

İŞLETME BİLİMİ DERGİSİ

THE JOURNAL OF
BUSINESS SCIENCE



JOBS

İşletme Bilimi Dergisi
2018
Cilt:6 Sayı:3



SAKARYA
ÜNİVERSİTESİ

JOBS

İşletme Bilimi Dergisi
The Journal of Business Science

Sakarya Üniversitesi / Sakarya University
İşletme Fakültesi / Sakarya Business School

i

Cilt/Volume : 6
Sayı/Issue : 3
Yıl/Year : 2018

ISSN: 2148-0737
DOI: 10.22139/jobs

İNDEKS BİLGİLERİ/ INDEXING INFORMATION



ii



Kurucu Sahip/Founder

Prof. Dr. Gültekin YILDIZ

İmtiyaz Sahibi / Owner

Prof. Dr. Kadir ARDIÇ

Editör / Editor

Doç. Dr. Mahmut AKBOLAT

Editör Yardımcıları / Assoc. Editors

Doç. Dr. Mustafa Cahit ÜNÇAN

Danışma Kurulu/Advisory Board

Prof. Dr. Ahmet Vecdi CAN	Sakarya Üniversitesi
Prof. Dr. Bülent SEZEN	Gebze Yüksek Teknoloji Enstitüsü
Prof. Dr. Dilaver TENGİLİMOĞLU	Atılım Üniversitesi
Prof. Dr. Erman COŞKUN	Sakarya Üniversitesi
Prof. Dr. Kadir ARDIÇ	Sakarya Üniversitesi
Prof. Dr. Mehmet BARCA	Ankara Sosyal Bilimler Üniversitesi
Prof. Dr. Nihat ERDOĞMUŞ	İstanbul Şehir Üniversitesi
Prof. Dr. Orhan BATMAN	Sakarya Üniversitesi
Prof. Dr. Recai COŞKUN	Sakarya Üniversitesi
Prof. Dr. Remzi ALTUNIŞIK	Sakarya Üniversitesi
Prof. Dr. Selahattin KARABINAR	İstanbul Üniversitesi
Prof. Dr. Sıdıka KAYA	Hacettepe Üniversitesi
Prof. Dr. Şevki ÖZGENER	Nevşehir Üniversitesi
Prof. Dr. Türker BAŞ	Galatasaray Üniversitesi
Doç. Dr. Surendranath Rakesh JORY	Southampton Üniversitesi

Yayın Kurulu / Editorial Board

Prof. Dr. Kadir ARDIÇ
Doç. Dr. Mahmut AKBOLAT
Doç. Dr. Mustafa Cahid ÜNĞAN

Sekreteryaya / Secreteria

Arş. Gör. Özgün ÜNAL
Arş. Gör. Mustafa AMARAT
Arş. Gör. Ayhan DURMUŞ

iv

Dergimize yayınlanmak üzere gönderilen makalelerin yazımında etik ilkelere uyulduğu ve yazarların ilgili etik kurulundan gerekli yasal onayları aldığı varsayılmaktadır. Bu konuda sorumluluk tamamen yazarlara aittir. İşletme Bilimi Dergisi'nde yer alan makalelerin bilimsel sorumluluğu yazara aittir. Yayınlanmış eserlerden kaynak gösterilmek suretiyle alıntı yapılabilir.

It is assumed that the articles submitted for publication in our journal are written in ethical principles and the authors have obtained the necessary legal approvals from the relevant ethics committee. The responsibility of this matter belongs to the authors. Scientific responsibility for the articles belongs to the authors themselves. Published articles could be cited in other publications provided that full reference is given.

İşletme Bilimi Dergisi; www.dergipark.gov.tr/jobs Sakarya Üniversitesi İşletme Fakültesi jobs@sakarya.edu.tr Esentepe Kampüsü 54187 Serdivan/SAKARYA

Bu Sayıda Katkıda Bulunan Hakemler *Reviewers List of This Issue*

İşletme Bilimi Dergisi
2018
Cilt:6 Sayı:3

Prof. Dr. Aykut Hamit Turan	Sakarya Üniversitesi
Prof. Dr. Remzi Altunışık	Sakarya Üniversitesi
Prof. Dr. Ruziye Cop	Abant İzzet Baysal Üniversitesi
Prof. Dr. Tamer Uğur	Atatürk Üniversitesi
Doç. Dr. Behçet Yalın Özkara	Eskişehir Osmangazi Üniversitesi
Doç. Dr. Burhanettin Zengin	Sakarya Uygulamalı Bilimler Üniversitesi
Doç. Dr. Faruk Anıl Konuk	Sakarya Üniversitesi
Doç. Dr. Hakan Murat Arslan	Düzce Üniversitesi
Doç. Dr. Harun Kırılmaz	Sakarya Üniversitesi
Doç. Dr. Mehmet Ali Alan	Cumhuriyet Üniversitesi
Doç. Dr. Musa Said Döven	Eskişehir Osmangazi Üniversitesi
Doç. Dr. Mustafa Cahid Ünğan	Sakarya Üniversitesi
Doç. Dr. Nevran Karaca	Sakarya Üniversitesi
Doç. Dr. Özgür Uğurluoğlu	Hacettepe Üniversitesi
Doç. Dr. Özlem BALABAN	Sakarya Üniversitesi
Doç. Dr. Ötügen Senger	Kafkas Üniversitesi
Doç. Dr. Sinan Esen	Sakarya Uygulamalı Bilimler Üniversitesi
Doç. Dr. Tansel Hacıhasanoğlu	Bozok Üniversitesi
Doç. Dr. Ümit Alnaçık	Kocaeli Üniversitesi
Dr. Öğr. Üyesi Ahmet Kar	Kırıkkale Üniversitesi
Dr. Öğr. Üyesi Ersin İrk	Gaziosmanpaşa Üniversitesi
Dr. Öğr. Üyesi Ferda Alper Ay	Cumhuriyet Üniversitesi
Dr. Öğr. Üyesi İbrahim Taha Dursun	Sakarya Üniversitesi
Dr. Öğr. Üyesi İsa Demirkol	Kırıkkale Üniversitesi
Dr. Öğr. Üyesi Osman Uslu	Sakarya Üniversitesi
Dr. Öğr. Üyesi Sema Akpınar	Sakarya Üniversitesi
Öğr. Gör. Dr. Metin Bayram	Sakarya Üniversitesi
Arş. Gör. Dr. Seda Aydan	Hacettepe Üniversitesi

Sizlerin özverili çalışmaları ve desteği ile İşletme Bilimi Dergisi'nin 2018 yılı üçüncü sayısı (Cilt 6, Sayı 3) huzurlarımızdayız. Önceki sayılarımızda olduğu gibi bu sayımızda da işletme biliminin turizm, sağlık yönetimi, muhasebe ve finans, örgütsel davranış, stratejik yönetimi ve sayısal yöntemler gibi farklı disiplinlerinden toplam 12 makaleye yer veriyoruz.

