



Gülser Acar DONDURMACI¹

Ayşe ÇINAR²

FİNANS SEKTÖRÜNDE VERİ MADENCİLİĞİ UYGULAMASI

Özet

Bu çalışmada İstanbul Menkul Kıymetler Borsası (İMKB)'nden elde edilen 1990-2010 yıllarını kapsayan döneme ait olmak üzere, 10 adet bankanın menkul kıymetlerinin kapanış fiyatları, menkul kıymet teknik analizlerinde yaygın biçimde kullanılan teknik göstergeler, altın fiyatlarındaki değişimler, dolar kurundaki değişimler ve bazı yurtdışı borsa göstergeleri göz önüne alınarak karar ağaçları oluşturulmuş ve bu ağaçlara dayalı olarak karar kuralları elde edilerek finans sektöründe bir veri madenciliği çalışması gerçekleştirilmiştir. Elde edilen veriler üzerinde çeşitli ön işlemler yapılarak veri madenciliği sürecinin ilk aşaması tamamlanmış ve daha sonra veri madenciliği modelleri içinde tahmin edici model olarak ele alınan sınıflandırma yöntemi uygulanmıştır. Analize dahil edilen niteliklerin sayısını azaltmak amacıyla "Çoklu Doğrusal Regresyon Modelleri" düzenlenmiş olup, analizin ilk aşamasında çalışma kapsamında incelenen bütün nitelikler modele alınmışken, ilerleyen aşamada, istatistiksel bakımdan anlamlı olmayan nitelikler modelden çıkarılmıştır. Bu deneysel işlemlerin sonucunda bazı menkul kıymetlere ait regresyon modellerinden elde edilen niteliklerin seçilmesiyle daha düşük ortalama hataya sahip regresyon ağaçlarının elde edilmiştir. Çalışmanın son aşamasında, incelenen veriler üzerinde regresyon ve sınıflandırma ağaçları (CART) ile sınıflandırma işlemi yapılarak, bir algoritma geliştirilmiş ve karar kuralları oluşturulmuştur.

Anahtar kelimeler: veri madenciliği teknikleri, veri hazırlama, CART, veri ambarı, sınıflandırma, karar kuralları, İMKB, menkul kıymetler

¹ Zirve Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü, gulser.dondurmaci@zirve.edu.tr

² Marmara Üniversitesi İşletme Fakültesi İngilizce İşletme Bölümü, acinar@marmara.edu.tr

A DATA MINING APPLICATION IN FINANCIAL SECTOR

Abstract

In this study a data mining process is performed in finance sector. During this process, considering closing prices of *securities* of 10 different banks which are obtained from Istanbul Stock Exchange (ISE) for the period 1990-2010 and some technical indicators which are widely used in some technical analysis of these securities, the changes of gold prices, exchange rate and also some international stock market indicators, some decision trees are built and then decision rules are obtained on the basis of these trees. Applying some data preprocessing operations on the data, the first stage of the process is completed. Afterward, a classification model, which is also called “predictive model” among other data mining models, is performed on the dataset. Besides, some Multiple Linear Regression Models are applied to reduce the number of attributes and thereby insignificant ones are removed in the following stages. Finally, selecting some attributes related with the securities’ regression models, some regression trees with the lowest standard error are handled in the last stages, building regression and classification trees (CART) on the data, an algorithm is developed and some decision rules are generated.

Key words: data mining techniques, data preparation, CART, data warehouse, classification, decision rules, ISE, securities

1. Veri Madenciliği Süreci

Karar destek sistemlerinin kullandığı araçlardan biri olan veri madenciliği, zamanla oluşmuş veri yığınları içerisinde, anlamlı ve faydalı ilişki ve modellerin keşif sürecidir. Bu süreç sonucunda elde edilen bilgiler karar verme işlemlerinde kullanılmaktadır. Günümüzde bilimsel araştırmalar, işletmeler, web teknolojileri ve kamusal alanda veri madenciliği tekniklerinin uygulanma oranı giderek artmaktadır.

Veri tabanlarında bilgi keşfi süreci, veri ambarı oluşturulması ve bu bilgilere dayanarak veri madenciliği işlemlerinin yerine getirilmesi aşamalarını kapsar (Mitra vd. 2003: 4).

1.1. Veri Madenciliği Sürecinin Adımları

Veri madenciliği algoritmalarının uygulanması öncesinde ham veri üzerinde bazı ön işlemlerin yerine getirilmesi söz konusu olabilir. Kurumların oluşturduğu birçok veritabanlarında bilgiler eksik, yanlış, tekrarlı ve gereksiz olabilir. Bundan dolayı ham veri, sırasıyla temizleme, bütünleştirme, indirgeme ve dönüştürme gibi işlemlerden geçtikten sonra, veri madenciliği algoritması uygulanarak sonuçlar elde edilmektedir.

