

VERİ MADENCİLİĞİ YÖNTEMLERİYLE AKILLI TELEFON FİYATLARININ ANALİZİ

Veri Madenciliği Projesi Raporu

İshak KUTLU
Bilişim Enstitüsü / Bilgisayar Bilimleri

Özet

Birer cep bilgisayarı olarak nitelendirilebilen akıllı telefonlar günümüzde artık zorunlu ihtiyaçlar olarak görülmektedir. Karmaşık özelliklere sahip akıllı telefonların, hem tüketiciler hem de üreticiler açısından, piyasa koşullarında fiyat aralıklarını öngörmek de oldukça zordur. Bu çalışmada akıllı telefon fiyatları üzerinde etkili olan teknik özellikler veri madenciliği yöntemleriyle analiz edilmiştir. Kaggle platformundan ulaşılan ilgili veri seti temizleme, gürültü giderme gibi veri ön işleme teknikleriyle modellenmeye hazır hale getirilmiştir. Akıllı telefonların özelliklerine göre fiyat sınıflarının orta, yüksek ve çok yüksek şeklinde tahmin edilmesi amacıyla, bir makine öğrenmesi algoritması olan KNN modelinden yararlanılmıştır. İlk aşamada veri setinde herhangi bir boyut indirgeme ve özellik çıkarımı işlemi yapılmadan, mevcut boyutlar/özellikler (dimesion/feature) kullanılarak fiyat sınıfları KNN ile tahmin edilmiştir. İkinci aşamada boyut indirgeme/özellik seçimi ve özellik çıkarımı gibi ilave veri ön işleme teknikleriyle boyutlar yeniden yapılandırılmış ve yapılandırılan yeni özellikler kullanılarak fiyat sınıfları KNN ile tekrar tahmin edilmiştir. Son aşamada ise oluşturulan iki modelin performansları karşılaştırılarak, veri ön işleme tekniklerinin, veriden faydalı bilginin (knowledge) çıkartılması (diğer bir ifadeyle modelleme başarımı) üzerindeki etkileri değerlendirilmiştir.

Python ile gerçekleştirilen projeye aşağıdaki bağlantıdan ulaşabilirsiniz.

<https://www.kaggle.com/code/shakkutlu/project-of-pad-1-knn-mobile-price-classification>

1. Giriş

Akıllı telefonlar günümüzde kişisel ya da profesyonel kullanımda önemli bir yer tutmaktadır. Bu makineler, iletişimin kapsamını karşılıklı konuşmanın çok ötesine taşımış olup boyutları ve karmaşık donanımlarıyla birer cep bilgisayarı olarak görülebilir [1]. Karmaşık yazılımsal ve donanımsal özelliklere sahip bu ürünler, hem tüketiciler hem de üreticiler tarafından fiyat belirsizliğine yol açmaktadır. Bir telefonun ait olduğu fiyat sınıfını belirlemek, ona ödeme yapmak isteyen ve faydalarını maksimize etmeyi amaçlayan tüketiciler tarafından önemli bir konudur. Diğer taraftan karlarını maksimize etmeyi amaçlayan üreticiler ise pazar koşullarını gözetmek koşuluyla, ürünlerinin mümkün olan en yüksek karlılık seviyesinde satışa sunmak isterler. Bu çerçevede tüketiciler ve üreticiler açısından herhangi bir akıllı telefonun ait olduğu fiyat sınıfını/düzeyini belirlemek önemli bir problem olarak ortaya çıkmaktadır [2, 3].

Bu çalışmada akıllı telefonların teknik özelliklerine dayalı olarak fiyat sınıflarının düşük, orta, yüksek, çok yüksek şeklinde tahmin edilmesi amaçlanmaktadır. Veri seti, açık kaynak bir veri platformu olan Kaggle'dan elde edilmiştir. Söz konusu veri seti, Python programlama dili kullanılarak veri temizleme, dönüştürme, gürültü tespiti, özellik ve örnek seçimi, boyut indirgeme, özellik çıkarımı gibi veri ön işleme teknikleriyle modelleme aşamasında kullanılmaya hazır hale getirilecektir.

Akıllı telefonların fiyatlarını etkileyebilen ram, işlemci hızı, batarya gücü, ekran boyutu, hafıza, çift sim kart yuvası gibi çok sayıda donanımsal ya da yazılımsal teknik özellik olabilir [4]. Söz konusu özellikler veri madenciliği teknikleriyle analiz edilerek bunların fiyat sınıfları üzerindeki etkileri değerlendirilecektir. Performans ölçüm kriterleri olarak “accuracy”, “precision”, “recall” ve “f1 score” metrikleri kullanılacaktır.