"Örgütsel Davranış Kongreleri Bildirilerinde "Metaforik Neolojizm" Sorunu Üzerine Eleştirel Bir İnceleme" başlıklı Hasan TUTAR'ın yazmış olduğu sayının ilk makalesi son dönemde örgütsel davranış literatüründe sıkça kullanılan "metaforik neoloji"lerin sorunsallaştırılması, bilimsel faaliyet adına yürütülen ve yabancı literatürde "neolojizm" olarak ifade edilen "yeni bir şeyler söyleme merakı"nın risklerine dikkat çekmektedir.

Bu sayımızın "Firmaların AR-GE Harcamalarının Aktif ve Özsermaye Karlılığına Etkisi: Bist Teknoloji Sektöründe Bir Uygulama" başlıklı ikinci makalesinde Hasan Hüseyin YILDIRIM ve Saşkir SAKARYA AR-GE harcamalarının firmaların finansal performansına etkisini araştırmaktadır.

vi

Hümeyra Töre BAŞAT ve Makbule AKGÜNDÜZ'ün kurumsal imaj kavramını ele aldıkları "Sektöre İlişkin Kurumsal İmaj Algılarının Kariyer Yapma İstekleri Üzerine Etkisi: Turizm Öğrencilerine Yönelik Bir Uygulama" başlıklı makale turizm eğitimi alan öğrencilerin sektöre yönelik imaj algılarının, kariyer yapma isteklerine etkisini belirlemeyi amaçlamaktadır.

"Trafik Kaza Desenlerinin Tanımlanmasında K-Means Kümeleme Algoritmasının Kullanılması: Sakarya İli Uygulaması" adlı makalenin yazarlığını Samet GÜNER, Keziban SEÇKİN CODAL, Hüseyin Serdar GEÇER ve Erman COŞKUN yapmıştır. Yazarlar makalede Sakarya ilinde meydana gelen trafik kazalarını incelemiş ve kazalar arasındaki benzerlikleri araştırmıştır. Araştırma sonuçları, Sakarya'da meydana gelen trafik kazalarının temel karakteristiklerini ortaya koymuştur.

Sayımızın beşinci makalesi Rifat YILMAZ'ın kaleme aldığı "Kültür Bağlamında Faaliyet Tabanlı Maliyetleme Yönteminin Türkiye'de Başarı Koşullarının Değerlendirilmesi" başlıklı makaledir. Yazar makalede Türk toplumunun kültür özelliklerinin faaliyet tabanlı maliyetleme yönteminin kabulüne etkisini ortaya koymaktadır.

Vildan ATEŞ ve Zafer KILIÇ tarafından yazılan “E-Müşterilerin Demografik Özelliklerine Göre Algı Ve Memnuniyetlerindeki Farklılıkların İncelenmesi: Gaziantep İli Örneği” adlı makale, e-müşterilerin cinsiyetinin, yaş grubunun ve okul türünün online alışverişe yönelik algılarında ve memnuniyetlerinde farklılık oluşturup oluşturmadığını araştırmaktadır. Sonuç olarak e-müşterilerin cinsiyetlerinin ve okul türünün müşteri algısında ve müşteri memnuniyetinde anlamlı farklılık oluşturduğunu bulunmuştur.

Son sayımızda yer alan “Girişimcinin Perspektifinden Kendi Girişimsel Başarı Değerlendirmesi Ve Bağlam Tartışması” başlıklı makalede Umut Sanem ÇİTÇİ, Oğuzhan ÖZTÜRK ve Metin DİNÇER öznel girişimsel başarı kavramına “bağlamın” etkisini ortaya koymayı amaçlamaktadır. Yazarlar araştırma sonucunda girişimsel başarı değerlendirmesinde üç farklı bağlam düzeyinin etkisinin olduğunu belirlemiş ve bu bağlamları “girişimciyi çevreleyen kurum temelli”, “girişimcinin temas halinde olduğu” ve “girişimcinin kendi ve firmasından kaynaklı” şeklinde tanımlamıştır.

Nuran AKŞİT AŞIK’ın yazdığı “İşe Yabancılaşma Boyutlarının Demografik Faktörler Açısından İncelenmesi” başlıklı makale otellerde çalışan bireylerin demografik özelliklerinin işe yabancılaşma boyutlarında fark oluşturup oluşturmadığını incelemektedir.

“Farklı Veri Setleri Üzerinde Smo Ve J48 Algoritmalarının Sınıflandırma Sonuçlarının Karşılaştırılması” başlıklı çalışmaları ile Mehmet Ali ALAN ve Cavit YEŞİLYURT, üç farklı veri seti ile TP-Oranı, FP-Oranı, Kesinlik, Duyarlık, F-ölçütü ve ROC analizi gibi çeşitli doğruluk ölçümlerini kullanarak, J48 ve SMO algoritmalarının sınıflandırma doğruluğu açısından performansını değerlendirmiştir.

Olgun Irmak ÇETİN “Etik Liderlerde Duygusal Emegin İşkolikliğe Etkisi” başlıklı çalışmada, iş yaşamında etik davranışlar sergileyen liderlerde duygusal emek algısının zamanla işkolikliğe dönüşeceği savını araştırmıştır. Çalışma sonucunda etik davranış gösteren liderlerde Derin Davranış ve Yüzeysel Davranış geliştiği ve Etik Liderlerde Derin Davranış olgusunun işkolikliğe yol açtığı bulunmuştur.

“Sağlık Profesyonellerinin Yalın Uygulamalara Direncini Belirlemeye Yönelik Bir Araştırma” başlıklı sağlık profesyonelleri üzerine yapılan araştırmayı Feryal BULUT ve Mehmet Selami YILDIZ kaleme almıştır.

Araştırmada yazarlar katılımcıların kurumlarında uygulanan yalın uygulamalarına karşı direnç durumlarının belirlenmesi amaçlanmaktadır.

“Endüstri 4.0 ve Çağrı Hizmetleri İlişkisi Üzerine Bir Araştırma” başlıklı makalede ise İsa DEMİRKOL ve Selami ÖZCAN, çağrı hizmetleri ile günümüzün gözde uygulamalarından olan endüstri 4.0 kullanımı arasındaki ilişkiyi tespit etmeye çalışmışlar ve endüstri 4.0’ın çağrı hizmetlerinde kullanımı ile çalışanların memnuniyeti, müşterilerin sorunlarını çözme, yöneticilerin davranışları, ekip çalışması ve çalışma ortamı anlamlı ve pozitif yönlü ilişkiler tespit etmişlerdir.

Görüldüğü gibi, dergimizin bu sayısı da işletmeciliğin farklı disiplinlerinde değerli bilim insanlarının kıymetli çalışmalarıyla oldukça zengin bir şekilde hazırlanmıştır. Dergi politikası olarak bundan sonraki sayılarımızda da işletme bilimine dayalı farklı disiplinlerden gelen çalışmaları yayınlamaya özen göstereceğiz. Bu sayımızda göndermiş oldukları makaleler ile dergimize katkı sağlayan tüm yazarlarımıza, dergimize gönderilen makalelerin değerlendirilmesi için kıymetli vakitlerini ayıran saygıdeğer hakemlerimize ve makalelerin dergide yayınlanmaya hazır hale gelmesi için yoğun bir gayret gösteren editör kurulumuz ve dergi sekreteryamıza teşekkürü borç bilirim. Dergimizin okurlarımız ve bilim insanlarına faydalı olması dileklerle sonraki sayılarımızda işletmeciliğin güncel çalışmalarını bilim dünyasının hizmetine sunmak için siz değerli bilim insanları ve araştırmacıların katkılarını bekliyoruz.

