Pamukkale Univ Muh Bilim Derg, 24(4), 713-719, 2018

Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi

Pamukkale University Journal of Engineering Sciences

Bir iplik üretim tesisinde nitelik seçimi ve sınıflandırma ile iplik kalitesinin belirlenmesi

Determining the yarn quality by feature selection and classification in a yarn production facility

Pınar Zarif TAPKAN[1*], *Tayfun ÖZMEN*[2]

[1]Endüstri Mühendisliği Bölümü, Mühendislik Fakültesi, Erciyes Üniversitesi, Kayseri, Türkiye.
pinartan@erciyes.edu.tr

[2]Orta Anadolu T.A.Ş., Osman Kavuncu Mah., Mensucat Cad. No:24, Melikgazi/Kayseri, Türkiye.
ozmentayfun@hotmail.com

Geliş Tarihi/Received: 09.05.2017, Kabul Tarihi/Accepted: 24.11.2017 doi: 10.5505/pajes.2017.37267
* Yazışılan yazar/Corresponding author
Araştırma Makalesi/Research Article

Öz

Günümüzde bilgisayar teknolojisi hızla ilerlemekte, bilgisayar kapasiteleri artmakta, bu da bilgi kaydı yapılan alan sayısını artırarak veriye ulaşmayı kolaylaştırmaktadır. Ancak üretilen ve kaydedilen veriler tek başlarına bir anlam ifade etmemekte, belli bir amaç doğrultusunda işlendiği zaman anlamlı hale gelmektedir. Ham veriyi anlamlı bilgiye dönüştürme işlemleri ise veri madenciliği ile yapılabilmektedir. Bu çalışmada bir iplik üretim tesisinde veri madenciliği yöntemlerinden biri olan sınıflandırma ile kural çıkarımı gerçekleştirilmiştir. Sınıflandırma öncesinde iplik kalitesine etki eden nitelikler belirlenmiş, Taguchi deneysel tasarım yöntemi ile etkin nitelikler tespit edilerek nitelik seçimi yapılmıştır. Kural çıkarımı aşaması ise hem yanlış sınıflandırma hataları sayısını en küçüklemeyi amaçlayan maliyete-duyarsız, hem de beklenen yanlış sınıflandırma maliyetini en küçüklemeyi amaçlayan maliyete-duyarlı sınıflandırma şeklinde uygulanmıştır. Sınıflandırma ile kural çıkarımı için Weka 3.8.1 ve MT-VeMa 1.0 paket programları kullanılmıştır. Elde edilen kurallar, kaliteli iplik üretimi için işletmeye yol gösterici özellikte olmuştur. Bu çalışma ile veri madenciliği ve deneysel tasarım uygulamalarının, bir tekstil şirketinde gerçek verilerle nasıl sonuca ulaştığı gösterilmiş ve ilgili sürece katkıda bulunulmuştur.

Anahtar kelimeler: Nitelik seçimi, Taguchi deneysel tasarım yöntemi, Veri madenciliği, Sınıflandırma

Abstract

Nowadays, computer technology is rapidly advancing, computer capacities are increasing, which makes it easier to reach the database by increasing the number of information recording areas. However, when the produced and recorded data are meaningless on their own, they become meaningful when processed for a certain purpose. Converting raw data to meaningful information can be done by data mining. In this study, rule extraction is realized in a yarn production facility by classification which is one of the data mining methods. Prior to classification, the features that affect the yarn quality are determined, and feature selection is realized by choosing the effective features by Taguchi experimental design method. Rule extraction phase is applied for both cost-insensitive classification that aims to minimize the number of misclassification errors and cost-sensitive classification that aims to minimize the expected misclassification cost. For rule extraction Weka 3.8.1 and MT-VeMa 1.0 package programs are used. The resulting rules guide the firm for producing qualified yarns. This study presents how data mining and experimental design applications at a textile firm have been achieved with actual data and the contributions to the processes of the firm.

Keywords: Feature selection, Taguchi experimental design, Data mining, Classification

1 Giriş

Bilgisayar sistemleri ile üretilen veriler tek başlarına bir anlam ifade etmezler, ancak belli bir amaç doğrultusunda işlendiklerinde anlamlı hale gelirler. Önemli olan geçmişe ait olaylara dair gizli bilgileri keşfetmek, ileriye yönelik durumsal öngörüler veren modeller ile önceden tedbir almayı sağlayacak bir yönetim anlayışına geçmek ve olası kayıpları öngörebilmektir. Bu yüzden büyük miktardaki verileri işleyebilen teknikler büyük önem kazanmaktadır [1]. Veri madenciliği, bu gibi durumlarda kullanılan büyük miktardaki veri kümelerinde saklı bulunan örüntü ve eğilimleri keşfetme işlemidir [2]. Diğer taraftan bilgisayar teknolojisindeki gelişmelerle birlikte donanımın ucuzlaması, verilerin uzun süre depolanmasına ve böylece büyük ölçekli veri tabanlarının oluşmasına neden olmuştur. Bu tür veri tabanları stratejik öneme sahip bilgileri gizlemektedir. Veri madenciliği, büyük veri tabanlarındaki gizli bilgi ve yapıyı açığa çıkarmak için çok sayıda veri analizi aracını kullanan bir süreçtir [3].

Ayrıca yaşanan hızlı değişim ortamı ve günümüz ekonomik koşullarında, tecrübe ve önsezilere dayanarak alınan kararların yanlış olma olasılığı oldukça fazladır. Yanlış karar alma riskini azaltmanın tek yolu bilgiye dayalı yönetimi öngören karar destek çözümleridir [4]. Klasik karar destek sistemi araçlarından farklı olarak, veri madenciliği ile çok daha kapsamlı analizler yapılabilmekte ve böylece işletmelerin daha etkin kararlar alması mümkün olabilmektedir [5].