Tahmin modeli olarak, makine öğrenmesi (machine learning, ML) yöntemlerinden biri olan KNN algoritması kullanılacak olup boyut indirgeme (reduce dimesion) ve özellik çıkarımı (feature extraction) yöntemlerinin fiyat sınıflarını tahmin etme üzerindeki etkileri değerlendirilecektir. İlk aşamada boyut indirgeme ve özellik çıkarımı “öncesi”, veri ön işleme sürecinden geçmiş veriler KNN algoritmasıyla modellenecektir. İkinci aşamada boyut indirgeme ve özellik çıkarımı “sonrası” yeniden yapılandırılan veri seti KNN algoritmasıyla modellenecektir. Üçüncü aşamada ise her iki modelin tahmin/sınıflandırma performansları, boyut indirgeme ve özellik çıkarımı süreçleri göz önüne alınarak analiz edilecektir.

2. LİTERATÜR TARAMASI

Anshori, M.Y. ve diğerleri geri yayımlı (backpropagation) yapay sinir ağları ve öğrenen vektör niceliği (Learning Vector Quantization, LVQ) modelleriyle ağırlık, dahili hafıza, ram, ön ve arka kamera ve marka girdilerinin akıllı telefonların fiyatları üzerinde etkilerini araştırmıştır. Bağımlı/hedef değişkeni temsil eden akıllı telefon fiyatları, belirli fiyat aralıkları baz alınarak 5 farklı (etiketli) sınıfta analiz edilmiştir. 1 no.lu sınıf en düşük fiyata sahip sınıfı, 5 no.lu sınıf ise en yüksek fiyata sahip sınıfı ifade etmektedir. Performans ölçümünde ise “accuracy”, “error rate”, “recall”, “specificity”, ve “precision” metrikleri kullanılmıştır. Verilerin %80’i eğitim, %20’si ise test için ayrılmıştır. İki gizli katmanın ve aşamalı azalma (gradient descent) yönteminin kullanıldığı geri yayımlı yapay sinir ağı modelinin maksimum adım (epoch) sayısı 1000 ve öğrenme katsayısı (learning rate) ise 0,2 olarak ayarlanmıştır. Öklit mesafesinin kullanıldığı LVQ modelinde ise maksimum adım sayısı 1000, alfa ve sigma öğrenme katsayıları ise sırasıyla, 0,6 ve 0,5 olarak ayarlanmıştır. Geri yayılım algoritmasının eğitim ve test veri seti için accuracy değerleri yüzde 88,75 ve 87,50 iken LVQ algoritmasının eğitim ve test veri seti için accuracy değerleri yüzde 81,25 ve 72,5 olarak gerçekleşmiştir. Söz konusu veri setine ilişkin sınıflandırma probleminin çözümünde geri yayılım algoritmasının LVQ algoritmasından daha başarılı olduğu sonucuna varılmıştır [4].

Bhattacharjee, D. ve diğerleri markalaşmış akıllı telefonların 13 farklı özelliğinin, genç Hintli tüketicilerin bireysel tercihleri üzerinde etkilerini analiz etmiş ve tüketicilerin markalaşmış akıllı telefonları satın alma eğilimlerini modellemiştir. Veriler, Google forms olarak bilinen Google’ın online anket aracı yoluyla elde edilmiştir. Ankete katılanlar 16-35 yaş aralığında olup yaş gruplarına göre dört sınıfa ayrılmıştır. Benzer şekilde veriler cinsiyete ve ailenin gelir seviyesine göre sınıflandırılmıştır. Veriler Microsoft Excel’de analiz edilerek hazırlanmıştır. Makine öğrenmesi modelleri ise Python ile oluşturulmuştur. Özellik seçiminde belirsiz hesaplama (uncertainty calculation), korelasyon (correlation) ve deneysel karar verme ve değerlendirme laboratuvarı (Decision-Making Trial and Evaluation Laboratory, DEMATEL) teknikleri kullanılmıştır. Seçilen özellikler bir makine öğrenmesi (machine learning, ML) tekniği olan K en yakın komşu (K nearest neighbour, KNN) ile modellenmiştir. Sonuçlar, diğer ML teknikleri olan C4.5, CART (Classification and Regression Trees) ve lojistik regresyon ile karşılaştırılmıştır. Performans ölçüm kriterleri olarak “precision”, “recall”, “f1 score” ve “support” metrikleri kullanılmıştır. Araştırmanın sonuçlarına göre kamera sayısının, markalaşmış telefonları satın alma kararları üzerinde bir etkiye sahip olabileceği ifade edilmiştir. Buna karşılık fiyatın ise markalaşmış telefonları satın alma kararları üzerinde herhangi bir rolü olmadığı sonucuna varılmıştır. Akıllı telefonlar gençler arasında ağırlıklı olarak sosyal medya paylaşımlarında kullanıldığı için kamera çözünürlüğü (mp) ve hafıza/depolama kapasitesi ile markalaşmış telefonları satın alma eğilimi arasında yüksek bir korelasyon tespit edilmiştir. Genç müşteriler, yüksek fiyatlı markalaşmış telefonlarda dayanıklılık aradığı için kullanılan metaryelin kalitesinin önemli olduğu sonucuna varılmıştır. KNN modeli, genç müşterilerin markalaşmış telefonları satın alma davranışlarını ve benzer telefon markaları arasında marka değiştirme eğilimlerini başarılı bir şekilde açıklayabilmiştir [2].