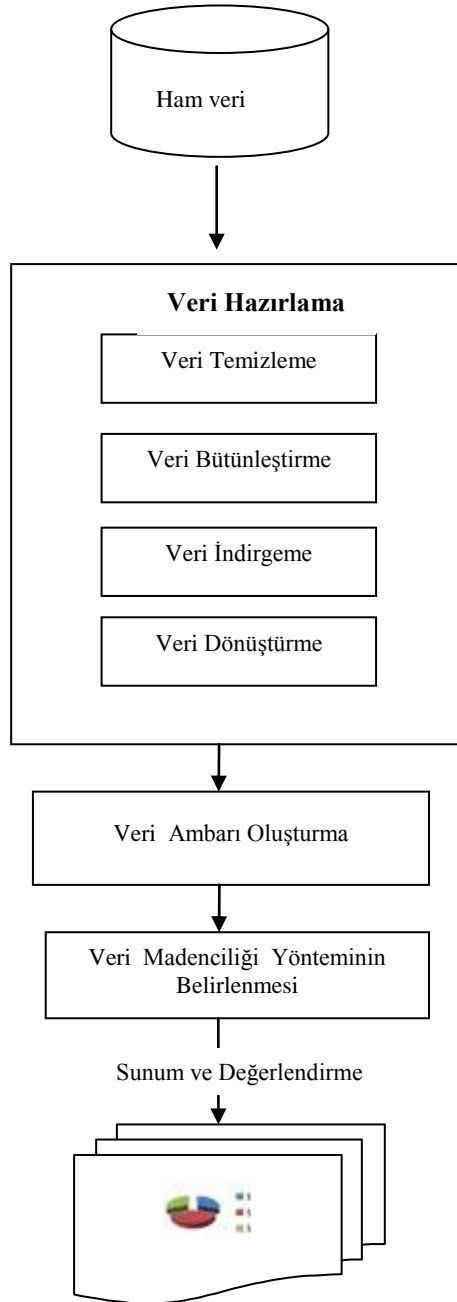
Eğer veri madenciliği verisi bir veri ambarından sağlanıyorsa, bu işlemlere gerek kalmayabilir. Bu tür işlemler veri ambarı hazırlanırken yerine getirilir.

Ön işleme işlemlerinin uygulanmasının ardından, veri analiz için hazır hale getirilir. Veri madenciliği yöntemlerinin bu işlenmiş veriye uygulanmasıyla sonuçlar elde edilir. Sonuçlar analiz edilerek veri içindeki örüntüler açığa çıkarılır.

Veri hazırlama ile ilgili işlemleri veri madenciliği kavramı içinde düşünürsek, veri madenciliğinin bir süreç olarak değerlendirmesi gerekir. Veri madenciliği sürecinin adımları Şekil 1.'de görülmektedir.

1.2. Veri Hazırlama

Veri tabanları içindeki verinin ve bu veriye dayalı olarak elde edilen veri madenciliği sonuçlarının kalitesinin artırılması, veriyi analize hazırlarken dikkat edilmesi gereken önemli noktadır. Veri madenciliği işlemlerini kolaylaştırmak ve verimliliği artırmak için veri tabanındaki veriler, bir “ön işleme” aşamasından geçirilir [2]. Veri madenciliği öncesinde verinin hazırlanması süreci olarak da kabul edilen bu işlemler özellikle veri tabanındaki bozuk değerleri ve veriler arasındaki tutarsızlıkları kaldırmayı amaçlamaktadır.



Şekil 1: Veri Madenciliği Sürecinin Adımları

1.2.1. Veri Temizleme

Veri temizleme aşaması, kayıp ya da eksik değerleri tamamlamak, aykırı değerleri belirleyerek gürültüyü azaltmak ve verilerdeki tutarsızlıkları gidermek için kullanılan birçok tekniği kapsamaktadır. Veri kümesini bu tür eksik yada kayıp değerlerden arındırmak için kullanılan yöntemlerden en önemlileri ve yaygın biçimde kullanılanları şunlardır (Han vd. 2006: 289).

Kayıp verilerin bulunduğu sorunlu kayıtların sayısı fazla değil ise, bu kayıtlar tümüyle yok edilir.

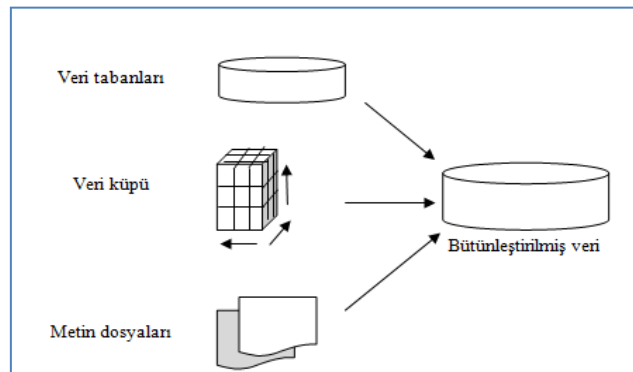
Kayıp değerler için genel bir sabit kullanılır. Bu yöntemde “BİLİNMIYOR” veya “∞” gibi sabit bir değer tanımlanabilir. Ancak veri madenciliği programları bu değeri ortak bir değer gibi algılayabilir. Bu nedenle çok tercih edilen bir yaklaşım değildir.

- Sayısal değerlere sahip nitelikler için hesaplanacak ortalama değer kayıp değerler yerine kullanılabilir.
- Kayıp değer mevcut verilere dayanarak en uygun değer kullanılarak tamamlanabilir. En uygun değer belirlenebilmesi için regresyon veya karar ağacı gibi teknikler kullanılabilir.

1.2.2. Veri Bütünleştirme

Veri madenciliği sürecinde kullanılan veri ambarının oluşturulması esnasında değişik kaynaklardan (veritabanları, veri küpleri, metin dosyaları, vb.) elde edilen veriler arasında uyum sağlamak için verilerin bütünleştirilmesi işlemlerinin yerine getirilmesi beklenir.

Bu işlem aşamasında, verinin bütünleştirilmesi için mevcut bütün veri kaynakları karakteristiklerine, özelliklerine, ve toplanma seviyelerine göre eşleştirilerek, tutarlılık sağlanır. Kurumun sahip olduğu birden fazla bilgi kaynaklarında kayıtlı olan verilerin biraraya getirilmesi aşamasında, aynı verinin farklı biçimlerde tutulması söz konusu olduğunda veri bütünleştirme işlemine başvurulur. Bir müşteriye ait aynı verinin farklı biçimlerde saklanmış olması buna örnek olarak verilebilir.



