

Kolektif Öğrenme Algoritmalarıyla Çocuklarda Obezite Hastalığına Yakalanma Olasılıklarının Hesaplanması

Beyza Nur KÖKCÜ¹

Rukiye Dilruba KÖSE²

Faruk BULUT³

M. Fatih AMASYALI⁴

^{1,2} Özel Büyükçekmece Fatih Fen Lisesi, Beylikdüzü, İSTANBUL

^{3,4} Bilgisayar Mühendisliği Bölümü, Elektrik-Elektronik Fakültesi
Yıldız Teknik Üniversitesi, Davutpaşa, İSTANBUL

Email: beyzanurkokcu@gmail.com

dilrubakose@gmail.com

f0110303@std.yildiz.edu.tr

mfatih@ce.yildiz.edu.tr

Özet

İleri yaşlarda meydana gelen obezite hastalığının temelini çocukluk yıllarındaki beslenme ve yaşam alışkanlıklarıyla ilgili olduğu bilimsel çalışmalarla tespit edilmiştir. Çalışmamız çocuklarda obeziteye yakalanma riskini hesaplayan bir erken uyarı sistemi üzerinedir. Makine öğrenmesi kolektif öğrenme algoritmaları (Ensemble Methods) kullanılarak yapay ve özgün bir klinik karar destek sistemi geliştirilmiştir. Bu amaçla çalışmamızda kullanılmak üzere hazırlanan bir anket ile devlet hastanelerinden ve okullarından veriler toplanarak bir eğitim seti oluşturulmuştur. Eğitim seti üzerinde kolektif öğrenme algoritmalarıyla yapılan öğrenme ve çapraz geçirme işlemlerinde başarılı sonuçlar elde edilmiştir.

1. Giriş

Dünyadaki yaygın hastalıklardan biri olan obezite, besinlerle alınan enerjinin harcanan enerjiden fazla olması ve fazla enerjinin vücutta yağ olarak depolanması sonucu ortaya çıkan bir hastalıktır. Dünya Sağlık Örgütü tarafından obezite, sağlıklı bozacak ölçüde vücutta aşırı yağ birikintisi olarak tanımlanmıştır [1]. Klinik olarak obeziteyi tanımlamak için vücut kitle indeksi (VKİ) ölçütü olarak kullanılır ve (1) numaralı formül ile hesaplanır. VKİ sonucuna göre bireyin obez olup olmadığı Tablo 1'deki verilere göre sınıflandırılır.

$$VKI = \frac{Ağırlık}{boy^2} \quad (1)$$

Tablo 1. VKİ sınıflandırması

Vücut Kitle İndeksi (VKİ)	Sonuç
18.5 kg / m ² 'den düşük	Zayıf
18,5-24,9 kg / m ² arasında	Normal kilolu
25-29.9 kg / m ² arasında	Fazla kilolu
30-39.9 kg / m ² arasında	Obez(şişman)
40 kg / m ² 'den büyük	İleri derecede obez

Çocukluk çağında başlayan obezitenin erişkin dönemde de devam etmesi ve sağlık için risk oluşturması söz konusudur. Çocuklukta obezite, 5-6 yaş arası ve gelişim döneminde artış göstermektedir. Obez çocukların yaklaşık %35'i, obez adolesanların ise %80'i erişkin yaşa ulaştıklarında da obez kalmaktadırlar. Diğer yandan erişkin yaşlarda görülen obezite vakalarının %30 kadarında başlangıcın çocukluk çağlarına dayandığı bilinmektedir [2]. Bu nedenle çocukluk ve ergenlik döneminde obeziteden korunma ve tedavi giderek önem kazanmaktadır. Gerekli tedbirler erken yaşta alınmadığında hastalığın ileri yaşlarda tedavi edilmesi ise daha da zorlaşmaktadır[3].

Bu nedenle çocuklarda obezite riskinin erken tespit ve tedavi edilmesi daha da önemli hale gelmektedir. Çalışmamız erken teşhis koyan bir klinik karar destek sistemi üzerinedir. Erken tespit ile alınacak önlemler ve yapılan tedaviler sonucunda çocuğun bir ömür boyu sağlıklı bir hayat geçirmesi amaçlanmıştır. Çalışmamızda ikinci bölümde oluşturulan modele, üçüncü bölümde kullanılan metotlara, dördüncü bölümde deneysel sonuçlara ve beşinci bölümde değerlendirilmelere yer verilmiştir.