Aggarwal, A. ve diğerleri tüketicilerin belirli bir bütçe aralığında akıllı telefon seçimine yardımcı olacak çok kriterli bir karar verme süreci (multi criteria decision making, MCDM) modellemiştir. Rekabetin yüksek olduğu akıllı telefon piyasası, farklı fiyat seviyelerinde (sınıflarında) geniş bir özellik çeşitliliğine sahiptir. Bu amaçla çok kriterli karar vermeye ilgili problemlerin çözümü için EDAS (The Evaluation Based on Distance from

Average Solution) tekniği önerilmektedir. EDAS, ortalama çözümden hesaplanan iki yönlü bir çözüm kriteri sunar. Söz gelimi belirli bir fiyat seviyesi için bir telefonun çözünürlüğü ortalamanın üstünde ise arzulanan, ağırlığı ortalamanın üstünde ise arzulanan bir durumu işaret eden iki yönlü bir çözüm kriterini ifade eder. Araştırmada arzulanan (pozitif) kriterler olarak kamera kalitesi, dahili depolama, batarya ömrü, ram, işlemci ve ekran performansı; arzulanan (negatif) kriterler olarak ağırlık ve fiyat kabul edilmiştir. Diğer taraftan belirli özelliklere sahip bir telefonun farklı markalar tarafından pazarlanan ürünleri ise tüketicilerin farklı alternatifler arasında geçiş yapabilmesini ifade eder. Dolayısıyla pozitif ve negatif karar tanımlama matrisleri, telefonun özelliklerini ifade eder. Bu çerçevede karar tanımlama matrislerinden hareketle, araştırmada kullanılan yedi telefon markasının, Hindistan tüketicileri tarafından tercih edilirliliği sıralanmıştır [1].

Singh, R. ve diğerleri insanların ihtiyaçlarına ve bütçelerine göre akıllı telefon seçiminde etkinliğin sağlanması ve ürünlerin özelliklerine göre fiyatlarının kategorize edilmesi üzerinde araştırma yapmıştır. Mevcut çalışma ile “aynı veri setinin” kullanıldığı söz konusu araştırma, keşifsel veri analizi ve model oluşturma şeklinde iki aşamalı olarak gerçekleştirilmiştir. Keşifsel veri analizinde veri ön hazırlık, tanımlayıcı istatistik ve çıkarımsal istatistik teknikleri; model oluşturmada ise özellik seçimi, KNN ve rassal orman modelleri ve karışıklık matrisi (confusion matrix) ile performans ölçümü tekniklerinden yararlanılmıştır. Veri ön hazırlık sürecinde veri seti Kaggle’den tedarik edilmiş ve null değerlere sahip satırlar silinmiş, gereksiz satır ve kolonlar kaldırılmıştır. Tanımlayıcı istatistik çerçevesinde (hedef değişkeni ya da etiketleri temsil eden) telefonların fiyatları düşük, orta, yüksek ve çok yüksek şeklinde dört sınıfa ayrılmıştır. Çıkarımsal istatistik aşamasında genelleme yapabilmek için örüntü belirleme ve korelasyon teknikleri kullanılmış; bu amaçla histogram, kutu, nokta, çizgi grafiği gibi çeşitli veri görselleştirme yöntemlerine başvurulmuştur. Böylece telefon fiyatları üzerinde en etkili ürün özellikleri tespit edilmeye çalışılmıştır. Özellik seçimi sürecinde “Select k Best” ve “Feature Importance” yöntemleri kullanılmıştır. Bu çerçevede kurulacak modelde yer alması gereken özelliklerin ram, batarya gücü, genişlik ve yükseklik çözünürlükleri, ağırlık ve hafıza olduğu belirlenmiştir. Python programlama dilinin kullanıldığı araştırmada, fiyat sınıflarının tahmin edilmesi için tembel öğrenen sınıflandırıcı (lazy learner classifier) olarak KNN ve istekli öğrenen sınıflandırıcı (eager learner classifier) olarak rassal orman (random forest) algoritması kullanılmıştır. “Accuracy” değeri KNN için 0,93, rassal orman için 0,81 olarak elde edilmiştir [3].

Mehrotra, D. ve diğerleri akıllı telefonlara yüklenen uygulamaların enerji tüketimlerinin tahmin edilmesi üzerine bir çalışma yapmıştır. Bu çerçevede 90 farklı Android uygulamasının 414 test senaryosunda enerji tüketimleri incelenmiş, telefonların ekran tipi, işlemcisi, wifi vb. teknik özelliklerinin enerji tüketimleri üzerindeki etkisi analiz edilerek veri seti deneysel yollarla oluşturulmuştur. Oluşturulan veri setleri etiketsiz olduğu için denetimsiz makine öğrenmesi, yani kümeleme algoritmalarıyla uygulamaların enerji tüketimleri düşük, orta ve yüksek şeklinde sınıflandırılmıştır. Daha sonra ise etiketli veri setleri kullanılarak, her bir uygulama bazında, telefonların teknik özelliklerinin enerji tüketimi üzerindeki etkileri denetimli makine öğrenmesi algoritmalarıyla modellenmiştir. R programlama dilinin kullanıldığı araştırmada, Bayesian ve KNN ile karar ağacı algoritmalarından olan rassal orman ve J48, kural ilanı (rule induction) ile problem modellenmiştir. Modellerin performansları karmaşıklık matrisinden elde edilen “accuracy” ve “correctness” metrikleri ile değerlendirilmiştir. Araştırmanın sonuçlarına göre beş model arasından rassal orman modeli, 0,97 ile en iyi “accuracy” değerine sahip model olarak sunulmuştur [5].

