

# YAPAY ZEKA YÖNTEMLERİ İLE DÜNYA DEPREMLERİNİN MODELLENEBİLİRLİĞİ

\*<sup>1</sup>Erkan ÖZHAN

<sup>1</sup>Namık Kemal Üniversitesi, Çorlu Meslek Yüksekokulu, 59860 Çorlu, TEKİRDAĞ

## Özet

Depremler, günümüzde insanoğlunun karakteristiğini tam olarak çözemediği nadir olaylardan biridir. Bu çalışmada dünya üzerinde meydana gelen depremlerin karakteristiği yapay zeka yöntemlerinden biri olan makine öğrenmesi ile modellenmeye çalışılmıştır. Modelde kullanılan veriler dünyamızda gerçekleşen 2013 yılına ait magnitüdü 3 ve üzeri olan depremlere aittir. Bu veriler veritabanına aktarılmış ve makine öğrenmesi algoritmalarını içerisinde barındıran Weka yazılımı ile analiz edilmiştir. Analizde makine öğrenmesi algoritmalarından tümü denenmiş ve en yüksek başarıyı ortaya koyan algoritmaların sonuçları ele alınmıştır. Son aşamada ise, 2014 yılına ait gerçekleşen depremler yer bilgisi verilmeden test verisi olarak kullanılmış ve elde edilen sonuçlar sunulmuştur. Depremlerin modellenmesi, o bölgede gerçekleştirilmesi planlanan veya hali hazırda yürüyen birçok işlemin planlanmasında deprem faktörünün etkisinin anlaşılabilirliğini sağlayarak meydana gelebilecek hasarları azaltabilir.

**Anahtar Kelimeler:** Yapay Zeka, Makine Öğrenmesi, Veri Madenciliği, Depremler.

## MODELLABILITY OF THE WORLD'S EARTHQUAKES WITH ARTIFICIAL INTELLIGENCE METHODS

## Abstract

Earthquake is one of the rare incidents characteristics of which humans have failed to develop a full insight into. In this study, the characteristics of the earthquakes happening all over the world were tried to be modelled with the use of machine learning, one of the methods of artificial intelligence. The data used in the model belong to the earthquakes which happened in 2013 and magnitude 3 and above on the Richter Scale. These data were transferred to the database and analysed with Weka software which houses machine learning algorithms. In the analysis, all the machine learning algorithms were tried and the results of the algorithms with the highest performance were dealt with. At the last stage, the earthquakes which happened in 2014 were used as the test data without giving the location information and the results obtained were presented. The modelling of earthquakes can decrease the amount of damage that is likely to happen by providing the comprehensibility of the effects of the earthquake factor in planning the things which are to be done or are currently being done in that region.

**Key words:** Artificial Intelligence, Machine Learning, Data Mining, Earthquakes,

## 1. Giriş

Günümüzde birçok disiplin genelde kendi alanında araştırma ve geliştirme faaliyetleri yürütse de disiplinlerarası işbirliği sağlanması, elde edilen sonuçların çeşitliliği ve verimliliği açısından önemli bir unsur haline gelmiştir. Bu bağlamda bu çalışmada veri madenciliği, yapay zekanın bir

\*Corresponding author: Erkan ÖZHAN Address: Çorlu MYO, Namık Kemal Üniversitesi, 59860, Tekirdağ/TÜRKİYE. E-mail address: eozhan@nku.edu.tr, Phone: +902822503521 Fax: +902822503599

alt kolu olan makine öğrenmesi yöntemleri ile deprem mühendisliğinden elde edilen veriler analiz edilmiştir.