Veri madenciliği modelleri, gördükleri işlevlere göre sınıflandırma ve regresyon, kümeleme, birliktelik kuralları olmak üzere temel olarak 3 grupta toplanır. Sınıflandırma ve regresyon modelleri tahmin edici, kümeleme ve birliktelik kuralları modelleri tanımlayıcı özelliktedir. Tahmin edici modeller ile sonuçları bilinen verilerden hareket edilerek bir model geliştirilmesi ve kurulan bu modelden yararlanılarak sonuçları bilinmeyen veri kümeleri için sonuç değerlerin tahmin edilmesi amaçlanmaktadır. Tanımlayıcı modeller ise karar vermeye rehberlik etmede kullanılabilecek mevcut verilerdeki örüntülerin tanımlanmasını sağlamaktadır. Tanımlayıcı özelliğe sahip kümeleme yöntemi, örnekler arasındaki yakınlık veya benzerliğin uygun niteliğine göre, örnek kümesinin belirli parça veya kümelere gruplanması

süreci olarak tanımlanır [6]. Birliktelik kuralları ise eşzamanlı gerçekleşen ilişkilerin tanımlanmasında kullanılır.

Taguchi deneysel tasarım yöntemi ise bir süreç en iyileme tekniği olarak geliştirilmiş olup kalite sistemini üretim öncesi ve üretim süreci olarak ikiye ayırarak bir ürünün kalitesinin ve müşteri memnuniyetinin, üretim öncesindeki aşamada tasarım ve geliştirmenin mükemmelliği ile yakından ilgili olduğunu göstermiştir [7]. Taguchi yöntemi farklı parametrelerin, farklı seviyeleri arasından en iyi kombinasyonu belirlemek için oldukça kullanışlı bir yöntemdir. Her parametrenin, her seviyesini içeren tüm kombinasyonlar için oldukça fazla deneysel çalışma yapılması gereken durumlarda Taguchi yöntemi ile dikey dizi tablosu kullanılarak çok daha az sayıda deneysel çalışmayla sonuca ulaşmak mümkündür [8].

Bu çalışmada bir iplik üretim tesisinde üretilen ipliğin kalitesi, nitelik seçimi ve sınıflandırma yöntemleri kullanılarak belirlenmiştir. Deney tasarımının amacı hedeflenen veya araştırılan sonuçları ve bu sonuçlardan meydana gelen sapmaları doğru bir şekilde analiz etmek için kullanılan deney grubunu etkili bir şekilde planlamaktır. Dolayısıyla bu çalışmada işletmeye esneklik kazandıracak sınıflandırma kuralları elde edilmeden önce Taguchi deneysel tasarım yöntemi ile nitelik seçimi gerçekleştirilmiş; ardından kaliteli iplik üretimini gerçekleştirebilmek için işletme tarafından kullanılabilecek kurallar, sınıflandırma yöntemi ile belirlenmiştir. Literatürde Taguchi deneysel tasarım ve sınıflandırma yöntemlerinin bir arada kullanımı incelendiğinde farklı uygulama alanları ile karşılaşılmaktadır. Chuang ve diğ. [9] bir hastanın kanser olma durumunu etkileyen genlerin oldukça fazla sayıda olduğu ancak literatürde bulunan kanser verilerinin hastanın kanser olup olmadığını belirleyemeyecek kadar küçük boyutta olduğu gerçeğinden yola çıkarak Taguchi yöntemini de kullanan bir nitelik indirgeme çalışması gerçekleştirmişler, ardından ise k-en yakın komşu algoritması ile sınıflandırma yapmışlardır. Elde edilen deneysel sonuçlar önerilen yöntemin etkin nitelikleri belirmede oldukça yararlı olduğunu ve yüksek doğrulukta sınıflandırma yapabildiğini göstermiştir. Kim ve diğ. [10] ise kalp ritim bozukluklarının tespiti için tam faktöriyel deney tasarımı yapmanın oldukça zaman alıcı olması sebebiyle olası kombinasyon/deney sayısını azaltmak için Taguchi deneysel tasarım yöntemini kullanan yeni bir sınıflandırma yöntemi önermişlerdir. Deneysel sonuçlar önerilen yöntemin önceki çalışmalarla rekabet edebilecek düzeyde olduğunu ve hesapsal karmaşıklık açısından önerilen yöntemin daha etkin olduğunu göstermiştir. Lee ve diğ. [11] internette arama yaparken çok fazla bilgi olması sebebiyle en faydalı bilgiyi en hızlı şekilde elde edebilmek için öncesinde Taguchi deneysel tasarım yöntemini kullanan basitleştirilmiş sürü optimizasyonuna dayalı bir sınıflandırma algoritması önermişlerdir. Elde edilen sonuçlar genetik algoritma, k-en yakın komşu algoritması ve bayes sınıflandırıcı ile karşılaştırılmış ve önerilen yöntemin daha iyi performans gösterdiği sonucuna ulaşılmıştır.

Mevcut çalışmanın 2. bölümünde iplik kalitesini etkileyen niteliklerden bahsedilmiş, Taguchi deneysel tasarım yöntemi ile bu niteliklerin önem sıraları belirlenerek nitelik seçimi gerçekleştirilmiştir. Bölüm 3'te nitelik seçimi sonrasında elde edilen veri kümesi çeşitli sınıflandırma algoritmalarına tabi tutularak iplik kalitesini etkileyen niteliklerden sınıflandırma kuralları elde edilmiştir. Sonuç bölümünde ise elde edilen bulgular değerlendirilmiş, işletmenin edindiği faydalar ortaya konmuştur.

2 İplik kalitesine etki eden niteliklerin Taguchi deneysel tasarım yöntemi ile belirlenmesi

Uygulama çalışması pamuklu denim giyim kumaşı üretimi yapan Orta Anadolu Tekstil'de gerçekleştirilmiştir. Firma faaliyet alanı içerisinde Türkiye'de birinci, Avrupa'da ikinci konumda bulunmakta olup üretiminin %70'ini başta Avrupa Topluluğu ülkeleri olmak üzere çeşitli dünya ülkelerine ihraç etmektedir. Diğer taraftan Orta Anadolu Tekstil dünya denim kumaş pazarının %1'ini elinde tutmaktadır. Firmanın iplik üretim bölümündeki işleyiş prosesi Şekil 1'deki gibidir.



Şekil 1: İplik üretim prosesi şematik çizimi.