3. METODOLOJİ

İki değişken arasındaki ilişkinin yönünü ve gücünü ölçen korelasyon katsayısı aşağıda gösterildiği gibi hesaplanır [2].

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (3.1)$$

Yukarıdaki denklemde r (Pearson) korelasyon katsayısını, x_i ve y_i değişkenlerin değerlerini, \bar{x} ve \bar{y} ise ilgili değişkenlerin ortalamasını ifade eder.

KNN, bir veri noktasını kendisine en yakın (k değeri ile temsil edilen) veri kümesine atama yapmak suretiyle sınıflandırma yapar. En yaygın kullanılan uzaklık ölçüleri “minkowski”, “euclidean”, “manhattan” olarak sayılabilir. Aşağıdaki denklem, veri noktaları arasındaki uzaklığı manhattan yöntemine göre ölçer.

$$d = \sum |x_i - y_i| \quad (3.2)$$

Yukarıdaki denklemde d veri noktaları arasındaki uzaklığı, x_i ve y_i değişkenlerin değerlerini ifade eder. KNN seçilen bir k değerine göre uzaklık ölçüsünü kullanarak bir veri noktasını, k elemandan oluşan kendisine en yakın kümeye atar [2].

Modelin performansını kıyaslamak amacıyla “accuracy” metriğine başvurulmuştur. Accuracy metriği, Tablo 3.11’de gösterilen karışıklık matrisi (confusion matrix) olarak bilinen bir tablodan türetilir [5].

		TAHMİN EDİLEN		TOPLAM
		C^+	C^-	
GERÇEK	C^+	TP True Pozitif (Hits)	FN False Negatif (Miss)	N^+ Gerçek Pozitif sayısı
	C^-	FP False Pozitif (Miss)	TN True Negatif (Hits)	N^- Gerçek Negatif sayısı
TOPLAM		\tilde{N}^+ Tahmin Pozitif sayısı	\tilde{N}^- Tahmin Negatif sayısı	N Toplam Örnek sayısı

Tablo 3.11. Karışıklık Matrisi

Tablo 3.11’de yer alan true positive (TP), gerçek değeri pozitif olup da pozitif değere sınıflandırılanların sayısını; false positive (FP), gerçek değeri negatif olup da pozitif değere sınıflandırılanların sayısını ifade eder. Gerçekte kovid bir insana kovid tanısının konması TP’ye, gerçekte kovid olmayan bir insana kovid tanısının konması ise FP’ye örnek gösterilebilir. Benzer şekilde true negative (TN), gerçek değeri negatif olup da negatif değere sınıflandırılanların sayısını; false negative (FN), gerçek değeri pozitif olup da negatif değere sınıflandırılanların sayısını temsil eder. Gerçekte kovid olmayan bir insana kovid değil tanısının konması TN’ye, gerçekte kovid bir insana kovid değil tanısının konması ise FN’ye örnek gösterilebilir. N^+ ve N^- ifadeleri ise sırasıyla, pozitif ve negatif gerçek değerlerin sayısını temsil eder. Accuracy, percision, recall ve f1-score metrikleri aşağıda gösterildiği şekilde hesaplanır [4].

$$\text{Accuracy} = \frac{TP+TN}{N^++N^-} \quad (3.3)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3.4)$$

$$\text{Recall} = \frac{TP}{N^+} \quad (3.5)$$

$$\text{f1-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.6)$$

Karışıklık matrisinden türetilen yukarıdaki metriklerin tümü test veri seti esas alınarak hesaplanır.

4. GERÇEKLEŞTİRİLEN ÇALIŞMA

Bu araştırma, kaggle platformu üzerinde, jupyter IDE'si kullanılarak python programlama dilinde gerçekleştirilmiştir. Veri ön işleme ve modelleme süreçlerinde pandas, numpy, sklearn; veri görselleştirme süreçlerinde ise matplotlib ve seaborn kütüphanelerinden yararlanılmıştır.

Çalışmanın yapıldığı bilgisayarın teknik özellikleri aşağıda listelenmiştir.

- CPU: AMD Ryzen 5 3500X 6-Core Processor, 3600 Mhz, 6 Core(s), 6 Logical Processor(s)
- GPU: NVIDIA GeForce RTX 3060, VRAM 12 GB
- RAM: 16 GB
- İşletim Sistemi: Windows 10

Python ile yapılan çalışmaya aşağıdaki bağlantıdan ulaşılabilir.