Saygılarımızla...

Doç. Dr. Mahmut AKBOLAT
Editör

İÇİNDEKİLER/CONTENTS

Yıl (Year) 2018 Cilt (Vol.) 6 Sayı (No) 3

İşletme Bilimi Dergisi

2018

Cilt:6 Sayı:3

Literatür Derlemesi/Literature Review

Örgütsel Davranış Kongreleri Bildirilerinde “Metaforik Neolojizm” Sorunu Üzerine Eleştirel Bir İnceleme

A Critical Review On The Problem Of “Metaphorical Neology” In The Proceedings Of Organizational Behavior Congresses

1-38

Prof. Dr. Hasan TUTAR

Araştırma Makaleleri/Research Articles

Firmaların Ar-Ge Harcamalarının Aktif Ve Özsermaye Karlılığına Etkisi: Bist Teknoloji Sektöründe Bir Uygulama

The Effect Of The Firms’ R&D Expenditures On Return On Assets And Return On Equity: Evidence From Bist Technology Sector

39-60

Dr. Öğr. Üyesi Hasan Hüseyin YILDIRIM, Prof. Dr. Şakir SAKARYA

Sektöre İlişkin Kurumsal İmaj Algılarının Kariyer Yapma İstekleri Üzerine Etkisi: Turizm Öğrencilerine Yönelik Bir Uygulama

The Effect Of Corporate Image Perceptions Related To Sector On The Willingness To Making A Career: An Application On Tourism Students

61-87

Dr. Öğr. Üyesi Hümevra TÖRE BAŞAT, Makbule AKAGÜNDÜZ

Trafik Kaza Desenlerinin Tanımlanmasında K-Means Kümeleme Algoritmasının Kullanılması: Sakarya İli Uygulaması

The Use Of K-Means Clustering Algorithm For Identifying The Traffic Accident Patterns: Case Of The Sakarya City

89-106

Dr. Öğr. Üyesi Samet GÜNER, Dr. Öğr. Üyesi Keziban SEÇKİN CODAL, Hüseyin Serdar GEÇER, Prof. Dr. Erman COŞKUN

Kültür Bağlamında Faaliyet Tabanlı Maliyetleme Yönteminin Türkiye’de Başarı Koşullarının Değerlendirilmesi

Evaluating The Success Conditions Of The Activity Based Costing Method In Turkey From The Cultural Context

107-125

Doç. Dr. Rifat YILMAZ

E-Müşterilerin Demografik Özelliklerine Göre Algı Ve Memnuniyetlerindeki Farklılıkların İncelenmesi: Gaziantep İli Örneği

An Examination Of Differences On Satisfaction And Perceptions Of E-Customers According To Demographic Characteristics: Gaziantep Province Case

127-151

Dr. Öğr. Üyesi Vildan ATEŞ, Uzman Zafer KILIÇ

Girişimcinin Perspektifinden Kendi Girişimsel Başarı Değerlendirmesi Ve Bağlam Tartışması

Discussion Of Subjective Success Evaluation And Context From Entrepreneurship Perspective

153-174

Doç. Dr. Umut Sanem ÇİTÇİ, Arş. Gör. Oğuzhan ÖZTÜRK, Dr. M. A. Metin DİNÇER

İÇİNDEKİLER/CONTENTS

Yıl (Year) 2018 Cilt (Vol.) 6 Sayı (No) 3

İşe Yabancılaşma Boyutlarının Demografik Faktörler Açısından İncelenmesi <i>Examining The Dimensions Of Work Alienation In Terms Of Demographic Factors</i>	175-197
<i>Nuran AKŞİT AŞIK</i>	
Farklı Veri Setleri Üzerinde SMO Ve J48 Algoritmalarının Sınıflandırma Sonuçlarının Karşılaştırılması <i>Comparison Of Classification Results Of SMO And J48 Algorithms On Different Data Sets</i>	199-213
<i>Doç. Dr. Mehmet Ali ALAN, Doç. Dr. Cavit YEŞİLYURT</i>	
Etik Liderlerde Duygusal Emegin İşkolikliğe Etkisi <i>The Effect Of Emotional Labour Of Ethical Leaders On Workaholism</i>	215-238
<i>Dr. Öğr. Üyesi Olgun Irmak ÇETİN</i>	
Sağlık Profesyonellerinin Yalın Uygulamalara Direncini Belirlemeye Yönelik Bir Araştırma <i>A Research For Determining The Resistance Of Health Professionals To Lean Applications</i>	239-272
<i>Feryal BULUT, Prof. Dr. Mehmet Selami YILDIZ</i>	
Endüstri 4.0 Ve Çağrı Hizmetleri İlişkisi Üzerine Bir Araştırma <i>A Research On The Factors Affecting The Industrial 4.0 And Call Services</i>	273-294
<i>Dr. Öğr. Üyesi İsa DEMİRKOL, Prof. Dr. Selami ÖZCAN</i>	

FARKLI VERİ SETLERİ ÜZERİNDE SMO VE J48 ALGORİTMALARININ SINIFLANDIRMA SONUÇLARININ KARŞILAŞTIRILMASI

Farklı Veri Setleri Üzerinde SMO ve J48 Algoritmalarının Sınıflandırma Sonuçlarının Karşılaştırılması

199

Doç. Dr. Mehmet Ali ALAN

Cumhuriyet Üniversitesi

alan@cumhuriyet.edu.tr

ORCID ID:0000-0001-8562-547X

Doç. Dr. Cavit YEŞİLYURT

Atatürk Üniversitesi

cavityesilyurt@atauni.edu.tr

ORCID ID: 0000-0001-9814-4085

Öz

Amaç: Veri madenciliği disiplinler arası bir alandır, sürekli gelişmekte ve kullanım alanları yaygınlaşmaktadır. Çeşitli tekniklerin ve algoritmaların kullanılmasıyla verilerin güvenilirliğinin sağlanmasına yardımcı olmaktadır. Sınıflandırma, araştırmacılar tarafından yaygın olarak kullanıldığı için önemli bir veri madenciliği tekniğidir.

Yöntem: Bu çalışmada, üç farklı öğrenci veri seti üzerinde SMO ve J48 algoritmalarının sınıflandırma sonuçları karşılaştırılmıştır. Çalışmada, üç farklı veri seti ile TP-Oranı, FP-Oranı, Kesinlik, Duyarlık, F-ölçütü ve ROC analizi gibi çeşitli doğruluk ölçümleri kullanılarak, J48 ve SMO algoritmalarının sınıflandırma doğruluğu açısından performansı değerlendirilmiştir.

Bulgular ve Sonuç: Yapılan testler sonucunda her üç veri setinde SMO algoritmasının sınıflandırma performansının daha iyi olduğu ortaya konmuştur.

