Gazi Üniv. Müh. Mim. Fak. Der.
Cilt 30, No 4, 585-596, 2015
DOI: 10.17341/gummfd.81654

Journal of the Faculty of Engineering and Architecture of Gazi University
Vol 30, No 4, 585-596, 2015

AĞIRLIKLANDIRILMIŞ KOŞULLU KARŞILIKLI BİLGİ İLE ÖZNİTELİK SEÇİMİ

Ceyhun ÇELİK, Hasan Şakir BİLGE
Gazi Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Ankara
ceyhun.celik@gazi.edu.tr, bilge@gazi.edu.tr

(Geliş/Received: 24.09.2014; Kabul/Accepted: 20.08.2015)

ÖZET

Büyük verilerin işlenmesi ve bu verilerden anlamlı bilgilerin çıkarılması veri madenciliğinin önemli konularından biridir. Uygulamada eldeki verilerin probleme uygun olup olmadığı bilinemez ve ilgisiz veriler oluşturulacak modelin karmaşıklığını arttırır. Daha basit ve düşük maliyetli modeller oluşturmak için problem parametreleri üzerinde boyut indirgeme yaklaşımları uygulanır. Bilgi teorisi tabanlı karşılıklı bilgi yaklaşımları boyut indirgemede yaygın olarak kullanılır. Bu yaklaşımlarda boyut indirgemesi yapılırken veri kümesinden elde edilecek alt kümede, girişlerin birbirleri ile minimum fazlalık ve çıkışla maksimum bağımlılığa sahip olması amaçlanmaktadır. Bu koşulu sağlamak için önerilen yaklaşımların özniteliklerin seçiminde kullandıkları sezgisel fonksiyonlar, bağımlılık ve fazlalık bilgileri arasındaki ilişkiyi problemden bağımsız ve sabit bir değişkenle kontrol eder. Bu çalışmada yeni bir karşılıklı bilgi yaklaşımı önerilmiştir. Önerilen yaklaşımda kullanılan sezgisel fonksiyon, fazlalığın seçim üzerindeki etkinliğini, özniteliklerin sınıfla olan karşılıklı bilgisi ve birbirleri ile olan karşılıklı bilgilerinin arasındaki ilişkiyi değerlendirerek ağırlıklandırır. Benzer şekilde maksimum bağımlılık için hem koşullu karşılıklı bilgi hem de karşılıklı bilgi hesaplanır. Böylelikle önerilen sezgisel fonksiyon, değişen problemler karşısında dinamik bir yaklaşım sergiler. Test sonuçları önerilen yaklaşımın başarısını göstermektedir.

Anahtar Kelimeler: Öznitelik seçimi, koşullu karşılıklı bilgi, maksimum bağımlılık, minimum fazlalık

FEATURE SELECTION WITH WEIGHTED CONDITIONAL MUTUAL INFORMATION

ABSTRACT

Huge data processing and extracting the meaningful information from those data is one of the important topics in data mining. In practice, it cannot be known whether present data are relevant to the problem, and irrelevant data increase the complexity of the prospective model. Dimensionality reduction approaches are applied to the problem parameters to build simpler and low cost models. Information theory based mutual information approaches are commonly used on dimensionality reduction. In these approaches, it is aimed to have the minimum redundancy and maximum dependency in outputs between inputs in the subset obtained from the data set in the execution of the size degradation. However the heuristic functions which are used in the proposed approaches to ensure this condition, control the relation between dependency and redundancy with the fixed parameter and problem-independent. In this study, a new mutual information approach is proposed. The heuristic function used in this proposed approach weights the effectiveness of redundancy on selection by evaluating the relationship between mutual information of features with class and mutual information of features among themselves. Similarly, both conditional mutual information and mutual information are calculated for maximum dependency. Thus the proposed heuristic function presents a dynamic approach to variety of problems. The obtained results of the tests point out the success of the proposed approach.

Keywords: Feature selection, conditional mutual information, maximum dependency, minimum redundancy

1. GİRİŞ (INTRODUCTION)

Günümüzde hemen her alanda veri havuzlarında çok büyük miktarlarda veri toplanmakta ve bu verilerin işlenerek faydalı/anlamlı bilgiye dönüştürülmesi zor ve önemli bir görev halini almaktadır. Anlamlı bilginin çıkarılması işleminde, probleme ait parametreler (öznitelikler) girişler/çıkışlar olarak

sınıflandırılır. Girişler, çıkışları temsil ettiği düşünülen birden fazla ölçüm ya da gözlemden oluşur. Bu ölçüm ya da gözlemlerin her biri öznitelik (feature) olarak adlandırılır. Verilerden anlamlı bilgilerin çıkarılması sürecinde, yapılması gereken ilk işlem öznitelikler ile çıkış verilerinin birbiri ile olan bağımlılığının ortaya konulmasıdır [1]. Bunun ilk sebebi oluşturulacak model ister sınıflandırma (classification) ister bağlanım (regression) için kullanılacak olsun, hesaplama maliyeti girişteki öznitelik sayısına bağlıdır [2]. Bu nedenle ilgisiz olduğu düşünülen öznitelikler model oluşturulmadan önce çıkarılabilir. İkinci sebebi görüntü işleme, gen sınıflandırma gibi problemlerde öznitelik sayısı oldukça fazlayken, örnek sayısı azdır. Modelin test sırasında çıkarım yapma karmaşıklığını azaltmak için yine ilgisiz olduğu düşünülen özniteliklerin çıkarılması gerekir [3]. Bu nedenle model oluşturulmadan önce boyut indirgeme işlemleri uygulanır. Boyut indirgeme ile ilgili iki temel yaklaşım vardır. Birinci yaklaşım Öznitelik Çıkarımı (Feature Extraction) olarak adlandırılır ve m adet öznitelikte bulunan bilgi yeni bir giriş uzayına dönüştürülerek d (d<m) adet öznitelikte toplanır [4]. Temel Bileşen Analizi (TBA) (Principal Component Analysis) ve Doğrusal Ayrıştırma Analizi (DAA) (Linear Discriminant Analysis) en popüler öznitelik çıkarımı yaklaşımlarıdır [4]. İkinci yaklaşım ise Öznitelik Seçimi (Feature Selection) olarak adlandırılır ve burada m öznitelik içerisinde veriyi en iyi şekilde ifade eden d alt öznitelikle ilgilenilir [5]. Altküme Seçimi (Subset Selection) bir öznitelik seçimi metodudur. Burada veriyi tanımlayacak en az sayıdaki özniteliğin bulunması amaçlanır [5]. Burada temel olarak iki arama stratejisi literatürde yaygın olarak kullanılmaktadır. Birinci yaklaşım İleriye Doğru Öznitelik Seçimi (İDÖS) (Forward Feature Selection) yöntemidir. Burada aday alt küme boş kümedir ve her bir yinelemede alt kümeye bir öznitelik eklenir. İkinci yaklaşım ise Geriye Doğru Öznitelik Seçimi (GDÖS) (Backward Feature Selection) yöntemidir. Burada ise aday alt küme tüm girişlere sahiptir ve her bir yinelemede bir öznitelik çıkarılır. Bu arama stratejilerinde, özniteliklerin eklenmesi ya da çıkarılması sırasında üç farklı model üzerinden özniteliklerin problem için ne kadar değerli oldukları değerlendirilir. Bu modeller literatürde filtreleme (filter), sarmalama (wrapper) ve karma (hybrid) model olarak adlandırılır [6]. Filtreleme modellerde özniteliklerin uygunluğunu değerlendiren ölçüt sınıflandırıcıdan bağımsızdır. Sarmalama modellerinde ise ölçüt doğrudan sınıflandırıcının kendisidir. Karma modellerde ise her iki ölçüt birlikte uygulanmaktadır. Bu üç model değerlendirildiğinde filtreleme modellerinin düşük maliyetle en etkili yaklaşım olduğu görülmektedir [6]. Filtreleme modellerinde yapılan işlem özniteliklerin ayrıştırılabilir olduğunun ortaya konulmasıdır. Dağılım Matrisi (Scatter Matrices), Iraksama (Divergence), Fisher'ın Ayıraç Oranı (Fisher's Discriminant Ratio) temel olarak kullanılan değerlendirme fonksiyonlarıdır [7]. Bu fonksiyonlar dışında bilgi teorisi tabanlı Karşılıklı Bilgi (KB) (Mutual Information - MI) yaklaşımları da oldukça popülerdir [8]. KB iki rastgele değişken arasındaki bilgiyi ölçebilmesi ile giriş ve çıkışlar arasındaki ilişkiyi etkili bir şekilde ifade eder [8]. KB'nin öznitelik seçimi işlemlerinde kullanılmaya başlanmasından bu yana konu ile ilgili birçok çalışma yapılmıştır. Bu yaklaşımlar temel olarak veri kümelerini Maksimum Bağımlılık (MB) (Maximum Dependency - MD) ve Minimum Fazlalık (mF) (Minimum Redundancy - mR) kavramları üzerinden değerlendirerek öznitelik seçimi stratejileri sunmuşlardır. Literatürde iki popüler KB tabanlı yaklaşım bulunmaktadır. Bu yaklaşımlar literatürde birçok farklı alanda uygulanmıştır [9-13]. Birinci yaklaşım KB tabanlı önerilen ilk yöntem olan Karşılıklı Bilgi Öznitelik Seçimi (KBÖS) (Mutual Information Feature Selection - MIFS) yöntemidir [14]. İkinci yaklaşım ise minimum Fazlalık Maksimum Uygunluk (mFMU) (minimum Redundancy Maximum Relevance - mRMR) yaklaşımıdır [15]. Yapılan analizler göstermiştir ki, KBÖS ve mFMU yaklaşımlarının başarısı uygulanan problemlere göre farklılıklar göstermektedir. Her iki yaklaşımda da kullanılan değerlendirme fonksiyonlarında fazlalığın seçim üzerindeki etkisi bir sabit ile kontrol edilir ve bu sabit problemden bağımsızdır. Bununla birlikte her iki yaklaşımda da yalnızca KB değerleri kullanılarak MB sağlanmaya çalışılmaktadır. Yakın zamanda yapılmış benzer bir çalışma da Dinamik Ağırlıklandırılmış Öznitelik Seçimi (DAÖS) (Dynamic Weighting-based Feature Selection - DWFS) yaklaşımıdır [16]. Tablo 1'de bu algoritmaların çalışma şekli açıklanmaktadır. Bu çalışmada MB ve mF arasındaki ilişkiyi değerlendirirken, fazlalığı özniteliklerin birbirleri ve sınıfla olan KB'lerine göre ağırlıklandıran ve bağımlılığı hesaplarken Koşullu Karşılıklı Bilgi'yi (KKB) (Conditional Mutual Information - CMI) de kullanan Ağırlıklandırılmış Koşullu Karşılıklı Bilgi Öznitelik Seçimi (AKKBÖS) (Weighted Conditional Mutual Information Feature Selection - WCMIFS) isimli yeni bir yaklaşım önerilmiştir. Önerilen yaklaşım ile literatürdeki çalışmalardan farklı olarak, hem KB hem de KKB bir arada bağımlılık ölçümünde kullanılarak, sınıfları en iyi tanımlayan özniteliklerin seçilmesi problemi çözülmeye çalışılmıştır. Buna ek olarak fazlalığın seçim üzerindeki etkinliği yine literatürdeki çalışmalardan farklı olarak, problemdeki özniteliklerin sınıfla ve birbirleri ile olan KB ölçümlerine göre ağırlıklandırılmıştır. Böylelikle öznitelikler tek başlarına iyi sınıflandırma yetenekleri sergilerken, bir arada da bu yeteneklerini koruyabilecekleri bir seçim algoritması geliştirilmeye çalışılmıştır. Önerilen yaklaşım 10 sınıflandırma problemi üzerinde test edilmiş ve sonuçlar KBÖS, mFMU ve DAÖS yaklaşımları ile karşılaştırılmıştır. Karşılaştırma işlemi için literatürde yaygın olarak

