Anadolu Üniversitesi Bilim ve Teknoloji Dergisi A- Uygulamalı Bilimler ve Mühendislik
Anadolu University Journal of Science and Technology A- Applied Sciences and Engineering

2016 - Cilt: 17 Sayı: 1
Sayfa: 97 - 109
DOI: 10.18038/btda.45787
Geliş: 18 Ocak 2016 Düzeltme: 02 Mart 2016 Kabul: 21 Mart 2016

ÇOK KATMANLI ALGILAYICILAR İLE DOĞRU MESLEK SEÇİMİ

Faruk BULUT[1, *], Dilara BOZYILAN[2]

[1]Bilgisayar Mühendisliği Bölümü, Gediz Üniversitesi, İzmir
[2]Kabataş Erkek Lisesi, Beşiktaş, İstanbul

ÖZET

Meslek seçimi güncelliğini hiçbir zaman yitirmeyen bir konudur ve insan hayatını etkileyen önemli bir karardır. Bireyin kişiliğine, isteklerine ve yeteneklerine uygun yapılan tercihler bireyde mutluluğu ve iş hayatında başarıyı doğurur. Yapılan araştırmalarda Türkiye'de ve dünyada meslek seçiminin yaygın olarak yanlış yapıldığı görülmektedir. Bu çalışma, meslek seçiminde karar aşamasında olan lise öğrencilerine yapay bir karar destek sistemiyle yardımcı olmayı amaçlamaktadır. Geliştirilen bu sistemde, hata toleranslı, eğitilebilen, deneyime dayalı özgün Çok Katmanlı Algılayıcılar (ÇKA) sınıflandırıcısı kullanılmıştır. Değişik sektörlerde çalışan ve mesleğinde memnun olan kişilerden elde edilen verilerle bir eğitim seti oluşturulmuştur ve bu eğitim seti özelleştirilen ÇKA sınıflandırıcısına aktarılarak sistem eğitilmiştir. Yapılan çapraz geçerleme işlemlerinde ve uygulamalarda diğer sınıflandırıcılara göre yüksek performanslı sonuçlar elde edilmiştir. Geliştirilen ÇKA tabanlı yapay karar destek sistemi sayesinde öğrenicilerin çok az bir hata oranıyla doğru meslek tercihi yapmaları sağlanmıştır. Sonuç olarak önerilen yapay model sayesinde üniversite sınavına girmek üzere olan öğrencilerin kendileri için uygun bir meslek tercihi yapabilmeleri mümkün kılınmıştır.

Anahtar Kelimeler: Çok Katmanlı Algılayıcılar, Sınıflandırma, Meslek seçimi

RIGHT CAREER CHOICE USING MULTI-LAYER PERCEPTRON

ABSTRACT

Choosing a career is an important decision that affects people's lives. The right choice fitting the needs, personality and capabilities of the individual results a happy life. In latest researches carried out in Turkey and in the world, it has been shown widely that the wrong career choice widely exists. In this study, a theoretical and empirical approach is suggested in order to help high school students to choose the right profession for their future life. With the help of the proposed system in this work, it enables to choose the right profession with a low error rate for the students. In the proposed artificial decision support system, there is a Multi-Layer Perceptron (MLP) mechanism that has a fault-tolerant, trainable, and a unique experience based system. The MLP classifier has been customized for that special study so as to achieve more accurate prediction results. A training dataset which includes the data obtained from the employees, who are satisfied and happy in various sectors and professions, has been created in order to use in the MLP prediction mechanism. This dataset is transferred into the system to train the MLP predictor. In the cross validation process and application by using MLP, high-performance in classification has been obtained. As a conclusion, this proposed artificial model provides high school students to choose a right carrier before the university placement exam.

Keywords: Multi-Layer Perceptron, Classification, Choosing a career

1. GİRİŞ

Meslek seçimi, hem dünyada hem Türkiye'de önemli bir konudur ve bireyin geriye kalan yaşamında çok önemli bir yere sahiptir. Kişi, yetenek, ilgi ve istekleri doğrultusunda doğru bir meslek seçimi yaparsa başarılı, verimli ve mutlu olur. Buna karşılık bir kişi, kişisel özelliklerini ve yeteneklerini göz önünde tutmadan meslek seçimi yaptığında başarısız, verimsiz ve mutsuz olur. Bu nedenle birey, meslek seçerken kendi özellikleri ile seçeceği mesleğin nitelikleri arasında uyum olmasına dikkat etmelidir.

* Sorumlu Yazar: faruk.bulut@gediz.edu.tr

Araştırma bulgularına göre, yetenek, beceri, ilgi ve ideallerine uygun meslek seçmiş olanların sayısı az; iş veya mesleğinden şikâyet edenlerin sayısı oldukça fazladır [1].
Uğur Kariyer Merkezi'nin 2004-2010 yılları arasında elli binden fazla öğrenci üzerinde yaptığı araştırmada, kariyer planlama sürecinin ilk adımlarından biri olan meslek seçiminin çoğu kez gerçek dışı beklentilerle ya da moda akımlarıyla yapıldığını ortaya koymuştur. Bu araştırmada insanların %85'inin hatalı meslek seçimi yaptıkları görülmüştür [1]. Ayrıca TRT'nin haftalık olarak hazırladığı "1000 kişiye Sorduk" programında yapılan bir anket çalışmasında da benzer sonuçlar elde edilmiştir. Bireylerin %32,9'unun kendilerine uygun olmayan bir meslekte çalıştıkları; %42,3'ünün ise mesleklerinden memnun olmadıkları saptanmıştır [2].

Yanlış meslek seçiminin üniversite sınavı öncesi yapıldığı da bir gerçektir. Bu nedenle bu çalışma, orta öğretim öğrencilerin seçimlerine yardımcı olacak bir karar destek sistemi (*Decision Support System*) üzerinedir.

Birey - meslek uygunluğu üzerine yapılmış bazı istatistiksel tabanlı çalışmalar vardır [3, 4]. Ayrıca bazı web sitelerinde kişiye en uygun mesleğin bulunması ile alakalı bir dizi çoktan seçmeli testlere verilen cevaplara göre meslek tavsiyesi yapılmaktadır. Bu tarz çalışmalarda, sorulara verilen cevaplara belirli puanların verilmesiyle elde edilen puan aralıklarına bakılarak yönlendirme yapılmaktadır. Bu durum belirli önyargıların ve kalıplaşmış fikirlerin puanlama sistemiyle sayısallaştırılmasından dolayı doğru sonuçlar vermediği düşünülmektedir [5]. Örneğin bir kişi Matematik dersinde başarılı ise mühendis olmalı, hayvanları seviyorsa veteriner olmalıdır düşüncesi ile örtüşmektedir. Fakat yapılan bu çalışmasında istatistiksel değerler ve bazı kalıp düşüncelere yer verilmemiştir, sınıflandırma işlemi salt olarak elde edilen veriler ile eğitilen sınıflandırıcı ile yapılmıştır. Ayrıca değişik meslek gruplarındaki bireylerden yararlanıldığı için güvenilirlik ve doğruluk oranı en üst düzeye çıkarılmaya çalışılmıştır.