2. Model Oluşturma ve Veri Toplama

Çalışmanın işleyişi temel olarak dört ana aşamadan oluşmaktadır. Bunlar gerekli verilerin elde edilebilmesi için anket sorularının hazırlanması, okul ve hastanelerden verilerin anket yardımıyla toplanması, algoritmaların uygulanması ve sonuçların elde edilip yorumlanması şeklindedir.

2.1. Anket Oluşturma ve Veri Seti Hazırlama

Çalışmada kullanılmak üzere özgün bir eğitim seti oluşturulması amaçlandı. Bu amaçla obezite riskini tespit etmek için hazırlanan anket soruları 6 ana bölüme ayrılmıştır:

1. Bireyin temel özellikleri ve VKİ değeri,
2. Psikolojik durumu,
3. Aile bilgileri,
4. Haftalık aktivite sıklığı,
5. Günlük öğün sıklığı,
6. Bazı besin gruplarını tüketim sıklığı.

2.1.1. Temel özellikler

Bu bölümde bireyin cinsiyeti, yaşı, kilosu ve boyu sorularak VKİ değeri hesaplanmıştır.

2.1.2. Psikolojik özellikler

Psikolojik sorunların yemek yeme alışkanlığını tetiklediği bilimsel çalışmalarda yer almaktadır [4-5]. Bundan dolayı bu bölümde 3 adet soru sorularak bireyin psikolojik durumu değerlendirildi. Psikolojik bir rahatsızlığın olup olmadığı, çok üzücü bir olay yaşanıp yaşanmadığı ve ailevi sorunların olup olmadığı soruldu.

2.1.3. Ailevi özellikleri

Bu bölümdeki sorular, çocukların genetik olarak obeziteye yatkın olup olmadığını incelemek amacıyla bilimsel çalışmalardan yararlanarak [4-5-6] hazırlanmıştır.

1. Ailede (şeker hastalığı) var mı?
2. Ailede obez olan var mı?
3. Ailede sigara içen var mı?
4. Anne boy ve kilo değerleri nedir?
5. Baba boy ve kilo değerleri nedir?
6. Aylık gelir durumunuz nedir?

2.1.4. Aktivite sıklığı

Hareketsiz yaşam ve elektronik cihaz karşısında geçirilen fazla süre obezite riskini arttırmaktadır [2]. Bireyin aktivite sıklığını incelemek amacıyla sportif ve bedensel aktivitelere ayrılan süre ile bilgisayar ve televizyon karşısında geçirilen süre sorulmuştur.

2.1.5. Öğün alışkanlıkları

Öğünlerin sıklık durumu, ara öğünlerde abur cubur yeme alışkanlığı, ayaküstü tarzı yiyeceklerin tüketim sıklığı obeziteye neden olduğu literatürde ispatlanmıştır [3,5] ve Tablo-2'deki sorular sorulmuştur.

Tablo 2. Öğün alışkanlıklarıyla ilgili sorular

	Sıklıkla Her Gün	Arada bir Normal	Hiç Çok az
1) Kahvaltı			
2) Öğle yemeği			
3) Ara öğün			
4) Akşam yemeği			
5) Gece yatarken			

2.1.6. Besin tüketim sıklığı

Bireyin Tablo-3'te bulunan besinleri tüketme sıklığı çocuklarda obezite oluşum riskini etkilediği yapılan çalışmalarla ortaya konmuştur [5-6].

Tablo 3. Besin tüketim sıklığı ile ilgili sorular

	Sıklıkla Her Gün	Arada bir Normal	Hiç Çok az
1) Süt ve süt ürünleri			
2) Et ve et ürünleri			
3) Tahıl ürünleri			
4) Reçel, bal Tereyağı			
5) Fast-food Hamur işi Cips patates Gazlı içecek Meyve suyu Şeker çikolata			
6) Meyve-Sebze			

2.2. Veri Toplama ve Eğitim Seti Oluşturma

Alınan resmi izin belgeleriyle İstanbul Üniversitesi Çapa Eğitim ve Araştırma Hastanesi, Şişli Etfal

Eğitim ve Araştırma Hastanesi bünyesindeki Çocuk Sağlığı ve Anabilim dalında Büyüme-Gelişme ve Pediatrik Endokrinoloji bölümlerinde anketler yapıldı. 7-15 yaş aralığındaki 60 kadar çocuk obez hastalardan toplanılan veriler eğitim setine eklendi. Ayrıca Büyükçekmece İlçe Milli Eğitim Müdürlüğünden alınan diğer bir resmi izinle Adem Çelik İlk Okulu ve Orta Okulunda, 7-15 yaş aralığındaki obez olmayan 40 öğrenciye anketler uygulanarak eğitim seti güçlendirildi.