Depremler can ve mal kaybı olarak büyük zararlara neden olan en yıkıcı doğal afetlerden biridir. Depremlerin ortalama her yıl 10.000 kişinin ölümüne neden olduğu ve yıllık ekonomik kayıpların milyarlarca dolara ulaştığı, dolaylı olarak ülkelerin gayri safi milli hasılasının büyük bir kısmını teşkil ettiği bilinmektedir [1]. Araştırmacılar depremlerin oluşmasını önceden tahmin etmek için çeşitli yollar denemişlerdir. Ancak, sonuca henüz ulaşılamamıştır. Bir fay hattı üzerinde gelecekte deprem olup olmayacağı kestirilebilir. Ancak tam olarak nerede ve ne zaman olacağını tespit etmek günümüzün bilimsel verileri ışığında olanaksızdır [2]. Deprem beklentilerinin modellenmesinde bölgesel farklılıklar, parametre belirsizlikleri ve fayların yapısal varyasyonları önemli bir bilimsel sorun haline gelmiştir. Sürekli tekrarlayan çok sayıda deprem olmasına rağmen zaman, büyüklük ve depremin merkezi hakkındaki belirsizlikler hala sürmektedir [3].

Veri madenciliği, verileri incelemeyi, model geliştirmeyi ve daha önce bilinmeyen örüntüler bulmayı sağlayan algoritma çıkarımları içeren veritabanı bilgi keşfi sürecinin çekirdeğidir. Model veriler, yeni verilerin analizi ile öngörü gerektiren olayların anlaşılması için kullanılır [4]. Veri madenciliği, bilgisayarlarda depolanan büyük miktarlardaki verilerin analizini ifade eder. Bunun yanında veri madenciliği genellikle büyük ölçekli veri setlerine uygulanan istatistik veya yapay zeka tekniklerinden faydalanılan analizleri içerir.

İnsan ve bilgisayar etkileşiminin kombinasyonu, grafik araçları ve tanımlamalarla beklenmedik örüntülerin keşfi yapılabilir [5]. Büyük miktardaki verilerin elle ayıklanması, analizinin gerçekleştirilmesi çok zordur. Bu problemlere çözüm bulmak amacıyla makine öğrenmesi yöntemleri geliştirilmiş ve geliştirilmeye de devam etmektedir. Makine öğrenmesi yöntemleri geçmişteki verileri kullanarak yeni verilere en uygun modeli bulmaya çalışır. Yeni gelen verileri de bu modele göre işlerler [6]. Sınıflandırma veri sınıflarını ayırt eden veya bağıntılarını açıklayan bir model (veya fonksiyon) bulma sürecidir. Model, eğitim verilerinin (sınıf etiketleri bilinmektedir yani verilerin sınıfları belirlidir) bir dizi analizine göre elde edilir ve sınıf etiketinin tahmini için kullanılır [7].

## 2. Materyal ve Metot

Analizlerde kullanılan deprem verileri, Amerika Birleşik Devletleri Jeolojik Araştırmalar (USGS-United States Geological Survey)<sup>1</sup> kurumunun web sitesinden alınmıştır. Veriler 2 Ocak 2013 ile 31 Aralık 2013 tarihleri arasında meydana gelen magnitüdü 3 ve üzeri olan dünyada meydana gelen tüm depremleri kapsamaktadır. 185 farklı bölgeye ait veri mevcuttur.

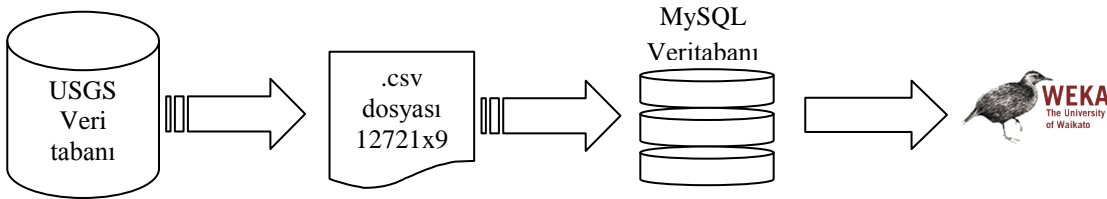
Veri formatı csv (virgüllerle ayrılmış metin belgesi)'dir. USGS tarafından sağlanan depremlere ait tüm öznitelikler alınmıştır. Toplamda 11 öznitelik barındıran ve yukarıda belirtilen tarih aralığına ait 12721 satırlık veri grubu analizlerde kullanılmıştır. Verilerin öznitelikleri ise Tablo 1'de gösterilmiştir. Bu özniteliklerden bazılarının analiz sonuçlarına etkileri bu çalışmanın analiz sonuçları bölümünde belirtilmiştir.