Karar destek sisteminde ÇKA ile sınıflandırma çalışması yapılarak birey için en uygun meslek seçilmeye çalışılmıştır. Değişik meslek gruplarında çalışan kişilerden elde edilen verilerle oluşturan eğitim seti ile ÇKA mekanizması eğitilmiştir. Bu çalışmada diğer sınıflandırıcılara göre ÇKA ile daha başarılı sonuçlar elde edilmiştir

Bu çalışma beş ana bölüm içermektedir. İkinci bölümde çalışmada kullanılacak olan veri setinin nasıl hazırlandığı anlatılmıştır. Üçüncü bölümde ÇKA'nın mimari yapısına, dördüncü bölümde elde edilen deneysel sonuçlara ve son değerlendirme bölümünden önceki bölümde de yapılabilecek ileriki çalışmalara yer verilmiştir.

2. VERİ SETİNİN HAZIRLANMASI

Bir tahmin mekanizmasının doğru sonuçlar vermesi için kullanılacak olan eğitim setinin güvenilir olması gerekir. Veri seti konu ile alakalı öznitelikleri içermeli ve yeteri sayıda örnek barındırmalıdır. Ayrıca veri setinin dengeli bir dağılıma sahip olabilmesi için hemen hemen her sınıftan eşit sayıda örnek bulundurulmalıdır. Aksi takdirde eğitim setinde önyargı oluşacaktır ve tahmin sonuçlarının hatalı çıkmasına neden olacaktır.

Bu konu ile ilgili sosyal bilimler alanında yapılan kaynak taramasında [3, 6, 7] bireyde doğru meslek tercihine etki eden 50 kadar muhtemel faktör tespit edilmiş ve hazırlanan ankete ilave edilmiştir. Makale sonunda Ek-1'de sunulan anket soruları bireyin kişisel özelliklerini irdeleyen ve mesleklere olan yatkınlığını analiz eden bir içeriktedir. Anket sorularında elde edilen veriler sayısallaştırılarak eğitim setine aktarılmıştır. 50 adet anket sorusundan ilk 43 tanesinin cevabı evet ve hayır şeklindedir. Eğitim setine evet cevabı 1 olarak; hayır cevabı ise 0 olarak kaydedilmiştir. 44-50 arasındaki soruların cevapları ise şıklıdır. Bu sorular veri setine nominal değerler olarak aktarılmıştır ve daha sonra ikili (*binary*) değerlere dönüştürülerek her bir nominal değer arasındaki uzaklığın (farklılığın) 1 olması sağlanmıştır. Anketteki soruların her biri ÇKA'nın giriş katmanı için öznitelik (*attribute*) olarak alınmıştır. Elde

edilen anket verileri eğitim setine aktarılırken kayıp değerler tolere edilmiştir. Verilerde normalizasyon işlemi uygulanmamıştır çünkü anket sorularının cevapları kategorilere ayrılmış bir yapıdadır. Ayrıca ankete katılan kişilerin verdikleri cevapların güvenilirliği, bazı sorulara verilen cevapların tutarlı olup olmadığı incelenerek anlaşılmaya çalışılmıştır. Dikkatsizce verilen cevaplardan ötürü çalışmanın güvenilirliğini sağlamak amacıyla bu veriler veri setine eklenmemiştir. Örneğin anketteki 11. ve 46. sorular anlam olarak benzerdir. Biri çocukluk çağında oynanan oyunlardaki liderliği, diğeri ise grup çalışmasındaki liderliği sormaktadır. Normalde her iki soruya da verilen cevabın aynı olması beklenir. Aynı şekilde 23.-26. ve 24.-25. soru ikilileri de benzer anlam içermektedir. Bu üç soru çiftine farklı cevap veren kişilerin anketleri çalışmamıza dâhil edilmemiştir.

Hazırlanan anket soruları en popüler beş meslek grubunda çalışan bireylere uygulanmıştır. Bunlar hukuk, tıp, mühendislik, sosyal bilimler ve medya sektörüdür. Daha çok meslek türü ve çeşidi olmasına rağmen çalışmada meslek türleri kısıtlanmıştır. Bunun nedeni çok sınıflı bir eğitim setinde sınıflandırma başarısının düşük çıkacak olmasındandır. Her kategoriden belirli kurumlarla görüşüp resmi izin alınmıştır. İzin alınan kurumlar arasında TRT İstanbul Radyosu, İstanbul Barosu, Büyükçekmece İlçe Milli Eğitim Müdürlüğü, İstanbul Bilgisayar Mühendisleri Odası ve ATV-Sabah Medya firması gibi kuruluşlar bulunmaktadır. Tablo 1'de görüldüğü gibi bahsi geçen bu kurumlarda çalışan toplam 260 kişiyle anket uygulanmıştır. Fakat eğitim setine sadece mesleğinde memnun 58 kişi dâhil edilmiştir. Mesleğinde memnun olmayan kişiler, karar destek sisteminin yanlış sonuçlar vermesini engellemek için eğitim setine alınmamıştır.

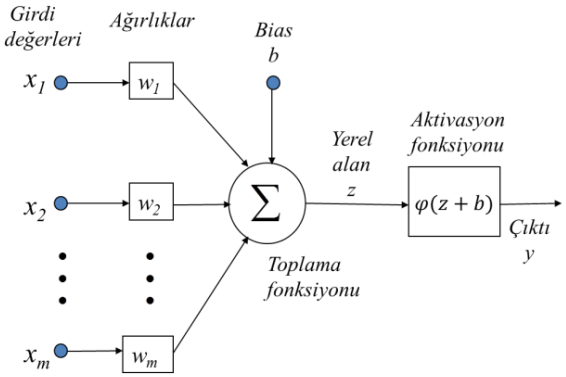
Tablo 1. Eğitim seti

Meslekler	Anket Yapılan Kişi Sayısı	Mesleğinden Memnun Olan Kişi Sayısı
Hukuk	40	13
Tıp	47	12
Mühendislik	63	11
Sanat	33	12
Medya	19	10

3. ÇKA OLUŞTURMA VE EĞİTME

Çalışmamızda ÇKA için, algılayıcılar kullanılarak denetimli bir öğrenme sistemi oluşturulmuştur. Sinir ağları ise geri yayılım (*back propagation*) yöntemi ile eğitilmiştir.

