# K-Ortalamalar Tabanlı Zaman Serisi Ağırlıklandırma Yöntemi ile EEG Sinyallerinden Epilepsi Tahmini

# K-Means Clustering Based Time Series weighting with Epileptic Seizure Detection

Muhammet Sinan Başarslan[1], Fatih Kayaalp[1]
Bilgisayar Mühendisliği Bölümü
Düzce Üniversitesi
Düzce, Türkiye
{muhammetsinanbasarslan, fatihkayaalp}@duzce.edu.tr

Kemal Polat[2]
Elektrik elektronik Mühendisliği Bölümü
Abant İzzet Baysal Üniversitesi
Bolu, Türkiye
kpolat@ibu.edu.tr

*Özetçe*—**Epilepsinin tespiti ve epileptik atakların sınıflandırılmasında, beynin elektriksel aktivitesi önemli bir veri kaynağı olarak kullanılmaktadır. Bu çalışmada, EEG sinyallerinden epilepsi hastalığının otomatik olarak sınıflandırılmasında veri kümesi UCI makine öğrenmesi veri deposundan alınmıştır. Bu veri kümesi ham zaman domeni EEG sinyallerinden oluşmaktadır. Epilepsi hastalığının yanı sıra dört farklı sınıf daha bulunmaktadır. Bu sınıflar, gözler açıkken kaydedilen EEG, gözler kapalı iken kaydedilen EEG, tümör bölgesine sahip kişilerden kaydedilen EEG ve sağlıklı kişilerden kaydedilen EEG olmak üzere toplam beş sınıftan oluşmaktadır. Epilepsi durumunu diğer sınıflardan ayırt etmek için zaman domeni EEG sinyallerinden herhangi bir öznitelik çıkarımı yapmadan sadece ham EEG sinyalleri üzerinden sınıflama yapılmıştır. Bu çalışmada, beş sınıflı epilepsi veri kümesini yüksek doğrulukla sınıflandırmak için k-ortalamalar kümele tabanlı zaman serisi ağırlıklandırma (KOKTZSA) yöntemi ön-işleme olarak ham EEG sinyallerine uygulanmış ve daha sonra ağırlıklandırılmış olan veri kümesini sınıflandırmak için Random Forest (rastgele orman) ve C4.5 karar ağacı sınıflama algoritmaları kullanılmıştır. Ham EEG sinyali, C4.5 karar ağacı sınıflama algoritması %0,705 sınıflama doğruluğu elde ederken, KOKTZSA ile ağırlıklandırılmış olan veri kümesi C4.5 karar ağacı ile %0,968 sınıflama doğruluğu elde etmiştir. Ham EEG sinyali, random forest (rastgele orman) sınıflama algoritması %0,81 sınıflama doğruluğu elde ederken, KOKTZSA ile ağırlıklandırılmış olan veri kümesi random forest sınıflandırma algoritması ile %0,993 sınıflama doğruluğu elde etmiştir. Elde edilen sonuçlar, önerilen hibrid model ile EEG sinyalinden herhangi bir öznitelik çıkarımı yapmadan yüksek bir sınıflama doğruluğu elde ettiğini göstermiştir. Bu durum da yüksek hesaplama maliyetini büyük bir oranda azaltmıştır.**

***Anahtar Kelimeler—epileptik nöbet tespiti; elektroensefalografi; k ortalamalar kümeleme tabanlı öznitelik ağırlıklandırma; sınıflandırma.***

*Abstract*— **In detecting epilepsy and classifying epileptic attacks, the electrical activity of the brain is used as an important data source. In this study, the data set for automatic classification of epilepsy from EEG signals was taken from the UCI machine learning data repository. This dataset consists of raw time domain EEG signals. Apart from epilepsy, there are four different classes. These classes consist of a total of five classes, EEG recorded when the eyes are open, EEG recorded when the eyes are closed, EEG recorded from people with the tumor zone, and EEG recorded from healthy people. To differentiate the epileptic condition from the other classes, only the raw EEG signals were categorized without any feature extraction from the time domain EEG signals. In this study, k-averages cluster-based time series weighting (KOKTZSA) method was applied to raw EEG signals as pre-processing to classify the five-class epilepsy data set with high accuracy and then used to classify the weighted data set in Random Forest and C4.5 decision tree classification algorithms have been used. The raw EEG signal obtained a classification accuracy of 70.55% for the C4.5 decision tree classification algorithm and 96.86% for the data cluster C4.5 decision tree weighted by the KOKTZSA. Raw EEG signal, random forest classification algorithm obtained 81% classification accuracy while data set weighted with KOKTZSA achieved 99.33% classification accuracy with random forest classification algorithm. The obtained results show that the proposed hybrid model achieves a high classification accuracy without extracting any features from the EEG signal. This has greatly reduced the high computational cost.**