<sup>1</sup> <http://earthquake.usgs.gov/earthquakes/search/> adresinden ulaşılabilir.

**Tablo 1.** Analizde kullanılan verilerin öznitelikleri

Öznitelik Adı	Açıklaması
date	Depremin meydana geldiği tarih . Yıl/Ay/Gün
time	Depremin oluş zamanı. Saat:Dakika:Saniye
latitude	Enlem (Derece)
longitude	Boylam (Derece)
depth	Derinlik (km)
mag	Magnütüd (Depremin Richter Ölçeğine göre büyüklüğü)
magType	Magnütüd türü.
gap	İstasyonla depremin meydana geldiği yerin derece olarak yatay hesaplanan göreceli konumu.
dmin	En yakın istasyonun depremin merkez üssüne derece olarak yatay uzaklığı (1 derece yaklaşık 111,2 km'dir).
Rms	Depremin konumunu hesaplamak için kullanılan hız modelinin değeridir (0,13-1,39 aralığındadır).

Bu veriler Şekil.1'de gösterildiği gibi önce “.csv” dosya formatında indirilmiştir. Daha sonra MySQL veri tabanına aktarılmış ve gerekli değişken tipi ayarları burada yapılmıştır. Verilerin analizinde makine öğrenmesi yöntemlerine ait çok sayıda algoritmayı bünyesinde bulunduran, Weka yazılımı kullanılmıştır. Weka, Yeni Zellanda'da Waikato Üniversitesi'nde geliştirilmiştir ve ismi Waikato Environment for Knowledge Analysis sözcüğünden türetilmiştir. Weka, Java programlama dili ile geliştirilmiş ve genel açık lisans (GPL-General public licence) ile internetten ücretsiz indirilebilmektedir [8]. Weka, makine öğrenmesi algoritmalarını ve veri ön işleme araçlarını organize eden ve bu araçları barındıran bir yazılımdır [4]. Son aşamada veriler, Weka yazılımı ile veritabanı arasında bağlantı kurularak Weka yazılımına aktarılmıştır.

**Şekil 1.** Verilerin analize hazırlık aşamaları

Çalışmada kullanılan deprem verileri, Weka yazılımına aktarıldıktan sonra, Bagging, ViaRegrassion, Decision Table, IBk, J48, Random Forest olmak üzere 6 sınıflandırıcı algoritma ile test edilmiştir. Bu algoritmaların her biri farklı sınıflandırma yöntemleri kullanmaktadır. Sınıflandırma tekniklerinin hepsinin denenmesi için bu algoritmalar tercih edilmiştir. J48 ve Random Forest algoritmaları karar ağacı tekniklerini, Bagging ve ViaRegression algoritmaları meta tekniklerini, Decision Table algoritması kural tabanlı yöntemleri, IBk algoritması ise lazy(tembel) sınıflandırma tekniklerini kullanmaktadır [7]. Veriler bu algoritmalara göre test edildikten sonra elde edilen sonuçlar kayıt altına alınmış ve bu çalışmanın sonuçlar bölümünde gösterilmiştir.

### 3. Analizde Kullanılan Algoritmalar ve Ölçekler

#### 3.1 Bagging Algoritması

Yapay zeka tekniklerinde bilgi kazanımı için Bagging, Boosting gibi makine öğrenmesi ve meta öğrenme tekniklerini kullanan algoritmalar kullanılabilir [9]. Bagging algoritması eğitim verisi setlerinin küçük boyutları ile çalışır. Özgün eğitim seti N adet alt kümelere bölünmüştür. Bu alt kümelerin her biri eğitim seti olarak kullanılır. Her bir alt küme aynı zamanda bir sınıflandırıcı oluşturur. Bu sınıflandırıcıları bir birleşik sınıflandırıcı bileştirir [10]. Bu nedenle Bagging (torbalama) ismi verilmiştir.