kullanılan iki popüler makine öğrenmesi algoritması olan Destek Vektör Makineleri (DVM) (Support Vector Machine - SVM) ve Yalın Bayes (YB) (Naive Bayes – NB) sınıflandırıcı k-kat çapraz geçerleme (k-fold cross validation) yöntemi ile eğitilmiştir.

Tablo 1. Öznitelik Seçimi Algoritmaları (Feature Selection Algorithms)

Öznitelik Seçimi Algoritmaları		
Kaynak	Öznitelik Seçim Yöntemi	Çalışma Şekli
[14]	Karşılıklı Bilgi Öznitelik Seçimi (KBÖS)	Karşılıklı bilgiye dayalı yapılan hesaplamalar ile öznitelikler seçilir.
[15]	minimum Fazlalık Maksimum Uygunluk (mFMU)	Minimum Fazlalık-Maksimum Uygunluk ölçütü tanımlanıp, bu ölçüte göre öznitelik seçimi yapılır.
[16]	Dinamik Ağırlıklandırılmış Öznitelik Seçimi (DAÖS)	Bilgi teorisi ölçütlerinden yararlanarak öznitelik ilişkisi, öznitelikler arası bağımlılık ve fazlalık analizleri yapılır. Sonra dinamik ağırlıklandırma ile öznitelikler seçilir.

2. KB GENEL BİLGİLERİ VE LİTERATÜRDE ÖNERİLMİŞ YAKLAŞIMLAR (BACKGROUND OF MI AND PROPOSED APPROACHES ON LITERATURE)

KB, bilgi teorisinin temel kavramlarından biridir. Temeli düzensizliğe (entropy) dayanır. Düzensizlik bir rastgele değişkendeki belirsizliği ölçer. Bu ölçüm Denklem 1'dedir.

$$H(x) = -\sum_{i=1}^{n} p(x_i)\log(p(x_i)) \quad (1)$$

Burada X ayrık verilerden oluşan rastgele bir değişken olsun. x_i bu değişkendeki farklı verileri temsil etsin. $p(x_i)$ bu farklı verilerin olasılığıdır. Düzensizlik 0 ve 1 arasında değer üreten bir ölçümdür. KB ise iki rastgele değişken arasındaki paylaşılan bilginin ölçümünü verir. Bu ölçüm Denklem 2'dedir.

$$I(X;Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right) \quad (2)$$

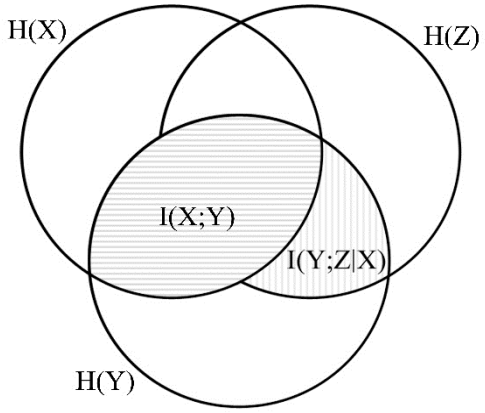
Burada değişkenlerin ayrık verilerden oluştuğu varsayılmaktadır. Sürekli veriler için ise KB hesabı Denklem 3'te gösterildiği gibidir.

$$I(X;Y) = \int_X\int_Y p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right)dxdy \quad (3)$$

Bazı durumlarda iki rastgele değişken arasındaki paylaşılan bilginin büyüklüğü üçüncü bir değişkene göre hesaplanmak istenebilir. Bu durumda KKB ölçümü kullanılır. KKB ölçümü Denklem 4'tedir.

$$I(X;Y\mid Z) = \sum_{z\in Z}\sum_{x\in X}\sum_{y\in Y} p(x,y,z)\log\left(\frac{p(z)p(x,y,z)}{p(x,z)p(y,z)}\right) \quad (4)$$

Üç rastgele değişken arasındaki bilginin paylaşımı bu ölçümlere göre Şekil 1'deki gibi görselleştirilebilir.



Şekil 1. Değişkenler arasındaki ilişki (Relation between variables)

KB ölçümü ile öznitelik seçimi yaklaşımını ilk olarak Battiti 1994 yılında önermiştir [14]. Battiti'nin yapmış olduğu ve sonraki yıllarda önerilmiş olan çalışmaların temel olarak yaptığı işlemler aşağıdaki şekildedir:

• Özniteliklerin sınıfla olan bağımlılıkları ve birbirleri ile olan fazlalıklarına göre analizi.
• MB ve mF arasındaki geçişi sağlayacak sezgisel bir fonksiyonun ortaya konması.
• Öznitelik alt kümesinin bulunması için kullanılacak arama stratejisinin ortaya konması.

