 60(8), pp.707-715.
- Barros, R.C., De Carvalho, A.C. and Freitas, A.A., 2015. Automatic design of decision-tree induction algorithms. *Springer*.
- Bishop, C.M., 2006. Pattern recognition and machine learning. *Springer Science+Business Media, LLC*.
- Blaschek, R., Hördt, A. and Kemna, A., 2008. A new sensitivity-controlled focusing regularization scheme for the inversion of induced polarization data based on the minimum gradient support. *Geophysics*, 73(2), pp. F45-F54.
- Brahmi, S., Baali, F., Hadji, R., Brahmi, S., Hamad, A., Rahal, O., and Hamed, Y. (2021). Assessment of groundwater and soil pollution by leachate using electrical resistivity and induced polarization imaging survey, case of Tebessa municipal landfill, NE Algeria. *arabian journal of geosciences*, 14, 1-13.
- Cao, X., Liu, Z., Hu, C., Song, X., Quaye, J.A. and Lu, N., 2024. Three-dimensional geological modelling in Earth science research: An in-depth review and perspective analysis. *Minerals*, 14(7), pp.686.

- Cao, J., Zhdanov, M., Zu, S., Wang, H., and Sun, J. (2023). Geophysical inversion and interpretation based on new generation artificial intelligence. *Frontiers in earth science*, 11, 1252072.
- Chambers, J. E., Loke, M. H., Ogilvy, R. D., and Meldrum, P. I. (2004). Noninvasive monitoring of DNAPL migration through a saturated porous medium using electrical impedance tomography. *Journal of contaminant hydrology*, 68(1-2), 1-22.
- Chen, Z., Jin, M., Deng, Y., Wang, J. S., Huang, H., Deng, X., and Huang, C. M. (2019). Improvement of a deep learning algorithm for total electron content maps: Image completion. *Journal of geophysical research: space physics*, 124(1), 790-800.
- Cracknell, M.J. and Reading, A.M., 2014. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Computers and geosciences*, 63, pp.22-33.
- Efron, B., 1992. Bootstrap methods: Another look at the jackknife. In: *breakthroughs in statistics: methodology and distribution*, New York, NY: Springer, pp.569-593.
- Flores-Orozco, A. and Bücke, M., 2022. Spectral induced polarization (SIP) imaging for the characterization of hydrocarbon contaminant plumes. In: *Instrumentation and measurement technologies for water cycle management, cham: springer international publishing*, pp.363-386.
- Florin, B. Ş., Alcaz, H. V., Trifonova, P., ve Nalan, Ü. (2014). *Karadeniz deprem güvenliği çalışma ağı projesi ESNET*.
- Gurin, G., Titov, K., Ilyin, Y., and Tarasov, A. (2015). Induced polarization of disseminated electronically conductive minerals: a semi-empirical model. *Geophysical journal international*, 200(3), 1555-1565.
- Işıkhan, E., 2020. Multi Layer Perceptron (MLP) Nedir? Available at: <<https://isikhanelif.medium.com/multi-layer-perceptron-mlp-nedir-4758285a7f15>> [Accessed 14 Jun. 2022].
- Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A., and Kumar, V. (2018). Machine learning for the geosciences: Challenges and opportunities. *IEEE transactions on knowledge and data engineering*, 31(8), 1544-1554.
- Kemna, A., Binley, A., Ramirez, A., and Daily, W. (2000). Complex resistivity tomography for environmental applications. *Chemical engineering journal*, 77(1-2), 11-18.
- Kemna, A., Räckers, E. and Binley, A., 1997. Application of complex resistivity tomography to field data from a kerosene-contaminated site. In: *3rd EEGS Meeting*. European association of geoscientists and engineers, pp.196–199.
- Kessouri, P., Johnson, T., Day-Lewis, F. D., Wang, C., Ntarlagiannis, D., and Slater, L. D. (2022). Post-remediation geophysical assessment: Investigating long-term electrical geophysical signatures resulting from bioremediation at a chlorinated solvent contaminated site. *Journal of environmental management*, 302, 113944.

- Kessouri, P., Furman, A., Huisman, J. A., Martin, T., Mellage, A., Ntarlagiannis, D., and Placencia-Gomez, E. (2019). Induced polarization applied to biogeophysics: recent advances and future prospects. *Near surface geophysics*, 17(6-Recent Developments in Induced Polarization), 595-621.
- Kish, L., 1965. Sampling organizations and groups of unequal sizes. *American sociological review*, pp.564-572.
- Kirmizakis, P., Kalderis, D., Ntarlagiannis, D., and Soupios, P. (2020). Preliminary assessment on the application of biochar and spectral-induced polarization for wastewater treatment. *Near surface geophysics*, 18(2), 109-122.
- Kuhn, M. and Johnson, K., 2013. Over-fitting and model tuning. In: *Applied predictive modeling*, pp.61-92.
- Li, Y. and Oldenburg, D.W., 2000. 3-D inversion of induced polarization data. *Geophysics*, 65(6), pp.1931-1945.
- Li, Z., Meier, M. A., Hauksson, E., Zhan, Z., and Andrews, J. (2018). Machine learning seismic wave discrimination: Application to earthquake early warning. *Geophysical research letters*, 45(10), 4773-4779.
- Lund, A. L., Slater, L. D., Atekwana, E. A., Ntarlagiannis, D., Cozzarelli, I., and Bekins, B. A. (2017). Evidence of coupled carbon and iron cycling at a hydrocarbon-contaminated site from time lapse magnetic susceptibility. *Environmental science and technology*, 51(19), 11244-11249.
- Ma, Z. and Mei, G., 2021. Deep learning for geological hazards analysis: Data, models, applications, and opportunities. *Earth-Science reviews*, 223, 103858.
- Marshall, D.J. and Madden, T.R., 1959. Induced polarization, a study of its causes. *Geophysics*, 24(4), pp.790-816.
- Martinho, E., 2023. Electrical resistivity and induced polarization methods for environmental investigations: An overview. *Water, air, and soil pollution*, 234(4), 215.
- Mensah, V., Boateng, C. D., Gidigas, S. S., Wemegah, D. D., Aryee, J. N., Osei, M. A. and Afful, S. K. (2024). Groundwater Exploration Methods in West Africa: A Review.
- Molua, C.O., 2024. Evaluating the contribution of geophysics to the assessment of natural hazards. *Nigerian journal of theoretical and environmental physics*, 2(2), pp.92-103.
- Moshe, S.B. and Furman, A., 2022. Real-time monitoring of organic contaminant adsorption in activated carbon filters using spectral induced polarization. *Water research*, 212, 118103.
- Moser, C., Binley, A. and Flores Orozco, A., 2023. 3D electrode configurations for spectral induced polarization surveys of landfills. *Waste management*, 169, pp.208-222.
- Nabiyev, V. V. (2012). *Yapay Zekâ: Seçkin Yayıncılık*. 776s, Ankara.
- Ntarlagiannis, D., Yee, N. and Slater, L., 2005. On the low-frequency electrical polarization of bacterial cells in sands. *Geophysical research letters*, 32(24).