***Keywords—epileptic seizure detection; electroencephalography; k averages clustering-based feature weighting; classification.***

## I. GİRİŞ

Beynimiz sinir hücresinden oluşmaktadır. Beynimizi oluşturan sinir hücreleri arasında hiç kesilmeyen elektiriksel bir iletişim vardır. Epilepside, Beyin sinir hücrelerinin normal çalışması dışında hücreler arası elektriksel iletişimde oluşan

978-1-5386-5135-3/18/$31.00 ©2018 IEEE

kontrol yayılım sonucu ortaya çıkan bir hastalıktır. Bu kontrolsüz yayılım sonucu, bir uyarı olmadan süregelen geçici bilinç kaybına sebep olur.

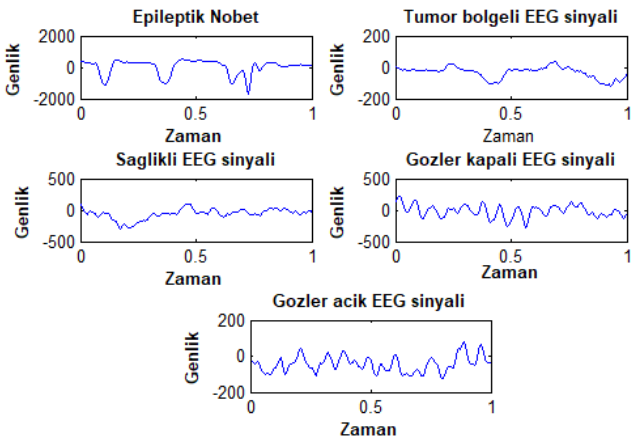
Elektroensefalografi (EEG), insan beyninin sinir hücreleri arası elektriksel iletişim ölçmede kullanılan bir yöntemdir. EEG yöntemi dünya nüfusunun %1'ini etkileyen epilepsi hastalığını teşhis etmede kullanılır [1]. İnsanlar günlük hayatlarını beyin içerisinde yer alan sinir hücrelerinin elektriksel iletişimi sayesinde devam ettiriler. Epilepsi hastalığı, beyinde yer alan sinir hücrelerinin anormal aktivitesinin sebep olduğu istemdışı vücut hareketleri, bilinç kaybı, kaslardaki kasılmalar, sıradışı duygusallık gibi kronik bir hastalık olarak tanımlanmaktadır [2]. EEG, bu hastalığı teşhiş etmede kullanılan en önemli yöntemlerden biridir [3].

Bu çalışmada, UCI makine öğrenmesi veri deposundan alınan Epileptik Nöbet Tanıma veri setinde yer alan (Epileptic Seizure Recognition Data Set) EEG sinyallerinden epilepsi hastalığını tahmin etmek için sınıflandırma işlemi yapılmıştır. Bu veri setinde epilepsi hastalığı, gözler açıkken kaydedilen EEG, gözler kapalı iken kaydedilen EEG, tümör bölgesine sahip kişilerden kaydedilen EEG ve sağlıklı kişilerden kaydedilen EEG olmak üzere toplam beş sınıfta yer almaktadır. Epilepsi durumunu diğer sınıflardan ayırt etmek için zaman domeni EEG sinyallerinden herhangi bir öznitelik çıkarımı yapmadan sadece ham EEG sinyalleri üzerinden sınıflama yapılmıştır.