Anahtar kelimeler: Veri Madenciliği, Sınıflandırma, SMO, J48

COMPARISON OF CLASSIFICATION RESULTS OF SMO AND J48 ALGORITHMS ON DIFFERENT DATA SETS

Abstract

Aim: Data mining is an interdisciplinary field, constantly developing and expanding its use. It helps to ensure the reliability of data by using various techniques and algorithms. Classification is an important data mining technique because it is widely used by researchers.

Method: In this study, the classification results of SMO and J48 algorithms were compared on the data of three arc students. The performance of J48 and SMO algorithms in terms of classification accuracy was evaluated using three different data sets and various accuracy measurements such as TP-Rate, FP-Ratio, Precision, Precision, F-criterion and ROC analysis.

Findings and Results: As a result of the tests, it was revealed that the classification performance of SMO algorithm was better in all three datasets.

Key Words: Data Mining, Classification, SMO, J48

I. GİRİŞ

Kurumların veri kaynakları, sosyal medya paylaşımları, web sitelerinde ve formlardaki yazılar büyük miktarlarda veri sağlamaktadır. Bu büyük miktarlardaki verileri klasik yollarla işlemek ve karar süreçlerinde kullanılmak üzere bilgi üretmek oldukça zordur. Bu bağlamda veri madenciliği, sunduğu gelişmiş tekniklerle, mevcut verilerden ihtiyaç duyulan bilgilerin üretilmesini sağlayabilir.

Veri tabanları, rasyonel karar almayı sağlayacak gizli bilgiler bakımından zengindir. Sınıflandırma ve tahmin, gelecek veri trendlerinin tahmini veya önemli veri sınıflarının açıklanmasında kullanılan iki önemli veri analiz tekniğidir. Bu analizler büyük miktarlardaki verilerin daha iyi anlaşılmasında kullanışlı olabilmektedir (Han ve Kamber, 2006:285).

Günümüzde kurumlar büyük miktarlarda veri üretmekte, ancak bu veriler içinde anlamlı ve yararlı bilgiyi ortaya çıkarmakta zorluklar yaşamaktadırlar. Geleneksel istatistik yöntemlerle büyük boyuttaki veriyi çözümlenmek kolay değildir. Bu nedenle verileri işlemek ve çözümlenmek için özel yöntemlere gereksinim duyulmuştur. Veri madenciliği yöntemleri bu gereksinimi karşılamak üzere ortaya çıkmıştır (Özkan, 2008: IV).

Bu çalışmanın amacı, veri madenciliğinin en yaygın kullanılan tekniklerinden sınıflandırma yönteminde kullanılan iki algoritmanın

performanslarını karşılaştırmaktır. Bu amaçla üç farklı öğrenci veri seti kullanılarak, veri madenciliği yapılmıştır.

Çalışma üç bölümden oluşmaktadır. Birinci bölümde veri madenciliği, sınıflandırma, SMO ve J48 algoritmaları hakkında bilgiler sunulmaktadır. İkinci bölümde konuyla ilgili yapılmış çalışmalar gözden geçirilmektedir. Üçüncü bölümde ise mevcut veriler üzerinde veri madenciliği yapılarak, algoritmaların performansları karşılaştırılmıştır.

II. Veri Madenciliği ve Kullanılan Algoritmalar

Veri madenciliği, hem yararlı hem de anlaşılabilir verilerle, alışılmamış yollarla, verileri özetleyen ve gizli ilişkileri ortaya koyan bir analiz yöntemidir (Larose, 2006). Bu yöntem, öncelikle bilinmeyen desenlerin ortaya konması amacıyla bilimsel ve teknik veri araştıran, veri tabanındaki bilgi keşfi süreçlerinden biridir (Rokach ve Maimon, 2005:2).

Veri madenciliği, reklamcılık, biyoinformatik, veritabanı pazarlama, dolandırıcılık tespiti, e-ticaret, sağlık, güvenlik gibi alanların içinde olduğu, farklı alanlarda uygulanabilen, değişik bakış açısı ve çalışması ile veri analizinden bilgi keşfetme süreci olarak bilinir (Jain et al., 2011).

Disiplinler arası nitelik taşıyan veri madenciliğini en yaygın kullanan bilim dalları; veritabanı sistemleri, istatistik, matematik, makine öğrenmesi, görselleme ve bilişim bilimleridir (Han ve Kamber, 2006: 29).

Veri madenciliği, verinin bütünü kullanması bakımından diğer istatistiksel verilerden ayrılmaktadır. Bu yöntemle, geleneksel yollarla elde edilmiş küçük verilerle çalışma yerine daha kolay değerlendirme yapabilecek, yeni bağımsız veriler tercih edilebilmektedir (Weiss ve Zhang, 2003: 426).

Gartner Group'a göre veri madenciliği, istatistiksel ve matematiksel yöntemler kadar desen tanıma teknolojilerinin kullanılmasıyla, muazzam miktarlardaki depolanmış verilerin elenmesi ile yeni anlamlı birliktelikler, desenler ve trendler keşfetme sürecidir (Larose, 2005: 2).

En yaygın veri madenciliği algoritmaları ve modelleri içinde karar ağaçları, sınıflandırma ağaçları olarak da adlandırılır (Bramer, 2007: 6); birliktelik kuralları, kümeleme, sınıflandırma, çoklu lineer regresyon, sıralı örüntüler ve zaman serileri tahmini, örüntü tanıma ve özelliklerinin belirlenmesi sayılabilir. Sınıflandırma, regresyon ve zaman serisi analizleri gizli örüntülerin ortaya çıkarılmasında ve şekillendirilmesinde uygun iken, birliktelik kuralları, kümeleme ve sırasal keşif yaklaşımları, hava tahmini ve şiddeti araştırmak ve tanımlamak için yararlı araçlar olabilir (Tadesse, 2009).

Sınıflandırma, günlük yaşamda çok sıklıkla başvurulan bir işlemdir. Sınıflandırma ile nesnelere bölünerek ayrıştırılır, yani karşılıklı olarak özel ya da genel kategorilerden her biri bir sınıf olarak atanabilir. Pek çok pratik karar verme işlemi, bir sınıflandırma problemi olarak formüle edilebilir. Örneğin kişiler ya da nesnelere birçok kategoriden biri olabilir (Bramer, 2007: 23).

Sınıflandırma, farklı sınıflardaki, değişik öğeleri ayırma sürecidir. Bu sınıflar, iş kuralları, sınıf sınırları veya bazı matematiksel fonksiyonlar olabilir. Sınıflandırma işlemi, sınıflandırılmış olan öğenin, bilinen bir sınıf değeri ile özellikleri arasındaki bir ilişki üzerine bina edilebilir. Bu sınıflandırma tipi, "denetimli öğrenme" olarak isimlendirilir. Eğer bir sınıfın bilinen örnekleri yoksa bu sınıflandırma denetimsizdir. En yaygın denetimsiz sınıflandırma yaklaşımı "kümeleme"dir. Kümeleme teknolojisinin en yaygın uygulamaları, perakende ürünlerde birliktelik analizi (market sepet analizi) ve dolandırıcılık tespitidir (Nisbet, et al., 2009: 235).

