

KİŞİ BAŞINA DÜŞEN MİLLİ GELİRE GÖRE ÜLKELERİN GIDA TÜKETİM EĞİLİMLERİNİN TESPİT EDİLMESİ – VERİ MADENCİLİĞİ UYGULAMASI

Fethi Ş. ÖZBEK*

ÖZET

Bu çalışmada, kişi başına düşen milli gelir (GNI) verileri ile günlük kişi başı gıda tüketim verileri kullanılarak ülkelere ait GNI gruplarına göre gıda tüketim eğilimleri, veritabanlarında bilgi keşfi (VTBK) sürecinin en önemli aşaması olan veri madenciliği uygulaması kullanılarak oluşturulan bir model yardımıyla tespit edilmiştir. Modelin oluşturulmasında, SPSS Modeller programı içerisinde yer alan karar ağacı algoritmalarından güçlü sonuçlar veren C5.0 algoritması kullanılmıştır. Oluşturulan model ile GNI gruplarına göre; et, şeker ve tatlandırıcılar, alkollü içecekler, yumurta, kuru baklagiller, sebzeler, diğer su ürünleri, balık, diğer nişastalı gıdalar ve meyveler tüketimi önemli değişkenler olarak bulunmuştur. Model sonuçlarına göre; GNI arttıkça, günlük kişi başı ortalama et, alkollü içecekler, bitkisel yağlar, balık, gıda benzerleri, hayvansal yağlar, meyveler, sakatat, süt, şeker, tatlandırıcılar ve yumurta tüketiminde artış; tahıl ve kuru baklagil tüketiminde ise düşüş olduğu tespit edilmiştir.

Anahtar kelimeler: Gıda tüketimi, GNI, Karar ağaçları, Veri madenciliği.

1. GİRİŞ

Bu çalışma ile ülkelere ait 2005, 2006 ve 2007 yıllarına ait kişi başına düşen milli gelir (GNI) verileri ile aynı ülkelere ve yıllara ait günlük kişi başı gıda tüketim verileri kullanılarak ülkelere ait GNI gruplarına göre gıda tüketim eğilimleri, veritabanlarında bilgi keşfi (VTBK) süreci kullanılarak tespit edilmiştir. Böylelikle; yüksek, orta ve düşük gelire sahip ülkelere ait gıda tüketim eğilimleri ortaya konmuştur.

Verilerden doğru, faydalı, yeni ve anlaşılır modeller ve kalıplar elde etmekte kullanılan VTBK; problemin tanımlanması, verilerin hazırlanması, modelin kurulması ve değerlendirilmesi, modelin kullanılması ve modelin izlenmesi aşamalarından oluşmaktadır. Verilerin hazırlanması aşaması; farklı kaynaklardan verilerin toplanması, zaman, kod, ölçü, vb. farklılıkları ortadan kaldırmak amacıyla verilere değer biçilmesi, farklı veri kaynaklarının birleştirilmesi, ihtiyaç duyulmayan verilerin temizlenmesi, modelde kullanılacak bağımlı ve bağımsız verilerin belirlenmesi, ihtiyaç duyulduğunda veri yığınının örnek seçilmesi ve veriler arasında gerekli dönüşümlerin yapılması kısımlarından oluşmaktadır (Batmaz ve Köksal, 2010; Çingir, 2010; Han ve Kamber 2006).

VTBK aşamalarından en önemlisini, modelin kurulması ve değerlendirilmesi kısımlarını içeren veri madenciliği aşaması oluşturmaktadır. Önceden bilinmeyen geçerli ve uygulanabilir bilginin veri yığınlarından dinamik bir süreç ile elde edilmesi olarak tanımlanan veri madenciliği, büyük veri yığınlarından oluşan veritabanları içerisinde gizli kalmış bilgilerin tespit edilmesine olanak sağlamaktadır (Baykal, 2006).

* TÜİK Uzmanı, Türkiye İstatistik Kurumu, e-posta: fethiozbek@tuik.gov.tr

Bu çalışmada, GNI gruplarına göre gıda tüketim eğilimlerinin tespit edilmesinde, veri madenciliği uygulamasında temel sınıflandırma tekniklerinden olan karar ağacı uygulaması kullanılmıştır. Karar ağacı ile veri kümesi, belirlenen girdi değişkenlerine göre alt gruplara ayrılmaktadır.