Literatürde yapılmış olan çalışmaların birçoğu arama stratejisi olarak açgözlü arama (greedy search) kullanmışlardır [8]. Buna göre buradaki en önemli adım MB ve mF arasındaki geçişi sağlayacak sezgisel fonksiyonun belirlenmesi işlemidir. Battiti önermiş olduğu çalışmada bu geçişi MB'ı sağlamak için Denklem 5'te gösterildiği şekilde yorumlamıştır.

$$G = I(C;f) - \beta\sum_{s\in S} I(f;s) \quad (5)$$

C veri kümesindeki çıkışları, f seçilmeye aday özniteliği ve s seçilmiş olan öznitelik kümesindeki bir özniteliği ifade eder. Burada MB ile mF arasındaki geçiş bir değişkenle düzenlenmektedir. $\beta = 0$ olduğunda özniteliklerin yalnızca sınıfla olan KB'si değerlendirilmektedir. Sun ve arkadaşları yapmış oldukları çalışmada bağımlılık, karşılıklı bağımlılık (interdependency) ve fazlalık kavramları ile ilgili yeni bir bakış açısı sunarak dinamik olarak ağırlıklandırılan bir öznitelik seçimi olan DAÖS yöntemini önermişlerdir [16]. Burada yeni seçilecek

olan özniteliğin sınıf ile olan bağımlılığı, bu özniteliğin daha önce seçilmiş olan öznitelikler ile olan fazlalığı ile ağırlıklandırılarak hesaplanmaktadır. Böylelikle sınıf ile en iyi bağımlılığa sahip olan öznitelikler seçilirken, birbirleri ile olan fazlalıklar değerlendirilmektedir. Ancak burada seçilen özniteliklerin DVM gibi ayırt edebilen (discriminative) modellerde, YB gibi olasılıksal modellere göre daha iyi sonuçlar verdiği görülmüştür. Dolayısıyla önerilen model kullanılacak sınıflandırıcılardan bağımsız bir seçim işlemi yapamamaktadır. Bir diğer çalışmada MB ve mF arasındaki geçiş fonksiyonu yorumlanırken, özniteliklerin birbirleri arasındaki ilgisizliğin sınıfa göre minimum olması gerektiği vurgulanmıştır [17]. Bu nedenle özniteliklerin birbirleri ile olan ilgisizliği KKB değerine göre belirlenmiştir. Ancak burada KKB Denklem 4'te gösterildiği şekilde kullanılmak yerine Denklem 6'daki gibi kullanılmıştır.

$$G = I(C;f) - \beta \sum_{s \in S} \frac{I(C;s)}{H(s)} I(f;s) \tag{6}$$

Peng ve arkadaşları önermiş oldukları mFMU yaklaşımında MB ve mF arasındaki geçişi ortalama mF kullanarak sağlamışlardır. Fonksiyon Denklem 7'dedir.

$$G = I(C;f) - \frac{1}{|S|} \sum_{s \in S} I(f;s) \tag{7}$$

Bir diğer çalışmada ikinci dereceden (quadratic) KB kullanılarak en uygun öznitelik seçimi KB (optimal feature selection MI) yaklaşımı önerilmiştir [18]. Estevez ve arkadaşları yapmış oldukları çalışmada geçiş fonksiyonu olarak normalleştirilmiş KB (normalized MI) kullanmışlardır. Bununla birlikte açgözlü arama yaklaşımının hızlı ve etkili olmasına rağmen, özniteliklerin sınıfı ayrı ayrı iyi nitelerken, birlikte daha kötü niteleme problemini çözememesinden dolayı Genetik Algoritma (GA) tabanlı bir arama stratejisi ile karma bir seçim yöntemi önermişlerdir [19]. Ancak oldukça fazla öznitelik olan problemlerde GA kullanılarak yapılacak olan bir arama stratejisi oldukça maliyetli olacaktır. 2012 yılında yapılmış olan çalışmada çekirdek kanonik ilinti analizi (kernel canonical correlation analysis), öznitelik ve sınıflar arasındaki doğrusal olmayan ilişkiyi bulmak için kullanılmıştır. Elde edilen fonksiyonu diğer yöntemlerin aksine özniteliklerin yerine mFMU yaklaşımında kullanmışlar ve böylece sınıfla ilişkili olmayan fazlalığı ortadan kaldırmayı amaçlamışlardır [20]. Yukarıdaki çalışmalar genel olarak sınıfla MB sağlamak için yalnızca KB ölçümünden yararlanmıştır [14-20]. Ancak literatürde sonraki dönemlerde yapılan çalışmalarda bu yaklaşımların eksikliği değerlendirilmiştir [21-23]. Burada ortaya konan problem özniteliklerin birbirleri arasındaki KB değerine bakılırken sınıfın göz önünde bulundurulmamasıdır. Sotoca ve Pla yapmış oldukları çalışmada KKB tabanlı bir yaklaşım önermiştir. Buradaki yaklaşımda MB kullanılarak sınıf ile öznitelikler arasındaki var olan bilginin sıkıştırılması amaçlanmıştır [21]. Bununla birlikte burada minimum ilgili olan özniteliklerin çıkarılması amaçlanmıştır. Bu nedenle dışarıda kalacak özniteliklerin sınıfla arasındaki bilgiye bakılırken seçilen özniteliklere göre KKB ölçümü yapılmıştır. Bu çalışmada seçim işlemi için literatürdeki açgözlü arama yöntemlerinden farklı olarak kümeleme yaklaşımı önerilmiştir. Bunun için literatürde önerilmiş olan K-ortalama (K-means) tabanlı bir yaklaşımla öznitelikler bir arada toplanarak en alakasız özniteliklerin dışarıda bırakılması sağlanmıştır. Ancak burada seçilecek özniteliklerin sınıfla olan KB ölçümleri değerlendirilmemiştir. 2011 yılında yapılan çalışmada yine önceki çalışmalarda olduğu gibi açgözlü arama yöntemi kullanan KKB tabanlı bir yaklaşım önerilmiştir [22]. Ancak çoklu rastgele değişkenler üzerinde KKB hesaplaması zor olduğu için önceki yaklaşımlarda olduğu gibi KKB için bir yaklaşık hesaplama yaklaşımı önerilmiştir. Yapılan çalışmada artımlı olarak her bir öznitelik seçildikten sonra, kalan özniteliklerin her birinin sınıfla arasındaki ilişkisi bu seçilmiş olan özniteliğe göre KKB ölçümü yapılarak değerlendirilir. Daha sonra kullanılan bir eşik değeri ile henüz seçilmemiş olan bu öznitelikler çıkarılır. Böylece sınıfla MB'a sahip olan öznitelikler seçilirken aynı zamanda sınıfla mF'a sahip öznitelikler de çıkarılmaya başlanır. Burada KB ölçümü seçim ölçütünde kullanılmamış, yalnızca mF oluşturmak için çıkarılan özniteliklerin belirlenmesinde bir eşik olarak kullanılmıştır. Bu durumda KKB'nin MB sağlamak için ayırt edici olmadığı durumlar göz ardı edilmiş olur. Zeng ve arkadaşları yapmış oldukları çalışmada KB ve KKB ölçümlerini literatürdeki yaklaşımların aksine farklı bir bakış açısı ile değerlendirmiş ve Kaba (Rough) KKB ile YB sınıflandırıcının kullanıldığı karma bir öznitelik seçimi önermişlerdir [23]. Ancak burada hem karma öznitelik seçimi yaklaşımının maliyeti hem de yalnızca YB sınıflandırıcıya olan bağımlılığı dezavantajdır. Şu ana kadar literatürde önerilen yaklaşımlardan özellikle mFMU yaklaşımı öne çıkmıştır. Daha önce bahsedildiği gibi literatürde farklı uygulama alanlarında kullanılmasının yanı sıra sezgisel yöntemlerle de karma olarak kullanılmıştır [24, 25]. Ancak bu karma yöntemlerde kullanılan sezgisel yöntemlerin uygunluk fonksiyonlarının, sınıflandırıcılar olması maliyetli bir yaklaşımdır. Az sayıda özniteliğe sahip bir problemde bu yaklaşımlar iyi sonuçlar verecek olsa da çok fazla özniteliğe sahip problemlerin çözümünde oldukça maliyetli olacaktır.

3. KB ANALİZİ VE AKKBÖS (MI ANALYSIS AND WCMIFS)

KB ölçümleri öznitelik seçimine uygulanırken karşılaşılan temel problem, ikiden fazla özniteliğin birlikte KB değerlerinin hesaplanması işlemidir [8,15]. Öznitelik seçim işlemi yeterince maliyetli bir işlem olduğu için, KB değerinin tam olarak ölçülmeye

çalışılması problem çözümünü imkânsız hale getirecektir. Bu nedenle literatürdeki çalışmalarda da görüldüğü üzere yaklaşık bir hesaplama kullanılmaktadır [23]. Her ne kadar birden fazla özniteliğin, sınıfla arasındaki bağımlılığının hesaplanmasında bu yaklaşık ölçüm algoritma performansını etkilemese de, seçilecek olan özniteliğin, seçilmiş olan özniteliklerle arasındaki fazlalığın yaklaşık ölçümü algoritma performansını etkilemektedir. Literatürdeki temel iki yaklaşım KBÖS ve mFMU algoritmalarının temel farkı, bu fazlalığın bağımlılık üzerindeki etkisinin belirlenmesidir. İlk algoritmada toplam fazlalık kullanılırken ($\beta = 1$ için), sonraki algoritmada ise ortalama fazlalık kullanılmaktadır. Bu iki temel KB kullanan öznitelik seçimi algoritmalarının başarıları problemden probleme değişmektedir. Bunun nedeni seçim işlemine fazlalığın etkisi sabit olarak verilmesi ve uygulanan problemdeki özniteliklerin bağımlılıkları ile fazlalıkları arasındaki ilişkinin göz ardı edilmesidir. Bu çalışmada, bu ilişki, öznitelik seçimi uygulanacak olan problemde seçilecek özniteliklerin fazlalık bilgileri çok az olduğu durumlarda özniteliklerin sınıfla olan bağımlılık bilgileri daha önemli; bunun tersi durumda ise özniteliklerin birbirleri ile olan fazlalık bilgileri daha önemli olacak şekilde bir ağırlıklandırma parametresi önerilmiştir. Bu parametre Denklem 8'dedir.