Veri madenciliğinde denetimli öğrenme kavramı, bir sınıflandırma ile bilinen veriler temelinde bir sınıflandırma fonksiyonu öğretmek ya da bir sınıflandırma modeli inşa etmektir. Bu fonksiyon ya da model, veri tabanındaki verileri hedef niteliklere dönüştürür, dolayısıyla yeni veriler sınıf tahmininde kullanılabilir (Dong-Peng, et al., 2008:36).

Veri madenciliği, disiplinler arası bir alandır, veri tabanları sistemleri, istatistikler, makine öğrenimi, görselleştirme ve bilgi bilimini de içeren bir dizi disiplinin birleşmesidir. Ayrıca, kullanılan veri madenciliği yaklaşımına bağlı olarak, diğer disiplinlerden gelen teknikler olabilir. Yapay sinir ağları, bulanık ve / veya kaba küme teorisi, bilgi gösterimi, endüktif mantık programlama veya yüksek performanslı hesaplama gibi teknikler uygulanabilir. Veri madenciliği sistemi, madencilik yapılacak veri türlerine veya belirli veri madenciliği uygulamasına bağlı olarak, uzamsal veri analizi, bilgi alımı, model tanıma, görüntü analizi, sinyal işleme, bilgisayar grafikleri, web teknolojisi, ekonomi, işletme, biyoenformatik veya psikoloji gibi alanlarla ilişkilidir (Han ve Kamber, 2006: 30).

2.1. SMO Algoritması

SMO (Sequential Minimal Optimization), herhangi bir ekstra matris depolaması olmaksızın ve sayısal QP optimizasyon adımlarını kullanmadan SVM QP problemini hızlıca çözebilen basit bir algoritmadır. SMO, her adımda mümkün olan en küçük optimizasyon problemini çözmeyi seçer. Standart SVM QP problemi için mümkün olan en küçük optimizasyon problemi iki Lagrange çarpanını içerir, çünkü Lagrange çarpanları lineer bir

eşitlik kısıtlamasına uymalıdır. Her adımda, SMO, ortaklaşa optimize etmek için iki Lagrange çarpanını seçer, bu çarpanlar için en uygun değerleri bulur ve SVM'yi yeni optimal değerleri yansıtacak şekilde günceller.

SMO'nun avantajı, iki Lagrange çarpanının çözümlenmesinin analitik olarak yapılabilmesi gerçeğinde yatmaktadır. Böylece, sayısal QP optimizasyonu tamamen önlenmiş olur. Algoritma sırasında daha fazla optimizasyon alt problemleri çözülsün de, her bir alt problem o kadar hızlıdır ki genel QP problemi çabucak çözülür. Ayrıca, SMO hiç bir ek matris depolama gerektirmez. Bu nedenle, çok büyük SVM eğitim sorunları, sıradan bir kişisel bilgisayarın veya iş istasyonunun hafızasına sığabilir. SMO'da hiçbir matris algoritması kullanılmadığı için sayısal hassasiyet problemlerine daha az duyarlıdır (Platt, 1998).

2.2. J48 Algoritması

J48, J. Ross Quinlan tarafından geliştirilen çok popüler C4.5 algoritması temeline dayanan bir karar ağacı algoritmasıdır. Karar ağaçları bir makine öğrenmesi algoritmasından bilgi temsil etmede klasik bir yoldur ve veri yapılarını ifade etmekte güçlü ve hızlı bir yol sunar. Bu algoritma verileri özyinelemeli olarak sınıflandırır. Bu işlem eğitim verilerinin maksimum doğruluğunu sağlar ama verilerin sadece belirli davranış özelliklerini tanımlayan aşırı kurallar oluşturabilir (Nizam ve Akın, 2014).

J48 Algoritması; Enformasyon Kazancı Teorisine (Information Gain Theory) dayanarak, verilerden ilgili özellikleri seçmek için otomatik işlem yeteneğine sahiptir. Enformasyon kazancının en iyi olduğu noktadan örnekleri bölen yinelemeli algoritmadır. IF-THEN kurallarına dayalı bir karar ağacı ve —üyelik fonksiyon kümeleri –membership functionsets-çıktısı verir. Ağaç yapısı, denekleri bölme ve ağacın en iyi kök değişkeninin seçilmesi süreci ile başlayıp yukarıdan aşağıya doğru inşası gerçekleştirilmektedir. J48, anlamlı olmayan diğer bir deyişle zayıf dalları kesmek için etkin bir budama işlemi yapabilmektedir. Bunun nedenlerinden biri, karar ağaçlarının amacının veri keşfetmek değil, veriler üzerinde basit bir sınıflandırma modeli oluşturmak olmasındandır (Daş ve Varol, 2013).

III. Literatür Özeti

Konuyla ilgili literatürde farklı veri setleri üzerinden yapılmış çalışmalar bulunmaktadır. Bunlardan Singaravelan et al (2015), Weka ara yüzüne desteklenen SMO ve J48 algoritmalarının sınıflandırma performanslarını, üç ayrı veri seti üzerinde test etmişler ve genellikle SMO'nun sınıflandırma performansının daha iyi olduğu sonucunu elde etmişlerdir. Kaur et al (2015), bir liseden alınan gerçek dünya verilerini

kullanarak, WEKA Açık Kaynak Aracı üzerinde, Multilayer Perception, Naïve Bayes, SMO, J48 ve REPTree algoritmalarını kullanarak test etmişler beş sınıflandırıcının karşılaştırılmasında da Multilayer Perceptron algoritmasının performansının daha iyi olduğunu ortaya koymuşlardır. Arora ve Sharma (2016), Üç ayrı tıbbi veri setini, naive bayes, J48 ve SMO algoritmalarını, veri madenciliği aracı WEKA'da kullanarak sınıflandırma performanslarını incelemişler ve en iyi performansı j48 algoritmasının gösterdiğini tespit etmişlerdir. Kamalakkannan ve Ramyachitra (2016), bir veri seti üzerinde (anneal dataset), Naïve Bayes, SMO, OneR ve J48 algoritmalarının verimliliğini test etmişler ve Naive Bayes algoritmasının sınıflandırma performansının daha iyi olduğu sonucuna ulaşmışlardır. Aharwal (2016), dört farklı veri seti üzerinde, J48, Multilayer Perceptron, Naïve Bayes ve SMO sınıflandırıcılarının performanslarını test etmiş ve SMO algoritmasının tüm veri setlerinde en başarılı sınıflandırıcı olduğunu ortaya koymuştur. Chaudhary ve arkadaşları (2015), dört farklı veri seti üzerinde, SMO, Random Forest ve J48 algoritmalarının sınıflandırma performanslarını Weka araçlarını kullanarak test etmişler ve Random Forest algoritmasının daha performanslı olduğunu vurgulamışlardır. Salama ve arkadaşları (2012), üç farklı veri seti üzerinde, Naive Bayes, MLP, J48, SMO ve IBK algoritmalarını, Weka veri madenciliği araçlarını kullanarak test etmişler ve farklı veri setlerinde farklı algoritmaların daha performanslı olduğunu ortaya koymuşlardır.