Daha önce yapılan çeşitli araştırmalarda, ekonomik gelişmişlik ile tüketilen gıda türleri arasındaki güçlü ilişki araştırmacılar tarafından ortaya konmuştur. Bu araştırmalardan Leenes (2006) tarafından yapılan çalışmada, gelişmiş ülkelerin hayvansal ürün tüketimlerinin ekonomisi güçsüz olan ülkelere göre daha fazla olduğu tespit edilmiştir. Mancini vd. (2009) tarafından yapılan çalışmada; Almanya, Belçika, Avusturya gibi GNI'si yüksek olan ülkelere göre kişi başına düşen hayvansal ürün tüketiminin, Romanya, Bulgaristan, Polonya gibi GNI'si düşük olan ülkelere göre daha yüksek; tahıl ve patates tüketiminin ise bu ülkelere göre daha düşük olduğu tespit edilmiştir.

2. MATERYAL VE YÖNTEM

2.1 Materyal

Bu çalışmanın ana materyalini, Dünya Tarım Örgütü (FAO) veritabanından elde edilen ülkelere ait kişi başı günlük gıda arzı verileri (FAO, 2010) ile Dünya Bankası (WB) veritabanından elde edilen ülkelere ait GNI verileri (WB, 2010b) oluşturmaktadır. FAO veritabanında yer alan gıda arzı verileri birincil bitkisel ve hayvansal ürün eşdeğerlerinden oluşmaktadır. Bu kısımda yer alan veriler, ülkelere ait ürün bazında oluşturulan gıda denge tablolarından elde edilmektedir. Belirli bir ürüne ait denge tablosundan elde edilen gıda arzı, birincil ürün ile bu üründen türetilen işlenmiş ürünlerin hammadde eşdeğeri toplamına karşılık gelmektedir. Örneğin, elma için gıda arzı yalnızca birincil ürün olarak tüketime sunulan elmayı değil elmanın işlenmesi ile elde edilen işlenmiş ürünlerdeki elma miktarını da kapsamaktadır.

Ünelere ait gıda tüketim eğilimlerinin tespit edilmesi amacıyla FAO veritabanında yer alan gıda tüketim gruplarının tamamı çalışma kapsamına dahil edilmiştir. Bu gruplar Tablo 1'de sunulmaktadır.

Tablo 1. Çalışma kapsamında kullanılan gıda tüketim grupları

Bitkisel ürünler		Hayvansal ürünler
Alkollü içecekler	Gıda benzerleri	Hayvansal yağlar
Tahıllar	Şeker ve tatlandırıcılar	Su ürünleri, diğer
Meyveler	Sert kabuklular	Yumurta
Yağlı tohumlar	Bitkisel yağlar	Balık
Kuru baklagiller	Sebzeler	Et
Baharatlar		Süt
Niştastalı gıdalar, diğer		Sakatat

Çalışmada, ilgili veritabanlarında tüm gıda gruplarına ait bitkisel ve hayvansal gıda verileri mevcut olan ülkeler kapsamıştır. Buna göre, yüksek GNI grubunda 26, yüksek orta GNI grubunda 23, düşük orta GNI grubunda 28 ve düşük GNI grubunda 13 olmak

üzere toplam 90 ülkeye ait 2005, 2006 ve 2007 verileri çalışma kapsamında kullanılmıştır.

2.2 Yöntem

Çalışmada kullanılan model, problemin tanımlanması, verilerin hazırlanması, modelin kurulması ve değerlendirilmesi, modelin kullanılması ve modelin izlenmesi aşamalarından oluşan VTBK süreci izlenerek oluşturulmuştur. VTBK sürecinde izlenen aşamalar şu şekildedir:

2.2.1 Problemin tanımlanması

VTBK sürecinin ilk aşaması olan bu aşamada, çalışma konusu; kişi başı milli gelir gruplarına göre ülkelerin gıda tüketim eğilimlerinin tespit edilmesi olarak belirlenmiştir.