İşletmeye gelen pamuk, pamuk ambarında stoklanır ve giriş testlerine tabi tutulur. Harman-hallaç makineleri olan Unifloc, Unimix, Erm ve Tarak makinelerinden sırasıyla geçirilerek temizlenip şerit haline getirilir. Kovalara yerleştirilen şeritler Cer-1 ve Cer-2 makinelerinde düzgünleştirilir. Bu şeritler kovalar halinde Fitil makinesine gelerek inceltilip büküm verilir ve fitil şeklinde makaralara sarılır. Daha sonra makaralar Ring İplik makinesine bağlanarak eğrilerek inceltilip iplik elde edilir. Bu süreç içerisinde iplik kalitesine birçok nitelik etki eder ancak Ring İplik makinesi, ipliği son haline getiren makine olması sebebiyle, iplik kalitesinde en çok etkiye sahiptir. Ring İplik makinesinde iplik kalitesine etki eden nitelik ve seviyeler işletme mühendisi ve tekstil makineleri literatürü baz alınarak belirlenmiş olup belirlenen 6 niteliğin her birine ait üçer seviye vardır. Bu nitelik ve seviyeler Tablo 1'deki gibidir.

Her makinenin 2.5 sa. 100 kg ve günlük yaklaşık 1 ton iplik ürettiği, iplik maliyetinin 5 $/kg olduğu düşünüldüğünde, 76 adet Ring İplik makinesi olan işletmede iplik kalitesinde oluşabilecek sorunlar ciddi maliyet kayıplarına neden olmaktadır. Dolayısıyla belirlenen 6 niteliğin her birinin iplik kalitesini ne kadar etkilediği ve hangi seviyesinde kaliteli iplik üretimine sebep olduğu cevabı mutlaka bilinmesi gereken sorulardır. İşletmenin bu soruları deneme yanılma ile çözmeye çalışması hem uzun sürmekte hem de maliyetli olmaktadır. Çünkü üretilen kalitesiz iplik hem para hem de zaman kaybına sebep olmaktadır. Bu çalışmada zaman ve para kaybını en az düzeye indirecek şekilde bir bilgi işleme sistematiği getirmek hedeflenmektedir. İşletmede iplik kalitesini ölçme testi, uluslararası bir standart olan Uster standartlarına göre değerlendirilmektedir. İşletme kalite testlerini en son teknoloji Uster Tester 5 cihazı ile yapmakta ve ince yer, kalın yer, tüylülük ve neps değerleri incelenmektedir. Ayrıca standardın bir ifadesi olan ve bahsedilen 4 değerin çeşitli katsayılarla çarpılması ile oluşturulan bir Uster değeri de bulunmaktadır. İşletme bahsedilen 5 değeri de inceleyerek ipliğin kalitesine karar vermektedir. Bu değerlerden en az bir tanesinin sonucunun olumsuz çıkması o ipliğin kalitesinin düşük olacağı anlamına gelmektedir.

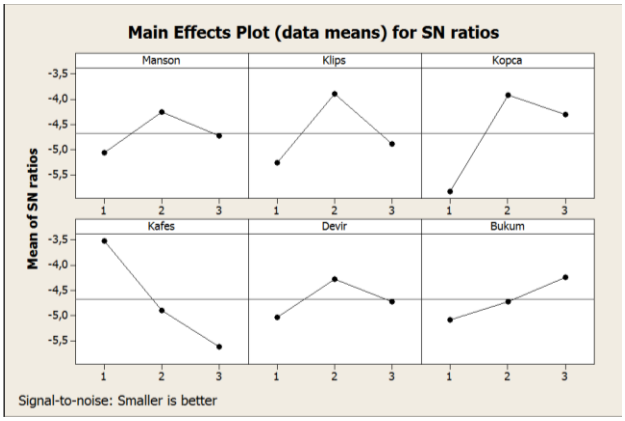
Tablo 1: İplik kalitesine etki eden nitelik ve seviyeleri.

Seviye/Nitelik	Manşon	Klips	Kopça	Kafes	Devir	Büküm
1	Yumuşak	3.5 mm	h2dr02	Dar	7500	581
2	Orta	4 mm	h2dr04	Mevcut	8000	618
3	Sert	4.5 mm	h2dr08	Geniş	8500	655

Bu çalışmada işletmede sürekli üretilen ve diğer iplikler konusunda da fikir sahibi edebilecek KLZ140NA (K: Karde, L: Likralı, Z: Z büküm, 140: 14 Ne, N: Normal iplik, A: İşletme versiyonu) iplik tipi üzerinde uygulama çalışması gerçekleştirilmiştir. Ring pamuk iplikleri temel olarak penye pamuk iplik ve karde pamuk iplik olarak ikiye ayrılmaktadır. Bu çalışmada penye pamuk iplik üretimine göre daha kısa pamuk elyafından üretilmiş ve tarama işlemi yapılmamış pamuk ipliği olan karde iplik kullanılmıştır. İpliğin likralı olması ise yüksek uzama yeteneğine sahip olduğunu göstermektedir. Diğer taraftan ipliklerin mukavemetleri belli bir miktar büküm uygulamak suretiyle sağlanmakta ve uygulanan burulmanın yönüne göre büküm S ve Z ile ifade edilir. Z büküm, dikey tutulan tek kat iplikte meydana gelen spirallerin Z harfinin ortasındaki doğru ile aynı yönde olduğunu göstermektedir. 140, ipliğin inceliğini, normal iplik ibaresi ise motordan gelen hareketin dişliler yardımıyla çekim silindirlerine hareket vermesi neticesinde üretimin gerçekleştiğini ifade etmektedir. Son olarak işletme versiyonu olan A ise farklı pamuk ve likra lotlarına göre işletme tarafından değiştirilen bir ifadedir. Testlerin sonucunun değişkenlik göstermesini engellemek için 75 nolu Ring İplik makinesi ve 1000-1005 iğleri kullanılmıştır. Diğer taraftan yukarıda verilen bilgiler ışığında 6 niteliğin ve 3 seviyesinin birbiri ile etkileşimini anlamak için 3^6=729 adet test yapılması gerekmektedir. Günde 4 adet test yapılabildiği göze alındığında bu gözlem için 182 gün gerekmektedir. Diğer taraftan testlerin olumsuz çıkması durumunda 1 ton/gün*5 $/kg=5000 $/gün para kaybı ortaya çıkacaktır. Test sayısını azaltmak için nitelik sayısının indirgenmesi ya da niteliklerin en etkili seviyelerinin belirlenmesi gerekmektedir. Bu bağlamda Taguchi deneysel tasarım yöntemi niteliklerin önem sıralarının ve en etkili seviyelerinin belirlenmesi için kullanılmıştır. Taguchi yönteminin gerek geniş kullanım alanına sahip olması, gerekse daha az deney yaparak hem zaman kazancı, hem de daha az maliyetle sonuçların elde edilmesine imkân sağlaması gibi avantajlar sunması sebebiyle mevcut çalışma için uygun bir yöntemdir. Mevcut çalışmada iplik kalitesine etki eden 6 parametre ve her parametrenin 3 seviyesi olduğundan en uygun Taguchi dikey dizisi L27'dir. Analizlerin sağlıklı olması açısından her test 6 kez tekrarlanmış, dolayısıyla 27*6=162 deney gerçekleştirilmiştir. L27 tasarımındaki nitelik seviyeleri kullanılarak Minitab 14 paket programında gerçekleştirilen deneyler sonucunda elde edilen *en küçük en iyi* durumu için hesaplanan S/N oranları Tablo 2'de yer almaktadır. Tablo 2'deki S/N oranı süreç değişkenliğinin bir ölçüsü olup S/N oranı kullanılarak sürecin optimize edilmesiyle, optimum süreç şartlarının dayanıklı ve kararlı olması yani mimimum süreç değişkenliğine sahip olması sağlanır. S/N oranı *en küçük en iyi* kriteri ile Eşitlik(1)'deki gibi hesaplanmakta ve y_i performans yanıtının *i.* gözlem değerini, *n* bir denemedeki test sayısını, $\hat{y}$ gözlem değerlerinin ortalamasını, S^2 gözlem değerlerinin varyansını ifade etmektedir.