Bu çalışmada, beş sınıflı epilepsi verilerini yüksek başarım ile sınıflandırmak için k-ortalamalar kümele tabanlı zaman serisi ağırlıklandırma (KOKTZSA) yöntemi, ön-işleme olarak ham EEG sinyallerine uygulanmıştır. Ham EEG sinyallerinin ağırlıklandırılmasından sonra epilepsi hastalığını tahmin etmek için sınıflandırma algoritmaları kullanılmıştır. Çalışmada kullanılan sınıflandırma yöntemleri; C4.5 karar ağacı ve Random Forest (rastgele orman) sınıflama algoritmaları kullanılmıştır. Oluşturulan bu sınıflandırma modellerinin performanslarını test etmek için 10 kat çapraz geçerleme yöntemi kullanılmıştır. Epileptik Nöbet Tanıma veri setinin 10 kat çapraz geçerleme ile eğitim ve test küme ayrımı sonrası sınıflandırma algoritmalarında en iyi sonucu ağırlıklandırma işlemine tabi tutulan modeler verdiği görülmüştür. Çalışma kapsamında ham veri kümesi ve ağırlıklandırılmış veri kümesi üzerinde oluşturulan sınıflandırma modellerinin performanslarına sonuçlar kısmında yer verilmiştir. Ham EEG sinyallerinden epilepsi durumunu tespit etmek için k-ortalamalar kümeleme tabanlı zaman serisi ağırlıklandırma yöntemi ile sınıflandırma algoritmaları birleştirilerek yeni bir hibrid model oluşturulmuştur.

## II. EPİLEPTİK NÖBET TESPİTİ VERİ SETİ

Çalışmada kullanılan EEG epilepsi veri kümesi, UCI makine öğrenmesi veri deposundan alınmıştır [4]. Bu veri seti toplam beş sınıftan oluşmaktadır ve her sınıf 23.6 saniye süreli 100 tek kanal EEG segmenti içermektedir. Veri setinde bulunan beş sınıf şunlardır; epilepsi hastalığı, gözler açıkken kaydedilen EEG, gözler kapalı iken kaydedilen EEG, tümör bölgesine sahip kişilerden kaydedilen EEG ve sağlıklı kişilerden kaydedilen EEG sınıflarıdır. Şekil 1' de EEG sinyallere ait zaman serisi görülmektedir.

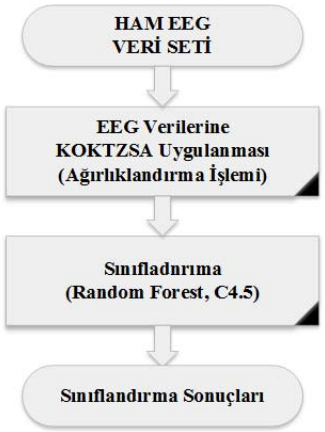



*Şekil 1*: EEG sinyallere ait zaman serisi

Beşinci ve dördüncü sınıf değeri sırasıyla gözler açıkken ve kapalıyken sağlıklı olan beş gönüllüden elde edilen yüzey EEG kayıtlarını içermektedir. Üçüncü sınıf değeri, hasta olan gönüllülerden nöbet öncesinde alınan sinyaldir. İkinci sınıf değeri epileptojenik bölgeden elde edilmiştir. Son sınıf olan birinci sınıf değeri ise hasta gönüllülerin kriz sırasındaki ölçümleridir. Her bir sınıftan 2300 adet veri bulunmaktadır. Toplam veri sayısı ise 11500 dir.

## III. METOT

Ham EEG sinyallerinden epilepsi durumunu tespit etmek için k-ortalamalar kümeleme tabanlı zaman serisi ağırlıklandırma yöntemi ile sınıflandırma algoritmaları birleştirilerek yeni bir hibrid model oluşturulmuştur. Önerilen hibrid modelin şematik olarak gösterimi Şekil 2' de verilmektedir.




*Şekil 2:* Önerilen hibrid modelin şematik gösterimi

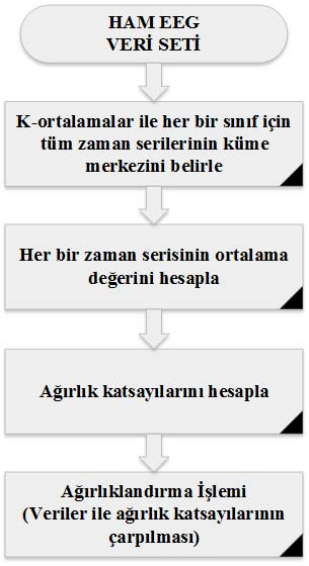
### A. K-ORTALAMALAR KÜMELE YÖNTEMİ

K-ortalamala kümele (KOK), en fazla kullanılan kümeleme yöntemlerden biridir. Veri madenciliği, nesne sınıflandırma gibi bilgisayar tabanlı uygulamalarda ayrıca iktisat, pazarlama, biyoinformatik gibi alanlarda kullanılır [5].