2.2.2 Verilerin hazırlanması

Verilerin hazırlanması aşaması; verilerin toplanması, verilere değer biçilmesi, birleştirme ve temizleme, seçim ve dönüştürme kısımlarından oluşmaktadır. Verilerin toplanması aşamasında belirlenen amaca ulaşmak için gerekli veriler ve bu verilerin toplanacağı veri kaynakları tespit edilmektedir. Çalışmada kullanılan veriler; ülkelere ait gıda tüketim verileri ve GNI olarak belirlenmiştir. Bu veriler, FAO ve WB veri tabanlarından elde edilmiştir.

Araştırmada kullanılan verilerde zaman, kodlama ve ölçü birimi farklılıkları olmaması nedeniyle verilere değer biçme işlemi uygulanmamıştır.

Birleştirme ve temizleme aşamasında, FAO veri tabanından elde edilen gıda tüketim verileri, WB veritabanından elde edilen GNI verileri ile birleştirilmiştir. Bu birleştirme sonrasında eksik verileri olan ülkelere ait kayıtlar veri kümesinden çıkartılmıştır.

Seçim aşamasında, kurulan modelde kullanılan hedef değişken GNI, girdi değişkenleri ise gıda tüketim grupları olarak belirlenmiştir. Modelde kullanılan veri kümesinden herhangi bir örnek seçimi yapılmamıştır.

Çalışmada kullanılan C5.0 algoritmasında sembolik tipte hedef değişken kullanılmaktadır. Bu nedenle dönüştürme aşamasında, sayısal tipteki hedef değişken değerleri gruplandırılarak sembolik hale dönüştürülmüştür. Hedef değişken grubu olan GNI; yüksek (H), yüksek orta (UM), düşük orta (LM) ve düşük (L) olarak sınıflandırılmıştır. Bu sınıflandırmada, Dünya Bankası'nın gelire göre ülke sınıflandırması değerleri kullanılmıştır (WB, 2010a).

2.2.3 Modelin kurulması ve değerlendirilmesi

Bazı sosyo-ekonomik değişkenlere göre ülkelerin gıda tüketim eğilimlerinin tespit edilmesinde SPSS Modeler programı içerisinde yer alan karar ağacı algoritmalarından güçlü sonuçlar veren C5.0 algoritması kullanılmıştır. Tahmin edici ve tanımlayıcı özelliklere sahip olan karar ağaçları, veri madenciliğinde kuruluşlarının ucuz olması, yorumlanmalarının kolay olması, veri tabanı sistemleri ile kolayca entegre

edilebilmeleri ve güvenilirliklerinin daha iyi olması nedenlerinden dolayı sınıflama modelleri içerisinde yaygın bir kullanım alanına sahiptir (Akpınar, 2000).

1998 yılında Ross Quinlan tarafından önerilen C5.0 algoritması, diğer bir karar ağacı algoritması olan C4.5 algoritmasının geliştirilmiş bir halidir. Bilgi kazanımı kavramını kullanarak ağaç oluşumu sağlayan bu algoritma; hız, etkin hafıza kullanımı, daha küçük ağaç oluşumu, budama, ağırlıklandırma ve daha kesin sonuçlar verme gibi özellikleri sayesinde C4.5 algoritmasına göre daha gelişmiş bir algoritmadır (Güntürkün, 2007).

Örneklerin kökten yapraklara doğru sıralanarak sınıflandırıldığı karar ağacı algoritmasında, ilk olarak ağacın kökünde hangi değişken ile test edilmesi belirlenmekte daha sonra yukarıdan aşağı doğru ağaç oluşturulmaktadır. Karar ağacı algoritmasında her düğümde hangi değişkenin yer alacağı, değişkenlere ait bilgi kazanımı (information gain) ölçüsü ile belirlenmekte olup en yüksek bilgi kazanımı değerine sahip değişken ağacın kökünde yer almaktadır. Bilgi kazanımı entropi kullanılarak hesaplanmaktadır. Entropi, örneklerin keyfi olarak toplanmasının kirliliğini tanımlamaktadır (Kök ve Kuloğlu, 2005). Bir değişkene ait entropinin yüksek olması; ilgili değişkenin tekdüze dağılım gösterdiği anlamına gelmektedir. Hedef değişkenin alabileceği değerleri içeren S topluluğunun entropi değeri (1)'de yer alan formül yardımıyla aşağıdaki gibi hesaplanmaktadır (Moore, 2003):

$$\text{Entropi}(S) = - \sum_{j=1}^m p_j \log_2 p_j = -p_1 \log_2 p_1 - p_2 \log_2 p_2 - \dots - p_m \log_2 p_m \quad (1)$$