Bu bölümde ÇKA'nın çalışma prensibine, yapısına ve kullanılan aktivasyon fonksiyonuna ait kısa açıklamalara yer verilmiştir.



Şekil 1. Bir algılayıcının (*perceptron*) işleyişi

ÇKA, hata üzerine dayandırılmış bir öğrenme algoritmasıdır. Temel iki işlevi olan öğrenme ve karar verme aşamalarını ağırlıklandırma, aktivasyon fonksiyonu ve *bias* sayesinde yapar. Ağırlık, her girdinin bir sonraki aşamaya gitmeden önce çarpıldığı katsayıdır. Tüm girdiler, kendilerine ait ağırlıklarla çarpılarak toplanır. Daha sonra aktivasyon fonksiyonuna bu değerin gönderilmesi sonucu ortaya çıkan cevap, sistemin kararı olur. *Bias* ise, kullanıcıdan kullanıcıya, mekanizmanın çalışma şekline veya amacına göre değişebilen, kullanıcı tarafından eklenen bir parametredir. Şekil 1'de yapay bir sinir ağının çalışması gösterilmiştir [8].

Eğitim setinde bulunan veri sayısı, diğer bir deyişle veri setinin yoğun veya seyrek oluşu, öğrenme ve sınıflandırma başarısını etkileyen bir faktördür. Etiketli verilerin toplanması esnasında karşılaşılan güçlüklerden ötürü az sayıda veri eğitim setine eklenebilmiştir. Sınıflandırma başarısını ve doğal olarak sistemin güvenilirliğini artırmak amacıyla ÇKA sisteminde bir takım değişiklikler yapılmıştır. Yüksek sınıflandırma başarısı elde etmek amacıyla ÇKA sisteminde yapılan özgünleştirme işlemleri şu şekildedir:

1. Katman sayısı
2. Her katmandaki algılayıcı sayısı
3. *Bias* (*b*) katsayısının değeri
4. Eğitimdeki iterasyon sayısı
5. En uygun aktivasyon fonksiyonu

En uygun katman sayısının birçok veri için 3 olarak alındığı bilinmektedir [9]. Bu uygulamada da katman sayısı 3 olarak alınmıştır. Ayrıca yapılan bir dizi pratik uygulamada ve makale taramasında [10] *bias* parametresi için genel olarak en ideal katsayın 0.7 olduğu görülmüştür. Çok katmanlı yapay sinir ağlarında her bir algılayıcı için *bias* değeri farklı olabilir. Bu değer veri setinin istatistiksel ve meta özelliklerine göre, eğitim sürecinde belirlenen diğer parametrelerin başlangıç değerlerine ve öğrenme algoritmasına göre değişiklik gösterebilir.

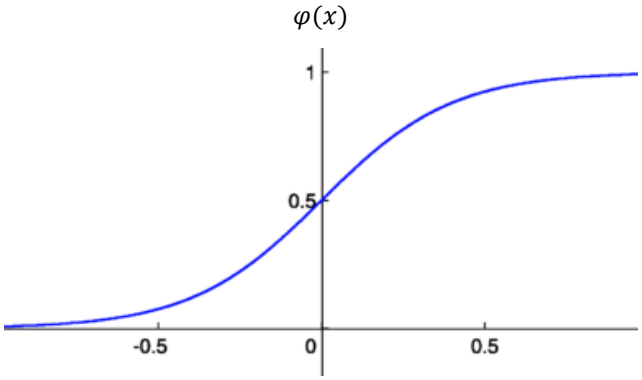
Şekil 1'de anket verilerinden elde edilen sorulara ait cevaplar girdi değerleri olarak algılayıcıya girilmektedir. Yapılan uygulamada toplam 50 anket sorusu bulunduğu için girdi adedi, $m = 50$ olarak alınmıştır. 1 numaralı formülde de görüldüğü gibi her bir x_i girdisi, belirli bir w_i ağırlık değeri ile çarpılmakta ve z toplama fonksiyonu ile lineer olarak birleştirilmektedir. Her bir giriş kendine ait bir ağırlığa sahiptir. Bir ağırlığa ait değerin büyüklüğü, o girişin yapay sinire güçlü bağlanması ya da önem derecesiyle doğru orantılıdır.

$$z = \sum_{i=1}^{m} w_i x_i \tag{1}$$

Literatürde tanımlanan birçok aktivasyon fonksiyonu vardı. Kullanılan eğitim setinin 50 boyutlu ve 5 sınıf etiketli seyrek (*sparse*) bir yapıdaki veri uzayı olduğu göz önüne alınırsa iyi bir seçim yapılması gerektiği ortaya çıkar. Doğru ve hassas sonuçların elde edilebilmesi için *Sigmoid* aktivasyon fonksiyonu tercih edilmiştir. Fonksiyonun çalışma prensibi Şekil 3'te, formülü ise 2 numaralı formülde görülmektedir [9, 10]. Bu nörona ait çıktının genliğini (*amplitude*) sınırlandıran *y* aktivasyon fonksiyonu ise 3 numaralı formülde görüldüğü gibidir.

$$\varphi(x) = \frac{1}{(1 + e^{-x})} \tag{2}$$

$$y = \varphi(z + b) \tag{3}$$

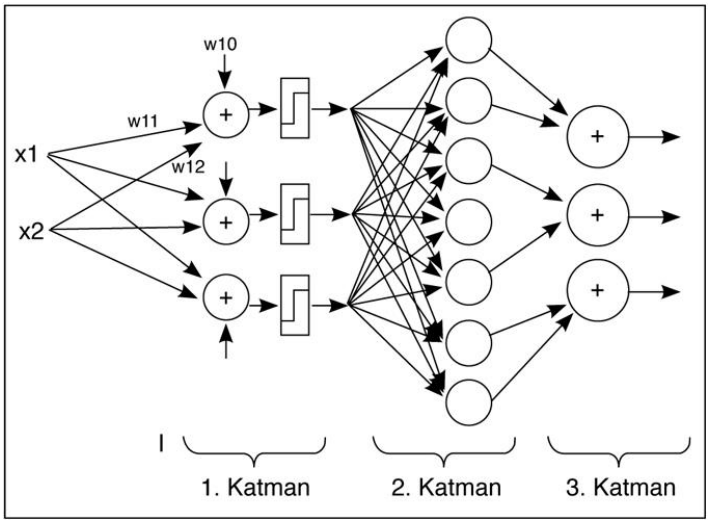


Şekil 2. Sigmoid aktivasyon fonksiyonu

Oluşturulan bir Çok Katmanlı Algılayıcının (ÇKA) çalışma prensibi, Şekil 3'te verilmektedir. ÇKA'nın düğümleri ve bağlantıları çok değişik biçimlerde bir araya getirilebilir. Ağ içerisindeki düğümler katmanlar halinde yerleştirilir. İlk katmandaki girişe verilen bilgi, ağ içinde ileriye doğru yayılır. Son katmandaki işaretler ağın çıkışıdır.