En çok tercih edilen gözetimsiz öğrenme yöntemlerinden biridir. K-means'in atama işleyişi her veri sadece bir kümeye sahip olması şeklindedir [6]. Bundan dolayı sert bir kümeleme algoritmasıdır. K-means algoritması n adet veri kümesinden oluşan veri setini, giriş parametresi olarak alınan k adet kümeye böler. Bunun amacı bölümleme işlemi sonucunda ortaya çıkan kümelerin kendi içlerinde benzerliklerini en yüksek seviyeye çıkararak kümeler arası benzerlikleri ise en alt seviyede tutmaktır [7].

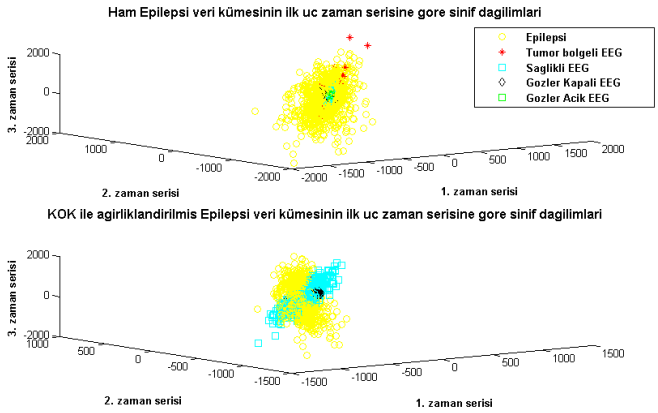
#### *1) K-Ortalamalar Kümele Tabanlı Zaman Serisi Ağırlıklandırma Yöntemi (KOKTZSA)*

Bu ağırlıklandırma yöntemi Gunes ve arkadaşları tarafından 2010 yılından önerilmiştir [8]. Araştırmacılar, bu önerilen metodu öznitelikleri elde edilmiş veri kümelerine uygulamışladır. Bu çalışmada ise, ham EEG sinyallerinden herhangi bir öznitelik çıkarımı yapmadan EEG zaman serisi verilerine KOKTZSA yöntemi uygulanmıştır. Yöntem kısaca şu şekilde çalışmaktadır: k-ortalama yöntemi kullanarak her bir zaman serisinin küme merkezi bulunmuştur. Daha sonra bu zaman serilerinin her birinin ortalama değeri hesaplanmıştır. Her bir sınıf için bütün zaman serilerindeki ortalama değerin küme merkezlerine oranı her bir zaman serisinin ağırlık katsayısı olarak alınmıştır. Bu hesaplanan ağırlık katsayıları her bir sınıftaki bütün zaman serileri ile çarpılarak veri kümesi ağırlıklandırılmıştır. Bu yöntemin akış şeması şekil 3'te görülmektedir.




*Şekil 3*: K-ortama kümeleme yöntemi ile ağırlıklandırma işlemi [8]

K-ortalama kümeleme işlemi öncesi ve sonrası Epilepsi veri kümesinin üç zaman serisine göre sınıf dağılımları şekil 4'te görülmektedir.




*Şekil 4: KOKTZSA* öncesi ve sonrası Epilepsi veri kümesinin üç zaman serisine göre sınıf dağılımları

### B. Çalışmada Kullanılan Sınıflandırma Algoritmaları

Ham EEG sinyallerinin ağırlıklandırılmasından sonra epilepsi hastalığını tahmin etmek için sınıflandırma algoritmaları kullanılmıştır. Çalışmada kullanılan sınıflandırma yöntemleri şunlardır; C4.5 karar ağacı ve Random Forest (rastgele orman) sınıflama algoritmalarıdır.

#### *1) Random Forest (rastgele orman) Algoritması*

Leo Breiman tarafından, sadece tek bir karar ağacı oluşturmak yerine her biri değişik eğitim veri kümelerinde eğitime tabi tutulmuş olan çok değişkenli ağacın kararlarını birleştirmek amacıyla geliştirilmiştir [6].

#### *2) C4.5 Karar Ağacı Algoritması*

C4.5 algoritması, Ross Quinlan tarafından 1993 yılında geliştirilmiştir. C4.5 karar ağacında, Kazanım Oranı (Gain Ratio) kullanılır. C4.5 algoritması hem kategorik hem de nümerik değerlikli nitelikler ile çalışabilmektedir [7].