V: X'in alabileceği değerler

$$p_1 = P(X=V_1)$$

$$p_2 = P(X=V_2)$$

$$p_3 = P(X=V_3)$$

...

$$p_m = P(X=V_m)$$

Bilgi kazanımı değeri ise (2)'de yer alan formül yardımıyla aşağıdaki gibi hesaplanmaktadır.

$$\text{KAZANIM}(S,A) = \text{Entropi}(S) - \sum_v (|S_v|/|S|) \times \text{Entropi}(S_v) \quad (2)$$

Kazanım (S,A): A değişkenine ait bilgi kazanımı değeri

v : A değişkeninin alabileceği tüm değerler

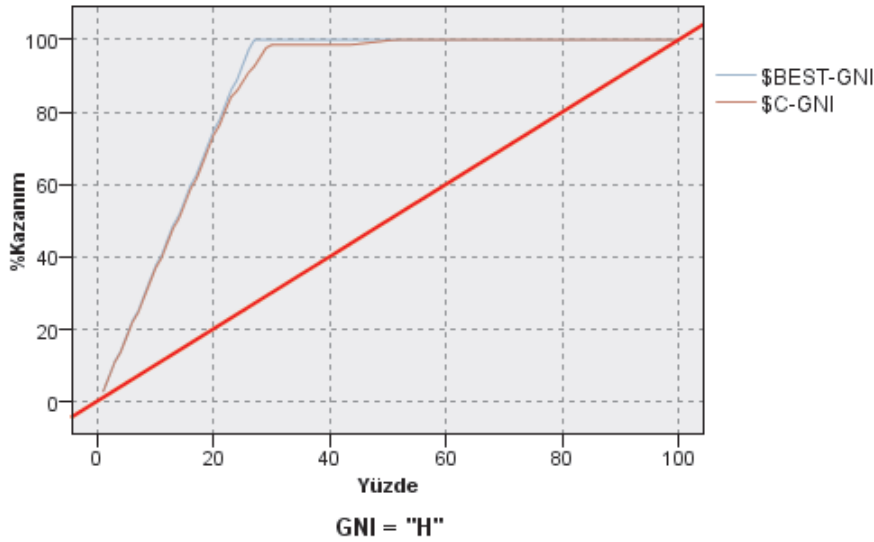
Karar ağacı algoritması kullanılarak kurulan modelin değerlendirilmesinde, değerlendirme grafikleri (kazanım grafiği), uygunluk matrisleri ve çapraz geçerlilik uygulamaları kullanılmıştır.

2.2.4 Modelin kullanılması

Model sonuçlarına göre GNI ile ülkelerin gıda tüketim eğilimleri tespit edilmiştir. Böylelikle, GNI gruplarına göre tüketimi fazla ve az gıda grupları belirlenmiştir. Oluşturulan model, gelir seviyesine göre tüketilen gıda gruplarının analiz edilmesi amacı ile kullanılabilir.

3. BULGULAR

SPSS Modeller kullanılarak C5.0 algoritması ile oluşturulan modelin doğruluğunun incelenmesi amacıyla oluşturulan değerlendirme grafiği Şekil 1'de sunulmaktadır. Grafikteki dikey eksen, hedef olarak seçilen değerlerin birikimli yüzdelerini gösterirken, yatay eksen model tahmini ve güven oranına göre sıralanmış yüzdeleri göstermektedir. Grafikte, köşegen doğrusu; hedef değişken rastgele bir seviyede tahmin edildiğinde beklenen temel oranı, en üstte bulunan eğri en iyi model veri kümesine uygulandığındaki sonuçları, ortadaki eğri ise model sonuçlarını göstermektedir. Grafikten de görüleceği gibi en iyi eğri ile model eğrisi birbirine oldukça yakındır, bu durum oluşturulan modelin doğruluğunun oldukça iyi olduğunu ifade etmektedir.