Şekil 2: Veri Bütünleştirme

1.1.1. Veri İndirgeme

Veri madenciliği sürecinin ön işlemleri aşamasında ele alınan veritabanında çok fazla kayıt ve gerekli olmayan nitelikler (değişkenler) olması durumunda veri indirgeme işlemi yapılır. Bu durumda verinin temel özelliklerini kaybetmeden, bir kısmını atarak azaltma

yoluna gidilip gidilmeyeceği sorunu ile karşılaşılır (Stepaniuk 2008: 43). Bu sorunu azaltmak için veri indirgeme yöntemine başvurulur.

Niteliklerdeki tutarsızlıklar elde edilen veri setinde fazlalıklara neden olabilir. Bu fazlalıkları tespit etmek için korelasyon analizinden yararlanılmaktadır. Örneğin korelasyon analizi sonucu, müşteri kimliği ile müşteri numarası isimli nitelikler arasında yüksek bir ilişki bulunursa bunlardan biri veri deposundan çıkarılarak indirgeme yapılır (Maimon vd. 2007: 95). Gereksiz boyut sayısı sorununun üstesinden gelmek için ise; nitelik (değişken) seçimi ve/veya kayıt seçimi işlemlerine başvurulabilir.

1.1.2. Veri Dönüştürme

Verilerin veri bütünlüğünü bozacak şekilde, farklı ölçek ya da kod ile kaydedildiği durumlarda başvuru bir yöntemdir. Buna örnek olarak; ağırlık ölçüsünün bir veritabanında kilogram ile kaydedilirken diğerinde paund ile kaydedilmiş olması verilebilir.

Veri dönüştürmede kullanılan yaklaşımlar aşağıda açıklanmıştır.

1.1.2.1. Ondalık Ölçekleme

Ondalık ölçekleme ile normalleştirmede ise, ele alınan değişkenin değerlerinin ondalık kısmı hareket ettirilerek normalleştirme gerçekleştirilir. Söz konusu ölçekleme, sayısal değerlerin -1 ile +1 arasında yer almalarını sağlayacak biçimde dönüştürülmesine karşılık gelir.

Hareket edecek ondalık nokta sayısı, değişkenin maksimum mutlak değerine bağlıdır. Ondalık ölçeklemenin formülü aşağıdaki şekildedir: Örneğin 900 maksimum değer ise, $n=3$ olacağından 900 sayısı 0,9 olarak normalleştirilir (Oğuzlar 2003:73).

1.1.2.2. Min-Max Normalleştirme

Min-Max normalleştirme ile orijinal veri üzerinde doğrusal bir dönüşüm yapılır. Bu yöntem aracılığıyla veriler genellikle 0-1 aralığına dönüştürülür.

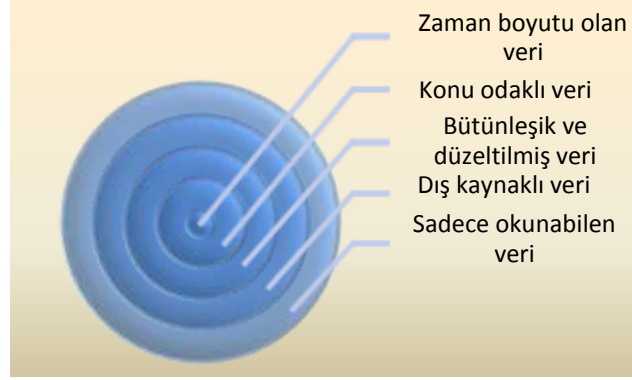
1.1.2.3. Z-Score Standartlaştırma

Verileri dönüştürmek amacıyla kullanılan bir diğer yöntem Z-Score standartlaştırma yöntemi olarak bilinir. İstatistiksel veri dönüştürme teknikleri arasında yer alan ve en yaygın biçimde kullanılan bu yöntem, ele alınan verinin ortalama ve standart sapma değerlerini kullanır.

Farklı tekniklerle gerçekleştirilebilen normalizasyon işlemi, veri boyutunun küçültülmesi amacıyla kullanılabilir gibi, verilerle gerçekleştirilecek işlemlerin uygun aralıklara normalize edilmiş değerlerle yapılarak işlemlerin daha hızlı gerçekleştirilip ve daha anlamlı ve kolay yorumlanabilir sonuçlar almak amacı ile de kullanılabilir (Khemka 2003:12-13).

Veri Ambarları

Veri madenciliği işlemlerini yerine getirirken genellikle karar destek sistemlerinin teknik altyapısını oluşturan veri ambarı olarak isimlendirilen özel bir veri tabanından yararlanılır.



Şekil 3: Veri Ambarlarının Özellikleri

Kurumların mevcut tüm verilerinin içerisinde çalışma kapsamındaki zaman aralığına ait verilerin saklandığı veri ambarları karar destek amacıyla, belirli bir zaman aralığı içerisindeki oluşmuş verileri içermektedir. Kuruma ilişkin eski verileri de saklaması karar destek sistemlerinin oluşturulmasında önemli bir rol oynamaktadır. Verilere hızlı erişimin çok önem kazandığı veri ambarları, büyük miktardaki veriyi hızlı biçimde sorgulayacak araçlara sahiptir. Aşağıda veri ambarlarının temel özellikleri görülmektedir.

1.2. Veri Madenciliği Yönteminin Belirlenmesi

Veri madenciliğinin bu kadar geniş bir alana yayılmasının sebebi ise veritabanı yönetimi, istatistik ve bilgisayar bilimi disiplinlerinin birleşmesinden oluşmasıdır. Dolayısıyla kullanılacak çok sayıda yöntem ve çok sayıda algoritma vardır. Bu durumda, uygulamada veri madenciliği yöntemlerinden hangisinin kullanılacağı belirlenmesi uygulamanın başarısı bakımından çok önemlidir. Örneğin uygulama için karar ağaçlarıyla analize karar verilmişse pek çok algoritma (ID3, C4.5, CART yöntemleri, Bayes ağları ya da Destek Vektör Makinaları, vb.) arasından amaca en uygun olanının seçilmesi esastır.