Böylece 100 veriden oluşan özgün bir eğitim seti kullanılan algoritmalar için hazır hale getirildi. Bölüm 2.1.'de 6 kategoriye ayrılan anket soruları oluşturulan eğitim setinde kullanılabilmesi için sayısal değerlere dönüştürüldü. Sorulara etki derecelerine göre puan verildi. Her bir kategori eğitim seti için bir özellik (*attribute*) olarak düşünüldü. Tüm veriler daha sonra normalize edildi.

3.Yöntemler

Sınıflandırma yöntemlerinden biri olan kolektif öğrenme (*Ensemble Learning*) metotları tekil öğrencilerin öğrenme başarısını artırmak için kullanılmaktadır. Çalışmamızda temel öğrenci (*base learner*) olarak örnek tabanlı öğrenciler tercih edilmiştir. C programlama dilinde yazılabilmesi ve özelleştirilebilmesinden dolayı tekil öğrenci olarak *k* en yakın komşuluk algoritması (*kNN*) seçilmiştir.

3.1. Tekil Öğrenciler

kNN sınıflandırıcısının farklı türü temel öğrenci olarak kullanıldı:

1. Normal *kNN*
2. $1/d^2$ uzaklık ağırlıklı *NN*

Örnekler arası uzaklığı hesaplamak için Öklid yöntemi kullanıldı. Öklid uzaklık ölçüsü kullanılarak *N* boyutlu uzayda *a* ve *b* noktaları arasındaki uzaklık (2) numaralı formül ile hesaplanır:

$$d(a, b) = \sqrt{\sum_{i=1}^N (x_{ai} - x_{bi})^2} \quad (2)$$

Bilindiği üzere *kNN* yöntemi, en yakın *k* adet komşunun sınıf etiketlerinin ortalaması ile bulunur. $1/d^2$ uzaklık ağırlıklı *kNN* yöntemi, örneklerin test örneğine olan uzaklıklarına bağlı olarak hesaplamadaki ağırlıklarını dikkate alarak sınıflandırma yapar. Bu teknikte yakındaki kayıtların

hesaplamadaki ağırlığı fazla; uzaktakilerin ağırlığı ise azdır [9-10]. Her bir kaydın hesaplamadaki ağırlığı şu formüller ile bulunur:

$$w_i = \frac{1}{d_i^2} \quad (3)$$

x test örneğinin sınıfı ise (4) numaralı formül ile bulunur:

$$f(x) = \frac{\sum_{i=1}^k w_i f(a_i)}{\sum_{i=1}^k w_i} \quad (4)$$

Formüldeki *k*, *kNN*'deki *k*'yı; *a_i* ise hesaplamaya katılan *k* adet örneğinin etiketini göstermektedir.

3.2. Kolektif Öğrenme Metotları

Kolektif öğrenmede kullanılan tekil öğrenciler genellikle bir karar ağacından oluşmaktadır; yerine göre *SVM (Support Vector Machine)* gibi başka bir sınıflandırıcı da olabilir. Tekil öğrencilerin bir araya gelerek öğrenci topluluğu (*ensemble*) oluşturmalarına kolektif öğrenme denir. Birçok uygulamada kolektif öğrenme başarısı tekil öğrenmeye göre yüksektir [7-8]. Örnek tabanlı tekil öğrencilere uygun olduğundan dolayı aşağıdaki kolektif öğrenme metotları kullanılmıştır:

1. Yerine koyarak örnekleme (*Bagging*)
2. Oylama (*Voting*)
3. Rastsal alt uzaylar (*Random Subspaces*)

3.2.1. Yerine koyarak örnekleme (*Bagging*)

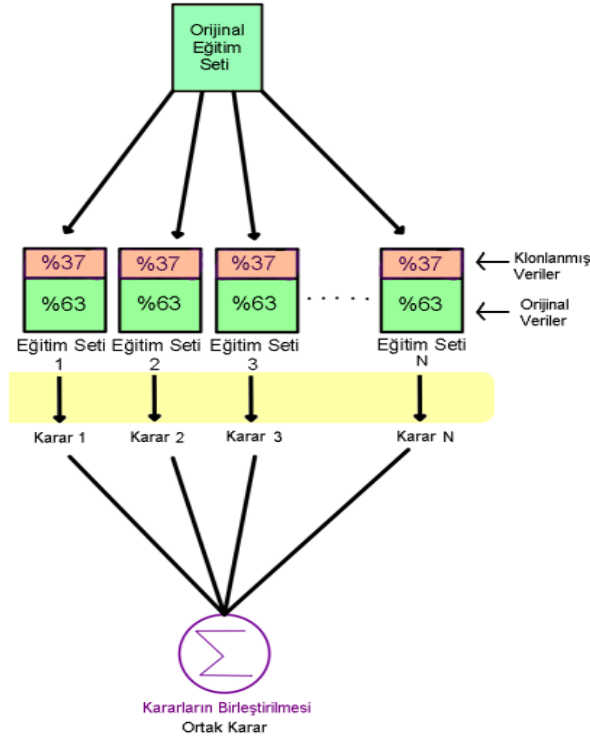
Bagging metodu (*Boot strapping AGGregating*), var olan bir eğitim setinden yeni eğitim setleri türeterek temel öğrenciyi yeniden eğitmeyi amaçlayan bir yöntemdir. Şekil-1'de yöntemin çalışma stili görülmektedir.

Eğitim setinden yaklaşık olarak %63,2 kadar orijinal örnek alınır ve alınan örneklerden bazıları çoğaltılarak eğitim setleri %100'e tamamlanır. Her defasında bu işlem yapılarak farklı eğitim setleri elde edilir. Her eğitim seti temel öğrenciye uygulanır ve alınan kararlar demokrasi usulüyle birleştirilir.

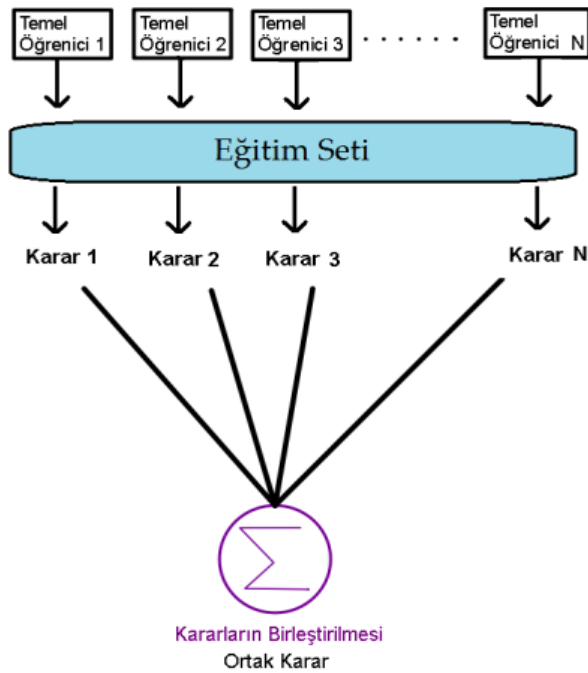
3.2.1. Oylama (*Voting*)

Oylama tekniğiyle sınıflandırmada en çok oy olan etiketin sonucu baz alınır. Basit oylamada tüm sınıflandırıcıların ağırlığı birbirine eşittir. Her bir

sınıflandırıcının tüm sınıf etiketleri hakkında verdikleri kararlar birleştirilir ve ortalamaları alınır. En yüksek oranın çıktığı sınıf etiketi test örneğine atanır. Çalışmamızda ikili sınıflandırma yani {0,1} sınıfları vardır. Şekil-2’de algoritmanın çalışma stili görülmektedir.