Şekil 1. Oluşturulan modele ait değerlendirme grafiği

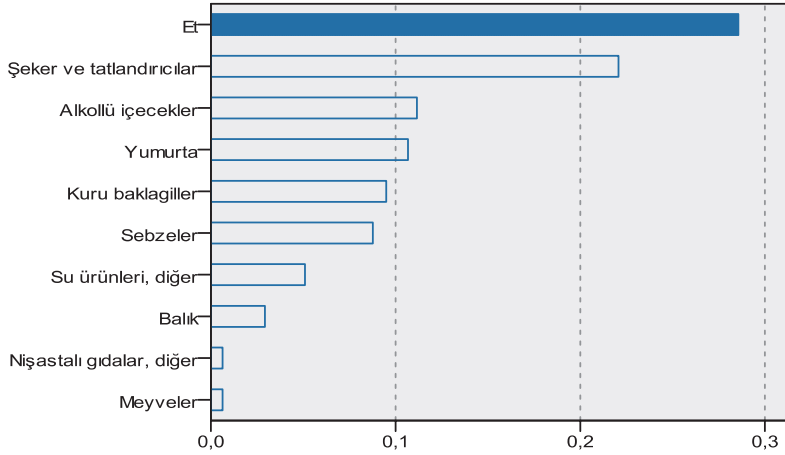
Analiz doğruluk oranı %97,28 olan modelin, gerçek değerlerle tahmin değerlerinin tutarlılığının kontrol edilmesi amacıyla oluşturulan çapraz tablo sonuçlarına göre yüksek ve düşük GNI gruplarının tamamı, yüksek orta ve düşük orta gruplarının ise tamamına yakını model tarafından doğru olarak tahmin edilmiştir. GNI gruplarına göre model tahmin sonuçları Tablo 2'de sunulmaktadır.

Tablo 2. GNI gruplarına göre model tahmin sonuçları (%)

GNI	H	L	LM	UM
H	100,0	0,0	0,0	0,0
L	0,0	100,0	0,0	0,0
LM	0,0	1,6	96,8	1,6
UM	2,7	2,7	2,7	91,9

Oluşturulan model ile GNI gruplarına göre et, şeker ve tatlandırıcılar, alkollü içecekler, yumurta, kuru baklagiller, sebzeler, diğer su ürünleri, balık, diğer nişastalı gıdalar ve meyve tüketimi önemli değişkenler olarak bulunmuştur. GNI gruplarına göre önem arz eden değişkenler Şekil 2’de sunulmaktadır.