Bu yöntemlerin bir çoğu istatistiksel tabanlıdır. Bazı yöntemler ise yapay zeka teknikleri arasındadır (Olson vd. 2008: 4).

Veri madenciliği yöntemleri temel olarak aşağıdaki şekilde gruplandırılabilir:

- Sınıflandırma
- Kümeleme
- Birliktelik kuralları

Tablo 1: Veri Madenciliği Modelleri

MODELLER		ÇALIŞMA ALANLARI		ÖRNEK UYGULAMALAR
Tanımlayıcı Modeller	⇒	<ul style="list-style-type: none"> Kavram Tanımlama Kümeleme 	⇒	<ul style="list-style-type: none"> Çapraz Satış İçin Müşteri Profili Pazar Segmentasyon
Tahmin Edici Modeller	⇒	<ul style="list-style-type: none"> Sınıflama Tahmin 	⇒	<ul style="list-style-type: none"> Hataların ve Hastalıkların Teşhisi Sözleşme Sürelerinin Tahmini
Açıklayıcı Modeller	⇒	<ul style="list-style-type: none"> Birliktelik Karakterizasyon 	⇒	<ul style="list-style-type: none"> Pazar Sepet Analizleri Web Sitelerinin Başarı Faktörleri

Kaynak: <http://www.sayisalyontemler.com/?q=taxonomy/term/58> (22.05.2012)

1.3. Sonuçların Sunum ve Değerlendirilmesi

Veri madenciliği yöntemlerinin uygulanmasının ardından elde edilen sonuçların karar vericilere sunulması gerekmektedir. Ayrıca sonuçların şekil veya uygun grafiklerle desteklenmesi beklenir.

2. Finans Sektöründe Bir Veri Madenciliği Uygulaması

2.1. Materyal Yöntem:

Çalışma kapsamında, Tablo.2’de yer alan finans kuruluşlarından veriler alınarak bu veriler üzerinde;

- çeşitli ön işlem teknikleri ve
- veri madenciliği yöntemleri arasından Sınıflandırma Yöntemi ele alınmış olup, hedef niteliğin sürekli değerlere sahip olması nedeniyle, bu kapsamda Sınıflandırma ve Regresyon Ağaçları Modeli (CART) uygulanmış olup, bu aşamada, Ağaç Yapılı Regresyon Modeli ve aynı zamanda Çoklu Doğrusal Regresyon Modeli ele alınarak, geliştirilen regresyon ağaçlarında hatayı azaltma işlemleri üzerine bir algoritma geliştirilmiştir.

Sonuçlar, XLMiner 3.2.6 yazılım programı kullanılarak elde edilmiştir.

2.1.1. Veri Üzerinde Ön İşlemler

Çalışmanın başlangıcında İMKB’den elde edilen 1990-2010 yıllarına ait menkul kıymetlerin kapanış verileri elde edilmiş olup, bu verilerin bir kesiti Tablo.3’de görülmektedir.

Söz konusu veriler üzerinde veri madenciliği sürecinin ilk ve önemli aşaması olan aşağıdaki ön işlemler yapılmıştır.

- Kayıp veriler içeren kayıtlar veri bütünlüğü bozulmadan silinmiştir.
- Z-Score Standartlaştırma işlemi yapılmıştır.
- Regresyon yöntemleri kullanılarak nitelik sayısı azaltılmıştır.

Tablo 2: Ele Alınan Sektör: Bankalar (1990-2010)

BANKA ADI	MENKUL KIYMET ADI
Akbank	AKBNK
AsyaBank	ASYAB
Garanti Bankası	GARAN
Halkbank	HALKB
İş Bankası	ISBTR
Şekerbank	SKBNK
Türkiye Ekonomi Bankası	TEBNK
Türkiye Sınai Kalkınma Bankası	TSKB
Vakıfbank	VKBNK
Yapı Kredi	YKBNK

Tablo 3: Menkul Kıymetlerin Kapanış Verileri (1990-2010)