#### 3.2 viaRegression Algoritması

Bu model, ağaç yaprakları gibi uç sınıflar yerine lineer regresyon fonksiyonlarını ile karar ağacının yapısını oluşturur. Lineer regresyon fonksiyonları karar ağaçlarının bir türüdür [11].

#### 3.3 Decision Table Algoritması

Bu algoritma sınıflandırma için bir karar tablosu oluşturur. Eğitim verilerinin öz niteliklerine göre ortaya çıkan bu karar tablosundan faydalanarak sınıflandırma yapar [12].

#### 3.4 IBk Algoritması

IBk, -kNN algoritması olarak ta bilinir. En yakın komşu yöntemine “tek bağlantı kümeleme yöntemi” ismi de verilmektedir. İlk anda tüm gözlem değerleri birer küme olarak değerlendirilir. Aşamalı olarak bu kümeler birleştirilerek yeni kümeler elde edilir. Bu yöntemde öncelikle gözlemler arasındaki mesafe belirlenir [13]. K-en yakın komşu (k-nearest neighbour - kNN) algoritması benzerlik yoluyla öğrenme teknikleri kullanan algoritmalarından biridir. Weka'da kNN algoritmasına IBk adı verilmiştir [14].

#### 3.5 J48 Algoritması

Bu algoritma karar ağacı tekniklerini kullanır. Yukarıdan aşağıya doğru böl ve yönet şeklinde bir ağaç yapısı oluşturur. Öznitelikler bu algorithmada düğüm noktası oluşturacak şekilde yerleştirilir ve eğitim verisine göre yapraklar meydana gelir. Yapraklar aynı zamanda sınıf etiketlerini belirtir [15].

#### 3.6 Random Forest Algoritması

Random Forest çok sayıda karar ağacı oluşturur. Karar ağacının ağırlığına göre algoritma bir çıkış üretir. Herbir ağaç, test verilerini sınıflandırırken bir oran verir. Algoritma, en büyük orana sahip olan karar ağacına göre test verisini sınıflandırma işlemini yapar [16].

### 3.7 Kappa Deęeri

Gözlemler arasındaki uyuşmayı ölçmek için kappa istatistięi (veya kapa katsayısı) en sık kullanılan yöntemdir. Kappa gözlemler arasındaki uyuşmanın büyüklüğünü nicel olarak göstermek için tasarlanmıştır. Kappa deęeri -1 ile 1 arasında deęişen bir ölçüdür [17]. 1 e yaklaştıkça uyuşma oranı artar.

### 3.8 Correctly Classified Instances

Bu deęer sınıflandırıcı algortimaların eğitim verilerini aynı zamanda test verisi olarak kullanıp yaptıklarını doęru ve yanlış sınıflandırmanın oranını gösteren deęerdir.