<https://www.kaggle.com/code/shakkutlu/project-of-pad-1-knn-mobile-price-classification>

4.1. Veri Ön İşleme

Kaggle platformundan elde edilen, akıllı telefonların teknik özelliklerinden ve fiyat sınıflarından oluşan veri setinde, Tablo 4.1'de gösterildiği üzere 2000 satır, 20 özellik (feature) ve dört kategorik değere sahip bir de sınıf (class) bulunmaktadır. Sınıf değerleri (price_range) 0, 1, 2 ve 3 kategorik değişkenlere sahip olup düşük, orta, yüksek ve çok yüksek fiyat düzeylerini temsil etmektedir.

	battery_power	blue	clock_speed	dual_sim	fc	four_g	int_memory	m_dep	mobile_wt	n_cores	...	px_height	px_width	ram	sc_h	sc_w	talk_time	three_g	touch_screen	wifi	price_range
0	842	0	2.2	0	1	0	7	0.6	188	2	...	20	756	2549	9	7	19	0	0	1	1
1	1021	1	0.5	1	0	1	53	0.7	136	3	...	905	1988	2631	17	3	7	1	1	0	2
2	563	1	0.5	1	2	1	41	0.9	145	5	...	1263	1716	2603	11	2	9	1	1	0	2
3	615	1	2.5	0	0	0	10	0.8	131	6	...	1216	1786	2789	16	8	11	1	0	0	2
4	1821	1	1.2	0	13	1	44	0.6	141	2	...	1208	1212	1411	8	2	15	1	1	0	1
...
1995	794	1	0.5	1	0	1	2	0.8	106	6	...	1222	1890	668	13	4	19	1	1	0	0
1996	1965	1	2.6	1	0	0	39	0.2	187	4	...	915	1965	2032	11	10	16	1	1	1	2
1997	1911	0	0.9	1	1	1	36	0.7	108	8	...	868	1632	3057	9	1	5	1	1	0	3
1998	1512	0	0.9	0	4	1	46	0.1	145	5	...	336	670	869	18	10	19	1	1	1	0
1999	510	1	2.0	1	5	1	45	0.9	168	6	...	483	754	3919	19	4	2	1	1	1	3

2000 rows x 21 columns

Tablo 4.1. Veri Seti

Aşağıda yer alan Tablo 4.2’de ise akıllı telefonların teknik özelliklerini ve fiyat sınıfını temsil eden boyutlara ilişkin açıklamalar bulunmaktadır.

features	Information about features
1 battery_power	Total energy a battery can store in one time measured in (mAh)
2 blue	Has bluetooth or not
3 clock_speed	Speed at which microprocessor executes instructions
4 dual_sim	Has dual sim support or not
5 fc	Front camera (Megapixels)
6 four_g	Has 4G or not
7 int_memory	Internal memory in (Gigabytes)
8 m_dep	Mobile depth in (Cm)
9 mobile_wt	Weight of mobile phone
10 pc	Primary camera (Megapixels)
11 px_height	Pixel resolution height
12 px_width	Pixel resolution width
13 ram	Random access memory in (Megabytes)
14 sc_h	Screen height of mobile in (Cm)
15 sc_w	Screen width of mobile in (Cm)
16 talk_time	Longest time that a single battery charge will last when you are constantly talking on the phone
17 three_g	Has 3G or not
18 touch_screen	Has touch screen or not
19 wifi	Has wifi or not
20 n_cores	Number of cores of processor
21 price_range	This is the Target variable with value of 0: (Low Cost), 1: (Medium Cost), 2: (High Cost), and 3: (Very High Cost)