KAP	KAP%	KAP(t-1)	RSI	STO	MACD	MOM	CCI	DOL	DOL%	DOL(t-1)	S&P	S&P%	S&P(t-1)	DOW	DOW%	DOW(t-1)	NIK	NIK%
12,2500		12,5000	45,4545	68,3333	1,2513	102,0833	-58,8258	0,0027		0,0027	316,5500		321,8600	2560,1500		2603,9600	25211,0000	
12,2500	0,0000	12,2500	45,4545	63,3333	1,2415	104,2553	-39,2172	0,0027	0,0000	0,0027	307,0600	-0,0309	316,5500	2483,4200	-0,0309	2560,1500	23738,0000	-0,0621
12,2500	0,0000	12,2500	50,0000	63,3333	1,2225	100,0000	-19,6086	0,0027	0,0000	0,0027	311,5100	0,0143	307,0600	2532,9200	0,0195	2483,4200	24166,0000	0,0177
12,5000	0,0200	12,2500	54,5455	73,3333	1,2585	100,0000	23,5303	0,0027	0,0000	0,0027	321,4400	0,0309	311,5100	2611,6300	0,0301	2532,9200	25142,0000	0,0388
12,5000	0,0000	12,5000	60,0000	80,0000	1,2640	96,1538	23,5303	0,0027	0,0000	0,0027	321,3400	-0,0003	321,4400	2614,8500	0,0012	2611,6300	25711,0000	0,0221
12,5000	0,0000	12,5000	60,0000	90,0000	1,2649	98,0392	15,6869	0,0027	0,0000	0,0027	324,1900	0,0088	321,3400	2632,4300	0,0067	2614,8500	24895,0000	-0,0328
12,7500	0,0196	12,5000	70,0000	100,0000	1,3025	100,0000	39,2172	0,0027	0,0000	0,0027	322,5600	-0,0051	324,1900	2614,3600	-0,0069	2632,4300	25978,0000	0,0417
12,5000	-0,0200	12,7500	55,5556	93,3333	1,2706	98,0392	7,8434	0,0027	0,0000	0,0027	323,0900	0,0016	322,5600	2613,1000	-0,0005	2614,3600	25420,0000	-0,0220
12,5000	0,0000	12,5000	50,0000	86,6667	1,2659	100,0000	3,9217	0,0027	0,0000	0,0027	323,0900	0,0000	323,0900	2613,3700	0,0001	2613,1000	24908,0000	-0,0206
12,2500	-0,0204	12,5000	28,5714	76,6667	1,2276	98,0000	-47,0606	0,0027	0,0000	0,0027	324,3900	0,0040	323,0900	2628,2200	0,0057	2613,3700	24078,0000	-0,0345
12,2500	0,0000	12,2500	33,3333	66,6667	1,2394	100,0000	-39,2172	0,0027	0,0000	0,0027	320,4600	-0,0123	324,3900	2596,2900	-0,0123	2628,2200	23812,0000	-0,0112
12,2500	0,0000	12,2500	33,3333	66,6667	1,2398	100,0000	-23,5303	0,0027	0,0000	0,0027	323,4000	0,0091	320,4600	2619,5500	0,0089	2596,2900	23962,0000	0,0063
12,2500	0,0000	12,2500	33,3333	73,3333	1,2397	100,0000	-11,7652	0,0027	0,0000	0,0027	321,6300	-0,0055	323,4000	2615,5900	-0,0015	2619,5500	25081,0000	0,0446
12,2500	0,0000	12,2500	40,0000	80,0000	1,2397	98,0000	-15,6869	0,0027	0,0000	0,0027	321,0400	-0,0018	321,6300	2612,6200	-0,0011	2615,5900	24605,0000	-0,0193
12,2500	0,0000	12,2500	40,0000	90,0000	1,2397	98,0000	3,9217	0,0027	0,0000	0,0027	322,5400	0,0047	321,0400	2625,7400	0,0050	2612,6200	25216,0000	0,0242
12,2500	0,0000	12,2500	50,0000	100,0000	1,2397	98,0000	3,9217	0,0027	0,0000	0,0027	318,6500	-0,0122	322,5400	2582,6700	-0,0167	2625,7400	25075,0000	-0,0056
12,5000	0,0200	12,2500	60,0000	100,0000	1,2772	98,0392	35,2955	0,0027	0,0000	0,0027	316,8300	-0,0057	318,6500	2564,1100	-0,0072	2582,6700	24897,0000	-0,0071
12,5000	0,0000	12,5000	60,0000	100,0000	1,2828	100,0000	27,4520	0,0027	0,0000	0,0027	317,7700	0,0030	316,8300	2567,3300	0,0013	2564,1100	24366,0000	-0,0218
12,5000	0,0000	12,5000	50,0000	100,0000	1,2837	100,0000	31,3737	0,0027	0,0000	0,0027	318,6000	0,0026	317,7700	2571,2900	0,0015	2567,3300	23885,0000	-0,0201
12,5000	0,0000	12,5000	50,0000	100,0000	1,2663	102,0408	3,9217	0,0027	0,0000	0,0027	316,6000	-0,0063	318,6000	2557,4300	-0,0054	2571,2900	23726,0000	-0,0067
12,5000	0,0000	12,5000	50,0000	100,0000	1,2651	102,0408	-3,9217	0,0027	0,0000	0,0027	311,4800	-0,0164	316,6000	2518,3200	-0,0155	2557,4300	23603,0000	-0,0052

2.1.2. Analizde Kullanılan Nitelikler

Çalışmada, menkul kıymet teknik analizlerinde kullanılan bazı göstergeler model nitelikler olarak ele alınmıştır. Söz konusu nitelikler Tablo 4'te sıralandığı gibidir.

Tablo 4: Araştırmada Kullanılan Nitelikler

NİTELİK	ANLAMI	EŞİTLİK
%KAP	İMKB ye dahil 10 adet Finans menkul kıymetleri günlük kapanış fiyatları	$\%KAP = \frac{KAP_t - KAP_{t-1}}{KAP_t}$
RSI ^v	Göreceli Güç İndeksi	$RSI = 100 - \frac{100}{1 + \frac{U}{D}} = \frac{100 U}{U + D}$
STO [*] Stokastik Osilatör	HD (Ham Değer)	$HD = \frac{100(\text{son günün kapanış değeri} - 20 \text{ gün kapanışlarının en düşüğü})}{(20 \text{ gün kapanışlarının en yükseği} - 20 \text{ gün kapanışlarının en düşüğü})}$
	K _t (HD'nin 3 günlük ortalaması)	$K_t = \frac{2}{3}K_{t-1} + \frac{1}{3}HD$
	D _t (K _t 'nin 3 günlük ortalaması)	$D_t = \frac{2}{3}D_{t-1} + \frac{1}{3}K_t$
MACD [®]	İki hareketli ortalamanın farkından yola çıkarak orta vadeli alım satım sinyalleri üreten ve piyasadaki fiyat hareketlerinin ne yöne doğru eğilimli olduğu sorusu ile ilgilenen bir gösterge.	MACD = (12 günlük üssel hareketli ortalama) - (26 günlük üssel hareketli ortalama)
Momentum [□]	Belli bir zaman dilimi içerisinde fiyatların ne yönde, ne miktarda ve ne şiddette hareket ettiğini açıklayan bir gösterge.	$MOM = \frac{\text{Son günün kapanışı}}{\text{n gün önceki kapanış}}(100)$
CCI [#]	Günün en yüksek, en düşük ve kapanış değerlerinin toplamalarının üçe bölünmesi ile bulunan ortalama fiyat.	$CCI = \frac{\text{Ortalama Fiyat} - \text{Ortalama Fiyatlar Hareketli ortalaması}}{(0,015)\text{Sapma Ortalaması}}$
%DOL	Dolar kurundaki değişme	$\%DOL = \frac{DOL_t - DOL_{t-1}}{DOL_t}$
%S&P	Standard & Poor uluslararası menkul kıymetler indeksindeki değişme	$\%S\&P = \frac{S\&P_t - S\&P_{t-1}}{S\&P_t}$
%DOW	ABD menkul kıymetler indeksi Dow Jones'taki değişme	$\%DOW = \frac{DOW_t - DOW_{t-1}}{DOW_t}$
%NIK	Japon menkul kıymetler indeksi Nikkei'deki değişme	$\%NIK = \frac{NIK_t - NIK_{t-1}}{NIK_t}$
%ALT	Altın fiyatlarındaki değişme	$\%ALT = \frac{ALT_t - ALT_{t-1}}{ALT_t}$