$$S/N = -10\, log\left(\frac{1}{n}\sum_{i=1}^{n} y_i^2\right) \quad (1)$$

S/N oranları baz alındığında %95 güven aralığında Kafes, Kopça ve Klips niteliklerinin üçünün de sonuçlar üzerinde etkili nitelikler olduğu ($p<0.05$); Büküm, Manşon ve Devir niteliklerinin ise etkisiz nitelikler olduğu görülmektedir. Ayrıca uyarlanmış R^2 değerinden modelin tahmin gücü %85.5 olarak elde edilmiştir. Niteliklerin birbirleri ile etkileşimlerine bakıldığında ise S/N oranları baz alındığında %95 güven aralığında Klips ile Devir arasında düşük ihtimalli de olsa bir etkileşim vardır. Ancak Tablo 2'deki Devir niteliğine ait p değerinin güven aralığı dışında olması ve önem sıralamasında son sırada yer alması sebebiyle bu etkileşim göz ardı edilmiştir. S/N oranları için elde edilen ve Şekil 2'de sunulan ana etki grafiğinden ve Tablo 2'deki koyu işaretlenmiş değerlerden görüldüğü gibi en iyi nitelik-seviye bileşkeleri Kafes-3, Kopça-1, Klips-1, Büküm-1, Manşon-1 ve Devir-1 olarak belirlenmiştir.



Şekil 2: S/N oranları için ana etkiler grafiği.

Böylece Taguchi deneysel tasarım yöntemi ile mevcut 6 niteliğin hangilerinin iplik kalitesi üzerinde etkili, hangilerinin etkisiz olduğu ve her niteliğin en iyi seviyeleri elde edilmiştir.

3 İplik kalitesine etki eden niteliklerden sınıflandırma ile kural çıkarımı

Bir veri madenciliği modeli olan sınıflandırma ile elde edilen kurallar sayesinde iplik kalitesini olumlu etkilediği bilinen belli bir nitelik seviyesinin kullanımının mümkün olmaması durumunda, iplik kalitesini istenen seviyeye getirebilmek için ilgili niteliğin mümkün olan seviyesi ile diğer nitelik seviyelerinin uygun kombinasyonları önceden bilinebilmekte, böylece işletmeye esneklik kazandırılmaktadır.

Örneğin işletmede 4.5 mm klips mevcut olmayıp yerine 3.5 mm klips kullanıldığında iplik kalitesinin olumlu neticelenebilmesi için kafes genişliği niteliğinin hangi seviyede tutulması gerektiği sınıflandırma kuralları ile belirlenebilecektir.

Tablo 2: Taguchi deneysel tasarım sonuçları özet tablosu.

Seviye	Manşon	Klips	Kopça	Kafes	Devir	Büküm
1	**-5.063**	**-5.258**	**-5.827**	-3.519	**-5.029**	**-5.084**
2	-4.245	-3.887	-3.911	-4.900	-4.276	-4.718
3	-4.726	-4.888	-4.296	**-5.614**	-4.728	-4.231
Sıra	5	3	2	1	6	4
p değeri	0.081	0.008	0.001	0.001	0.105	0.070

Sınıflandırma ile kural çıkarımı aşamasına geçmeden önce Taguchi deneysel tasarım yöntemi ile iplik kalitesine etkinliği tespit edilen Kafes, Klips ve Kopça niteliklerinin tüm seviyelerinin değerlendirileceği, etkisi olmayan Büküm, Manşon ve Devir niteliklerinin ise en iyi seviyelerinde sabit tutulacağı yeni bir test kümesi oluşturulmuştur. Böylece etkin olan 3 niteliğin birbirleri ile etkileşimlerini değerlendirecek $3^3 = 27$ test yapılması yeterli olmuştur. Ancak, sınıflandırma ile kural çıkarımı aşamasında analizlerin sağlıklı olması açısından her test yine 6 kez tekrarlanmış, dolayısıyla 27*6=162 adet test gerçekleştirilmiştir. Netice olarak sınıflandırma yapılacak veri kümesinin boyutu etkisiz niteliklerin en iyi seviyelerinde olduğu, etkili niteliklerin ise olası bütün seviyelerinin ele alındığı 162 veriyi içermektedir.