ÇKA'nın çalışma prensibini açıklamak amacıyla bir örnek verelim. Giriş katmanı olan ilk katmana, özniteliklerden oluşan bir *X* vektörü test girdisi olarak verilmiş olsun. Her bir öznitelik farklı *w* ağırlığıyla çarpılarak toplanır. Elde edilen değer *bias* katsayısı ile çarpılarak aktivasyon fonksiyonuna ve oradan da gizli katmana (*hidden layer*) iletilir. Daha sonra çıkış katmanına iletilen değer sınıflandırma sonucu olarak belirlenir [9].

Çalışmamız 5 meslek türü üzerinde olduğu için ÇKA sistemi 5 sınıf etiketi üzerinde olasılıksal çıktı vermektedir. Ağırlıklı oylama yöntemi ile en yüksek değerlikli C_i sınıfı, *X* test girdisi için etiket olarak atanmaktadır. Başarılı bir sınıflandırma yapabilmek için veri setinin istatistiksel özelliklerine en uygun gelen 3 değeri, katman sayısı olarak alınmıştır. Yapılan testlerde hesaplama süresi ve sistem performansı açısından en uygun katman sayısının 3 olduğu gözlemlenmiştir.



Şekil 3. ÇKA katmanları ve *Bias*

Sistemin sınıflandırma başarısının artırılması amacıyla algılayıcılara gelen her girdi için ideal ağırlığın verilmesi gerekmektedir. Danışmalı (*supervised*) öğrenmede ÇKA'ya cevabı bilinen girdiler verilir. Bu

yolla ağ, tahmininin doğru olup olmadığını kontrol edebilir. Eğer ÇKA tarafından yapılan tahmin yanlış ise, ağırlıklar hataya göre yeniden düzenlenir. Ağırlıkları düzenleme süreci belirli bir iterasyon sayısınca (*epoch*) şu şekilde gerçekleştirilir:

1. Algılayıcıya eğitim setinden cevabı bilinen vektörel girdiler verilir.
2. Algılayıcıdan bir cevap tahmin etmesi istenir.
3. Ortalama Mutlak Hata (MAE, *Mean Absolute Error*) hesaplanır.
4. Hataya göre ağırlıklar azaltılır, artırılır veya sabit bırakılır.
5. Adım 2' ye dönülüp tekrar edilir.

4. DENEYSEL SONUÇLAR

Deneysel sonuçların elde edilebilmesi için Weka [11] uygulamasından ve ayrıca ÇKA'nın özgünleştirilebilmesi için C'de yazılan kodlardan yararlanılmıştır.

Bir sınıflandırıcının belirli bir veri seti üzerindeki performansının analiz edilebilmesi için değişik çapraz geçerleme (*Cross Validation*) yöntemleri vardır. 5×2 ve 10 Kat Çapraz Geçerleme yöntemleri yerine çalışmamızda Birini Dışarıda Bırak (*Leave One Out, LOO*) yöntemi tercih edilmiştir. Etiketli verinin bulunmasının çok zor olduğu uygulamalarda birini dışarıda bırakma işlemi literatürde tavsiye edilen bir yöntemdir [12]. Bu yöntemde sırasıyla eğitim setindeki her bir örnek dışarıda bırakılarak ÇKA eğitimi yapılır. Dışarıda bırakılan etiketli verinin sınıfı ÇKA sınıflandırıcısına sorularak doğru tahmin yapılıp yapılmadığı anlaşılmaya çalışılır. Yapılan doğru tahminlerin, toplam tahmin adedine oranı sınıflandırma performansını belirler.

Tablo 2. Karışıklık matrisi

		Tahmini değerler				
		Hukuk	Tıp	Mühendislik	Sanat	Medya
Gerçek değerler	Hukuk	12	0	0	1	0
	Tıp	1	10	1	0	0
	Mühendislik	0	1	10	0	0
	Sanat	1	0	1	10	1
	Medya	1	0	0	1	8

Ayrıca bu yöntem ile yapılan çapraz geçerleme işleminden Tablo 2'deki Karışıklık Matrisi (Confusion *Matrix*) tablosu elde edilmiştir. Satırlarda bulunan her bir sınıfa ait değerler gerçekte var olanlar; sütunlarda bulunanlar ise sistem tarafından tahmin edilen değerlerdir. Bu tabloda yatay olarak değerler toplandığında gerçekte var olan her bir sınıfın (*actual class*) kaçar adet örneği olduğu görülür. Dikey olarak bakıldığında ise kullanılan ÇKA sınıflandırıcısının her bir meslek grubu için öngördüğü kararın (*predicted class*) toplamını gösterir. Kısaca bu tabloda her bir satır ilgili meslek grubundaki bireylerin gerçek sayısını, her bir sütun ise ÇKA sisteminin her bir meslek grubu için yaptığı tahmin sonuçlarını göstermektedir. Örneğin Hukuk satırındaki değerler toplandığında veri setinde toplam 13 kişi olduğu görülür. Fakat ÇKA, 12 kişinin gerçek mesleğini bilmesine rağmen gerçekte hukukçu olan 1 kişiye sanatçı tahmininde bulunmuştur. Benzer bir şekilde Hukuk sütunu incelendiğinde ÇKA sistemi toplam 15 kişinin hukukçu olduğunu söylemiştir. Gerçekte tıp, sanat ve medya alanında çalışan birer kişinin de hukukçu olduğu iddia edilmiştir. Eğitim setinde bulunan 58 kişiden 13 tanesi hukuk, 12 tanesi tıp, 11 tanesi mühendis, 12 tanesi sanat bölümünde, 10 tanesi de medya sektöründe çalışan kişilerden oluşmaktadır.

Doğruluk (*accuracy*) oranı formül 4'te görüldüğü gibi doğru tahminlerin veri sayısına oranı ile bulunur. 500 çevrim ile eğitilen ÇKA sisteminin verdiği bu karmaşıklık tablosundan yararlanarak elde edilen başarı oranı 0.8474'dir.

$$Acc = \frac{Doğru\ Tahmin\ Edilenler}{Veri\ sayısı} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

Yukarıdaki formüllerde kullanılan *TP, TN, FP, FN* kısaltmaları sırasıyla *True Positive*, *True Negative*, *False Positive* ve *False Negative* anlamına gelmektedir. Sınıflandırıcının her bir sınıf etiketi üzerindeki başarısı için Netlik (*Presicion*), Hassasiyet (*Recall, Sensitivity*) ve F1 Skoru metrikleri kullanılmıştır. Bu kavramlar kısaca şu şekilde açıklanmaktadır:

Netlik: 5 numaralı formülde görüldüğü üzere netliği hesaplamak için doğru tahmin edilenler tüm tahminlerin sayısına bölünür. Her bir sınıf etiketi için ayrı ayrı hesap yapılmalıdır. Bunun amacı sınıflandırıcının her bir sınıftaki netliğini hesaplayabilmektir.