IV. Veri Seti ve Yöntem

Bu çalışmada üniversite öğrencilerine ait üç farklı veri seti kullanılmıştır. Veriler, Excel makroları kullanılarak, gerekli düzenlemelere tabi tutulmuş ve veri ambarları hazırlanmıştır. Gerekli dönüşümler yapıldıktan sonra veriler "iibf1.arff", "iibf2.arff" ve "myo.arff" adlı metin dosyasına yazdırılmıştır. Veri setlerinin, Örnek sayıları, Nitelik sayıları ve Class sayıları Tablo 1'de sunulduğu gibidir.

**Tablo 1.
Veri Setleri ve Özellikleri**

Veri Setleri	Örnek Sayıları	Nitelik Sayısı	Class Sayısı
<i>iibf1</i>	2597	7	2
<i>iibf2</i>	2235	7	2
<i>Myo</i>	3165	6	2

V. Uygulama

Yapılan çalışmada Waikato Üniversitesince geliştirilmiş olan WEKA Programının (Waikato Environment for Knowledge Analysis) 3.7.2 sürümü

kullanılmıştır (Wekafull paket). WEKA Programı, açık kaynak kodlu bir yazılımdır. Bu program pek çok sınıflandırma, kümeleme ve birliktelik kurallarına ait algoritmayı desteklemektedir. WEKA, metin tabanlı pek dosya tiplerinin yanı sıra, veri tabanlarını ve verilerin olduğu URL adreslerini de desteklemektedir.

Her bir veri seti için, öğrencinin cinsiyeti, ili, aile gelir düzeyi, ebeveynlerin yaşayıp yaşamadığı, kardeş sayısı, okuyan kardeş sayısı ve giriş puanı nitelik olarak alınmıştır. Giriş puanının derecesi sınıf tanımlarında kullanılmıştır.

Mevcut veri setleri ile yapılan 10 kat çapraz doğrulama ve tam eğitilmiş set kullanılarak yapılan uygulama sonucunda her bir veri seti için SMO ve J48 algoritmalarına ait doğruluk değerleri izleyen tablolarda sunulmuştur.

Tablolarda yer alan; TP birinci sınıftaki doğru olarak sınıflandırılmış kayıtların sayısını veren TP (True Positive) oranını, FP (False Positive) birinci sınıfta sınıflandırılmış, ikinci sınıftaki kayıtların oranını vermektedir. Tabloda yer alan Precision (Kesinlik), Recall (Duyarlılık) ve F-Measure (F-ölçütü) aşağıdaki gibi hesaplanmaktadır:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

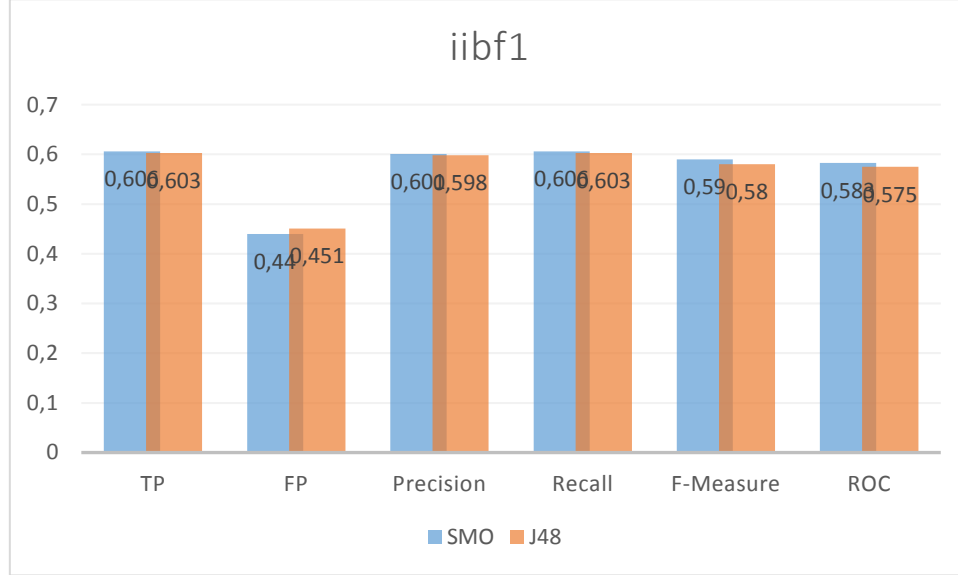
$$F - Measure = \frac{2 \times Recall \times Precision}{Recall + Precision}$$

Kesinlik ve Duyarlılık arasındaki dengeyi değerlendirmek için ROC (Receiver Operating Characteristics) eğrisi ve altında kalan alan kullanılır. ROC puanı 1'e yaklaştıkça pozitifler daha iyi bir şekilde negatiflerden ayrılır (Akçetin ve Çelik, 2014).

Tablo 2'de iibf1 veri seti için doğruluk değerleri sunulmuştur.

Tablo 2.
İİBF1 Veri Seti İçin Sonuçlar

S.No.	Parametreler	SMO	J48
1	TP	0,606	0,603
2	FP	0,44	0,451
3	Precision	0,601	0,598
4	Recall	0,606	0,603
5	F-Measure	0,59	0,58
6	ROC Area	0,583	0,575



Şekil 1.
İİBF1 Veri Setinin Grafiği

Hem Tablo 2, hem de Şekil 1' deki sonuçlara göre iibf1 veri seti için SMO algoritmasının bütün değerlerinin J48 algoritmasına göre daha olumlu olduğu anlaşılmaktadır.

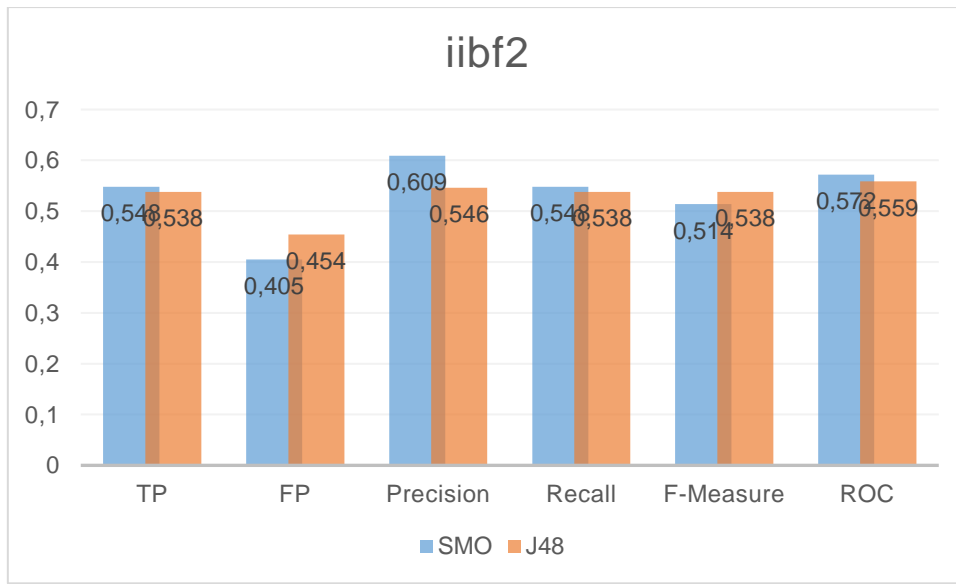
Tablo 3'de iibf2 veri seti için doğruluk değerleri sunulmuştur.