Tablo 4.2. Özellik ve Sınıf Açıklamaları

Veri setinin null değerlere sahip örnekler içerip içermediği ve veri tipleri incelenmiş ve Tablo 4.3'te gösterildiği gibi herhangi bir null değere sahip satıra ya da uygun olmayan veri tipine rastlanmamıştır.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2000 entries, 0 to 1999
Data columns (total 21 columns):
#   Column                Non-Null Count  Dtype
---  -
0   battery_power         2000 non-null   int64
1   blue                  2000 non-null   int64
2   clock_speed           2000 non-null   float64
3   dual_sim              2000 non-null   int64
4   fc                    2000 non-null   int64
5   four_g                2000 non-null   int64
6   int_memory            2000 non-null   int64
7   m_dep                 2000 non-null   float64
8   mobile_wt             2000 non-null   int64
9   n_cores               2000 non-null   int64
10  pc                    2000 non-null   int64
11  px_height              2000 non-null   int64
12  px_width               2000 non-null   int64
13  ram                   2000 non-null   int64
14  sc_h                  2000 non-null   int64
15  sc_w                  2000 non-null   int64
16  talk_time              2000 non-null   int64
17  three_g               2000 non-null   int64
18  touch_screen          2000 non-null   int64
19  wifi                  2000 non-null   int64
20  price_range           2000 non-null   int64
dtypes: float64(2), int64(19)
memory usage: 328.2 KB
```

Tablo 4.3. Null Değerlerin Araştırılması ve Veri Tipleri

Veri setinin gürültülerden arındırılması amacıyla, Tablo 4.4'te gösterildiği gibi özelliklerin alabileceği değer aralıkları araştırılmıştır.

	count	mean	std	min	25%	50%	75%	max
battery_power	2000.0	1238.51850	439.418206	501.0	851.75	1226.0	1615.25	1998.0
blue	2000.0	0.49500	0.500100	0.0	0.00	0.0	1.00	1.0
clock_speed	2000.0	1.52225	0.816004	0.5	0.70	1.5	2.20	3.0
dual_sim	2000.0	0.50950	0.500035	0.0	0.00	1.0	1.00	1.0
fc	2000.0	4.30950	4.341444	0.0	1.00	3.0	7.00	19.0
four_g	2000.0	0.52150	0.499662	0.0	0.00	1.0	1.00	1.0
int_memory	2000.0	32.04650	18.145715	2.0	16.00	32.0	48.00	64.0
m_dep	2000.0	0.50175	0.288416	0.1	0.20	0.5	0.80	1.0
mobile_wt	2000.0	140.24900	35.399655	80.0	109.00	141.0	170.00	200.0
n_cores	2000.0	4.52050	2.287837	1.0	3.00	4.0	7.00	8.0
pc	2000.0	9.91650	6.064315	0.0	5.00	10.0	15.00	20.0
px_height	2000.0	645.10800	443.780811	0.0	282.75	564.0	947.25	1960.0
px_width	2000.0	1251.51550	432.199447	500.0	874.75	1247.0	1633.00	1998.0
ram	2000.0	2124.21300	1084.732044	256.0	1207.50	2146.5	3064.50	3998.0
sc_h	2000.0	12.30650	4.213245	5.0	9.00	12.0	16.00	19.0
sc_w	2000.0	5.76700	4.356398	0.0	2.00	5.0	9.00	18.0