Taguchi deneysel tasarım yöntemi ile elde edilen, işletmeye ait veriler hem maliyete-duyarsız hem de maliyete-duyarlı sınıflandırmaya tabi tutularak kurallar elde edilmiştir. Maliyete-duyarsız sınıflandırma, yanlış sınıflandırma hataları sayısını en küçüklemeyi; maliyete-duyarlı sınıflandırma ise beklenen yanlış sınıflandırma maliyetini en küçüklemeyi amaçlamaktadır. Maliyete-duyarsız sınıflandırmada yanlış sınıflandırma maliyetleri arasında bir fark yokken, maliyete-duyarlı sınıflandırmada belli bir sınıfa ait bir örneği başka bir sınıfta tahmin etmenin maliyeti, tahmin edilen sınıfa göre farklılık göstermektedir. Diğer bir deyişle, maliyete-duyarsız sınıflandırma, maliyete-duyarlı sınıflandırmanın özel bir şeklidir [12]. Pozitif ve negatif olarak adlandırılan iki sınıf içeren ikili-sınıflandırmaya dayalı örnek bir maliyet matrisi Tablo 3'te verilmiştir [13].

Tablo 3: İkili-sınıflandırma için maliyet matrisi örneği.

	Tahminlenen Negatif	Tahminlenen Pozitif
Gerçek negatif	C(0.0) TN	C(1.0) FP
Gerçek pozitif	C(0.1) FN	C(1.1) TP

Örnek maliyet matrisinde FP yanlış pozitif yani gerçekte negatif ancak pozitif olarak tahmin edilme, FN yanlış negatif yani gerçekte pozitif ancak negatif olarak tahmin edilme, TP doğru pozitif yani gerçekte pozitif ve pozitif olarak tahmin edilme ve TN doğru negatif yani gerçekte negatif ve negatif olarak tahmin edilme durumlarını göstermektedir. *C(i,j)* aslında sınıfı *i* olan bir örneği *j* sınıfı olarak tahmin etmenin maliyetini; *C(i,i)* değerleri ise fayda olarak nitelendirilmekte olup doğru tahmin edilen örnekleri temsil etmektedir. Verilen örnekte 1 pozitif sınıfı, 0 ise negatif sınıfı temsil etmektedir. Verilen bu maliyet matrisine bağlı olarak eldeki örnekler minimum beklenen maliyet değerlerine göre sınıflandırılacaktır. *x* örneğini *j* sınıfı olarak sınıflandırmanın beklenen maliyeti Eşitlik (2)'deki gibi tanımlanmaktadır.

$$R(j/x) = \sum_i P(i/x) C(i,j) \qquad (2)$$

Verilen eşitlikte *P(i/x)*, *x* örneğini *i* sınıfı olarak sınıflandırmanın olasılık tahminini göstermektedir. Mevcut çalışmanın maliyete-duyarlı sınıflandırılabilmesi için oluşturulan maliyet matrisi Tablo 4'teki gibidir.

Tablo 4: Mevcut çalışmaya ait maliyet matrisi.

	Tahminlenen Negatif	Tahminlenen Pozitif
Gerçek Negatif	0	2200
Gerçek Pozitif	500	0

Ring İplik makinesi 1 takımda 100 kg iplik üretmektedir. Üretilen ipliğin satış fiyatı 5 $/kg'dır. Üretilen 100 kg iplik ile 666 m kumaş üretilmektedir. Kumaşın satış fiyatı 3.30 $/m'dir. Sonuç olarak eğer gerçek sonuç pozitif iken, negatif değerlendirilirse 100 kg iplik hurdaya ayrılmakta ve 500 $ maliyete neden olmaktadır. Eğer gerçek sonuç negatif iken pozitif kabul edilirse, 666 m kumaş hurdaya ayrılmakta, bu da 2200 $ maliyete neden olmaktadır.

Belirlenen maliyet matrisi ile Taguchi deney tasarımı neticesinde oluşturulan veri kümesi, maliyete-duyarsız ve maliyete-duyarlı olarak Weka 3.8.1 ve MT-VeMa 1.0 veri madenciliği paket programları kullanılarak sınıflandırılmıştır. WEKA (Waikato Environment for Knowledge Analysis), Waikato Üniversitesi tarafından geliştirilmiş bir makine öğrenme ve veri madenciliği yazılımıdır. Akademik araştırmalar, eğitim ve endüstriyel uygulama alanlarında kullanım yeri olan WEKA, veri analizi ve tahminleyici modelleme için geliştirilmiş algoritma ve araçların görsel bir birleşimini içerir. Geliştirilen yazılımın temel avantajları geniş veri önişleme ve modelleme tekniklerine sahip olması, grafiksel kullanıcı ara yüzü sayesinde kullanımının kolay olması ve Java programlama dili ile uygulandığından herhangi bir platformda kullanılabilmesi yani taşınabilir olmasıdır [14]. MT-VeMa 1.0 ise metasezgisel yöntemler içeren BEE-miner, MEPAR-miner ve DIFACONN-miner algoritmalarının ortak bir platformda kullanımına imkan vermektedir. İlgili algoritmalar hem maliyete-duyarsız hem de maliyete-duyarlı sınıflandırma kabiliyetine sahiptir.

Maliyete-duyarsız sınıflandırma için WEKA 3.8.1 veri madenciliği yazılımındaki, literatürde en çok kullanılan 6 maliyete-duyarsız sınıflandırıcı değerlendirmeye alınmıştır. Bu algoritmalar PART [15], C4.5 [16], SimpleCART [17], RIPPER [18], NBTree [19] ve DecisionTable [20]'dır. PART, sınıflandırma kuralları üretmekte yetenekli bir kural öğrenme yapısıdır. C4.5, SimpleCART ve RIPPER algoritmaları, uygulamada yaygın olarak kullanılan karar ağacı algoritmalarıdır. NBTree algoritması, birçok makine öğrenme araştırmacısı tarafından çalışılmış bir algoritmadır ve Bayes tekniklerini kullanır. DecisionTable algoritması, basit bir karar tablosu oluşturup kullanan bir sınıflandırıcıdır. Maliyete-duyarlı sınıflandırma için ise CostSensitiveClassifier ve MetaCost algoritmaları kullanılmıştır. Bahsedilen algoritmalar temel maliyete-duyarsız sınıflandırıcıları maliyete-duyarlı hale getirmektedirler. CostSensitiveClassifier'a maliyet fonksiyonu

kazandırabilmek için her bir sınıfa atanan toplam maliyete göre eğitim örnekleri yeniden ağırlıklandırılmakta veya minimum beklenen yanlış sınıflandırma maliyetleriyle sınıf tahmin edilmektedir. MetaCost ise eğitim örneklerini tahminlenen minimum maliyet sınıflarına göre tekrar etiketleyerek oluşturduğu yeni eğitim kümesini, maliyete-duyarsız bir sınıflandırma algoritması ile çalıştırarak ilgili algoritmayı maliyete-duyarlı hale dönüştürmektedir [21].