Tablo 3.
İİBF2 Veri Seti İçin Sonuçlar

S.No.	Parametreler	SMO	J48
1	TP	0,548	0,538
2	FP	0,405	0,454
3	Precision	0,609	0,546
4	Recall	0,548	0,538
5	F-Measure	0,514	0,538
6	ROC Area	0,572	0,559

Farklı Veri Setleri Üzerinde SMO ve J48 Algoritmalarının Sınıflandırma Sonuçlarının Karşılaştırılması

207



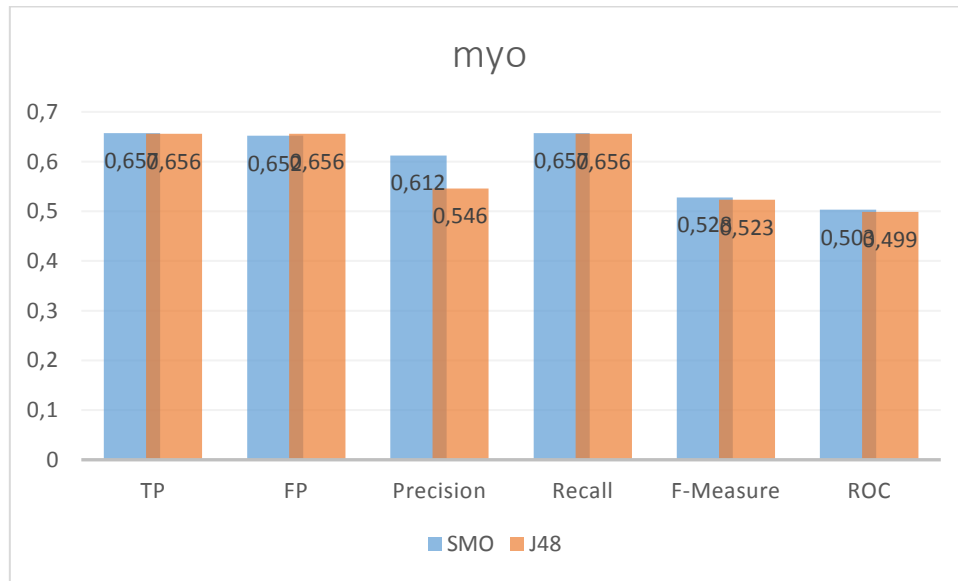
Şekil 2.
İİBF2 Veri Setinin Grafiği

Hem Tablo 3, hem de Şekil 2' deki sonuçlara göre iibf1 veri seti için SMO algoritmasının bütün değerlerinin J48 algoritmasına göre daha olumlu olduğu anlaşılmaktadır.

Tablo 4'te myo veri seti için doğruluk değerleri sunulmuştur.

Tablo 4.
MYO Veri Seti İçin Sonuçlar

S. No.	Parametreler	SMO	J48
1	TP	0,657	0,656
2	FP	0,652	0,656
3	Precision	0,612	0,546
4	Recall	0,657	0,656
5	F-Measure	0,528	0,523
6	ROC Area	0,503	0,499



Şekil 3.
MYO Veri Setinin Grafiği

Hem Tablo 4, hem de Şekil 3' teki sonuçlara göre iibf1 veri seti için SMO algoritmasının bütün değerlerinin J48 algoritmasına göre daha olumlu olduğu anlaşılmaktadır.

VI. Sonuç

Veri madenciliği, gizli, önemli, önceden bilinmeyen, yararlı bilgileri ortaya koyan bir veri analiz tekniğidir. Bu teknikte en yaygın kullanılan

yöntemlerin başında sınıflandırma analizi gelmektedir. Sınıflandırma analizinde ise sınıflandırma derecesi ve bu sınıfların doğruluğunun test edilmesi son derece önemlidir.

SMO ve J48 algoritmaları, veri madenciliğinin sınıflandırma yönteminde kullanılan en popüler algoritmalarındandır. Kullanılan veri setine bağlı olmakla beraber, aynı veri setinde hangi algoritmanın daha iyi sonuç ürettiği önemlidir. Bu çalışmada, üç farklı veri seti ile TP-Oranı, FP-Oranı, Kesinlik, Duyarlık, F-ölçütü ve ROC analizi gibi çeşitli doğruluk ölçütleri kullanılarak, J48 ve SMO algoritmalarının sınıflandırmadaki doğruluk değerleri karşılaştırılmıştır. Yapılan testler sonucunda her üç veri setinde de SMO algoritmasının sınıflandırma sonuçlarının, J48 algoritmasına göre daha iyi olduğu ortaya konmuştur.

Analizde kullanılan veri setlerine göre, SMO algoritmasının sınıflandırmadaki başarı derecesinin daha yüksek olması, bu algoritmayı kıyaslanan J48 algoritmasına göre daha güvenilir kılmaktadır.

KAYNAKÇA

- Aharwal, R. P. (2016). Evaluation Of Various Classification Techniques Of Weka Using Different Datasets. *International Journal of Advance Research and Innovative Ideas in Education*, 2(2): 558-552.
- Akçetin, E. ve Çelik, U. (2014). İstenmeyen Elektronik Posta (Spam) Tespitinde Karar Ağacı Algoritmalarının Performans Kıyaslaması, *İnternet Uygulamaları ve Yönetimi*. (5/2): 43-56.
- Arora, M. ve Sharma, A. (2016). Chronic Kidney Disease Detection by Analyzing Medical Datasets in Weka. *International Journal of Computer Application* 6(4): 20-26.
- Bramer, Max (2007). *Principles of Data Mining*, Springer, London.
- Chaudhary, N., Mehta ve G., Bajaj, K. (2015). Comparison Of Classification Algorithms And Design Of A Percentage-Split Based Method For Data Classification, *International Journal of Computer Science & It*, 2(5):1-6.
- Daş, B. ve Varol, A. (2013). 2D:4D Sayısal Parmak Oranına Göre Bireylerin Kişilik Durumlarının Sınıflandırılması, *International Symposium on Digital Forensics and Security (ISDFS'13)*.
- Dong-Peng, Y., Li Jin-Lin, L. R. ve Chao Z. (2008). Applications of Data Mining Methods in the Evaluation of Client Credibility, *Applications of Data Mining in E-Business and Finance* C. Soares et al. (Eds.), IOS Press, Amsterdam, 35-43.
- Han, J. ve Kamber, M. (2006). *Data Mining: Concepts and Techniques*, Second Edition, Morgan Kaufmann Publications, San Francisco.