Okamoto, S. and Yugami, N., 2003. Effects of domain characteristics on instance-based learning algorithms. *Theoretical computer science*, 298(1), pp.207-233.

Oldenburg, D.W. and Li, Y., 1994. Inversion of induced polarization data. *Geophysics*, 59(9), pp.1327–1341.

Paasche, H. and Tronicke, J., 2007. Cooperative inversion of 2D geophysical data sets: A zonal approach based on fuzzy c-means cluster analysis. *Geophysics*, 72(3), pp.A35-A39.

Painuli, S., Elangovan, M. and Sugumaran, V., 2014. Tool condition monitoring using K-Star algorithm. *Expert Systems with Applications*, 41(6), pp.2638-2643.

Placencia-Gómez, E., and Slater, L. D. (2014). Electrochemical spectral induced polarization modeling of artificial sulfide-sand mixtures. *Geophysics*, 79(6), EN91-EN106.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat, F. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743), 195-204.

Revuelta, M.B., 2017. *Mineral resources: From exploration to sustainability assessment*. Springer.

Rosbach, S., Beaver, C. L., Atekwana, E. A., Enright, A. M., Ntarlagiannis, D., Lund, A., Slater, L. D. (2016, December). Multi-Year analysis of hydrocarbon-degrading microbial communities at the petroleum-contaminated site in Bemidji, Minn. *In AGU fall meeting abstracts* (vol. 2016, pp. ns24a-02).

Saneiyan, S., Ntarlagiannis, D., Werkema Jr, D. D., Ustra, A. (2018). Geophysical methods for monitoring soil stabilization processes. *Journal of applied geophysics*, 148, 234-244.  
Seigel, H.O., 1959. Mathematical formulation and type curves for induced polarization. *Geophysics*, 24(3), pp.547-565.

Shi, W., Rodi, W., Morgan, F.D., 1998. 3-D induced polarization inversion using complex electrical resistivities. In: *11th EEGS symposium on the application of geophysics to engineering and environmental problems*. european association of geoscientists and engineers.

Wemegah, D. D., Fiandaca, G., Auken, E., Menyeh, A., Danuor, S. K. (2014, September). Time-domain spectral induced polarization and magnetics for mapping municipal solid waste deposits in Ghana. *In near surface geoscience 2014-20th european meeting of environmental and engineering geophysics* (Vol. 2014, No. 1, pp. 1-5). European association of geoscientists and engineers.

Wong, J., 1979. An electrochemical model of the induced-polarization phenomenon in disseminated sulfide ores. *Geophysics*, 44(7), pp.1245-1265.

Woodhead, J. and Landry, M., 2021. Harnessing the power of artificial intelligence and machine learning in mineral exploration—opportunities and cautionary notes. *SEG Newsletter*, (127), pp.19-31.

Yang, Y., Zhang, G., Yao, C., Deng, Z., Ren, Z., Li, C. (2023). Application of Induced Polarization Method in Mineral Resource Exploration. *Sustainability*, 15(4), 3840.

Vinegar, H.J., Waxman, M.H., 1984. Induced polarization of shaly sands. *Geophysics*, 49(8), pp.1267-1287.

Yijun, L., Yizhou, Q., Jiarong, L., 2023. Artificial intelligence for satellite communications and geophysics: Current and future trends. *Journal of artificial intelligence practice*, 6(7), pp.43-51.

Yokus, S. Z; Slater, L; Ntarlagiannis, D; Falzone, S, 2022. Using geophysical methods to understand, monitor, and assess a hydrocarbon-contaminated area. Thesis (Master) Rutgers University; Graduate School – Newark.

Zhang, T., Wang, C. J., Liu, S. Y., Zhang, N., Zhang, T. W. (2020). Assessment of soil thermal conduction using artificial neural network models. *Cold regions science and technology*, 169, 102907.

Zhao, T., Wang, S., Ouyang, C., Chen, M., Liu, C., Zhang, J., Wang, L. (2024). Artificial intelligence for geoscience: Progress, challenges and perspectives. *The Innovation*.

## İNTİHAL RAPORU İLK SAYFASI