$$K = \text{logsig}\left(\frac{\sum_{s\in S} I(f;s) + mF_S}{I(f;C) + MB_S}\right) \tag{8}$$

Burada mF_S seçilmiş olan özniteliklerin fazlalık miktarını, MB_S ise bağımlılığını göstermektedir. Fazlalığın bağımlılığa oranından elde edilecek değeri 0 ve 1 arasında normalleştirmek için bir s-biçimli (sigmoid - sig) fonksiyon olan logaritmik (logarithmic - log) s-biçimli fonksiyon kullanılır. Buna göre bağımlılık ve fazlalık arasındaki geçiş fonksiyonu Denklem 9'daki gibi olacaktır.

$$G = I(C;f) - K\sum_{s\in S} I(f;s) \tag{9}$$

Burada $I(C;f)$ seçilecek özniteliğin sınıfla olan bağımlılığının ölçümüyken, $\sum_{s\in S} I(f;s)$ seçilmiş öznitelik kümesi ile seçilecek öznitelik arasındaki fazlalığın ölçümüdür. Bu bağımlılık ve fazlalık arasındaki geçiş Denklem 8'de tanımlanan parametre ile ağırlıklandırılır. Öznitelik seçiminde başarıyı etkileyen etkenlerden bir diğeri de kullanılan bağımlılıktır. KB ölçümü kullanılarak hesaplanan bağımlılıklarda Şekil 1'de de görüleceği üzere, seçilecek özniteliğin seçilmiş özniteliklerle arasında sınıfa göre bir bilgi paylaşımı olduğu görülmektedir. Bu nedenle KKB ölçümü yapılarak kullanılacak bir geçiş fonksiyon kullanımı daha uygun olacaktır. Bu şekilde MB aranırken, KKB ölçümü ile sınıfla yüksek KB ilişkisine sahip öznitelikler arasındaki paylaşılan bilgi ayırt edici duruma gelir. KKB kullanılarak tanımlanan bir geçiş fonksiyonu Denklem 10'dadır.

$$G = \sum_{s\in S} I(\mathrm{C};f \mid s) - K\sum_{s\in S} I(f;s) \tag{10}$$

Ancak sınıfla yüksek bağımlılığa sahip özniteliklerin, sınıfa göre birbirleri ile olan bilgisinin çok düşük olduğu problemlerde, öznitelik seçimi yapılırken KKB ölçümü kullanmak sınıfla daha az bağımlılığa sahip özniteliklerin seçilmesine neden olacak ve bu durum genel MB yaklaşımına ters düşecektir. Yapılan analizler sonucunda MB ve mF arasındaki geçişi sağlayacak olan bir fonksiyon şu özellikleri sağlamalıdır.

• MB ve mF arasındaki geçişi düzenlerken uygulanan problemi de göz önüne almalıdır.
• MB'ı sağlarken seçilmeye aday öznitelik ile seçilmiş öznitelik kümesi arasındaki paylaşılan bilgiyi ayırt edici bir etken olarak kullanmalıdır.
• MB'ı sağlarken seçilmeye aday öznitelik ile seçilmiş öznitelik kümesi arasındaki paylaşılan bilginin ayırt edici olmadığı problemleri de göz önüne almalıdır.

Bu özellikleri taşıyan bir geçiş fonksiyonu Denklem 11'de tanımlanmıştır.

$$G = I(C;f) + \sum_{s\in S} I(\mathrm{C};f \mid s) - 2K\sum_{s\in S} I(f;s) \tag{11}$$

Bu geçiş fonksiyonu öncelikle MB'ı sağlamak için KB ve KKB ölçümlerini birlikte kullanmaktadır. KKB ölçümünde aday özniteliğin sınıfla paylaşılan bilgisi tüm seçilmiş öznitelik alt kümesine göre alındığı için, burada seçilmiş öznitelik sayısı arttıkça KKB'nin değeri dolayısıyla MB sağlamadaki etkinliği artacaktır. Öznitelik seçimi işleminin başlangıcında ise KB değerleri kullanılarak KKB'nin sağladığı ayırt ediciliğin kullanılamadığı problemlerde ise yüksek bağımlılığa sahip öznitelikler seçilmiş olacaktır. mF sağlanırken ise MB bölümündeki KB ve KKB için iki fazlalık ölçümü kullanılmaktadır. Bununla birlikte Denklem 8'de tanımlanmış katsayı ile MB ve mF arasındaki geçiş probleme göre ağırlıklandırılır. Geçiş fonksiyonu önerilen AKKBÖS algoritmasında kullanılacaktır. Bu algoritmanın adımları Şekil 2'dedir.

4. TEST VERİLERİNİN ANALİZİ (ANALYSIS of TEST DATA)

Bu çalışmada önerilen öznitelik seçimi yaklaşımının başarısı UCI Machine Learning Repository veri tabanında bulunan ve öznitelik seçimi yaklaşımlarında sıklıkla kullanılan Arrhythmia, Hill-Valley, Libras, Optdigits, Parkinson, Soybean, Spambase ve Vehicle veri kümeleri ile test edilmiştir [26]. Bu problemlere ek olarak 2 adet tümör sınıflandırma problemi olan Leukemia [27] ve Colon [28] veri kümesi test için

kullanılmıştır. Test için kullanılan veri kümeleri ile ilgili genel bilgiler Tablo 2'dedir. Tüm veri kümeleri için eksik değere sahip olan öznitelikler ortalama yöntemi ile doldurulmuştur.

C : Sınıf;
S : Seçilmiş alt öznitelik kümesi; $S = \{\varnothing\}$
F : Öznitelik kümesi;
d : Seçilecek öznitelik sayısı;
K : Katsayı;
MB_S : Seçilmiş öznitelik alt kümesinin maksimum bağımlılığı; $MB_S = 0.001$
mF_S : Seçilmiş öznitelik alt kümesinin minimum fazlalığı; $mF_S = 0.001$

1. F öznitelik kümesindeki her bir f özniteliğin C sınıfla olan $I(f;C)$ değerini hesapla $KB_f = I(f;C)$
2. En büyük KB değerine sahip özniteliği ilk öznitelik olarak seç $F \leftarrow F - \{f_{\max KB}\}$; $S \leftarrow \{f_{\max KB}\}$
3. F öznitelik kümesindeki her bir f özniteliğin S seçilmiş alt öznitelik kümesindeki her bir öznitelik ile KB değerini hesapla $mF_f = \sum_{s \in S} I(f;s)$
4. F öznitelik kümesindeki her bir f özniteliğin S seçilmiş alt öznitelik kümesindeki her bir öznitelik ile $\sum_{s \in S} I(f;C|s)$ değerini hesapla $KKB_f = \sum_{s \in S} I(f;C|s)$
5. $MB_f = KB_f + KKB_f$
6. $K_f = \text{logsig}\left(\frac{\sum_{s \in S} I(f;s) + mF_S}{I(f;C) + MB_S}\right)$
7. MB_f ile mF_f arasındaki geçişi sağlayacak sezgisel fonksiyonu çalıştır ve en uygun özniteliği seç $\vec{f} = MAX(MB_f - 2K_f mF_f)$; $F \leftarrow F - \{\vec{f}\}$; $S \leftarrow \{\vec{f}\}$
8. $MB_S = MB_S + MB_{\vec{f}}$
9. $mF_S = mF_S + mF_{\vec{f}}$
10. Eğer |S|==d ise algoritmayı durdur; Değilse 3'e dön

Şekil 2. AKKBÖS algoritması (WCMIFS algorithm)

Tablo 2. Veri Kümesi Bilgileri (Data Set Information)

Veri Kümeleri				
Veri Kümeleri	Veri Sayısı	Öznitelik Sayısı	Sınıf Sayısı	Veri Tipi
Arrhythmia [26]	452	279	16	Karışık
Colon [28]	62	2000	2	Sürekli
Hill-Valley [26]	606	100	2	Sürekli
Leukemia [27]	72	7129	2	Sürekli
Libras [26]	360	90	15	Sürekli
Optdigits [26]	5620	64	10	Kategorik
Parkinsons [26]	195	23	2	Sürekli
Soybean [26]	307	35	19	Kategorik
Spambase [26]	4601	57	2	Sürekli
Vehicle [26]	846	18	4	Sürekli