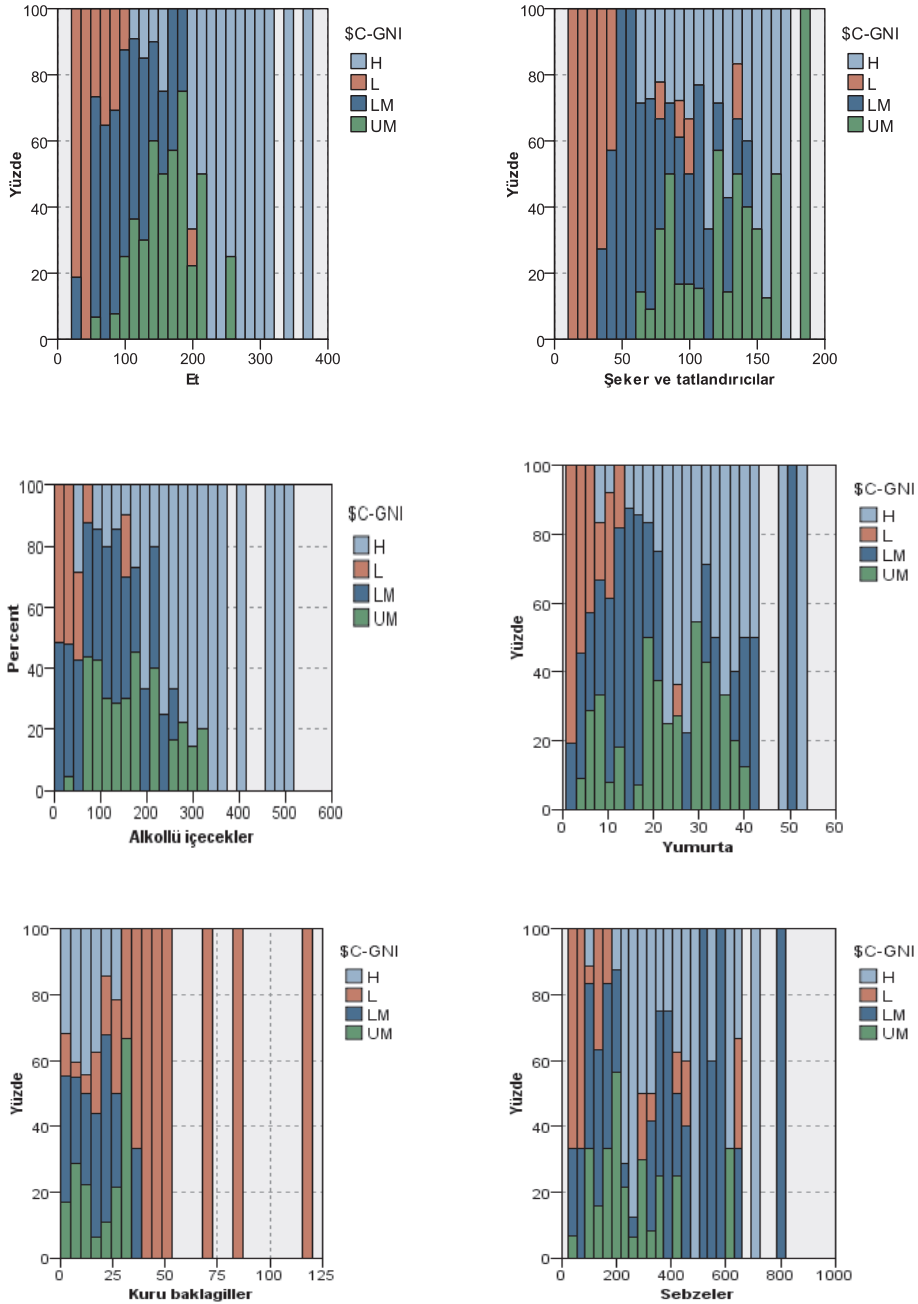
Önemli değişkenler



Şekil 2. GNI gruplarına göre tüketimi önem arz eden değişkenler

Oluşturulan ağaçta ilk bölünme “Et” değişkeninde olmaktadır. Et tüketim miktarının belirli bir değer üstünde olması durumunda sonraki bölünmeler sert kabuklular ve diğer su ürünleri şeklinde olmaktadır. Et tüketim miktarının belirli bir değer altında olması durumunda ise sonraki bölünmeler şeker ve tatlandırıcılar, kuru baklagiller, diğer su ürünleri ve diğer nişastalı gıdalar olarak devam etmektedir. Model çıktılarının yer aldığı Şekil 3’den de görüldüğü gibi günlük kişi başı 195 gramın üstünde et tüketimi olan ülkeler, yüksek ya da yüksek orta GNI grubunda yer almaktadır. Günlük kişi başı et tüketimi 195 gramın altında olan ülkelerden günlük kişi başı şeker ve tatlandırıcı tüketimi 34,76 gramın altında olan ülkeler düşük GNI grubunda yer almaktadır. Günlük kişi başı şeker ve tatlandırıcı tüketimi 34,76 ve 43,79 arasında olan ülkeler ise günlük diğer nişastalı gıdalar tüketimine göre düşük veya düşük orta GNI grubunda yer almaktadır. Günlük kişi başı şeker ve tatlandırıcı tüketimi 43,79 gramın üzerinde olan ülkeler ise kuru baklagiller ve diğer su ürünleri tüketimine göre gruplandırılmaktadır.

GNI'ye göre gruplanan ülkelerde önemli gıda tüketim değişkenlerinden olan et, şeker ve tatlandırıcılar, alkollü içecekler, yumurta, kuru baklagiller ve sebze tüketiminin GNI gruplarına göre histogram grafikleri Şekil 4'de sunulmaktadır. Grafikten de görüleceği gibi yüksek GNI gruplarında et, şeker ve tatlandırıcılar, alkollü içecekler, yumurta, tüketimi fazla, kuru baklagiller tüketimi ise düşüktür.