talk_time	2000.0	11.01100	5.463955	2.0	6.00	11.0	16.00	20.0
three_g	2000.0	0.76150	0.426273	0.0	1.00	1.0	1.00	1.0
touch_screen	2000.0	0.50300	0.500116	0.0	0.00	1.0	1.00	1.0
wifi	2000.0	0.50700	0.500076	0.0	0.00	1.0	1.00	1.0
price_range	2000.0	1.50000	1.118314	0.0	0.75	1.5	2.25	3.0

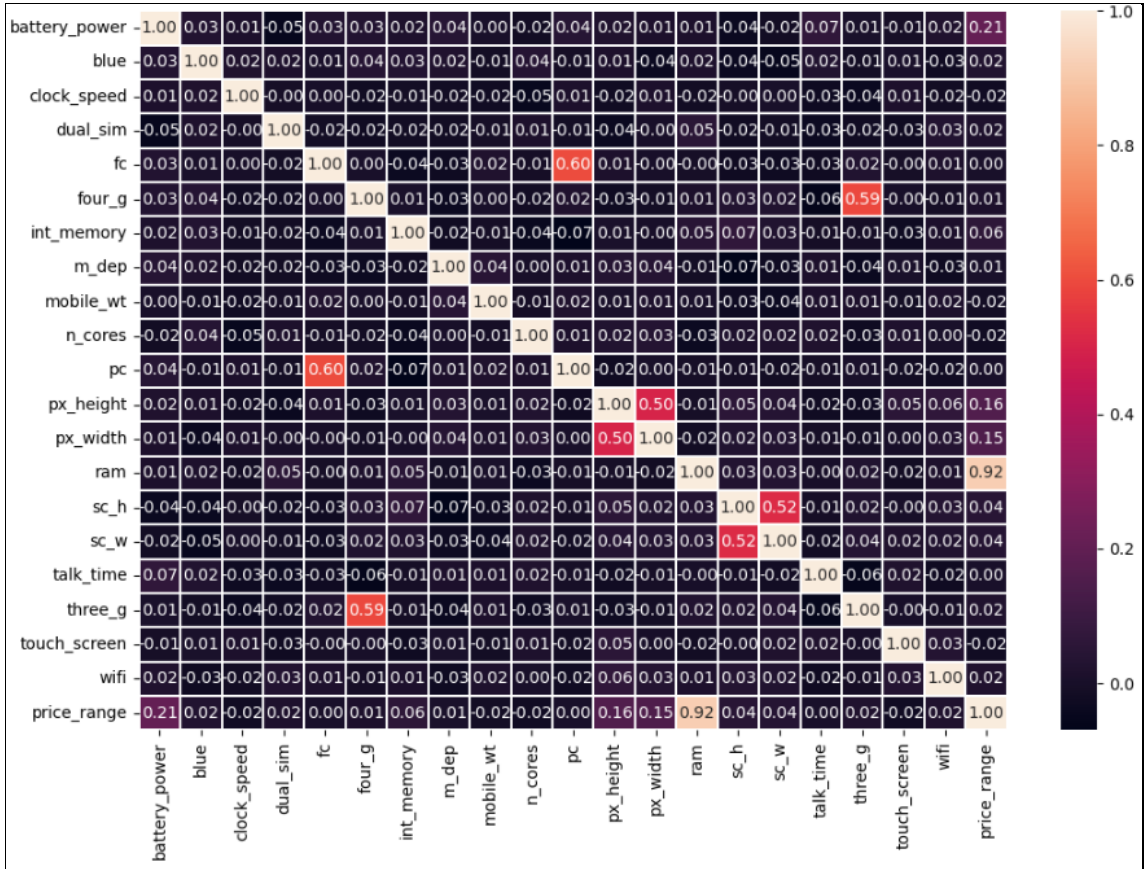
Tablo 4.4. Özelliklerin Alabileceği Değer Aralıklarının Araştırılması

Tablo 4.4'ten hareketle, piksel yüksekliğini ifade eden "px_height" ve ekran genişliğini temsil eden "sc_w" özelliklerinde tespit edilen 0 değerleri silinmiştir. Benzer şekilde ön (front) ve arka/birincil (primary) kameranın mb cinsinden görüntü kalitesini ifade eden "fc" ve "pc" boyutlarında tespit edilen 0 değerleri de silinmiştir. Veri setinin gürültülerden arındırılma süreci sonunda, Tablo 4.5'te gösterildiği üzere 1393 örnek kalmıştır.

battery_power	1393
blue	1393
clock_speed	1393
dual_sim	1393
fc	1393
four_g	1393
int_memory	1393
m_dep	1393
mobile_wt	1393
n_cores	1393
pc	1393
px_height	1393
px_width	1393
ram	1393
sc_h	1393
sc_w	1393
talk_time	1393
three_g	1393
touch_screen	1393
wifi	1393
price_range	1393
dtype:	int64

Tablo 4.5. Gürültülerden Arındırma Sonrası Örnek Sayısı

Aşağıda yer alan Şekil 4.1’de, keşifsel veri analizi kapsamında boyutların korelasyon matrisi oluşturulmuştur.



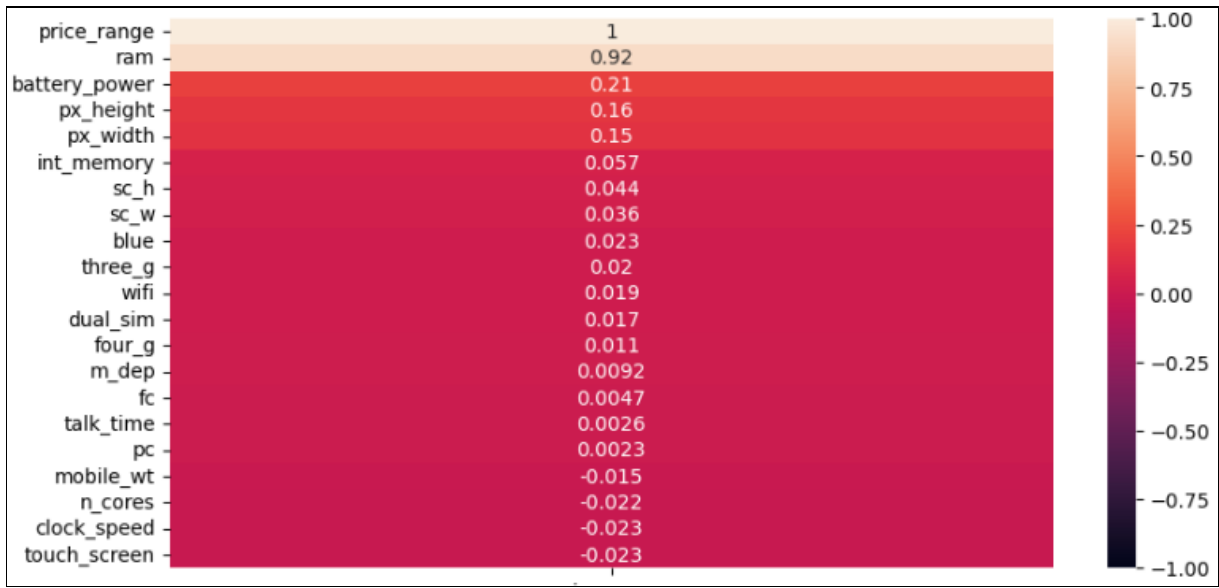
Şekil 4.1. Tüm Boyutların Korelasyon Matrisi

Özellik çıkarımı sürecinde bir fikir vermesi amacıyla, korelasyon matrisi dikkatlice incelenmiştir. Bu çerçevede piksel yüksekliği ve genişliğini ifade eden “px_height” ve “px_width” boyutları arasında nispeten önemli

sayılabilecek 0,50 değeriinde aynı yönlü bir korelasyon tespit edilmiştir. Benzer şekilde ekran yüksekliği ve genişliğini ifade eden “sc_h” ve “sc_w” boyutları arasında 0,52 değeriinde aynı yönlü başka bir korelasyon tespit edilmiştir. 3G ve 4G özellikleri ile ön ve arka kamera özellikleri arasında da korelasyonlar tespit edilmesine rağmen, bunların, sınıf özelliği “price_range” ile düşük bir korelasyona sahip olması sebebiyle üzerinde durulmamıştır.

4.2. Boyut İndirgeme ve Özellik Çıkarımı Öncesi KNN

Aşağıda bulunan Şekil 4.2’de ise özellik çıkarımı öncesi, tüm boyutların fiyat sınıfı “price_range” ile korelasyonları gösterilmektedir. Akıllı telefon fiyatları ile “ram” özelliği arasında 0,92 değeriinde aynı yönlü güçlü bir korelasyon tespit edilmiştir. Telefon fiyatları ile batarya gücü (“battery_power”), dikey ve yatay piksel sayısı (“px_height” ve “px_width”), dahili hafıza (“int_memory”) özellikleri arasında da nispeten zayıf korelasyonlar olduğu görülmektedir.