Diğer taraftan MT-VeMa 1.0 paket programından ise BEE-miner [12], MEPAR-miner [22] ve DIFACONN-miner [22]'ın maliyete-duyarsız ve maliyete-duyarlı versiyonlarına yer verilmiştir. BEE-miner iki/çok sınıflı yapılara uygulanabilen ve doğadaki arıların yiyecek arama davranışından esinlenen bir metasezgisel olan arı algoritmasına [23] dayanan maliyete duyarsız/duyarlı bir sınıflandırıcı özelliği taşımaktadır. Geliştirilen algoritma doğrudan bir yöntem olma özelliği taşıdığından girdi ya da çıktı üzerinde yapılacak güncellemeler yerine, maliyet bileşeni algoritmanın çalışma prensibi içine dâhil edilmiştir. MEPAR-miner ise sembolik regresyona yönelik geliştirilen çoklu denklem programlama yaklaşımının, sınıflandırma kuralları türetmek üzere geliştirilmesi ile ortaya konmuştur. Etkin bir kromozom gösterimi ile genetik programlamanın avantajlarını içerisinde barındıran ve değişken uzunlukta sınıflandırma kuralları türeten MEPAR-miner yaklaşımı, genetik operatörlerle uygun olmayan bireylerin türetilmesi, kural boyutunun aşırı büyümesi gibi dezavantajları da ortadan kaldırmıştır. DIFACONN-miner ise ileri beslemeli yapay sinir ağlarından sınıflandırma kuralları çıkarımına yönelik bir algoritmadır. Yapay sinir ağlarının eğitimi için diferansiyel gelişim algoritmasını, kural çıkarımı için ise tur atan karınca koloni optimizasyonu algoritmasını kullanmaktadır.

Sınıflandırma algoritmaları ile kural çıkarmada veri kümesinin eğitim ve test kümesi olarak ikiye ayrılması gerekmektedir. Eğitim kümesi, algoritmalara iyi öğrenme kabiliyeti kazandırmak için kullanılırken, test kümesi algoritmaların performansını değerlendirmek için kullanılmaktadır. Eğitim ve test kümelerinin oluşturulmasında yüzde bölme yöntemi kullanılmış olup veri kümesinin %66'sı eğitim, kalan kısmı ise test kümesini oluşturmaktadır. BEE-miner, MEPAR-miner ve DIFACONN-miner algoritmalarının parametre seçim aşamasında, ilgili çalışmalarda önerilen parametre kombinasyonları kullanılmıştır. Her algoritmanın 30 koşma sonucu değerlendirilmiş ve ortalama (ort.), minimum (min), maksimum (mak) ve standart sapma (ss) değerleri sunulmuştur. Weka 3.8.1 ve MT-VeMa 1.0 algoritmaları Intel Core 2 Duo, 800 MHz., 2.0 Ghz., 2 MB Cache Bellek, T6400 sahip bilgisayarda çalıştırılarak maliyete-duyarsız sınıflandırma sonuçları Tablo 5'teki, maliyete-duyarlı sınıflandırma sonuçları ise Tablo 6'daki gibi elde edilmiştir. İlgili tablolarda test doğruluğu, kural sayısı, hesapsal zaman ve yanlış sınıflandırma maliyeti değerleri verilmiştir. Tablo 6'da sunulan BEE-miner, MEPAR-miner aşağı ve yukarı örnekleme, DIFACONN-miner aşağı ve yukarı örnekleme algoritmalarının 30 kez çalıştırılması sonucunda elde edilen minimum/maksimum değerler, minimum/maksimum yanlış sınıflandırma maliyetine karşılık gelen test doğruluğu, kural sayısı ve CPU bilgilerini içermektedir.

Tablo 5'te görüldüğü gibi, test doğruluğu açısından Weka algoritmaları %83.63 değeri ile en yüksek test doğruluğuna ulaşmışlardır. Diğer algoritmaların maksimum değerlerine baktığımızda, BEE-miner %78.18, MEPAR-miner ve DIFACONN-miner %78 doğrulukla Weka algoritmalarına yakın değerler elde etmişlerdir. Ortalama değer açısından, BEE-miner %78.18 ile aynı kalmasına karşın, MEPAR-miner %54 ve DIFACONN-miner %72 değerlerine düşmüştür. Kural sayısı açısından bakıldığında ise Weka algoritmalarından Decision Table, C4.5 ve NBTree 1 kural ile hem en yüksek doğruluk hem de en az kural oluşturmuşlardır. BEE-miner 2, MEPAR-miner ve Ripper 3, Difaconn-miner 4 kuralda kalmışlardır. Hesapsal zaman açısından en iyi değere 1sn ile test doğruluğu ve kural sayısı da en iyi olan C4.5 algoritması ulaşmıştır. Diğer Weka algoritmaları süre açısından birbirine yakındır. DIFACONN-miner ikili yapısından dolayı en iyi doğruluğu 655 sn. MEPAR-miner ise 108 sn'de bulabilmiştir. Weka algoritmalarına en yakın hesapsal süre değerini 10sn ile BEE-miner algoritması vermiştir.

Tablo 6'da görüldüğü üzere tüm algoritmaların maliyete duyarlı kalabilmek adına test doğruluklarında bir düşüş olmuştur. Yanlış sınıflandırma maliyetleri dikkate alındığında en düşük yanlış sınıflandırma maliyetini 176.36 değeri ile CostSensitiveClassifier(PART) vermektedir. BEE-miner, MEPAR-miner ve DIFACONN-miner 256.36 değeri ile aynı maliyet değerine ulaşmışlardır. Test doğruluğu açısından en iyi değeri %76.36 ile MetaCost(PART) vermiştir. En iyi yanlış sınıflandırma maliyetine sahip olan CostSensitiveClassifier(PART) ise %70.9 doğruluk elde etmiştir. BEE-miner, MEPAR-miner ve DIFACONN-miner algoritmaları ise %67 doğruluğa ulaşmışlardır. Kural sayısı açısından ise BEE-miner 4 kural, MEPAR-miner 3 kural ve DIFACONN-miner 4 kural elde etmişlerdir. En iyi yanlış sınıflandırma maliyetine sahip CostSensitiveClassifier (PART) 6 kural, en iyi test doğruluğuna sahip MetaCost(PART) ise 7 kural elde etmiştir. Hesapsal zaman açısından Weka algoritmaları 1 ile 49 sn. arasında değerler almaktadır.