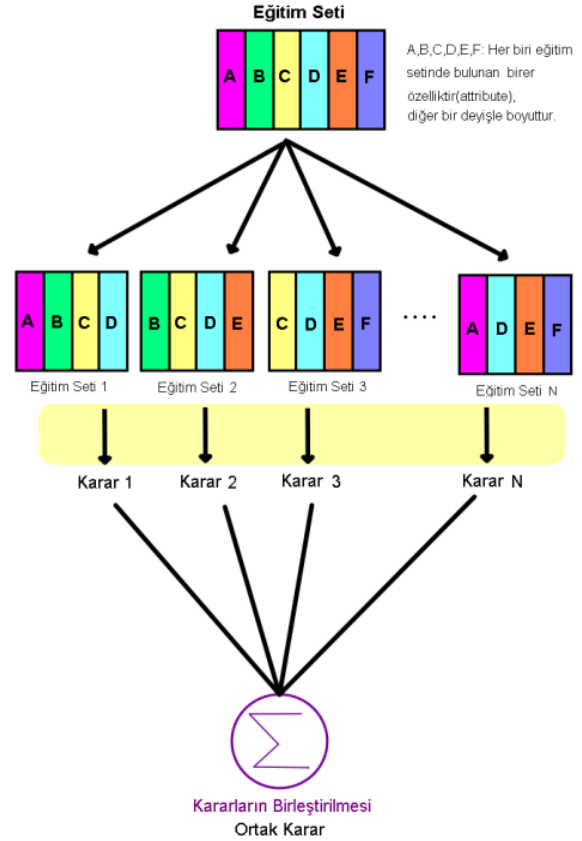
Şekil 1. Bagging yöntemi.



Şekil 2. Oylama (voting) yöntemi.

3.2.1. Rastsal altuzaylar (Random subspaces)

Bu yöntem, öğrenci topluluğundaki her bir öğrenciyi aynı eğitim seti ile eğitmek yerine aynı eğitim setinin bazı özelliklerini hariç tutup yeni eğitim setleri türeterek öğrencileri eğitmeyi amaçlamaktadır. Her bir öğrenci rastgele seçilen özellikler grubu ile oluşturulmuş yeni eğitim setleri ile eğitilir. Öğrencilerin farklı kararları birleştirilerek komite kararı oluşturulur [9]. Şekil-3’de bu tekniğin çalışma stili görülmektedir.



Şekil 3. Rastsal alt uzaylar yöntemi.

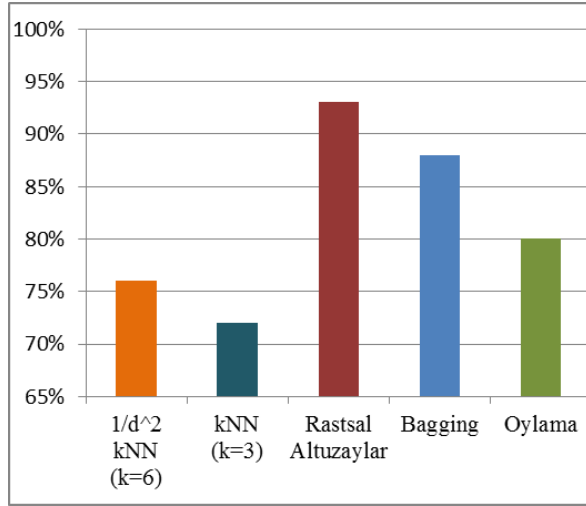
4. Deneysel Sonuçlar

Deneysel sonuçlar üç ana başlık altında incelenmiştir. Bunlar öğrencilerin doğruluk oranları, özelliklerin sınıflandırmadaki bilgi kazançları ve özelliklere ait korelasyon matrisi şeklindedir.

4.1. Doğruluk Oranları

Çalışmamızda, kNN ve uzaklık ağırlıklı kNN yöntemleri için değişik k parametreleriyle 21 adet birbirinden farklı temel öğrenici oluşturuldu. Temel öğrenciler eğitim seti üzerine uygulandı. Temel öğrencilerin eğitim seti üzerindeki doğruluk oranları

%70 ile %76 arasında değişmektedir. 21 farklı sınıflandırıcının doğruluk oranlarındaki farklılık 1 ile 15 arasında alınan değişik k parametreleriyle sağlanmaya çalışılmıştır. Fakat kolektif öğrenmenin başarılı olabilmesi için öğrencilerin kararları arasında farklılık (*diversity*) olmalıdır. kNN sınıflandırıcısının farklı k parametrelili türleri farklılığı sağlayamamıştır. Bu nedenle 21 temel öğrencinin doğruluk oranları birbirine çok yakındır.