Önerilen yaklaşımın başarısı KBÖS, mFMU, DAÖS algoritmaları ile karşılaştırılmıştır. Karşılaştırma işlemi için DVM ve YB sınıflandırıcı k-kat çapraz geçerleme yöntemi ile eğitilmiştir. k-kat çapraz geçerleme yönteminde veri kümesi k adet parçaya bölünür. Bu parçalardan bir tanesi test, k-1 tanesi ise eğitim için kullanılır. Bu işlem k defa tekrarlanır ve her seferinde test için kullanılan parça eğitime eklenirken, test edilmemiş bir parça test edilmek için eğitimden çıkarılır ve eğitim süreci gerçekleştirilir. Bu çalışmada yapılan test işlemlerinde k=10 olarak belirlenmiştir. DVM algoritması için LibSVM [29] araç kutusunun Matlab sürümü kullanılmıştır. DVM çekirdek fonksiyonu (kernel function) olarak araç kutusundaki varsayılan fonksiyon Dairesel Taban Fonksiyonu (DTF) (Radial Basis Function – RBF) ve C parametresi 100 olarak alınmıştır. Test sonuçlarını gösteren grafikler Matlab kullanılarak elde edilmiştir. Sürekli verilere sahip veri kümeleri üzerinde KB tabanlı öznitelik seçimi yaklaşımı uygulamak için literatürde iki popüler yaklaşım vardır. Bu yaklaşımlardan birisi Parzen Penceresi (Parzen Window) kullanımıdır [30]. Bir diğer yaklaşım ise sürekli verilerin kategorik (categorical) verilere dönüştürülmesidir. Bu işlemde her bir özniteliğin ortalama (μ) (mean) ve standart sapması (σ) (standard deviation) değerine göre istenen sayıda kategoriye ayrılır. Bu çalışmada ikinci yaklaşım kullanılmıştır ve her bir sürekli öznitelik 3 kategoriye ayrılmıştır. $\mu - \sigma/2$ 'den küçük olan değerler 0, $\mu - \sigma/2$ ve $\mu + \sigma/2$ arasında olan değerler 1, $\mu + \sigma/2$ 'den büyük olan değerler 2 olarak kategorize edilmiştir. Bahsedilen stratejilere göre elde edilen sınıflama performanslarının test sonuçları Tablo 3 ve Tablo 4'te gösterilmektedir. Tablo 3'te tüm alt kümelere göre sınıflandırıcılardan elde edilen doğruluk oranlarının ortalaması, Tablo 4'te ise en yüksek doğruluk oranları gösterilmektedir. Öznitelik seçiminde hem ortalama hem de en yüksek sınıflandırıcı doğruluğu iki önemli başarı ölçütüdür. Tüm alt küme durumları için yapılan seçim işlemi bir

tür sıralama işlemidir. Bu nedenle seçilen alt kümelerin sınıflandırma doğruluğu ortalamalarının yüksek olması özniteliklerin sınıfla ilişkisini iyi kurduğunu gösterir. İkinci kıstas olarak en yüksek doğruluk oranı ise olası en iyi niteliklerin seçildiğini gösterir.

Tablo 3 incelendiğinde önerilen yöntem AKKBÖS YB sınıflandırıcı için Arryhthmia ve Hill-Valley veri kümeleri hariç tüm veri kümelerinde en yüksek ortalamaya sahiptir. DVM algoritması için ise Hill-Valley, Libras, Parkinsons, Spambase ve Vehicle veri kümeleri dışında diğer veri kümeleri için en yüksek ortalamaya sahiptir. Hill-Valley veri kümesi için KBÖS ve mFMU yöntemlerinden daha iyi sonuç vermiştir. Libras ve Parkinsons veri kümeleri için mFMU ve DAÖS, Spambase için KBÖS ve DAÖS, Vehicle veri kümesi için ise KBÖS ve mFMU algoritmalarından daha yüksek ortalamaya sahiptir.

Tablo 4 incelendiğinde ise önerilen algoritma AKKBÖS YB sınıflandırıcı için Arrhythmia, Colon, Libras, Soybean ve Vehicle veri kümeleri için en yüksek doğruluk oranlarını vermiştir. Leukemia ve Hill-Valley veri kümeleri için tüm yaklaşımlar eşit doğruluk oranı verirken, Parkinsons veri kümesi için DAÖS yaklaşımı dışında tüm seçim yaklaşımları eşit doğruluk oranı bulmuştur. Opt. Digits ve Spambase veri kümelerinde ise KBÖS yaklaşımı mFMU, AKKBÖS ve DAÖS algoritmalarından daha iyi sonuç vermiştir. DVM algoritması testlerinde ise Arrhythmia, Colon, ve Soybean veri kümeleri için AKKBÖS algoritmasının en yüksek doğruluk oranı verdiği görülmektedir. Hill-Valley veri kümesi için DAÖS en iyi sonucu verirken diğer üç yaklaşım eşit doğruluk oranı vermiştir. Leukemia için önerilen algoritma DAÖS algoritması ile birlikte en iyi sonucu vermiştir. Libras veri kümesi için DAÖS en iyi sonucu verirken önerilen algoritma KBÖS algoritmasından daha iyi, mFMU algoritması ile aynı doğruluk oranına sahiptir. Opt. Digits veri kümesi için AKKBÖS, mFMU ile birlikte en iyi doğruluk oranını verirken, Parkinson veri kümesinde ise KBÖS ile en iyi doğruluk oranını vermiştir. Spambase veri kümesi için önerilen algoritma KBÖS algoritmasından daha iyi sonuç vermiştir. Son olarak Vehicle veri kümesi için KBÖS ve mFMU yaklaşımlarından daha iyi sonuç verdiği görülmektedir.

Tablo 3. Ortalama Sınıflandırma Doğruluğu (Mean Classification Accuracy)

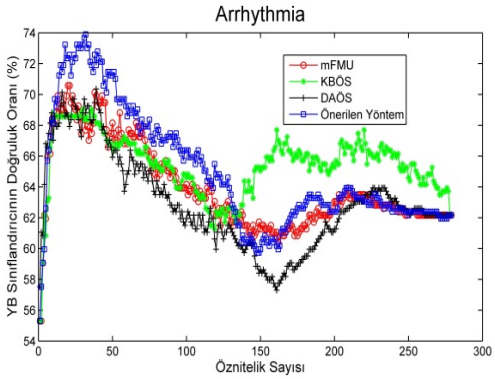
Test Sonuçları								
Algoritmalar / Veri Kümeleri	YB Sınıflandırıcı				DVM			
	KBÖS	mFMU	AKKBÖS	DAÖS	KBÖS	mFMU	AKKBÖS	DAÖS
Arrhythmia	65,47	63,94	65,00	62,96	64,78	68,62	68,74	68,58
Colon	88,13	88,74	90,35	88,37	79,83	81,33	86,01	81,04
Hill-Valley	50,91	50,91	50,92	51,04	51,89	51,96	51,99	55,38
Leukemia	95,05	99,69	99,92	99,30	90,59	97,01	99,20	96,42
Libras	55,61	52,98	57,31	50,76	85,78	83,54	85,69	79,29
Optdigits	85,15	85,34	85,72	84,95	92,89	94,12	94,26	93,97
Parkinsons	83,64	83,15	83,96	81,33	91,42	90,91	91,33	89,37
Soybean	79,82	78,61	81,39	76,77	84,60	85,70	86,53	84,04
Spambase	86,80	87,36	87,90	87,45	90,62	91,32	91,31	91,30
Vehicle	50,28	49,82	51,23	46,05	69,98	70,34	71,13	77,04

Tablo 4. En iyi Sınıflandırma Doğruluğu (Maximum Classification Accuracy)

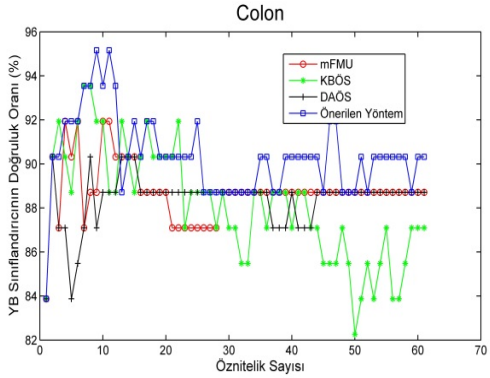
Test Sonuçları								
Algoritmalar / Veri Kümeleri	YB Sınıflandırıcı				DVM			
	KBÖS	mFMU	AKKBÖS	DAÖS	KBÖS	mFMU	AKKBÖS	DAÖS
Arrhythmia	69,03	70,58	73,89	70,35	71,90	72,79	73,23	72,79
Colon	93,55	91,94	95,16	90,32	85,48	85,48	91,94	85,48
Hill-Valley	51,32	51,32	51,32	51,32	56,77	56,77	56,77	58,42
Leukemia	100	100	100	100	94,44	98,61	100	100
Libras	59,17	59,44	61,39	58,61	88,89	89,44	89,44	90,00
Opt. Digits	91,35	90,73	90,73	90,48	98,92	99,32	99,32	99,31
Parkinsons	87,69	87,69	87,69	84,62	95,90	94,87	95,90	92,31
Soybean	86,97	85,99	88,60	85,67	92,18	93,16	93,49	91,86
Spambase	89,81	89,37	89,79	89,76	93,18	93,24	93,20	93,28
Vehicle	53,31	53,19	54,02	48,46	83,10	83,10	83,57	86,41

Şekil 3 – Şekil 10 grafikleri Arrhythmia, Colon, Leukemia ve Soybean veri kümeleri üzerinde uygulanan KBÖS, mFMU, AKKBÖS ve DAÖS öznitelik seçimi algoritmalarına göre elde edilen öznitelik alt kümelerinin YB sınıflandırıcı ve DVM üzerinde vermiş olduğu doğruluk oranlarını göstermektedir. X ekseni öznitelik sayısını gösterirken, Y ekseni ise kullanılan sınıflandırıcının doğruluk oranını göstermektedir.