DIFACONN-miner ikili yapısından dolayı en düşük yanlış sınıflandırma maliyetine aşağı örneklemede 87 sn. MEPAR-miner ise aşağı örneklemede 66 sn'de ulaşmıştır. Weka algoritmalarına en yakın değeri 14 sn. ile BEE-miner algoritması vermiştir. En iyi yanlış sınıflandırma maliyetine sahip CostSensitiveClassifier (PART) 3sn, en iyi test doğruluğuna sahip MetaCost(PART) ise 14 sn. çözüme ulaşmıştır.

En düşük yanlış sınıflandırma maliyetine sahip CostSensitiveClassifier(PART) algoritması sonuçları incelendiğinde test doğruluğu %70.9 çıkmış ve 6 kural oluşmuştur. Kurallar şu şekildedir (1: İplik kaliteli; 0: İplik kalitesiz):

1. Kural: Kopca > 1 AND Kafes <= 2 AND Kopca <= 2: 1
2. Kural: Kopca <= 2 AND Kopca <= 1: 0
3. Kural: Kafes > 2 AND Klips > 1: 0
4. Kural: Kopca > 2 AND Klips > 2 AND Kafes <= 1: 1
5. Kural: Klips <= 2 AND Kafes > 1: 1
6. Kural:Diğer ihtimaller: 0

Örneğin 4. kural incelendiğinde iplik kalitesinin iyi olması için Kopça h2dr08, Klips 4.5 mm. ve Kafes dar olduğunda iplik kaliteli olacaktır. İşletmenin elinde 4.5 mm. klips yoksa 4 mm. veya 3.5 mm. klips mevcutsa, 5. kurala uygun olarak kafes mevcut veya geniş yapılarak yine kaliteli iplik üretilebilir. Bu durum işletmeye çok büyük esneklik sağlamaktadır.

Tablo 5: Maliyete-duyarlı olmayan sınıflandırma sonuçları.

		Test Doğruluğu	Kural Sayısı	CPU (sn)
BEE-Miner	ort.	78.18	2.03	9.03
	s.s	0	0.18	0.49
	min.	78.18	3	10
	mak.	78.18	2	8
MEPAR-Miner	ort.	54	3	112.6
	s.s	15	0	4
	min.	36	3	109
	mak.	78	3	108
DIFACONN-Miner	ort.	72	6.87	637.3
	s.s.	5	1.53	13.47
	min.	56	8	649
	mak.	78	4	655
PART		83.63	8	2
Ripper		83.63	3	2
DecisionTable		83.63	1	3
C4.5 (J48)		83.63	1	1
SimpleCART		83.63	12	6
NBTree		83.63	1	7

Tablo 6: Maliyete-duyarlı sınıflandırma sonuçları.

		Yanlış Sınıflandırma Maliyeti	Test Doğruluğu	Kural Sayısı	CPU (sn)
BEE-miner	ort.	366.18	63.03	4.37	13.97
	s.s	88.28	13.37	0.67	1
	min.	256.36	67.27	4	14
	mak.	550.91	45.45	4	15
MEPAR-miner Aşağı Örnekleme	ort.	372.36	58	3	68.3
	s.s	102.19	9	0	2.44
	min.	256.36	67	3	66
	mak.	474.55	55	3	69
MEPAR-miner Yukarı Örnekleme	ort.	293.45	53	3	131.73
	s.s	32.99	13	0	4.86
	min.	256.36	67	3	124
	mak.	321.82	42	3	134
DIFACONN-miner Aşağı Örnekleme	ort.	377.76	48	5.13	88.2
	s.s	97.32	19	0.97	1.54
	min.	256.36	67	4	87
	mak.	620	25	5	89
DIFACONN-miner Yukarı Örnekleme	ort.	336.36	67	5.83	14179.87
	s.s	95.74	6	1.23	263.31
	min.	256,36	67	5	13627
	mak.	572.73	47	6	14311
MetaCost(PART)		180	76.36	7	14
MetaCost(Ripper)		220	74.54	5	15
MetaCost(DecisionTable)		321.81	41.81	1	11
MetaCost(C4.5)		189.09	74.54	9	8
MetaCost(SimpleCART)		229.09	72.72	5	42
MetaCost(NBTree)		267.27	52.72	1	49
CostSensitiveClassifier (PART)		176.36	70.90	6	3
CostSensitiveClassifier (Ripper)		207.27	70.90	3	3
CostSensitiveClassifier (DecisionTable)		207.27	70.90	4	3
CostSensitiveClassifier (C4.5)		185.45	69.09	8	1
CostSensitiveClassifier (SimpleCART)		220	74.54	9	6
CostSensitiveClassifier (NBTree)		207.27	70.90	1	7