* STO- (Aybars 1997: 67).

® MACD- (Person 2004: 143).

□ MOM- (Weir 2006: 135).

#CCI- (Özekşi 2002: 238).

¥ RSI - (Aybars 1997: 58).

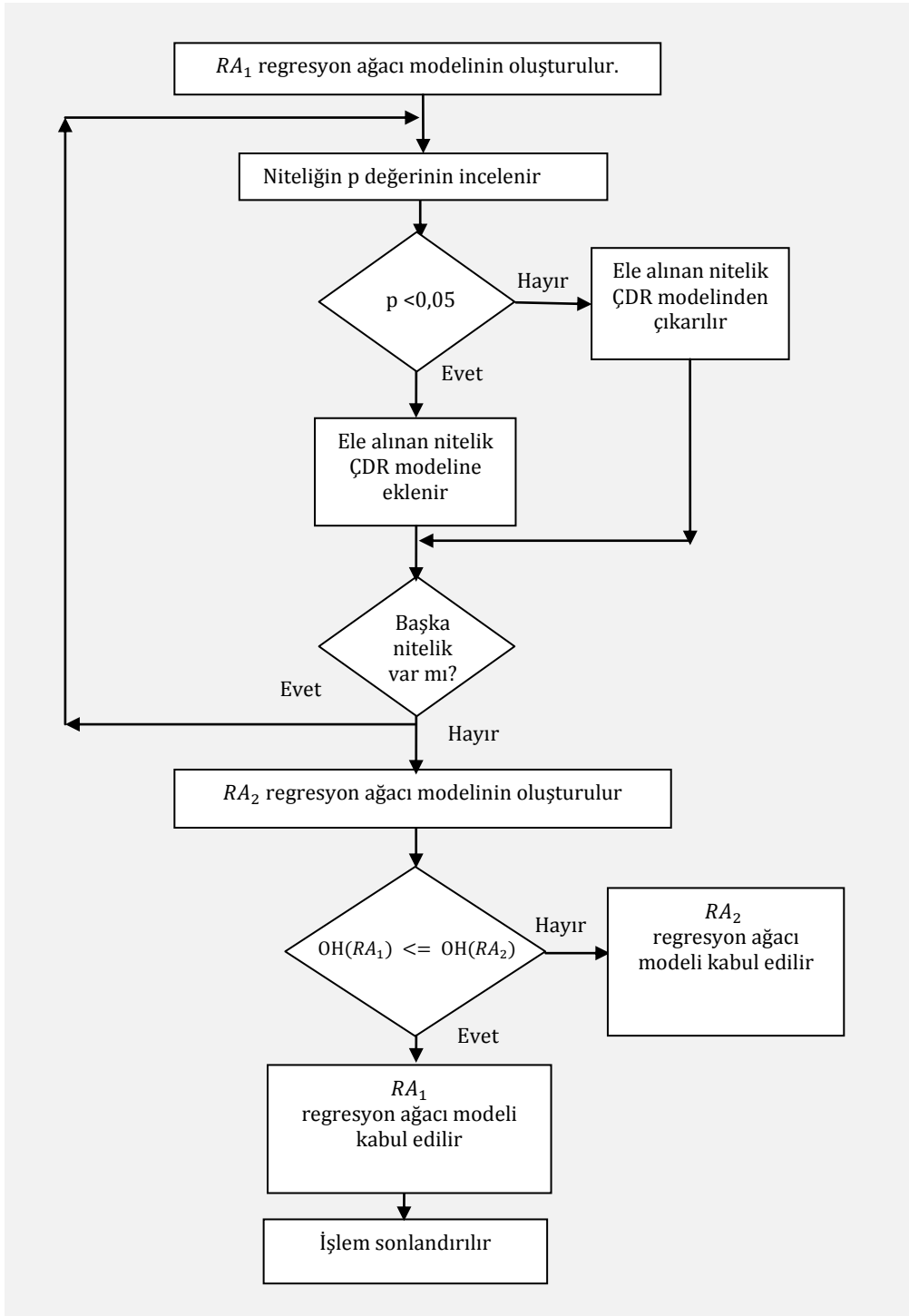
Analizde kullanılacak olan hedef değişken kapanış fiyatlarını temsil eden KAP isimli değişken olup, söz konusu değişkenin sayısal değerler içermesi nedeniyle bu çalışmada Sınıflandırma ve Regresyon Ağaçları Modeli (CART) uygulanmıştır. Bu bağlamda, toplanmış olan veriler üzerinde ön işlemler yapılmış olup, elde edilen nihai veriler ikiye ayrılarak analize katılmıştır. Holdout Yöntemi gereğince verilerin %66,66'sı eğitim, %33,34'ü test verisi olarak ele alınmıştır [2]. Çalışmanın bu aşamasında regresyon ağaçlarında ortalama hatayı azaltma işlemini için 6 adımdan (Şekil 5) oluşan bir algoritma geliştirilmiştir.