### 3.9 Root Mean Squared Error(RMSE)

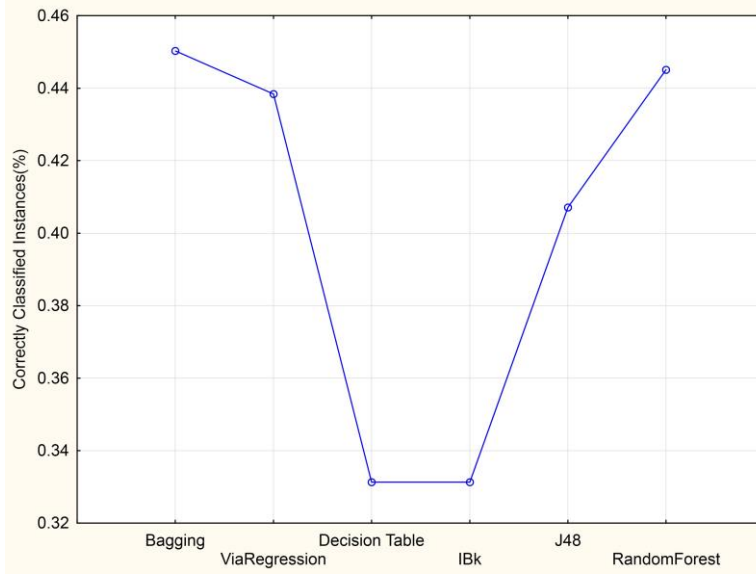
Ortalama karesel hatanın karekökü olarak tanımlanan bu deęer gözlenen deęerler ile gerçek deęerler arasındaki hata oranını ifade etmek için kullanılır [8]. Literatürde root mean square deviation olarak da bilinmektedir. Bu deęerin küçülmesi istenir çünkü sınıflandırıcın gözlenen deęerlerle gerçek deęerleri doęru sınıflandırması bu deęerin azalması ile aęırlık kazanır.

### 3.10 F-Ölçüsü

$F_1$  skoru veya F skoru olarak da adlandırılır. Verilerin sınıflandırılmasında hassasiyet(Precision) ve getirim(recall)'in harmonik ortalamasıdır [7]. Bu deęer maksimum 1 olabilir. İlgili sınıfın, sınıflandırıcı algoritma tarafından ne derece doęru sınıflandırma tahmini yapabileceğini gösterir. Deęer 1'e yaklaştıkça ilgili sınıfın algoritma tarafından öğrenilmesi o derece başarılı anlamına gelir.

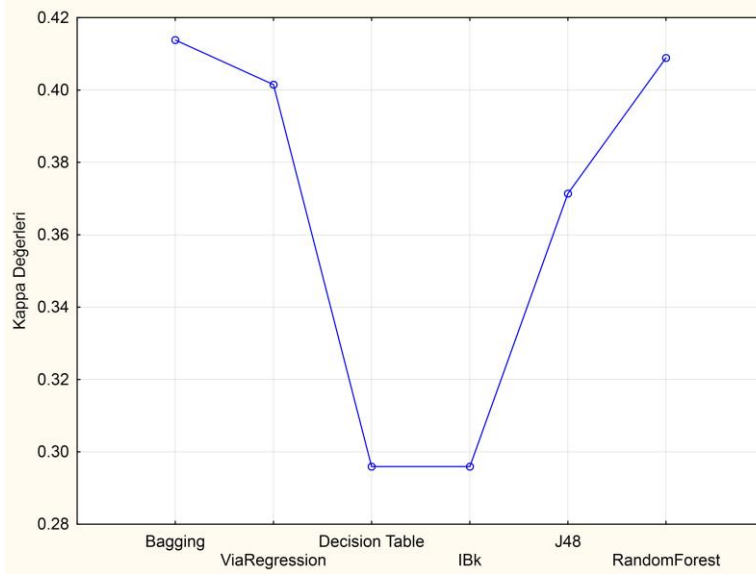
## 4. Analiz Sonuçları

Kullanılan 6 sınıflandırıcı algoritmanın Correctly Classified Instances deęerleri Şekil 2'de gösterilmiştir. Şekil 2'deki sonuçlara bakıldığında Bagging ve RandomForest algoritmalarının en yüksek doęru sınıflandırılan örnek oranına sahip oldukları görülür. Decision Table ve IBk algoritmaları ise en düşük doęru sınıflandırma oranına sahiptir. ViaRegression ve J48 algoritmaları ise üçüncü ve dördüncü sırada doęru sınıflandırma oranına sahip algoritmalar olarak görülmektedir.