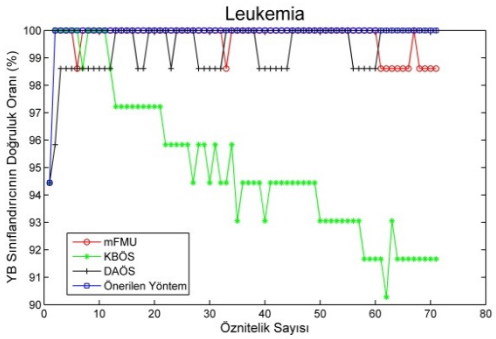
Grafikler incelendiğinde önerilen AKKBÖS algoritmasının KBÖS, mFMU ve DAÖS algoritmalarına göre daha iyi olduğu görülmektedir. Buna ek olarak Colon ve Soybean veri kümeleri için elde edilen sonuçlar incelendiğinde YB sınıflandırıcı için KBÖS algoritması daha iyi sonuç verirken, DVM sınıflandırıcı için ise mFMU algoritması daha iyi sonuç vermektedir. DAÖS algoritmasının ise yine bu iki farklı sınıflandırıcıya göre elde ettiği sonuçlar diğer yöntemlere göre farklılık göstermektedir. Önerilen yöntem AKKBÖS her iki veri kümesinde de en iyi sonucu vermiştir. Böylece önerilen yaklaşımın MB ve mF arasındaki ilişkiyi iyi kurduğu ve sınıflandırıcıdan bağımsız bir seçim yaptığı görülmektedir.



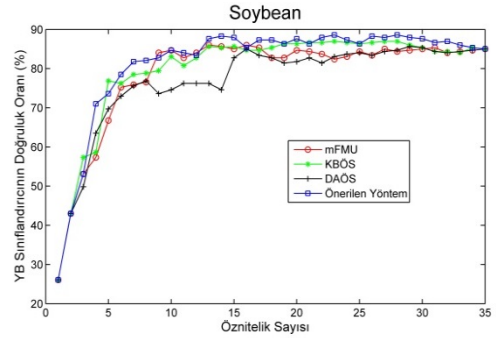
Şekil 3. Arrhythmia veri kümesi için YB sınıflandırıcı ile karşılaştırmalı sonuçlar (Compared results with NB classifier for Arrhyhmia data set)



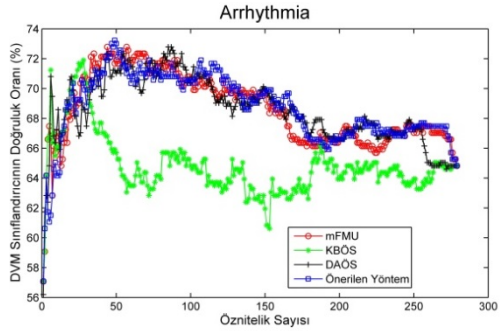
Şekil 4. Colon veri kümesi için YB sınıflandırıcı ile karşılaştırmalı sonuçlar (Compared results with NB classifier for Colon data set)



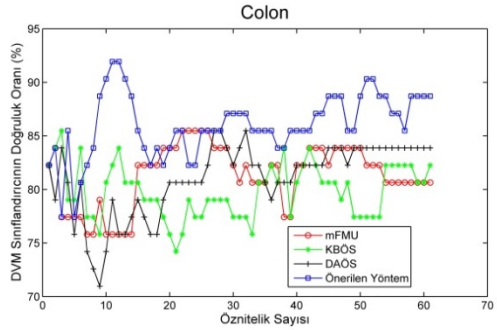
Şekil 5. Leukemia veri kümesi için YB sınıflandırıcı ile karşılaştırmalı sonuçlar (Compared results with NB classifier for Leukemia data set)



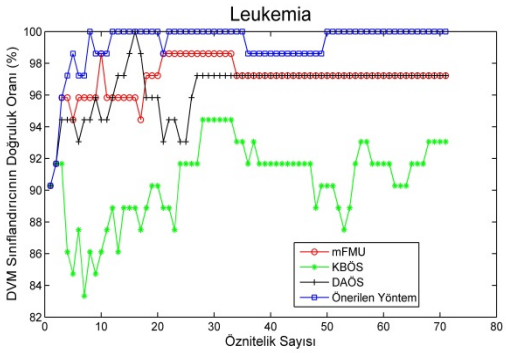
Şekil 6. Soybean veri kümesi için YB sınıflandırıcı ile karşılaştırmalı sonuçlar (Compared results with NB classifier for Soybean data set)



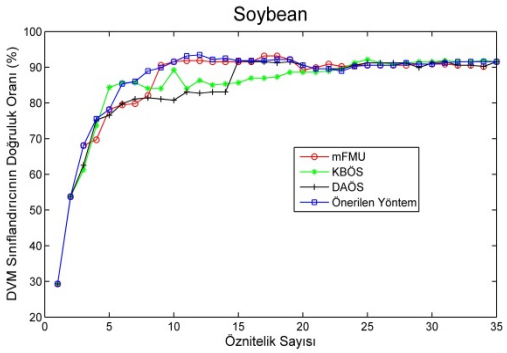
Şekil 7. Arrhythmia veri kümesi için DVM ile karşılaştırmalı sonuçlar (Compared results with SVM for Arrhyhmia data set)



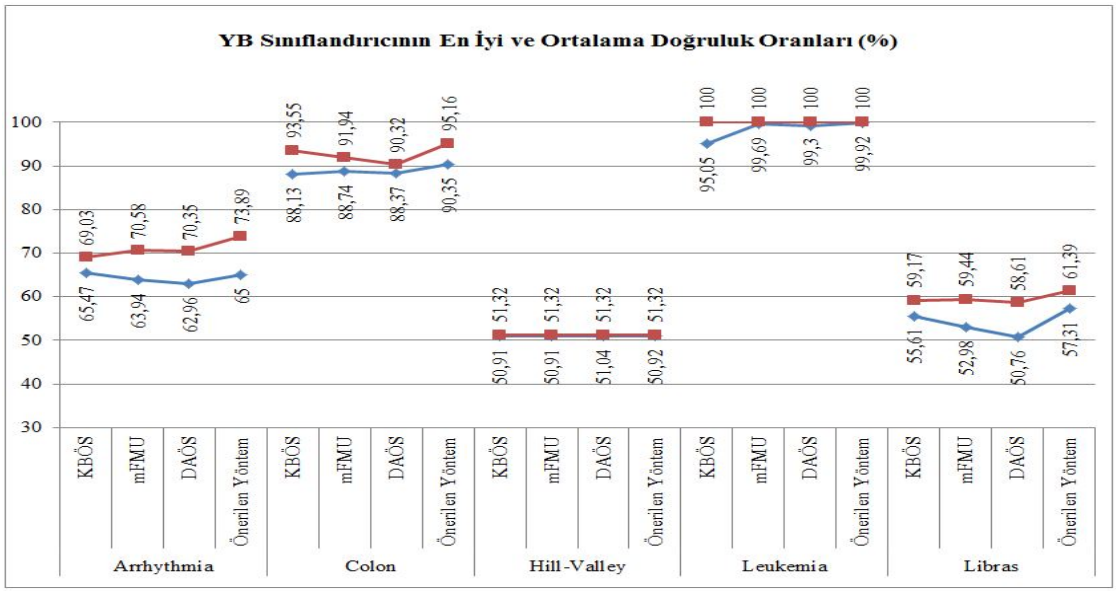
Şekil 8. Colon veri kümesi için DVM ile karşılaştırmalı sonuçlar (Compared results with SVM for Colon data set)