Şekil 4. DeneySEL sonuçlar

Şekil-4’de görüldüğü üzere oylama ve yerine Oylama ve yerine koyarak örnekleme (*bagging*) yöntemlerinde de sırasıyla %80 ve %88 başarı elde edilmiştir. Bu durum doğruluk açısından bir miktar gelişme sağlandığını göstermektedir. Güncel hayatta da alınan ortak kararın çoğu zaman doğru sonuçlar verdiği ve güvenilir olduğu bilinmektedir.

Tekil öğrenci olarak seçilen örnek tabanlı öğrencilerin kolektif öğrenmede kullanılmasının başarıyı istenilen düzeyde artırmadığı görülmüştür. Rastsal alt uzayların %93 gibi çok yüksek bir oranda başarı elde etmesinin nedeni türetilen yeni eğitim setlerinde farklılığın (*diversity*) kendiliğinden oluşması olarak görülmüştür. 6 boyutlu eğitim setinden 4 boyutlu olarak seçilen 10 farklı eğitim seti kombinasyonu çoğul öğrenme başarısını artırmıştır. Eğitim setindeki VKİ ve beslenme alışkanlıklarının bulunduğu özellikler rastsal alt uzaylar uygulamasında silinmediğini önemle vurgulamak isteriz. Bu özellikler obezite riskinin belirlenmesinde büyük bir öneme sahiptir. Bu özelliklerin değişik kombinasyonlarda bulunmaması durumunda obeziteyi tespit etmek zorlaşır ve bu da doğruluk oranının düşük çıkmasına sebep olur.

4.2. Bilgi Kazancı

Weka yazılımı kullanılarak oluşturduğumuz eğitim setinin elde edilen bilgi kazançları (*information gain*) Tablo-4’de olduğu gibidir.

Elde edilen bu değerlere göre obezitenin birinci nedeni VKİ çıkmıştır. VKİ, obeziteyi tanımlayan bir faktör olduğu için bir neden olarak görülmemeli, VKİ-Obezite arasında yüksek bir korelasyon gibi değerlendirilmelidir. 2. sırada yer alan bazı besin gruplarının tüketim sıklığı obeziteyi tetikleyen en önemli faktör olarak düşünülmelidir. Çok az bir farkla 3. sırada yer alan bireyin içinde bulunduğu psikolojik durum da diğer bir etmendir.

Tablo 4. Özelliklerin bilgi kazançları

Sıra	Özellik	IG
1	VKİ	0.2894
2	Bazı besin gruplarını tüketim sıklığı	0.1947
3	Psikolojik durumu	0.1828
4	Günlük öğün sıklığı	0.1548
5	Aktivite	0.1236
6	Aile bilgileri	0.0704

4.3. Korelasyon Matrisi

Tablo-5’te eğitim setinde bulunan özelliklerin matris korelasyonları hesaplanmıştır. Özelliklerin kısaltmaları şu şekildedir:

- Öz-1. Bireyin temel özellikleri ve VKİ değeri,
- Öz-2. Psikolojik durumu,
- Öz-3. Aile bilgileri,
- Öz-4. Haftalık aktivite sıklığı,
- Öz-5. Günlük öğün sıklığı,
- Öz-6. Bazı besin gruplarını tüketim sıklığı.

Tablo 5. Özelliklerin matris korelasyonu

	Öz-1	Öz-2	Öz-3	Öz-4	Öz-5	Öz-6
Öz-1	1	0.17	0.10	0.06	0.05	0.46
Öz-2		1	0.65	0.25	0.20	0.22
Öz-3			1	0.17	0.07	0.42
Öz-4				1	0.16	0.47
Öz-5					1	0.78
Öz-6						1

Bilindiği üzere -1 değerlikli korelasyon, iki değişken arasında tam ters ilişkiyi; 0, ilişkisizliği; +1 ise tam lineer düz ilişkiyi göstermektedir. Genel olarak 0.80 ve üzeri değerler anlamlı bir korelasyonun göstergesidir. Öz-5 (*günlük öğün sıklığı*) ile Öz-6 (*bazı besin gruplarını tüketim sıklığı*) arasındaki 0.78'lik değer arada güçlü bir korelasyon olduğunu göstermektedir. Yani ankete katılan öğrencilerin gün içindeki öğün sayıları ile bazı besin gruplarını tüketme sıklıkları arasında doğru orantı oluşmuştur.