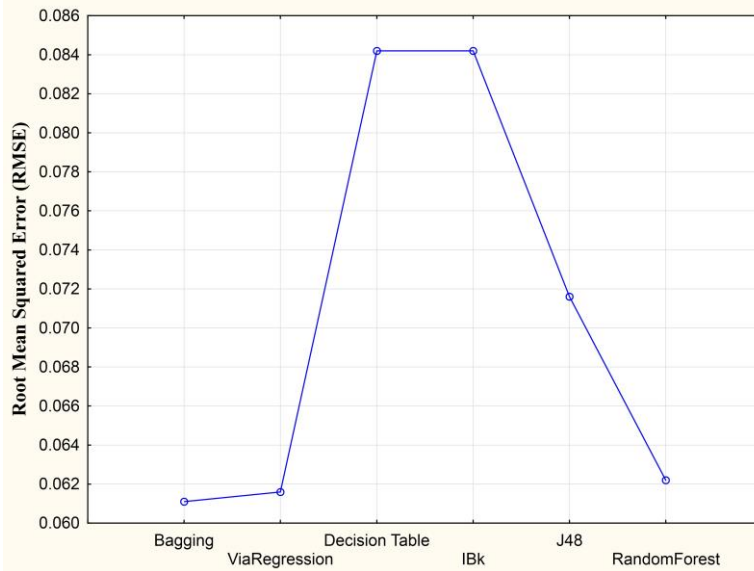
Şekil 2. 6 sınıflandırıcı için Correctly Classified Instances değerleri

Analizlerden elde edilen bir diğer sonuç ise kappa değerleridir. Şekil 3'te kappa değerleri gösterilmiştir.



Şekil 3. 6 sınıflandırıcı için kappa değerleri

Şekil 3'e bakıldığında Correctly Classified Instances değerlerine paralel bir görünüm olduğu farkedilir. Kappa değeri doğru sınıflandırılan örnek sayısı oranı ile doğru orantılıdır. Şekil 3'te görüldüğü gibi Bagging ve RandomForest algoritmaları kappa değeri en yüksek çıkan algoritmalarıdır. Decision Table ve IBk algoritmalarının kappa değerleri ise en düşüktür.



Şekil 4. 6 sınıflandırıcı için RMSE değerleri

Analizlerden elde edilen sonuçlardan biride RMSE (Root Mean Squared Error) değerleridir. Bu değerler Şekil 4'te gösterilmiştir. Şekil 4'te görüldüğü gibi RMSE değeri en düşük çıkan algoritma Bagging algoritmasıdır. Ancak RMSE değeri diğer değerlerde olduğu gibi RandomForest algoritması için beklendiği gibi çıkmamıştır. RandomForest algoritması en düşük RMSE oranına sahip üçüncü algoritma olmuştur. En düşük RMSE oranına sahip ikinci algoritma ViaRegression algoritması olmuştur.

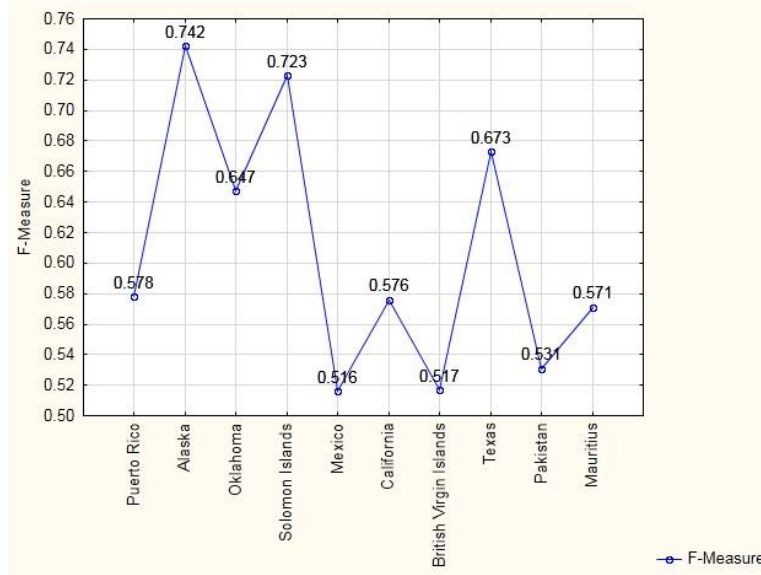
Bu sonuçlara göre deprem verilerinin analiz edilmesinde ve modellenmesinde kullanılabilir en uygun algoritmanın Bagging algoritması olduğu sonucuna varılmıştır. Bu algoritmaya göre deprem verileri analiz edilmiştir. Veriler için sınıf olarak location(yer) bilgisi seçilmiştir. 185 adet sınıf bulunmaktadır. Ancak bu seçim, verilerin öznelikleri arasında yer alan latitude (enlem) ve boylam (longitude) niteliklerinin location niteliği ile doğrudan ilişkili olduğu F-ölçüsü değerlerinden anlaşılmış ve analizlerden çıkarılmıştır. Bagging algoritmasına göre deprem verileri analiz edildiğinde sınıfların F-ölçüsü değerleri elde edilmiştir. Bu değerler 185 sınıf için alınmış ve yaklaşık 7 sayfa tutmaktadır. Bu nedenle sadece F-Measure değeri 0.5 ve üzeri olan depremler ve bunların sayıları Tablo 2'de gösterilmiştir.