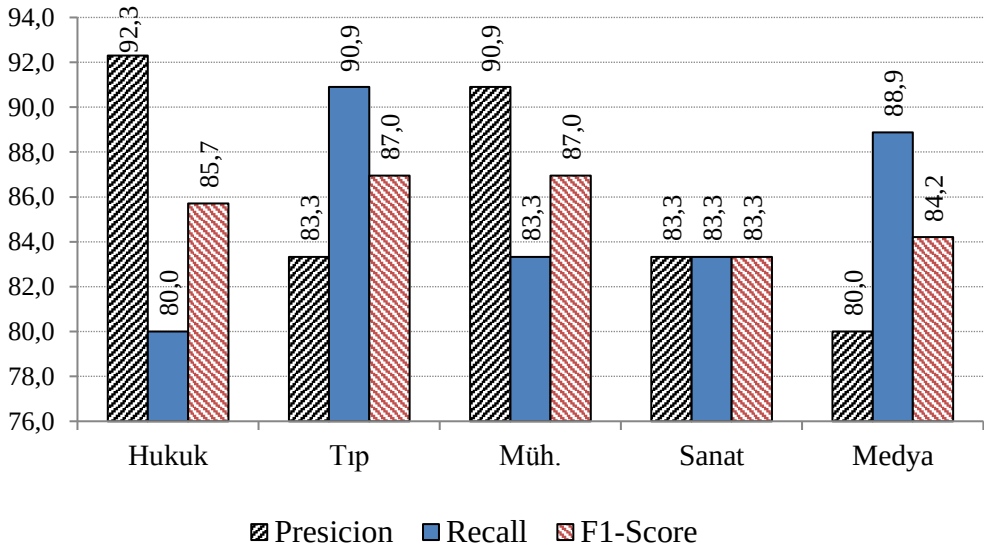
$$P = \frac{Doğru\ Tahmin\ Edilenler}{Tahmin\ Sayısı} = \frac{TP}{TP + FP} \tag{5}$$

Hassasiyet: Doğru tahmin edilenlerin gerçek değere oranıdır ve 6 numaralı formül ile bulunur.

$$R = \frac{Doğru\ Tahmin}{Gerçek\ Değer} = \frac{TP}{TP + FN} \tag{6}$$

*F*1 Skoru: Netlik ve hassasiyet değerlerinin harmonik ortalamasıdır ve 7 numaralı formül ile bulunur.

$$F1 = 2 \times \frac{P \times R}{P + R} \tag{7}$$



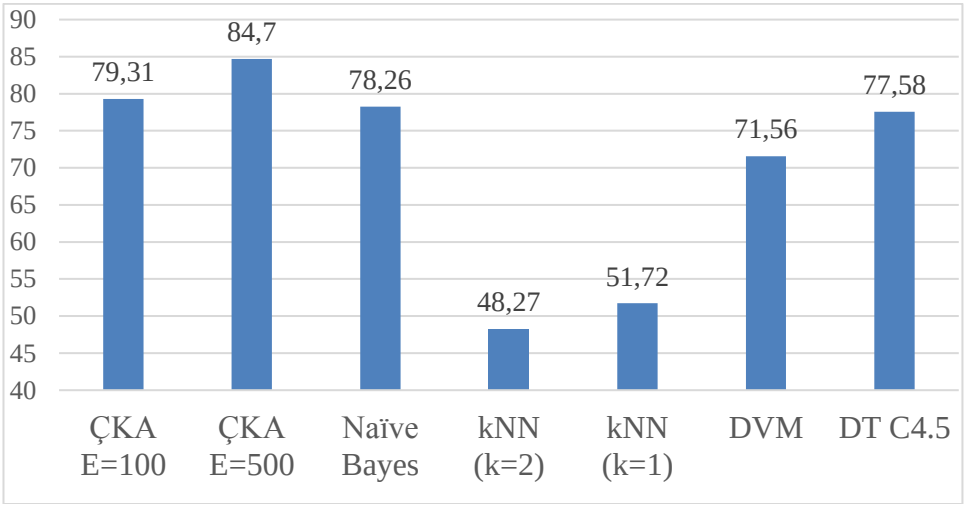
Şekil 4. Sınıf etiketlerinin P, R ve F1 değerleri

Her bir sınıf için hesaplanan P, R ve F1 değerleri Şekil 4'te verilmiştir. Görüldüğü üzere ÇKA sistemi bütün sınıf etiketleri üzerinde başarılı bir sınıflandırma işlemi yapmaktadır. Hukuk sınıfı üzerinde

ÇKA'nın yüksek bir netliğe sahip olasına karşın aynı sınıf etiketi üzerinde hassasiyeti zayıftır. Bu iki ölçütün harmonik ortalama değeri şekilde kırmızı ile gösterilmiştir ve diğer sınıf etiketlerinden elde edilen F1 değerlerine yakındır.

Şekil 5'te ÇKA yönteminin diğer sınıflandırma yöntemleriyle karşılaştırılması yapılmıştır. Bunlar Destek Vektör Makinesi (DVM), Karar Ağacı (Decision Tree), k En Yakın Komşuluk algoritması (k-NN) ve Naive Bayes sınıflandırıcısıdır. k En Yakın Komşuluk algoritması için k parametresi sırasıyla 1 ve 2 olarak alınmıştır. Bilindiği üzere DVM sınıflandırıcısı iki sınıflı eğitim setleri üzerinde kullanılabilir. Uygulamamızda çoklu sınıflar üzerinde çalışabilen DVM sınıflandırıcısı kullanılmıştır. DVM için C karmaşıklık parametresi 1, kernel fonksiyonu olarak da en sık kullanılan polinom ve radyal tabanlı Pearson VII (PUK) tercih edilmiştir. PUK için omega ve sigma (ω, σ) değerleri de 1 olarak alınmıştır.

Ayrıca yapılan çalışmalar sonucunda, çevrim sayısını artırdığımızda doğruluk oranının arttığı görülmüştür. Veri seti ile yapılan eğitimde çevrim sayısının belirli bir düzeye kadar artırılmasının sistemin doğruluğunu doğru orantılı bir şekilde etkilediği gözlemlenmiştir. 500'den sonra artırdığımız değerler için doğruluk oranında belirgin bir artış olmamıştır. Bu nedenle çevrim sayısı 500'e sabitlenmiştir. Çevrim sayısının düşük tutulmamasının ÇKA sisteminin kendini tam olarak eğitememesine sebep olduğu anlaşılmıştır.