### C. Model Performans Değerlendirme Kriterleri

Sınıflandırma algoritmaları ile oluşturulan modelin değerlendirilmesi çeşitli yöntemlerle yapılır. Bu yöntemlerden biri karışıklık matrisidir (confusion matrix) [8]. Gerçek değerler ve sınıflandırma algoritması ile tahmin edilen değerler Tablo 1'de gösterilmiştir. Tablo 1'ye göre sınıflandırma algoritmalarına yönelik performans değerlendirme ölçütleri aşağıda verilmiştir [8-10].

*Tablo 1*: %60 Karışıklık matrisi

| | | TAHMİN | |
|---|---|---|---|
| | | DOĞRU | YANLIŞ |
| Gerçek | DOĞRU | DD | DY |
| | YANLIŞ | YD | YY |

Tablo 1'e göre sınıflandırma algoritmaları ile oluşturulan modelin doğruluk Denklem (1)'de görülmektedir [10].

$$Doğruluk\ (Acc) = \frac{DD + YY}{DD + DY + YD + YY} \quad (1)$$

Tablo 1'e göre sınıflandırma algoritmaları ile oluşturulan modelin kesinlik ve duyarlılık değeri sırasıyla Denklem (2) ve Denklem (3)'te görülmektedir [10].

$$Kesinlik\ (Kes) = \frac{DD}{DD + YD} \quad (2)$$

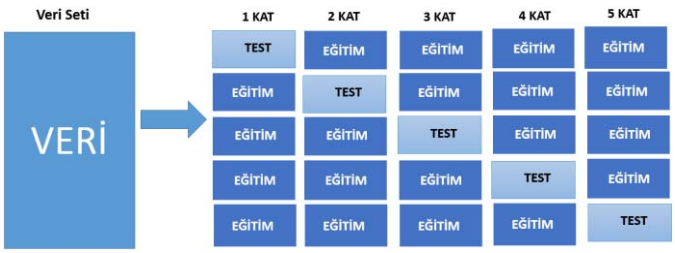
$$Duyarlılık\ (Duy) = \frac{DD}{DD + DY} \quad (3)$$

Tablo 1'e göre sınıflandırma algoritmaları ile oluşturulan modelin özgünlük ve F-ölçü değeri değeri sırasıyla Denklem (4) ve Denklem (5)'da görülmektedir [10].

$$Özgünlük\ = \frac{YY}{YY + YD} \quad (4)$$

$$F - ölçü\ (F) = \frac{2 \times Duyarlılık \times Kesinlik}{Duyarlılık + Kesinlik} \quad (5)$$

Veri setini üzerine tahmin işlemi yapmak için sınıflandırma algoritmaları modeller oluşturulur. Oluşturulan sınıflandırma modellerinin performanslarını görmek için sınıflandırma modelleri eğitim ve test verisi olarak ikiye bölünür. Bu bölme işlemi için çeşitli yöntemler geliştirilmiştir. Bu çalışmada bu yöntemlerden 10-çapraz geçerleme yöntemi kullanılmıştır. k-kat çapraz geçerlemede veri seti, k adet eşit parçaya ayrılarak her defada bir tanesi test, k-1 tanesi de eğitim için kullanılır. Bu yöntemin çalışması, Şekil 5'te yer almaktadır.




*Şekil 5:* 5 kat çapraz geçerleme yöntemi ile test ve eğitim küme ayrımı

## IV. SONUÇLAR

Bu çalışmada, beş sınıflı epilepsi verilerini yüksek başarım ile sınıflandırmak için k-ortalamalar kümele tabanlı zaman serisi ağırlıklandırma (KOKTZSA) yöntemi ön-işleme olarak ham EEG sinyallerine uygulanmıştır. Ham EEG sinyallerinin ağırlıklandırılmasından sonra epilepsi hastalığını tahmin için sınıflandırma algoritmaları ile sınıflandırma modelleri oluşturulmuştur. Çalışmada kullanılan sınıflandırma yöntemleri; C4.5 karar ağacı ve Random Forest sınıflama algoritmaları kullanılmıştır. Oluşturulan bu sınıflandırma modellerinin performanslarını test etmek için 10 kat çapraz geçerleme yöntemi kullanılmıştır. Sınıflandırma algoritmalarının performanslarını değerlendirmek için; sınıflama doğruluğu (SA), kesinlik (K), hassasiyet (H) ve f-ölçümü (F) değerleri kullanılmıştır. Ayrıca bu performans değerlendirmesi için eğitim ve test veri seti 10 kat çapraz geçerlemeyle ayrılmıştır. Bu performans değerlendirme ölçütlerinin sonuçları Tablo 2'de 10 kat çapraz geçerleme performansı görülmektedir.