4 Sonuç

Özellikle zaman içinde verinin azlığının değil, çokluğunun bir sorun olması ve bilgisayarların veri saklama ve işleme hızındaki artış sonucunda, yığın halindeki veriler arasındaki gizli örüntülerin çıkarılmasına yarayan veri madenciliğinin güncelliği her geçen gün artmaktadır. Veri madenciliği yöntemleri içinde ise veri tabanlarında gizli olan bilgiyi kullanıcıların anlayabileceği kural listeleri veya karar ağaçları formunda sunan sınıflandırma dikkat çekmektedir. Bu çalışmada bir iplik üretim tesisinde maliyete-duyarlı ve maliyete-duyarsız sınıflandırma ile kural çıkarımı gerçekleştirilmiş, ancak öncesinde iplik kalitesine etki eden nitelikler belirlenerek işletmede yapılacak testlerin sayısını anlamlı bir şekilde azaltabilmek için Taguchi deneysel tasarım yöntemi ile nitelik seçimi yapılmıştır. Taguchi deneysel tasarım yöntemi ile iplik kalitesi üzerinde hangi niteliklerin daha etkili olduğu ve her bir niteliğin en iyi seviyeleri belirlenmiştir. Elde edilen kurallar sayesinde iplik kalitesini olumlu etkilediği bilinen belli bir nitelik seviyesinin kullanımının mümkün olmaması durumunda, iplik kalitesini istenen seviyeye getirebilmek için ilgili niteliğin mümkün olan seviyesi ile diğer nitelik seviyelerinin uygun kombinasyonları önceden bilinebilmekte, böylece işletmeye esneklik kazandırılmaktadır. Ayrıca maliyete-duyarlı sınıflandırma ile elde edilen kuralların etkin kullanımı ile işletmenin olası maliyet kayıpları da azaltılmıştır. Bu uygulama çalışmanın işletmede yaygın olarak üretilen bir iplik türü üzerinde yapılması sayesinde farklı iplik türleri için de işletmenin bir öngörüye sahip olması sağlanmıştır. Bahsedilen çalışmanın işletmede uygulanması neticesinde, farklı iplik türlerine ait kalite testlerinin mevcut 6 niteliğin farklı seviyelerine göre işletme mühendisinin tecrübelerine dayalı olarak belirlenmesi yerine Taguchi deneysel tasarım yöntemi ile belirlenen etkin niteliklere göre daha az sayıda test gerçekleştirilebilmiştir. Bu da işletmenin hem maddi kaybını hem de üretim yapmak yerine testler için ayrılan zamanı azaltmıştır. Diğer taraftan işletmede mevcutta bulunmayan nitelik seviyeleri için kaliteli iplik üretimini sağlayacak diğer seviyeleri tespit edebilmesine imkân vermiştir. Netice olarak bu çalışma ile veri madenciliği ve deneysel tasarım uygulamalarının, bir tekstil şirketinde gerçek verilerle nasıl sonuca ulaştığı gösterilmiş, ilgili sürece katkıda bulunulmuş ve elde edilen bulgular işletmeye yol gösterici nitelikte olmuştur.

5 Teşekkür

Orta Anadolu AŞ'ye bu çalışmada kullanılan verilere ulaşmada sağlamış oldukları destekten dolayı teşekkür ederiz.

6 Kaynaklar

[1] Kalikov A. Veri Madenciliği ve Bir e-ticaret Uygulaması. Yüksek Lisans Tezi, Gazi Üniversitesi, Ankara, Türkiye, 2006.

[2] Thuraisingham BM. *Web Data Mining and Applications in Business Intelligence and Counter Terrorism*. Boca Raton, FL, USA, CRC Press LLC, 2003.

[3] Zhou Z. "Three perspectives of data mining". *Artificial Intelligence*, 143(1), 139-146, 2002.

[4] Savaş S, Topaloğlu N, Yılmaz M. "Veri madenciliği ve Türkiye'deki uygulama örnekleri". *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 21, 1-23, 2012.

[5] İnan O. Veri Madenciliği. Yüksek Lisans Tezi, Selçuk Üniversitesi, Konya, Türkiye, 2003.

[6] Ye N. *The Handbook of Data Mining*. 1st ed. Mahwah, NJ, USA, Lawrence Erlbaum, 2003.

[7] Ranjit KR. *A primer on the Taguchi method.* New York, USA, Van Nostrand Reinhold, 1990.

[8] Güral G. Gaz kaynağında proses parametrelerinin optimizasyonu. Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi, İzmir, Türkiye, 2003.

[9] Lee JH, Yeh WC, Chuang MC. "Web page classification based on a simplified swarm optimization". *Applied Mathematics and Computation*, 270, 13-24, 2015.

[10] Chuang LY, Yang CH, Wu KC, Yang CH. "A hybrid feature selection method for DNA microarray data". *Computers in Biology and Medicine*, 41, 228-237, 2011.

[11] Kim YJ, Heo J, Park KS, Kim S. "Proposition of novel classification approach and features for improved real-time arrhythmia monitoring". *Computers in Biology and Medicine*, 75, 190-202, 2016.

[12] Tapkan P, Özbakır L, Baykasoğlu A, Kulluk S. "A cost sensitive classification algorithm: BEE-Miner". *Knowledge-Based Systems*, 95, 99–113, 2016.

[13] Zadrozny B, Elkan C. "Learning and making decisions when costs and probabilitiesare both unknown". *7th International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 26-29 August 2001.

[14] Tapkan P, Özbakır L, Baykasoğlu A. "Weka ile veri madenciliği süreci ve örnek uygulama". *Endüstri Mühendisliği Yazılımları ve Uygulamaları Kongresi*, İzmir, Türkiye, 30 Eylül -1/2 Ekim 2011.

[15] Frank E, Witten IH. "Generating accurate rule sets without global optimization". *15th International Conference on Machine Learning*, Madison, Wisconsin, USA, 24-27 July 1998.

[16] Quinlan JR. *C4.5: Programs for Machine Learning*. San Francisco, CA, USA, MorganKaufmann, 1993.

[17] Breiman L, Friedman J, Stone CJ, Olshen RA. *Classification and Regression Trees*. Monterey, California, USA, Wadsworth, 1984.

[18] Cohen W. "Fast effective rule induction". *12th International Conference on Machine Learning*, Tahoe City, California, USA, 9-12 July 1995.

[19] Kohavi R. "Scaling up the accuracy of naive-bayes classifiers: A decision-tree hybrid". *2nd International Conference on Knowledge Discovery and Data Mining*, Portland, Oregon, USA, 2–4 August 1996.

[20] Kohavi R. "The power of decision tables". *8th European Conference on Machine Learning*, Heraclion, Crete, Greece, 25-27 April 1995.

[21] Domingos P. "MetaCost: A general method for making classifiers cost-sensitive". *5th International Conference on Knowledge Discovery and Data Mining*, San Diego, CA, USA, 15-18 August 1999.

[22] Kulluk S, Özbakır L, Baykasoğlu A, Tapkan P. "Cost-sensitive meta-learning classifiers: MEPAR-miner and DIFACONN-miner". *Knowledge-Based Systems*, 98, 148–161, 2016.

[23] Pham DT, Ghanbarzadeh A, Koç E, Otri S, Rahim S, Zaidi M. "The bees algorithm-A novel tool for complex optimisation problems". *Proceedings of Innovative Production Machines and Systems Virtual Conference*, Cardiff, UK, 3–14 July 2006.