Tablodaki diğer verilerin düşük olması ankette sorulan soruların birbirlerinden bağımsız olduğunu göstermektedir.

5. Değerlendirme ve ileri uygulamalar

Çalışmamızda sınıflandırma başarısını artırmak için kullanılan kolektif öğrenme yöntemleri farklı bir alana uygulandı. Orijinal, güvenilir ve gerçek bir eğitim seti kullanılarak kolektif öğrenciler ile çağın büyük bir sorununa karşı *linik karar destek sistemi* geliştirildi. Oluşturulan bu erken uyarı sistemi sayesinde gelişmekte olan bir çocuğun obezite riski çocuktan alınan bilgilerle tahmin edilmeye çalışılabilir. Çıkan sonuçlara göre yüksek risk yüzdesine sahip kişiler haberdar edilip ilgili kurumlara yönlendirilebilir.

6. Teşekkürler

Anket çalışması yapabilmemiz için bize resmi izin veren, destekleyen, hoş gören İstanbul İl Milli Eğitim Müdürü Dr. Muammer YILDIZ'a, İstanbul İl Sağlık Müdürlüğüne, Şişli Etfal Eđt. Ve Arşt. Hastanesi ile Çapa Eđt. Ve Arşt. Hastanesi yönetimine ve çalışanlarına, B.Çekmece Adem Çelik İlkokulu yönetici ve öğretmenlerine, İstanbul Çapa Eđt. Ve Arşt. Hastanesi Büyüme-gelişme ve Pediatrik Endokrinoloji bilim dalı Öğretim Üyesi Prof.Dr. Nurçin SAKA'ya, Şişli Etfal Eđt. Ve Arşt. Çocuk Endokrinoloji Doktoru Uz. Dr. Mehmet BOYRAZ'a teşekkürlerimizi bir borç bilir saygı ve sevgilerimizi sunarız.

7. Kaynaklar

- [1] Erken Yaşta Görülen Obezite: Nedenleri ve Tedbirle, Doç. Dr. A. ONAY Beşikçi, Mised Dergisi; Sayı:23-24, s. 82-85, 2010.
- [2] A. Parlak, Ş. Çetinkaya, Çocuklarda Obezitenin Oluşumunu Etkileyen Faktörler, Fırat Sağlık

Hizmetleri Dergisi; Cilt:2, Sayı:5, s. 24-35, (2007).

- [3] F. Şimşek, B.Ulukol, M. Berberoğlu, S.B. Gülnar, Ankara'da bir ilköğretim okulu ve lisede obezite sıklığı, Ankara Üniversitesi Tıp Fakültesi Mecmuası; s. 163-166, 2005.
- [4] Çocukluk Çağında Obezite, Dr. K. Babaoğlu, Dr. Ş. Hatun, Kocaeli, Sted Dergisi; Cilt 11, 2002.
- [5] İlköğretim Çağındaki Çocuklarda Obezite Prevalansının Belirlenmesi ve Risk Faktörlerinin Araştırılması, S. Öztora, S. Hatipoğlu, M.B. Barutçugil, B. Salihoğlu, R. Yıldırım, E.Şevketoğlu, Bakırköy Tıp Dergisi; Cilt 2, 2006.
- [6] Çocukluk Döneminde Obezite, Prof. Dr. P.Cinaz, Sağlık&İnsan Dergisi; s. 32-34, 2012.
- [7] Ethem Alpaydın, Yapay Öğrenme kitabı, Boğaziçi Üniversitesi Yayınevi, İstanbul, 2010
- [8] M. Fatih Amasyalı, "Yeni Makine Öğrenmesi Metotları ve İlaç Tasarımına Uygulanması", Doktora Tezi, Yıldız Teknik Üni., İstanbul, 2008.
- [9] Tin Kam Ho, The Random Subspace Method for Constructing Decision, Pattern Analysis and Machine Intelligence, IEEE Transactions, Lucent Tech no 1., Volume: 20 , Issue: 8, pp: 832 - 844 , AT&T Bell Labs., Murray Hill, NJ , 1998