*Tablo 2:* Epilepsi veri seti 10 kat çapraz geçerleme ayrımlarına ilişkin Doğruluk, Kesinlik, F ölçü değerleri.

| **Algoritmalar** | **C4.5** | | | | **Random Forest** | | | |
|---|---|---|---|---|---|---|---|---|
| **Performans Ölçütleri** | **SD** | **K** | **H** | **F** | **SD** | **K** | **H** | **F** |
| **Ham Veri** | 0,705 | 0,699 | 0,700 | 0,70 | 0,81 | 0,794 | 0,81 | 0,773 |
| **Ağırlıklandırma Sonrası** | **0,968** | **0,969** | **0,988** | **0,988** | **0,993** | **0,993** | **0,994** | **0,993** |

Tablo 2'de görüldüğü gibi ağırlıklandırma işlemine tabi tutulan ham veri iki algoritmada da daha iyi sonuç vermiştir. Ayrıca Random Forest algoritması hem ham veride hem de KOKTZSA uygulanmış veride daha başarılı sonuç vermiştir

## V. TARTIŞMA

Bu çalışmada, ham EEG sinyallerinden beş sınıflı epilepsi verilerini yüksek başarım elde etmek için yeni bir hibrid model önerilmiştir. Bu önerilen sistem, gerçek zamanlı olarak çalışabilme potansiyeline sahiptir. İşlem yükü ve hesaplama maliyeti diğer modellerden az olduğu için uygulama başarısı yüksek olabilir.

## KAYNAKÇA


[1] Fanntool,"Yapay sinir ağları ile epilepsi için otomatik eeg analizi", ttp://fanntool.blogspot.com.tr/2013/03/yapay-sinir-aglar-ile-epilepsi-icin.html.

[2] R. Tekin, Y. Kaya, ve M.E. Tağluk, "K-means ve YSA temelli Hibrit Bir Model ile Epileptik EEG İşaretlerinin Sınıflandırılması," *Elektrik Elektronik Bilgisayar Semp.* Elazığ, *2011*.

[3] A. Subasi, "Epileptic seizure detection using dynamic wavelet network," *Expert Systems with Applications*, vol. 29, pp. 343–355, 2005.

[4] Q. Wu, E. Fokoue, "Epileptic Seizure Recognition Data Set", UCI Machine learning Repository, Access: https://archive.ics.uci.edu/ml/datasets/Epileptic+Seizure+Recognition.

[5] Z. Cebeci, F. Yıldız, G.T. Kayaalp, "K-Ortalamalar Kümelemesinde Optimum K Değeri Seçilmesi", *2. Ulusal Yönetim Bilişim Sistemleri Kongresi*, s. 231-242, 2015

[6] Dinçer, E. 2006. Veri Madenciliğinde K-Means Algoritması ve Tıp Alanında Uygulanması. Yüksek Lisans Tezi, Kocaeli Üniversitesi, Fen Bilimleri Enstitüsü, Kocaeli, 101s.

[7] Evans, S., Lloyd, J., Stoddard, G., Nekeber, J., Samone, M. 2005. Risk Factors For Adverse Drug Events. The Annals of Pharmacotherapy, 39, 1161-1168.

[8] S. Güneş, K. Polat, Ş. Yosunkaya "Efficient sleep stage recognition system based on EEG signal using k-means clustering based feature weighting," *Expert Systems with Applications*, vol. 37, no. 12, s. 7922-7928, 2010.

[6] Liaw A., (16 October 2012). "Documentation for R package randomForest" (PDF). Retrieved 15 March 2013.

[7] Witten I.H., Eibe Frank; Mark A. Hall (2011). "Veri Madenciliği: Pratik makine öğrenme araçları ve teknikleri, 3. Baskı" . Morgan Kaufmann, San Francisco. s. 191.

[8] N. Japkowicz, "Performance evaluation for learning algorithms," International Conference on Machine Learning, Edinburg, Scotland, 2012.

[9] M. Clark, "An Introduction to machine learning with Applications in R," Lecture Notes, University of Notre Dame, 2015.

[10] P. Flach, "The many faces of ROC analysis in machine learning," Lecture Notes, University of Bristol, 2004.