<p>Adım 1: Tüm nitelikleri kullanarak RA_1 regresyon ağacını oluşturun.</p> <p>Adım 2: Tüm nitelikler için ÇDR modelini oluşturun.</p> <p>Adım 3: ÇDR modelinde $p < 0.05$ koşuluna uygun nitelikleri seçin.</p> <p>Adım 4: ÇDR modelinden seçilen nitelikleri kullanarak RA_2 regresyon ağacını oluşturun.</p> <p>Adım 5: Eğer (RA_1 ortalama hatası (OH (RA1)) < RA_2 ortalama hatası (OH (RA2))) ise RA_1 regresyon ağacı modelini, aksi takdirde RA_2 modelini kabul edin.</p> <p>Adım 6: Algoritmayı sonlandırın.</p>

Şekil 5. Ortalama Hatayı Azaltma İşlemleri

2.1.3. Regresyon Ağaçları İle Sınıflandırma Yöntemi

Bir iş akış diyagramı halinde (Şekil 6)'da görülen bu algoritma aracılığıyla analizde kullanılacak olan bir sonuç tablosu (Tablo 5) elde edilmiştir. Söz konusu algoritmanın girdisi olarak, menkul kıymet kapanış fiyatlarındaki değişimleri ortaya koyan değerler ele alınmış ve çıktı olarak ortalama hatanın en düşük olduğu bir adet regresyon ağacı elde edilmiştir.



Şekil 6. Regresyon Ağaçlarında Ortalama Hatayı Azaltan Akış Diyagramı

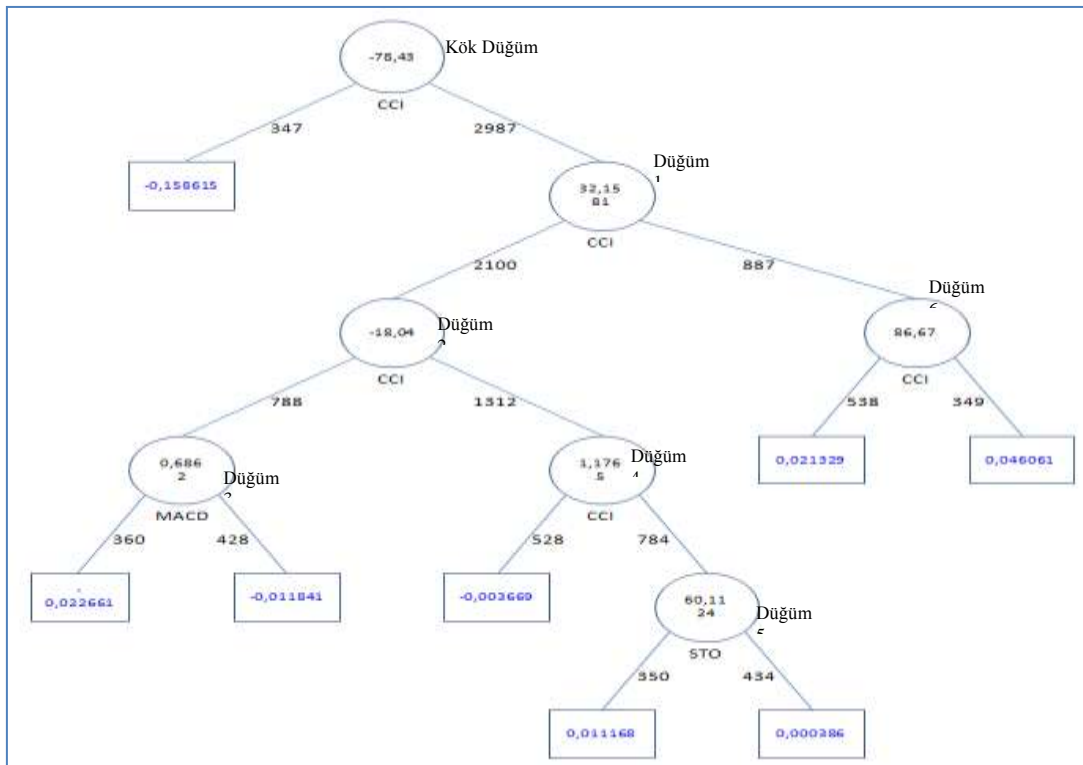
Tablo 5: Algoritma Sonuçları

Menkul Kıymet	Nitelikler	Eğitim Verisi		Test Verisi	
		Ortalama Hata	RMS Hata	Ortalama Hata	RMS Hata
AKBNK	CCI, MACD, STO	8,69826E-09	0,338316411	0,00921957	0,144009861
ASYAB	CCI, MACD, DOW%	-2,7248E-09	0,061912484	0,004681016	0,0245632
GARAN	CCI, STO	-3,59928E-09	0,198603326	0,006203455	0,112176101
HALKB	CCI, MACD, DOL%, DOW%	-1,76056E-09	0,026745338	-0,00097506	0,026470658
ISBTR	CCI, MOM, STO	-5,12357E-09	0,49594771	-0,044651386	2,236403563
SKBNK	CCI, STO	-2,74157E-08	0,120674552	-0,005559353	0,162722976
TEBNK	CCI, STO, MACD	-8,39832E-09	0,338196068	0,003068805	0,244864695
TSKB	CCI, MACD	1,38387E-08	0,437800235	-0,01195646	0,718476877
VKBNK	CCI, NIK%, DOL%, DOW%	-9,82801E-09	0,023697388	-0,002603027	0,05197582
YKBNK	CCI, MACD	-2,30885E-08	0,285801666	0,007151257	0,11890213

2.1.4. Karar Kuralları

Çalışma kapsamında ele alınan bankalara ilişkin regresyon ağacı modelinden yararlanarak karar kuralları elde edilmiştir (Dondurmacı 2011: 110-140).