Şekil 4. Gıda tüketimlerinin GNI gruplarına göre histogram grafikleri

Ülkelerin gıda tüketim gruplarının GNI grupları ayrımında ortalamaları Tablo 3'de sunulmaktadır. Buna göre et, alkollü içecekler, bitkisel yağlar, balık, gıda benzerleri, hayvansal yağlar, meyveler, sakatat, süt, şeker, tatlandırıcılar ve yumurtada GNI azaldıkça günlük kişi başı ortalama tüketim azalmaktadır. GNI azaldıkça, tahıl ve kuru baklagil tüketiminde ise artış gözlemlenmektedir. Düşük GNI grubunda günlük kişi başı ortalama diğer su ürünleri, hayvansal yağlar, sert kabuklular, alkollü içecekler, gıda benzerleri, yumurta, süt, et ve meyve tüketimi; yüksek GNI grubundaki tüketime göre sırasıyla %95, %92, %92, %89, %87, %86, %79, %78 ve %77 azalma göstermektedir.

Tablo 3. Ülkelerin gıda tüketim gruplarının GNI grupları ayrımında ortalamaları (g/gün/kişi)

Değişken adı	H	UM	LM	L
Alkollü içecekler	273,857	160,438	78,175	28,878
Baharatlar	1,880	3,368	3,154	0,895
Balık	77,931	52,576	53,213	19,501
Bitkisel yağlar	47,466	36,006	25,798	21,934
Et	241,541	147,958	97,111	53,118
Gıda benzerleri	24,766	13,672	9,471	3,133
Hayvansal yağlar	30,295	12,085	7,302	2,288
Kuru baklagiller	9,648	13,400	13,183	27,813
Meyveler	313,808	230,566	210,939	70,976
Niştastalı gıdalar, diğer	183,421	156,530	200,529	248,133
Sakatat	11,168	7,730	7,585	6,210
Sebzeler	325,058	223,399	252,421	125,675
Sert kabuklular	13,077	3,959	5,782	1,073
Su ürünleri, diğer	0,600	0,172	0,119	0,031
Süt	607,766	390,595	256,063	128,126
Şeker ve tatlandırıcılar	119,071	108,989	83,121	34,650
Tahıllar	322,386	377,939	391,670	432,508
Yağlı tohumlar	13,214	16,876	43,770	26,009
Yumurta	30,228	22,634	16,195	4,238

4. TARTIŞMA VE SONUÇ

Bu çalışmada kullanım alanı gün geçtikçe yaygınlaşan veri madenciliği uygulaması kullanılarak karar ağacı algoritması ile oluşturulan bir model yardımıyla ülkelere ait GNI gruplarına göre gıda tüketim eğilimleri tespit edilmiştir. Analiz doğruluk oranı %97,28 olan modelde, hedef değişkenin tamamına yakını doğru olarak tahmin edilmiştir.

Oluşturulan modelde; şeker ve tatlandırıcılar, alkollü içecekler, yumurta, kuru baklagiller, sebzeler, diğer su ürünleri, balık, diğer nişastalı gıdalar ve meyve tüketimi önemli değişkenler olarak bulunmuştur. GNI'si yüksek olan ülkelerde, aynı zamanda karar ağacının ilk bölünmesi olan et tüketiminin belirli bir değerin üstünde olduğu tespit edilmiştir.

Model sonuçlarına göre; GNI'si düşük ülkelerde, günlük kişi başı ortalama et, alkollü içecekler, bitkisel yağlar, balık, gıda benzerleri, hayvansal yağlar, meyveler, sakatat, süt, şeker, tatlandırıcılar ve yumurta tüketiminin düşük; tahıl ve kuru baklagil tüketiminin ise yüksek olduğu tespit edilmiştir. Çalışmadan elde edilen bu sonuçlar, ayrıntıları "Giriş" kısmında verilen Leenes (2006) ve Mancini vd. (2009) tarafından yapılan çalışmalardan elde edilen sonuçlarla tutarlılık göstermektedir.

Bu sonuçlar ışığında, gelir seviyesi düşük olan ülkelerde yüksek fiyatları nedeniyle et, süt, meyve, yumurta gibi besleyici yönü güçlü gıdaların tüketiminin düşük olduğu, tahıl ve kuru baklagil gibi fiyatı düşük ürünlerin tüketiminin ise yüksek olduğu anlaşılmaktadır. Bu durum, gelir seviyesi düşük olan ülkelerde beslenme yetersizliğinden kaynaklı sorunlara neden olmaktadır.