Tablo 2. F değeri 0.5 ve üzeri olan depremler ve veri seti sayıları

Sınıf adı (Class name)	Veri Seti Sayısı	F-Değeri
Alaska	1781	0.742
British Virgin Islands	311	0.517
California	579	0.576
Mauritius	5	0.571
Mexico	413	0.516
Oklahoma	301	0.647
Pakistan	94	0.531

Puerto Rico	419	0.578
Solomon Islands	685	0.723
Texas	55	0.673

Tablo 2’de görülen F ve örnek sayıları arasındaki ilişkiye bakıldığında ise veri seti sayısı ile F değeri arasında doğrudan bir ilişki olmadığı görülmüştür. Örneğin 55 veri setine sahip Texas sınıfı için F değeri 0.673 iken, 579 veri setine sahip California sınıfının F değeri 0.576 çıkmıştır. Şekil 5’te F değerlerinin dağılımı gösterilmektedir.



Şekil 5. F değeri 0.5 ve üzeri olan sınıflar.

Şekil 5’te görüldüğü gibi F değeri en yüksek çıkan sınıf Alaska sınıfıdır. En düşük F değerine sahip sınıf ise British Virgin Islands’dir.

Bagging algoritmasına göre üretilen model kaydedilmiş ve 2014 yılına ait olan ancak eğitim verileri içerisinde yer almayan Alaska sınıfına ait 100 test verisi ile test edilmiştir. Verilen test verilerinin hepside Alaska sınıfına aittir ve 64 tanesi başarılı bir şekilde sınıflandırılmıştır. Test verilerinin %64’ü başarılı bir şekilde sınıflandırılmıştır.

## 5. Sonuç ve Öneriler

Sonuç olarak deprem verilerinin bazılarının yapay zekanın bir alt kolu olan makine öğrenmesi yöntemleri ile modellenilebileceği görülmüştür. Bu modellenme için Kappa, Correctly Classified Instances ve RMSE değerine göre en uygun algoritma Bagging olabilir. F- değeri 0.5 ve üzeri olan sınıflar(ki depremin yeridir)’ın modellenilebilir oldukları söylenebilir. Buradan hareketle modellenilebilen depremlerin belirli bir karakteristiğe sahip oldukları söylenebilir. Bir yıllık veriler ışığında bu çalışmada Alaska, Solomon Islands, Texas, Oklahoma, Puerto Rico, California, Mauritius, Pakistan, British Virgin Islands, Mexico bölgelerindeki depremlerin modellenilebileceği



görülmüştür. Bu sayede ilgili bölgelerde gerçekleştirilecek faaliyetlerin planlanmasında bu durum dikkate alınarak yerel-genel yönetimlerin faaliyetleri, insanların en az zararı görebilmeleri için alınacak önlemlerin planlanması gibi birçok açıdan fayda sağlanabilir.