Şekil 5. Sınıflandırıcıların doğruluk oranlarının karşılaştırması

Görüldüğü üzere ÇKA modeli diğer sınıflandırıcılara göre daha yüksek oranda başarı elde etmiştir. Diğer sınıflandırıcılara göre doğruluk oranının daha yüksek çıkmasına karşın ÇKA'ya ait eğitim süresinin daha uzun sürdüğü bir gerçektir. Ayrıca ÇKA'nın eğitilmesinde katman sayısının belirlenmesi, *bias* değerinin atanması, aktivasyon fonksiyonun seçilmesi gibi kullanıcı destekli hiper parametrelerin fazlalığı diğer bir tartışma konusudur. Fakat belirtmek isteriz ki bu parametreler sayesinde ÇKA yapısı özgünleştirilmiş ve diğer sınıflandırıcılara göre yüksek performanslı sonuçlar elde edilmiştir. Karşılaştırılan diğer sınıflandırıcı türleri birçok farklı parametrelerle eğitilmiş olmasına rağmen ÇKA'dan daha yüksek sonuçlar elde edilememiştir. Bu arada vurgulanması gereken önemli bir konu da her veri türü için yüksek oranda performans gösteren bir sınıflandırıcı türünün olmadığıdır [13].

Ancak veri setinin seyrek bir yapıda olması, sınıflandırma işleminde istenilen düzeyde performans alınamamasına neden olmaktadır. *Occam Razor* felsefesinde de belirtildiği gibi veri setindeki özellik (*feature*) sayısının fazla olması seyrek bir yapıyı neticelendirmekte ve aynı zamanda sınıflandırma başarısını düşürmektedir. Bu nedenle özellik sayısının, PCA (*Principal Component Analysis*) [14], CFS

(*Correlation-based Feature Subset Selection*) [15] ve Faktör Analizi gibi yöntemler kullanılarak azaltılması sonucunda sınıflandırma başarısının artırılabileceği düşünülmüştü. CFS ile yapılan boyut indirgeme işleminde veri seti 50 boyuttan 21 boyuta indirgenerek tekrar çapraz geçerleme işlemi yapılmıştır. 21 boyutlu veri seti ile yapılan sınıflandırma işleminde yaklaşık olarak %2 oranında daha fazla başarılı sonuçlar elde edilmiştir. Ayrıca PCA işlemi ile boyut indirgeme yapılmış ve sınıflandırma işleminde benzer sonuçlar elde edilmiştir. Daha az boyutlu bir veri seti ile aynı sonuçların elde edilmesi, zaman karmaşıklığının (*time complexity*) ve hesaplama zamanının (*execution time*) azalması nedeniyle tercih sebebidir.

Ayrıca veri setindeki her bir özniteliğin sınıflandırma işlemindeki kıymeti (*worth*) Weka'da One-R [16] sınıflandırıcısı ile LOO çapraz geçerleme yapılarak hesaplanmıştır. Bu işlem için OneRAttributeEval sınıfı kullanılmıştır. Aşağıdaki Tablo 3'te ankette bulunan 50 sorudan sadece en önemli olan 15 tanesi önem derecelerine göre sıralanarak gösterilmiştir. Çapraz geçerleme işlemi sonucunda ortalamaları elde edilen tablodaki "*average merit*" ve "*average rank*" ölçütleri sırasıyla her bir özniteliğin kıymetini (önemini) ve derecelendirilmesini sayısal olarak göstermektedir. Öznitelikteki yüksek değer, yüksek öneme işaret etmektedir. "±" işaretinden sonra gelen değerler ise standart sapma değerleridir. Bu tabloya göre sınıflandırma işleminde en belirleyici üç öznitelik "*sanat severlilik*", "*matematik veya edebiyata olan ilgi*" ile "*grupsal-bireysel çalışma isteği*" olduğu görülmektedir. Bu özellikler sırasıyla ilgili meslek türlerinin seçiminde en önemli ayırıcı faktörler olduğu görülmektedir. Fakat burada belirtmek isteriz ki ankette yer alan sorulara verilen cevaplara göre bireye en uygun meslek tavsiyesi sınıflandırıcı tarafından yapılmaktadır. Ayrıca sorulara verilen cevaplar da Tablo 3'teki sayısal değerlerin değişmesine neden olacaktır.

Tablo 3. Bilgi kazancı tablosu

Öznitelik	Öznitelik sorusu	Average Merit	Average Rank
Soru 38	Sanatsever misiniz?	39.583 ± 1.04	1.1 ± 0.35
Soru 45	Matematik mi Edebiyat mı?	36.879 ± 2.142	2.5 ± 1.51
Soru 46	Grup çalışması mı bireysel çalışma mı?	35.417 ± 1.018	3.4 ± 0.89
Soru 39	Bir işi bitirmeden başka bir işe başlar mısınız?	34.752 ± 2.229	4.3 ± 1.93
Soru 13	Çekingen misinizdir?	31.605 ± 2.814	6.9 ± 2.38
Soru 44	Hangi oyuncaklarla oynardınız?	30.629 ± 4.466	9.1 ± 9.54
Soru 40	Canlıları incelemeyi sever misiniz?	28.945 ± 2.668	9.6 ± 2.19
Soru 31	Bilim-teknik türü dergiler okur musunuz?	29.743 ± 4.397	11.6 ± 12.5
Soru 6	Farklı ve yeni fikirlere açık mısınız?	29.433 ± 4.804	12.7 ±12.59
Soru 42	Uzun süre masa başında çalışabilir misiniz?	26.02 ± 2.348	12.9 ± 3.34
Soru 11	Oynadığınız oyunlarda lider miydiniz?	27.704 ± 5.111	15.3 ±13.43
Soru 15	Kararsız mısınız?	24.557 ± 3.607	18.1 ±12.01
Soru 17	Kişisel gelişim kitapları okur musunuz?	25.177 ± 3.77	18.3 ±13.37
Soru 36	Olumsuz eleştirilere tahammülünüz?	25.975 ± 4.894	19.1 ±14.29

Bireyin kendine en uygun mesleği seçebilmesi için olası tüm faktörler çalışmanın başında ele alınmış ve 50 adet soru şeklinde ankete yerleştirilmişti. Fakat görüldüğü üzere kişiye uygun ve en iyi mesleği belirleyen faktörlerin sadece bir kısmı yukarıdaki tabloda önem sırasına göre verilmiştir.