Çalışma sonuçlarına göre gelir seviyesi yüksek olan ülkelerde; şeker, tatlandırıcılar, alkollü içecekler, bitkisel ve hayvansal yağ tüketiminin yüksek olduğu tespit edilmiştir. Bu gıda gruplarının fazla tüketiminin gelir seviyesi yüksek olan ülkelerde başta obezite olmak üzere dengesiz beslenmeden kaynaklanan çeşitli sorunlara yol açtığı bilinmektedir. Ayrıca, obezite ve fazla alkol tüketiminin, gelişmiş ülkelerde en sık rastlanan ölüm nedeni olan kardiyovasküler hastalıklar için risk faktörleri arasında yer aldığı bilinmektedir (Samur ve Yıldız, 2008) .

5. KAYNAKLAR

Akpınar, H., 2000. Veri tabanlarında bilgi keşfi ve veri madenciliği. İ.Ü. İşletme Fakültesi Dergisi, 29, 1-22.

Batmaz, İ., Köksal, G., 2010. "Overview of knowledge discovery in databases process and data mining for surveillance technologies and early warning systems", in *Surveillance Technologies and Early Warning Systems: Data Mining Applications for Risk Detection*. Koyuncugil, A. S., Özgülbaş, N. (editors), IGI Global Publisher, (basım aşamasında).

Baykal, A., 2006. Veri madenciliği uygulama alanları. D.Ü. Ziya Gökalp Eğitim Fakültesi Dergisi, 7, 95-107.

Çingir, H., 2010. Veri madenciliğine giriş-ders notları, <http://yunus.hacettepe.edu.tr/~hcingir/ist376a/6Bolum.doc>, 15 Ağustos 2010.

FAO, 2010. <http://faostat.fao.org/site/345/default.aspx>, 01 Ağustos 2010.

Güntürkün, F., 2007. A comprehensive review of data mining applications in quality improvement and a case study. O.D.T.Ü., Ankara (in English).

Han, J. Kamber M., 2006. Data Mining: Concepts and Techniques. 2nd ed., Morgan Kaufmann Publishers, USA.

Kök, B. V., Kuloğlu, N., 2005. Sollama esnasında taşıt ve yol ile ilgili faktörlerin karar ağacı yöntemiyle irdelenmesi. Erciyes Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 21, 180-188.

Leenes, P. W., 2006. Natural resource use for food: land, water and energy in production and consumption systems. Groningen, Holland.

Mancini, L., Lettenmeier M., Rohn, H., Liedtke C., 2009. Using MIP methodology to evaluate food-farming systems sustainability. http://fromGNItoWellbeing.univpm.it/doc/papers/GNI2WB_034.pdf, 03 Eylül 2010.

Moore, A. W., 2003. Information gain. <http://www.autonlab.org/tutorials/infogain11.pdf>, 15 Eylül 2010.

Samur, G., Yıldız, E., 2008. Obezite ve kardiyovasküler hastalıklar/Hipertansiyon. http://www.beslenme.saglik.gov.tr/content/files/yayinlar/kitaplar/obezite_bilgi_serisi/D6.pdf, 08 Aralık 2010.

WB, 2010a. Ülke sınıflamaları. <http://data.worldbank.org/about/country-classifications>, 01 Eylül 2010.

WB, 2010b. <http://databank.worldbank.org/ddp/home.do?Step=2&id=4&DisplayAggregation=N&SdmxSupported=Y&CNO=2>, 14 Eylül 2010.

DETERMINING COUNTRIES' FOOD CONSUMPTION PATTERNS WITH RESPECT TO NATIONAL INCOME PER CAPITA - A DATA MINING APPLICATION

ABSTRACT

Food consumption patterns of countries by gross national income (GNI) groups are determined with a model by using GNI per capita and daily food consumption per capita data. The model was formed with data mining application, the most important stage of the knowledge discovery in databases. C5.0, one of the decision tree's algorithms giving powerful results in SPSS Modeler, was used to form the model. According to the model, the consumption of meat, sugar and sweeteners, alcoholic beverages, eggs, pulses, vegetables, other aquatic products, fish, other starchy foods and fruits are determined to be important variables by GNI groups. It is also determined that when GNI increases, daily average consumptions per capita of meat, alcoholic beverages, vegetable oils, fish, stimulants, animal fats, fruits, offal, milk, sugar and sweeteners and eggs increase; while those of cereals and pulses decrease.

Keywords: Data mining, Decision trees, Food consumption, GNI.