Gelecekte araştırmacılar, daha geniş bir zaman dilimini içine alan veriler ile bu analizleri tekrarlayabilirler. Modellenen depremlerin yerel ve daha geniş verilerine ulaşarak o bölge için modelin kapsamını ve sunabileceği imkanları artırabilirler.

### **Teşekkür**

Deprem verilerinin elde edilmesinde sağladıkları online hizmetlerden dolayı Amerika Birleşik Devletleri Jeolojik Araştırmalar (USGS- United States Geological Survey) Kurumuna teşekkür ederim.

### **Referanslar**

- [1]. Elnashai AS, Sarno LD. Fundamentals of Earthquake Engineering İngiltere: John Wiley&Sons, Ltd.; 2008.
- [2]. Canbolat BB, Erberik A, Tüydeş H, Karaesmen E. Deprem Ankara: ODTÜ Yayıncılık; 2008.
- [3]. Ozerdem MS, Ustundag B, Demirer MR. Self-organized maps based neural networks for detection of possible earthquake precursory electric field patterns. Advances in Engineering Software. 2006 Nisan; 37(4).
- [4]. Maimon O, Rokach L, editors. Data Mining and Knowledge Discovery Handbook. 2nd ed. London,: Springer Science+Business Media; 2010.
- [5]. Olson DL, Delen D. Advanced Data Mining Techniques. 1st ed. Berlin, Almanya: Springer-Verlag; 2008.
- [6]. Amasyalı FM. Yeni Makine Öğrenmesi Metotları ve İlaç Tasarımına Uygulamaları. Doktora Tezi. İstanbul: Yıldız Teknik Üniversitesi, Bilgisayar Mühendisliği; 2012.
- [7]. Han J, Kamber M. Data Mining Concepts and Techniques. 3rd ed. San Francisco,: Morgan Kaufmann Publishers is an imprint of Elsevier; 2006.
- [8]. Witten IH, Frank E, Hall MA. Data Mining Practical Machine Learning Tools and Techniques. 3rd ed. Burlington, USA,: Morgan Kaufmann Publishers is an imprint of

Elsevier; 2011.

- [9]. Kolandaivel P. A Novel Earthquake Mitigation Information expert System: EMIS. In Electrical and Computer Engineering, 2006. CCECE '06. Canadian Conference on; 2006; Ottawa. p. 337-340.
- [10]. Machova K, Barcak F, Bednar P. A Bagging Method using Decision Trees in the Role of Base Calssifiers. Acta Polytechnica Hungarica. 2006; 3(2).
- [11]. ARDIL E. Esnek Hesaplama Yaklaşımı İle Yazılım Hatası Kestirimi. Yüksek Lisans Tezi. Edirne: Trakya Üniversitesi, Bilgisayar Mühendisliği; 2009.
- [12]. Gandhi MG, Srivatsa SK. Classification Algorithms in Comparing Classifier Categories to Predict the Accuracy of the Network Intrusion Detection – A Machine Learning Approach. Advances in Computational Sciences and Technology. 2010; 3(3).
- [13]. Özkan Y. Veri Madenciliği Yöntemleri. 1st ed. Çölkesen R, editor. İstanbul, Türkiye: Papatya Yayıncılık; 2008.
- [14]. Güner ES. Türkçe İçin Derlem Tabanlı Bir Anafor Çözümleme Çalışması. Tech. rep. Edirne,: Trakya Üniversitesi Fen Bilimleri Enstitüsü; 2008.
- [15]. Kargupta H, Han J, Yu PS, Motwani R, Kumar V. Next Generation of Data Mining Minnesota: Chapman & Hall/CRC; 2008.
- [16]. Dua S, Du X. Data Mining and Machine Learning in Cybersecurity. 1st ed. Boca Raton, FL, USA: Auerbach Publications Taylor & Francis Group; 2011.
- [17]. Viera AJ, Garrett JM. Understanding Interobserver Agreement: The Kappa Statistic. Society of Teachers of Family Medicine. 2005 Mayıs; 37(5): p. 360-363.