5. İLERİ ÇALIŞMALAR

İleri aşamada yapılabilecek çalışmaları dört ana gruba ayırabiliriz. Bunlar; eğitim setinin zenginleştirilmesi, çalışmanın bir komite tarafından yürütülmesi, karar destek sistemimin performansı en yüksek sınıflandırma modeliyle gerçekleştirilmesi ve son olarak da mesleğinden memnun olmayan kişilerden elde edilen verilerle farklı bir modelin oluşturulması ile mevcut karar destek sistemine ilave edilmesi şeklinde olabilir.

Öncelikle, denetimli öğrenme için etiketli bir veri seti oluşturmak zor, zaman alan ve masraflı bir iştir. Eğitim setinde yeteri kadar bulunan örneklerle daha güvenilir bir sınıflandırma yapılacağı bir gerçektir. Üzerinde çalışılan bir veri setinde öznitelik ve sınıf etiketi sayısına bağlı olarak yeteri düzeyde örnek olması gerekmektedir. Fakat üzerinde çalışılan bu projede bir sınıflandırma uygulaması değil bir model ve prototip geliştirme çalışması yapılmıştır.

İleri uygulamalarda daha güvenilir bir model ortaya koyabilmek için bu konu üzerinde çalışan bilimsel bir komite oluşturulabilir. Uygulanacak anket ve sınıflandırma algoritmaları üzerinde çalışan bu komitede değişik meslek gruplarından temsilciler, eğitimciler, psikologlar, sosyologlar, PDR (Psikolojik Danışma ve Rehberlik) uzmanları, bilişim uzmanları ve konu ile alakalı kişiler de bulunabilir. Bu sayede bir lise öğrencisi için yükseköğrenime adım atmadan önce en sağlıklı tercihin yapılması sağlanabilir.

Ayrıca ileri çalışma olarak, yeni geliştirilen kolektif öğrenme (ensemble learning) ve başka sınıflandırma metotları da tahmin mekanizmasında kullanılarak yüksek başarı elde edilmesi hedeflenebilir.

Son olarak, mesleğinden memnun olmayan kişilerden elde edilen verilerle farklı bir sınıflandırma modeli oluşturulabilir. Bu veriler bir eğitim sistemine sayısallaştırılarak aktarılabilir ve bir sınıflandırıcı ile sistem eğitilebilir. Elde edilen bu sınıflandırıcı ile teste katılan bireyin hangi mesleği seçmemesi gerektiği bulunabilir. Çıktı olarak elde edilen meslek türünün, ilgili birey tarafından seçilmemesi önerilerek kümülatif olarak mevcut karar destek sistemi güçlendirilmiş olabilir.

Tüm bu önerilerden farklı olarak elde edilen verilerle oluşturulan eğitim seti psikolojik, sosyolojik ve istatistik disiplini altında incelenebilir, öznitelikler arası korelasyonlar çıkarılabilir, değişik istatistik testleri yardımıyla faklı çıkarımlar ve yorumlar yapılabilir. Bu çalışmadaki ana gayenin bu olmadığını bir kez daha belirtmek isteriz.

6. SONUÇLAR

İnsana ait psikolojik özelliklerin, yeteneklerin ve yaşamsal alışkanlıkların sayısallaştırılması herkesin kabul ettiği gibi zor bir iştir. Yeteneklerin, mental özelliklerin, alışkanlıkların, tutumların ve yönelimlerin tam sayı değerlere dönüştürülmesi neredeyse imkânsızdır. Bu çalışmada, bahsedilen zorlukları aşmaya çalışarak ileri düzey yapay bir karar destek sisteminin bir prototip modeli geliştirilmeye çalışıldı. Bu sistemde değişik sektörlerde çalışan ve mesleğinden memnun bireylerden elde edilen verilerle yeni ve özgün bir ÇKA sistemi ile sınıflandırma modeli önerilmiştir. Önerilen sistem ile diğer sınıflandırıcıların performansları kıyaslanmıştır. Projemizin amacı, kıyaslama yapılabilen bir veri seti üzerinde değişik sınıflandırıcıların performanslarını karşılaştırmak ve içlerinden en iyisini önermek değil, meslek seçiminde karar vermek üzere olan kişilere kendilerine en uygun mesleği seçmelerinde yardımcı olacak bir modeli öne sürmektir. Aynı zamanda amacımız, psikometri ve istatistiksel çalışmalardan farklı olarak ortaya bir sınıflandırma modeli ile karar destek sistemi sunmaktadır. Kişiye en uygun meslek önerisi yapan istatistiksel yöntemlerden farklı olarak bu çalışmada iki farklı yaklaşım vardır. Birincisi ortaya konulan yeni bir modeldir, diğeri ise yüksek sınıflandırma performansı için önerilen modele uygun özelleştirilebilen ÇKA mekanizmasındır.

TEŞEKKÜR

Anket çalışması yapabilmemiz için bize resmi izin veren, destekleyen ve hoş gören İstanbul İl Milli Eğitim Müdürlüğüne, İstanbul Büyükçekmece Anadolu Lisesi'ne, İstanbul İl Sağlık Müdürlüğüne, İstanbul Çapa Eğitim ve Araştırma Hastanesine, İstanbul Bilgisayar Mühendisleri Odasına, İstanbul Baro'suna, İstanbul TRT Radyosuna, ATV Ana Haber çalışanlarına teşekkürlerimizi bir borç bilir, saygılarımızı sunarız.

KAYNAKLAR

[1] Uğur Kariyer Merkezi, "Meslek Seçimi Araştırması, İstanbul" 2010. URL: http://www.ugurkariyermerkezi.net/, [Erişildi: 10/08/2015].

[2] TRT Çalışması, 1000 Kişiye Sorduk, "Türkiye'de ve Dünya'da Yanlış Meslek Seçimi", URL: http://www.trthaber.tv/p/1000_kisiye_sorduk/, [Erişildi: 10/08/2015].

[3] Berdan, İ. E., "Ortaöğretim Kurumlarındaki Öğretmen ve Öğrencilerin Yeteneklerinin Alan ve Meslek Seçimine Etkileri" Beykent Üniversitesi, Sosyal Bilimler Enstitüsü, İşletme Yönetimi Anabilim Dalı, cilt Yüksek Lisans Tezi, 2008.

[4] Yelken, K., Ortaöğretim son sınıf öğrencilerinin üniversite tercihlerini ve meslek seçimini etkileyen faktörler "Sakarya il merkezi örneği", Sakarya Üniversitesi / Sosyal Bilimler Enstitüsü / Sosyoloji Bölümü / Sosyoloji Anabilim Dalı / Sosyoloji Bilim Dalı, Yüksek Lisans Tezi, 2008.