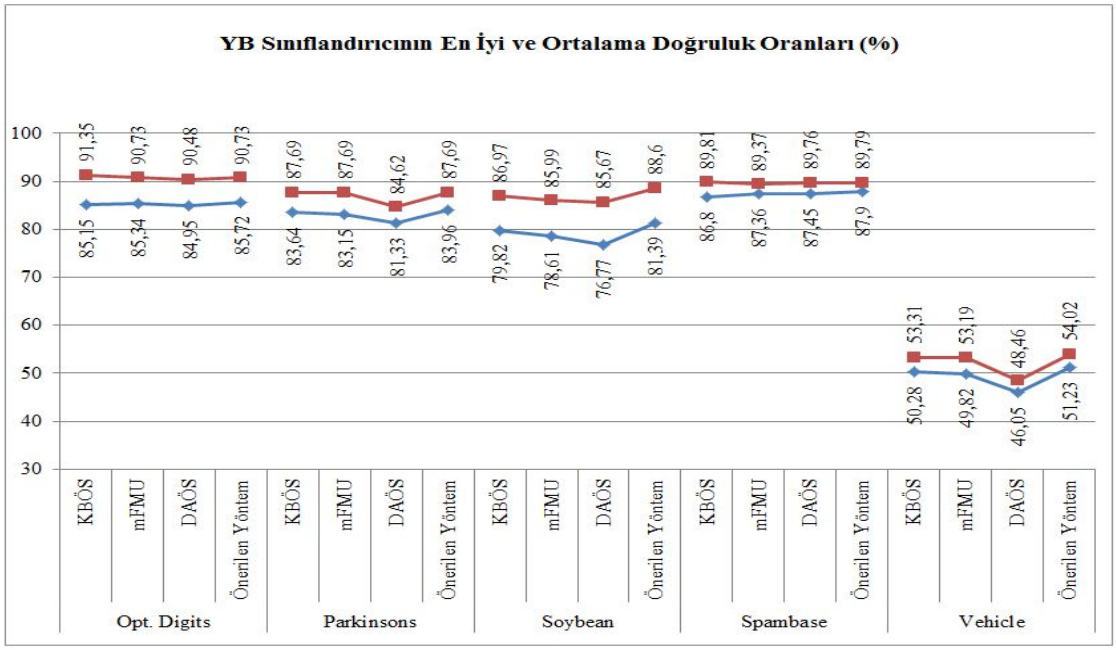
Şekil 9. Leukemia veri kümesi için DVM ile karşılaştırmalı sonuçlar (Compared results with SVM for Leukemia data set)



Şekil 10. Soybean veri kümesi için DVM ile karşılaştırmalı sonuçlar (Compared results with SVM for Soybean data set)

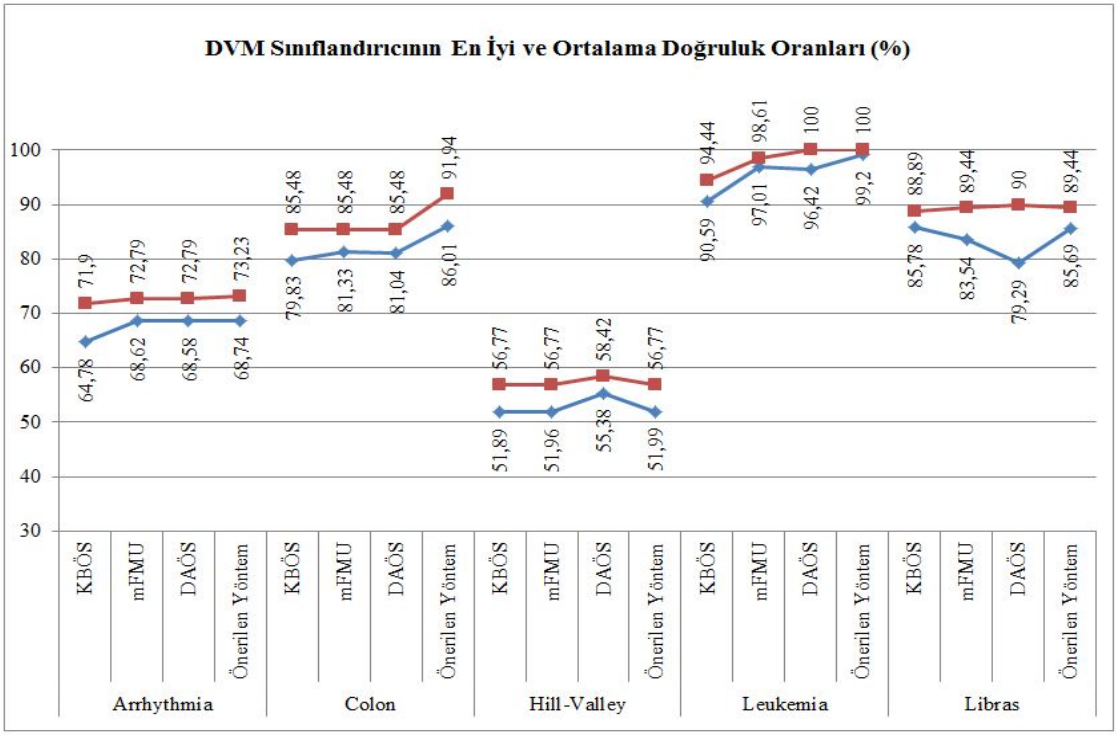


Şekil 11. Arrhythmia, Colon, Hill-Valley, Leukemia ve Libras veri kümeleri için YB sınıflandırıcı ile karşılaştırmalı sonuçlar (Compared results with NB classifier for Arrhythmia, Colon, Hill-Valley, Leukemia and Libras data set)

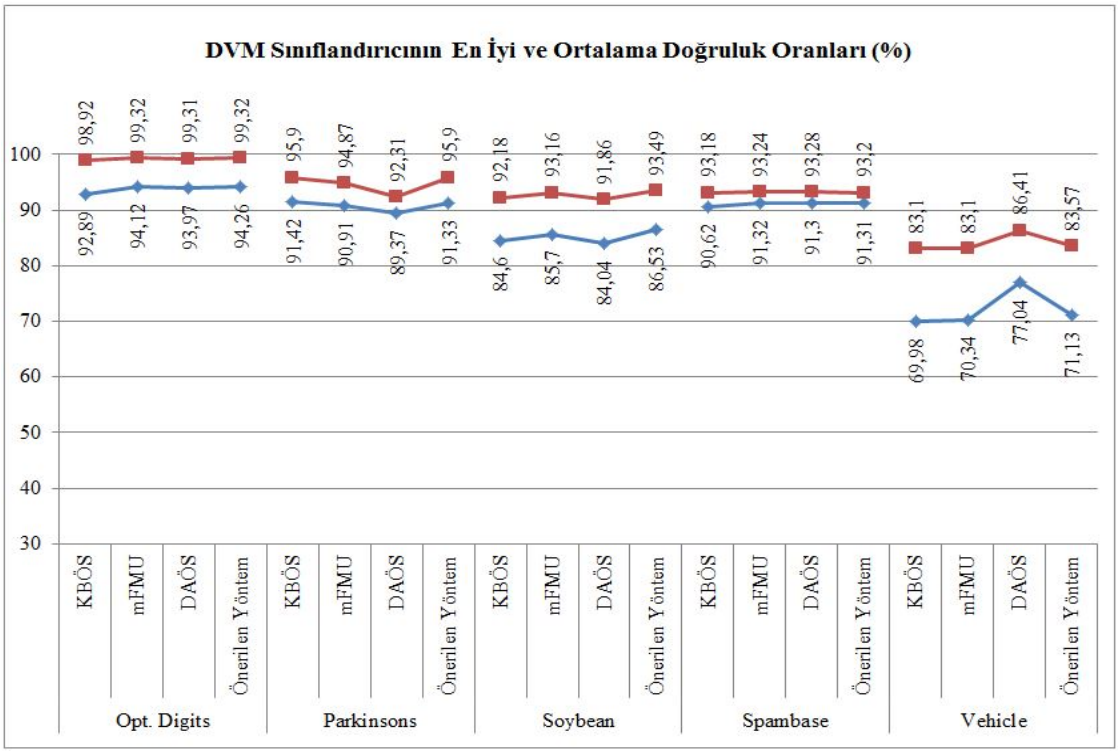


Şekil 12. Opt. Digits, Parkinsons, Soybean, Spambase ve Vehicle veri kümeleri için YB sınıflandırıcı ile karşılaştırmalı sonuçlar (Compared results with NB classifier for Opt. Digits, Parkinsons, Soybean, Spambase and Vehicle data set)

Şekil 11 - Şekil 14 grafiklerinde ise tüm test veri kümeleri için elde edilen en iyi ve ortalama doğruluk oranları hem YB sınıflandırıcı hem de DVM algoritması için verilmiştir.