**Farklı Veri
Setleri Üzerinde
SMO ve J48
Algoritmalarını
n Sınıflandırma
Sonaçlarının
Karşılaştırılması**

210

- Jain, Y. K., Yadav, V. K. ve Panday, G. S. (2011). An Efficient Association Rule Hiding Algorithm for Privacy Preserving Data Mining. *International Journal On Computer Science And Engineering*, 3(7): 2792-2798.
- Kamalakkannan ,V.ve Ramyachitra, D.(2016). Analysis of Different Classification Algorithms Applied to Anneal Dataset Using Data Mining Techniques, *International Journal of Future Innovative Science and Engineering Research (IJFISER)* , 2(1): 27-134.
- Kaura, P., Singhb, M.ve Josan, G. S. (2015). Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector, *3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015)*, Procedia Computer Science 57: 500-508.
- Larose, D. T. (2005). *Discovering Knowledge In Data*, Wiley Publication, New Jersey.
- Nisbet, R., Elder, J. ve Miner, G. (2009). *Handbook of Statistical Analysis and Data Mining Applications*, Elsevier Inc, Burlington.
- Özkan, Y. (2008). *Veri Madenciliği Yöntemleri*, Papatya Yayınları, İstanbul.
- Platt, J. C. (1998). Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines, *Technical Report MSR-TR-98-14*, <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/tr-98-14.pdf> (01.12.2018).
- Rokach, L. ve Maimon, O. (2008). *Data Mining with Decision Trees*, World Scientific, New Jersey.
- Salama, G., Abdelhalim, M. B. ve Zeid, M. A. (2012). Experimental Comparison of Classifiers for Breast Cancer Diagnosis, *978-1-4673-2961-3/12 ©2012 IEEE*, DOI: 10.1109/ICCES.2012.6408508:180-185.
- Singaravelan, S., Murugan, D. ve Mayakrishnan, R. (2015). Analysis of Classification Algorithms J48 and Smo on Different Datasets, *World Engineering & Applied Sciences Journal*, 6(2): 119-123.
- Tadesse, T., Wardlow, B. ve Hayes, M.J. (2009). The Application of Data Mining for Drought Monitoring and Prediction, *Data Mining Applications for Empowering Knowledge Societies*, Edited by Hakikur Rahman, Information Science Reference, New York, 280-291.
- Weiss, S. M. ve Zhang, T. (2003). Performance Analysis and Evaluation, *The handbook of Data Mining*, Edited by. Nong Ye, Lawrence Erlbaum Associates Publishers. London, 436-439.

COMPARISON OF CLASSIFICATION RESULTS OF SMO AND J48 ALGORITHMS ON DIFFERENT DATA SETS

Farklı Veri Setleri Üzerinde SMO ve J48 Algoritmalarının Sınıflandırma Sonuçlarının Karşılaştırılması

EXTENDED ABSTRACT

The data sources of institutions, social media shares, articles on websites and forms provide large amounts of data. It is very difficult to process large amounts of data in traditional ways and to produce information for use in decision processes.

In this context, data mining can provide the production of the information needed from the available data with the advanced techniques that it offers.

Databases are rich in confidential information that will enable rational decision-making. Classification and estimation are two important data analysis techniques used for estimating future data trends or explaining important data classes. These analyzes can be useful in better understanding of large amounts of data. Today, institutions produce large amounts of data, but they have difficulties in revealing meaningful and useful information within these data. It is not easy to analyze large data with traditional statistical methods. Special methods are therefore required to process and analyze data. Data mining methods have emerged to meet this requirement.

The aim of this study is to compare the performances of the SMO and J48 algorithms used in the classification of data mining. For this purpose, data mining was performed by using three different student data sets.

Data mining is an analysis method that summarizes data and exposes hidden relationships with both useful and understandable data, in unusual ways. This method is one of the processes of knowledge discovery in the database, which first explores scientific and technical data to reveal unknown patterns. Classification is a process that is frequently used in daily life. By classification, the objects are split and separated, that is, each of the mutually exclusive or general categories can be assigned as a class. Many practical decision-making processes can be formulated as a classification problem. For example, people or objects can be one of many categories. Classification is the process of assigning different elements in different classes. These classes may be business rules, class boundaries, or some mathematical functions. The classification process can be constructed on a relationship between a class of the classified element and a known class value and properties. This type of classification is called "supervised

211

Farklı Veri Setleri Üzerinde SMO ve J48 Algoritmalarının Sınıflandırma Sonuçlarının Karşılaştırılması

212

learning". If there are no known examples of a class, this classification is unsupervised. The most common uncontrolled classification approach is clustering. The most common applications of clustering technology are retail basket analysis and fraud detection.

The concept of controlled learning in data mining is to teach a classification function on the basis of known data with a classification or to construct a classification model. This function or model converts data from the database into target attributes, so new data can be used in class estimation. The data mining system relates to areas such as spatial data analysis, information retrieval, model recognition, image analysis, signal processing, computer graphics, web technology, economics, business, bioinformatics or psychology, depending on the types of data to be mining or the specific data mining application.

SMO (Sequential Minimal Optimization) is a simple algorithm that can quickly solve the SVM QP problem without any extra matrix storage and without using numerical QP optimization steps. SMO chooses to solve the smallest possible optimization problem at every step. The smallest possible optimization problem for the standard SVM QP problem involves two Lagrange multipliers because the Lagrange multipliers must comply with a linear equality constraint. At each step, the SMO selects two Lagrange multipliers to jointly optimize it, finds the most appropriate values for these multipliers and updates the SVM to reflect the new optimal values. The advantage of SMO lies in the fact that the analysis of two Lagrange multipliers can be done analytically. Thus, numerical QP optimization is completely prevented. Although more optimization sub-problems are solved during the algorithm, each sub-problem is so fast that the general QP problem is solved quickly. Furthermore, SMO does not require any additional matrix storage. Therefore, very large SVM training problems can fit into the memory of an ordinary personal computer or workstation. SMO is less sensitive to numerical sensitivity problems since no matrix algorithm is used.

J48 is a decision tree algorithm based on the very popular C4.5 algorithm developed by J. Ross Quinlan. Decision trees are a classic way of representing information from a machine learning algorithm and provide a powerful and fast way to express data structures. This algorithm classifies the data recursively. This ensures the maximum accuracy of the training data, but it can only create extreme rules that define the specific behavior characteristics of the data. J48 Algorithm; Based on the Information Gain Theory, it has the ability to automatically process the data to select the relevant properties. It is the iterative algorithm that divides the samples from

the point where information gain is the best. The tree structure starts with the process of dividing the subjects and selecting the best root variable of the tree and building it from top to bottom. The J48 is able to perform an effective pruning process to cut weak branches, which is not meaningful. One of the reasons is that the purpose of decision trees is not to discover data, but to create a simple classification model on the data.

In this study, three different data sets of university students were used. The data were subjected to the necessary regulations using Excel macros and data warehouses were prepared. After making the necessary conversions, the data is printed in the text file "iibf1.arff", "iibf2.arff" and "myo.arff". In the study, the WEKA Program (Waikato Environment for Knowledge Analysis) version 3.7.2 developed by the University of Waikato was used. For each data set, the student's gender, province, family income level, the number of siblings, number of siblings studying, and entry point were taken as qualifications. The degree of entry score is used in the class definitions.

According to the data results, the success rate of the SMO algorithm in the classification is higher compared J48 algorithm, making this algorithm more reliable.

Key Words: Data Mining, Classification, SMO, J48

**Farklı Veri
Setleri Üzerinde
SMO ve J48
Algoritmalarının
Sınıflandırma
Sonuçlarının
Karşılaştırılması**

213