[5] Erdem, A.Y., "Development of the irrational beliefs relating to career choice scale for high school students", Hacettepe Üniversitesi / Sosyal Bilimler Enstitüsü / Eğitim Bilimleri Anabilim Dalı / Psikolojik Danışmanlık ve Rehberlik Bilim Dalı, Yüksek Lisans Tezi, 2006.

[6] Kıyak, S., "Genel Lise Öğrencilerinin Meslek Seçimi Yaparken Temel Aldığı Kriterler" Yeditepe Üniversitesi Sosyal Bilimler Enstitüsü, cilt Eğitim Yönetimi Ve Denetimi Yüksek Lisans Programı, 2006.

[7] Malkoçoğlu, C., "Duyu Sal Mizaç ve Diğer Faktörlerin Türkiye'deki Bir Grup Lise Son Sınıf Öğrencisi Ergenin Meslek Seçimine Etkileri: Bir Betimleyici, Kesitsel Alan Çalışması" İstanbul Bilgi Üniversitesi, Sosyal Bilimler Enstitüsü Yüksek Lisans Tezi, 2009.

[8] Elmas, P. D., Yapay Zeka Uygulamaları, Seçkin Yayıncılık, İstanbul, 2010. pp:50-87.

[9] Nebiyev, V., Yapay Zeka Problemler Yöntemler Algoritmalar, İstanbul, Seçkin Yayıncılık, 2013. pp:100-110.

[10] Russell, S., Artificial Intelligence: A Modern Approach, 3. baskı, Prentice Hall, 2009, pp: 71-72.

[11] Hall, M.A., ve Smith, L.A., "Practical feature subset selection for machine learning" Springer Computer Science '98 Proceedings of the 21st Australasian Computer Science Conference ACSC'98, Berlin, 1998.

[12] Alpaydın, E., Introduction to Machine Learning, 2. baskı., Massachusetts: The MIT Press, 2010. pp: 487-490.

[13] Labatut, V. ve Cherifi, H., "Evaluatıon Of Performance Measuresfor Classifiers Comparison" Special Issue of ICIT Conference, 2011.

[14] Jolliffe, I.T., "Principal Component Analysis" Springer Series in Statistics, John Wiley & Sons, Ltd, 2002.

[15] Hall, M. A. ve Smith, L. A., "Practical feature subset selection for machine learning", Springer Computer Science '98 Proceedings of the 21st Australasian Computer Science Conference ACSC'98, Berlin, 1998.

[16] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H., The WEKA data mining software: an update. ACM SIGKDD explorations newsletter, 2009; 11(1): 10-18.

EK-1: Anket

Uygulanan anket soruları aşağıdaki gibidir. İlk 43 sorunun cevapları "Evet" ve "Hayır" şeklindedir.

1. İnsanların sorunlarına çözüm aramayı sever misiniz?
2. İnsanlara karşı empati yapabilir misiniz?
3. Sabırlı mısınız?
4. Size dertlerini anlatan insanları dinlerken onlara karşı dürüst olur musunuz?
5. Başkalarını dinlemeyi sever misiniz?
6. Çevrenizde fikirleri kabul edilen biri misiniz?
7. Anlatmayı sever misiniz?
8. Yapılan bir haksızlık karşısında müdahale eder misiniz?
9. Etkileyici konuşma yeteneğiniz var mı?
10. Olayları farklı bakış açılarıyla yorumlar mısınız?
11. Küçüklüğünüzde oynadığınız oyunlarda lider olur muydunuz?
12. Sorumluluk alır mısınız?
13. Çekingen misinizdir?
14. Disiplinli misiniz?
15. Kararsız mısınız?
16. Hafızanız kuvvetli midir?
17. "Nasıl başarılı olunur" gibi kitaplar okur musunuz?
18. Hayatınızı plan program dâhilinde yaşar mısınız?
19. Ekonomik gelişmeleri takip eder misiniz?
20. Gazete veya haberleri takip eder misiniz?
21. Hayatınızda hiç satış yaptınız mı? (limonata, kurabiye, oyuncaklarınız vs.)
22. Etrafınızda olup biten olayların aslını öğrenmek için garip bir merak duyar mısınız?
23. Makine ve malzemeler ile çalışmaktan hoşlanır mısınız?
24. Etrafınızda gördüğünüz nesnelerin nasıl çalıştığını merak eder misiniz?
25. Elektronik aletlere merak duyar mısınız?
26. Yeni bir nesne gördüğünüzde bu nesneyi detaylı inceler misiniz?
27. İnsanları yönetmeyi sever misiniz?
28. Kitap okumayı sever misiniz?
29. Tarih kitapları okumaktan hoşlanır mısınız?
30. İnsan ve evrenin niçin yaratıldığını düşünür müsünüz?
31. Bilim-teknik türü dergiler okur musunuz?
32. Bulmaca çözmeyi sever misiniz?
33. Harita üzerinde farklı ülkelerin yerlerini bulmayı sever misiniz?
34. Gördüğünüz ilginçliklerin fotoğraflarını çeker misiniz?
35. Rekabetten hoşlanır mısınız?
36. Aldığınız olumsuz eleştiriler sizi kızdırır mı?
37. Bilmediğiniz bir kelime için sözlüğe bakar mısınız?
38. Sanat sergilerini gezmeyi sever misiniz?
39. Bir işi bitirmeden başka bir işe başlar mısınız?
40. Canlıların fizyolojik işleyişini incelemeyi sever misiniz?
41. Hayvanları sever misiniz?
42. Uzun süre masa başında çalışmaya sabredebilir misiniz?
43. Doğayla iç içe olmaktan hoşlanır mısınız?
44. Küçüklüğünüzde aşağıdaki oyuncaklardan hangisiyle daha fazla zaman geçirdiniz?
    a) Top
    b) Pelüş hayvan
    c) Uzaktan kumandalı araba
    d) Müzik aletleri
    e) Kukla
45. Matematik soruları mı, edebiyat metinleri mi?
    a) Matematik soruları
    b) Edebiyat metinleri

46. Grup çalışması mı, bireysel çalışmalar mı?
    a) Grup çalışması
    b) Bireysel çalışma
47. Bilginin mi deneyimin mi gücüne inanırsınız?
    a) Bilgi
    b) Deneyim
48. Mantığınız mı sizin için ön plandadır, duygularınız mı?
    a) Mantık
    b) Duygular
49. Size ait hangi özelliğinizle öne çıkmak istersiniz?
    a) Zekâmla
    b) Paramla
    c) Sanatımla
50. Ne sıklıkla hayal kurarsınız?
    a) Genellikle kurarım
    b) Arada bir kurarım
    c) Hiç kurmam