Şekil 13. Arrhythmia, Colon, Hill-Valley, Leukemia ve Libras veri kümeleri için DVM ile karşılaştırmalı sonuçlar (Compared results with DVM for Arrhythmia, Colon, Hill-Valley, Leukemia and Libras data set)



Şekil 14. Opt. Digits, Parkinsons, Soybean, Spambase ve Vehicle veri kümeleri için DVM ile karşılaştırmalı sonuçlar (Compared results with SVM for Opt. Digits, Parkinsons, Soybean, Spambase and Vehicle data set)

5. SONUÇLAR VE TARTIŞMA (CONCLUSIONS AND DISCUSSIONS)

Bu çalışmada KB tabanlı yeni bir öznitelik seçimi yaklaşımı önerilmiştir. Yaklaşım literatürdeki yaklaşımlara benzer olarak açgözlü arama stratejisi ile ileri yönlü öznitelik seçimi yaparak en iyi öznitelik alt kümesini seçmeyi hedeflemektedir. Önerilen yaklaşımın literatürdeki yaklaşımlardan farkı mF ve MB hesaplaması ve bu iki ölçüm arasındaki geçişi sağlama biçimidir. MB hesaplaması yapılırken KKB ve KB ölçümleri birlikte kullanılmıştır. Böylelikle KB ölçümleri yüksek olan problemler için öznitelikler arasında birbirleri ile en az bilgi paylaşan özniteliklerin seçilmesi için KKB dikkate alınmış, KB ölçümleri düşük olan problemler için ise sınıfla en yüksek bilgiyi paylaşan özniteliklerin seçilmesi için başlangıçtaki özniteliklerin seçiminde KB, seçilen öznitelik sayısı arttıkça KKB dikkate alınmıştır. Buna ek olarak MB ve mF arasındaki geçişi sağlamak için problem bağımlı bir katsayı kullanılmıştır. Kullanılan katsayı yaklaşımın uygulandığı problemdeki özniteliklerin sınıfla olan bağımlılığı ve birbirleri ile olan fazlalığını dikkate almaktadır. Böylelikle MB ve mF arasındaki geçiş problem ve o anki öznitelik durumuna bağlı dinamik olarak sağlanmaktadır. Önerilen yaklaşım literatürde birçok alanda uygulaması bulunan iki popüler algoritma ve yakın zamanda önerilmiş olan benzer bir çalışma ile karşılaştırılmış ve başarılı olduğu görülmüştür.

Yapılan literatür incelemesi sonucunda, önerilen KB tabanlı yaklaşımların çoğunlukla açgözlü arama stratejisi kullandığı görülmektedir. Gelecek çalışmalarda önerilen yaklaşımın farklı arama stratejileri ile kullanılması planlanmaktadır. Buna ek olarak KB ölçümlerinin analizlerine devam edilerek yeni geçiş fonksiyonları üzerinde çalışılacaktır.

KAYNAKLAR (REFERENCES)

1. Kahraman, H. T., Sagiroglu, S., Colak, I., "Developing Intuitive Knowledge Classifier and Modeling of Users' Domain Dependent Data in Web", **Knowledge Based Systems**, Cilt 37, 283-295, 2013.
2. Dash, M., Liu, H., "Feature Selection for Classification", **Intelligent Data Analysis**, Cilt 1, 131-156, 1997.
3. Chandrashekar, G., Sahin, F., "A Survey on Feature Selection Methods", **Computers & Electrical Engineering**, Cilt 40, No 1, 16-28, 2014.
4. Alpaydin, E., **Introduction to Machine Learning,** The MIT Press, 2010.
5. Guyon, I., "An Introduction to Variable and Feature Selection", **Journal of Machine Learning Research**, Cilt 3, 1157-1182, 2003.
6. Liu, H., Yu, L., "Toward Integrating Feature Selection Algorithms for Classification and Clustering", **IEEE Transactions on Knowledge and Data Engineering**, Cilt 17, 491-502, 2005.
7. Theodoridis, S., Koutroumbas, K., **Pattern Recognition**, Elsevier, 2008.
8. Vergara, J., Estevez, P., "A Review of Feature Selection Methods Based on Mutual Information", **Neural Computing and Applications**, Cilt 24, No 1, 175-186, 2014.
9. Cai, Y., Huang, T., Hu, L., Shi, X., Xie, L., Li, Y., "Prediction of Lysine Ubiquitination with mRMR Feature Selection and Analysis", **Amino Acids**, Cilt 42, No 4, 1387-1395, 2012.
10. Abdi, M. J., Hosseini, S. M., Rezghi, M. A., "Novel Weighted Support Vector Machine Based on Particle Swarm Optimization for Gene Selection and Tumor Classification", **Computional and Mathematical Methods in Medicine**, Cilt 2012, 1-7, 2012.
11. Hejazi, M. I., Cai, X., "Input Variable Selection for Water Resources Systems Using A Modified Minimum Redundancy Maximum Relevance (mRMR) Algorithm", **Advances in Water Resources** , Cilt 32, No 4, 582-593, 2009.
12. Wang, Y., Miao, Q., Ma, E., Tsui, K. L., Pecht, M., "Online Anomaly Detection for Hard Disk Drives Based on Mahalanobis Distance", **IEEE Transactions on Reliability**, Cilt 62, No 1, 136-145, 2013.
13. Bulling, A., Ward, J., Gellersen, H., Troster, G., "Eye Movement Analysis for Activity Recognition Using Electrooculography", **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Cilt 33, No 4, 741-753, 2011.
14. Battiti, R., "Using Mutual Information for Selecting Features in Supervised Neural Net Learning", **IEEE Transactions on Neural Networks**, Cilt 5, No 4, 537-550, 1994.
15. Peng, H., Long, F., Ding, C., "Feature Selection Based on Mutual Information Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy", **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Cilt 27, No 8, 1226-1238, 2005.
16. Sun, X., Liu, Y., Xu, M., Chen, H., Han, J., Wang, K., "Feature Selection Using Dynamic Weights for Classification", **Knowledge-Based Systems** , Cilt 37, 541-549, 2013.
17. Kwak, N., Choi, C. H., "Input Feature Selection for Classification Problems", **IEEE Transactions on Neural Networks**, Cilt 13, No 1, 143-159, 2002.
18. Chow, T. W. S., Huang, D., "Estimating Optimal Feature Subsets Using Efficient Estimation of High-Dimensional Mutual Information", **IEEE Transactions on Neural Networks**, Cilt 16, No 1, 213-224, 2005.
19. Estevez, P., Tesmer, M., Perez, C., Zurada, J., "Normalized Mutual Information Feature Selection", **IEEE Transactions on Neural Networks**, Cilt 20, No 2, 189-201, 2009.

20. Sakar, C. O., Kursun, O., Gurgen, F., "A Feature Selection Method Based on Kernel Canonical Correlation Analysis and The Minimum Redundancy-Maximum Relevance Filter Method", **Expert Systems with Applications**, Cilt 39, No 3, 3432-3437, 2012.

21. Sotoca, J. M.i Pla, F., "Supervised Feature Selection by Clustering Using Conditional Mutual Information-Based Distances", **Pattern Recognition**, Cilt 43, No 6, 2068-2081, 2010.

22. Cheng, H., Qin, Z., Feng, C., Wang, Y., Li, F., "Conditional Mutual Information-Based Feature Selection Analyzing for Synergy and Redundancy" , **ETRI Journal**, Cilt 33, 210-218, 2011.

23. Zeng, Z., Zhang, H., Zhang, R., Zhang, Y., "Hybrid Feature Selection Method Based on Rough Conditional Mutual Information and Naive Bayesian Classifier", **ISRN Applied Mathematics**, Cilt 2014, 1-11, 2014.

24. El Akadi, A., Amine, A., El Ouardighi, A., Aboutajdine, D., "A Two-Stage Gene Selection Scheme Utilizing mRMR Filter and GA Wrapper", **Knowledge and Information Systems**, Cilt 26, No 3, 487-500, 2011.

25. Huang, H., Xie, H. B., Guo, J. Y., Chen, H. J., "Ant Colony Optimization - Based Feature Selection Method for Surface Electromyography Signals Classification", **Computers in Biology and Medicine**, Cilt 42, No 1, 30-38, 2012.

26. Bache, K., Lichman, M., UCI Machine Learning Repository, University of California, Irvine, School of Information and Computer Sciences, 2013.

27. Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., "Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring", **Science**, Cilt 286, 531-537, 1999.

28. Alon, U., Barkai, N., Notterman, D. A., Gish, K., Ybarra, S., Mack, D., Levine, A. J., "Broad Patterns of Gene Expression Revealed by Clustering Analysis of Tumor and Normal Colon Tissues Probed by Oligonucleotide Arrays" **Proceedings of the National Academy of Sciences**, Cilt 96, No 12, 6745-6750, 1999.

29. Hsu, C.W., Lin, C.J., "A Comparison of Methods for Multiclass Support Vector Machines", **IEEE Transactions on Neural Networks**, Cilt 13, No 2, 415-425, 2002.

30. Kwak, N., Choi, C. H., "Input Feature Selection by Mutual Information Based on Parzen Window", **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Cilt 24, No 12, 1667-1671, 2002.