Bu kurallar aynı zamanda sınıflayıcı işlevini de yerine getirmekte olup, Şekil 6'da AKBNK (Akbank menkul kıymeti) için elde edilen bir Regresyon Ağacı örneği verilmiştir.



Şekil 6: AKBNK İçin Regresyon Ağacı.

Söz konusu Regresyon Ağacı göz önüne alınarak 8 adet değişik kural edilmiştir.

Örneğin belirlenen bir tarihte,

$$CCI = 30, \quad MACD = 3,2$$

olarak hesaplanmış ise, yukarıdaki ağaca göre; diğer kurallar arasından

Kök düğüm -> Düğüm1 -> Düğüm2 -> Düğüm3

yolunu takip eden aşağıdaki kuralın kullanılabilmesi anlaşılmaktadır.

EĞER CCI > -78,43 İSE VE

$$CCI \leq 32,1581 \text{ İSE VE}$$

$$CCI \leq -18,04 \text{ İSE VE}$$

$$MACD > 0,6862 \text{ İSE}$$

$$KAP\% = -0,011841$$

Kural içinde yer alan koşullar sınıflandırma işlemi yerine getirmektedir. Bu kurala göre kapanış fiyatlarının belirlenmesinde CCI ve MACD nitelikleri etkili olmaktadır. İlk üç koşul CCI ile ilgilidir. Bu koşullar göz önüne alındığında,

$$CCI = 30 \text{ için} \quad -78,43 \leq CCI \leq 32,1581 \text{ ve } MACD > 0,6862$$

koşullarının geçerli olduğu görülmüştür.

Bu koşulların tümünü karşılayan kapanış fiyatlarındaki değişim yüzdesi -0,011841 olarak belirlenmiştir. Regresyon ağaçlarında karar niteliği sürekli değerlerden oluştuğu için sınıf değeri de sürekli değer olarak kestirilmiştir.

Ele alınan konu kapsamında yapılan literatür taraması sonucu, CCI ve MACD niteliklerine karşın diğer ülkelerde farklı niteliklerin kapanış fiyatlarının belirlenmesinde etkin rol oynadığı saptanmıştır. Bu araştırmaların sonuçlarına göre, Kore'de ROC (SHIN vd. 1998: 4), Kuveyt'te ROC ve Momentum (AL-QAHERI vd. 2008: 15), isimli niteliklerinin etkin olduğu görülmüştür

3. Sonuç ve Değerlendirme

İMKB 30 şirketlerinin tümünde regresyon ağaçlarının CCI niteliğini içerdiği ortaya çıkmıştır. Teknik analizlerde menkul kıymetlerin kapanış fiyatlarının tahmininde kullanılan CCI niteliğinin bu şekilde tüm modellere girmesi beklenen bir sonuç olmuştur.

Bu çalışma kapsamında incelenen menkul kıymetlerin sahip olduğu nitelikler ele alındığında; Tablo.3'de görüldüğü gibi, tümü CCI içerirken, STO'nun 5, MACD'nin 6, DOW'un 3, DOL'un 2, NIK ve MOM'un sadece birer menkul kıymette yer aldığı belirlenmiştir. Bunun yanı sıra, %S&P ve %ALT niteliklerinin regresyon ağacı modellerine giremediği ve böylece sınıflandırma ve kestirim sürecinde etkili olmadığı anlaşılmıştır.

KAYNAKÇA

- MITRA , S.; ACHARYA, T.(2003) Data Mining Multimedia, Soft Computing and Bioinformatics, John Wiley & Sons, Inc., ABD
- HAN, J.;KAMBER, M. (2011) Data Mining Concepts and Techniques,; Third Edition, Morgan Kaufman Elsevier Inc, ABD
- STEPANIUK, J. R. (2008) Granular Computing in Knowledge Discovery and Data Mining, Springer.
- MAİMON, O.; ROKACH, L.(2007) Data Mining And Knowledge Discovery Handbook, Springer Science+Business Media Inc., New York.
- OĞUZLAR, A. (2003) “Veri Ön İşleme”, Ege Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, Sayı 21, Temmuz-Aralık 2003, s.73, <http://iibf.erciyes.edu.tr/dergi/sayi21/aoguzlar.pdf> , (07.04.2012.)
- KHEMKA, A., (2003) A Collaborative Predictive Data Mining Model, Yayınlanmamış Yüksek Lisans Tezi, Faculty of University of Missouri-Kansas City, Missouri
- OLSON, D. L. ; DELEN, D. (2008) Advanced Data Mining Techniques, Springer, New York.
- PERSON, J. L. (2004) A Complete Guide to Technical Trading Tactics, John Wiley & Sons, Inc., ABD
- WEIR , D. J. (2006) Timing The Market : How To Profit In The Stock Market Using The Yield Curve, Technical Analysis, And Cultural Indicators, John Wiley & Sons, Inc., ABD
- ÖZEKŞİ, A. (2005) Kazanma Sanatı, Alfa Yayınevi, İstanbul.
- AYBARS, A. (1997) İMKB Uygulamalı Modern Teknik Analiz, Ekonomi Dizisi:2, Globus Dünya Basımevi, İstanbul.
- DONDURMACI, G. (2011) Veri Madenciliği'nde Regresyon Ağaçları İle Sınıflandırma Ve Bir Uygulama, Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü, Yayınlanmamış Doktora Tezi, İstanbul.
- <http://www.sayisalyontemler.com/?q=taxonomy/term/58>, (22.05.2012)
- SHIN,K., KIM,K.; HAN,I. (1998) Financial Data Mining Using Genetic Algorithms Technique: Application to KOSPI200, s.4, http://afis.kaist.ac.kr/download/inner_con053.pdf, (27.09.2012)
- AL-QAHERI, H.; HASSANIEN A.E.; ABRAHAM, A. (2008), Discovering Stock Price Prediction Rules Using Rough Sets, s.15, www.softcomputing.net/abonn.pdf, (27.09.2012)