Şekil 4.2. Özellik Çıkarımı Öncesi: Tüm Boyutların Özellik Sınıfı “price_range” ile Korelasyonu

1393 örnekten oluşan veri setinin yüzde 30’u test, yüzde 70’i ise eğitim için kullanılmıştır. Oluşturulan KNN modelinde, deneme-yanılma yöntemiyle yapılan araştırmalar sonucunda, uzaklık ölçüsü olarak “manhattan”, ağırlık (weights) yöntemi olarak ise uzaklıklara göre veri noktalarını eşit bir şekilde ağırlıklandıran “uniform” parametresi kullanılmıştır [6].

Herhangi bir boyut indirgeme ve özellik çıkarımı işlemi uygulanmayan ve Tablo 4.6’da temsil edilen veri seti, KNN algoritmasına girdi olarak verilmiştir.

	battery_power	blue	clock_speed	dual_sim	fc	four_g	int_memory	m_dep	mobile_wt	n_cores	pc	ram	sc_solution	sc_dim	talk_time	three_g	touch_screen	wifi	price_range
0	842	0	2.2	0	1	0	7	0.6	188	2	2	2549	756.0	11.0	19	0	0	1	1
2	563	1	0.5	1	2	1	41	0.9	145	5	6	2603	2131.0	11.0	9	1	1	0	2
4	1821	1	1.2	0	13	1	44	0.6	141	2	14	1411	1711.0	8.0	15	1	1	0	1
5	1859	0	0.5	1	3	0	22	0.7	164	1	7	1067	1935.0	17.0	10	1	0	0	1
6	1821	0	1.7	0	4	1	10	0.8	139	8	10	3220	1087.0	15.0	18	1	0	1	3
...
1992	674	1	2.9	1	1	0	21	0.2	198	3	4	1180	1898.0	7.0	4	1	1	1	0
1994	858	0	2.2	0	1	0	50	0.1	84	1	2	3978	1511.0	23.0	3	1	1	0	3
1997	1911	0	0.9	1	1	1	36	0.7	108	8	3	3057	1848.0	9.0	5	1	1	0	3
1998	1512	0	0.9	0	4	1	46	0.1	145	5	5	869	750.0	21.0	19	1	1	1	0
1999	510	1	2.0	1	5	1	45	0.9	168	6	16	3919	895.0	19.0	2	1	1	1	3

1393 rows × 19 columns

Tablo 4.6. Boyut İndirgeme ve Özellik Çıkarımı Öncesi: Veri Seti

En uygun k değerini belirlemek amacıyla maksimum 20 iterasyondan oluşan bir döngüden ve “accuracy” metriğinden yararlanılmıştır.

k	test	train
1	0.892	1.0
2	0.883	0.945
3	0.907	0.952
4	0.9	0.941
5	0.909	0.943
6	0.904	0.934
7	0.911	0.942
8	0.895	0.933
9	0.895	0.95
10	0.892	0.937
11	0.892	0.939
12	0.895	0.941
13	0.888	0.943
14	0.89	0.941
15	0.892	0.939
16	0.89	0.934
17	0.892	0.933
18	0.895	0.932
19	0.89	0.933
20	0.902	0.937

Tablo 4.7. Boyut İndirgeme ve Özellik Çıkarımı Öncesi: Accuracy Metriğiyle En Uygun K Değerinin Tespit Edilmesi

İterasyon sonuçlarının yer aldığı Tablo 4.7’ye göre boyut indirgeme ve özellik çıkarımı öncesi, 0,91 (0,911) “accuracy” değeriyle en uygun k değeri 7 olarak tespit edilmiştir. Test ve eğitim setlerine ilişkin “accuracy” metriklerinin birbirine çok yakın olması (0,911’e 0,942), modelde aşırı öğrenme/ezberleme (overfitting/memorization) probleminin olmadığına işaret etmektedir. Diğer taraftan test ve eğitim veri setlerine ilişkin performans metriklerinin (somut probleme göre değişmekle birlikte) yüzde 90 üzerinde gerçekleşmesi, modelin yetersiz öğrenme (underfitting) problemine maruz kalmadığını ifade etmektedir.

4.3. Boyut İndirgeme ve Özellik Çıkarımı Sonrası KNN

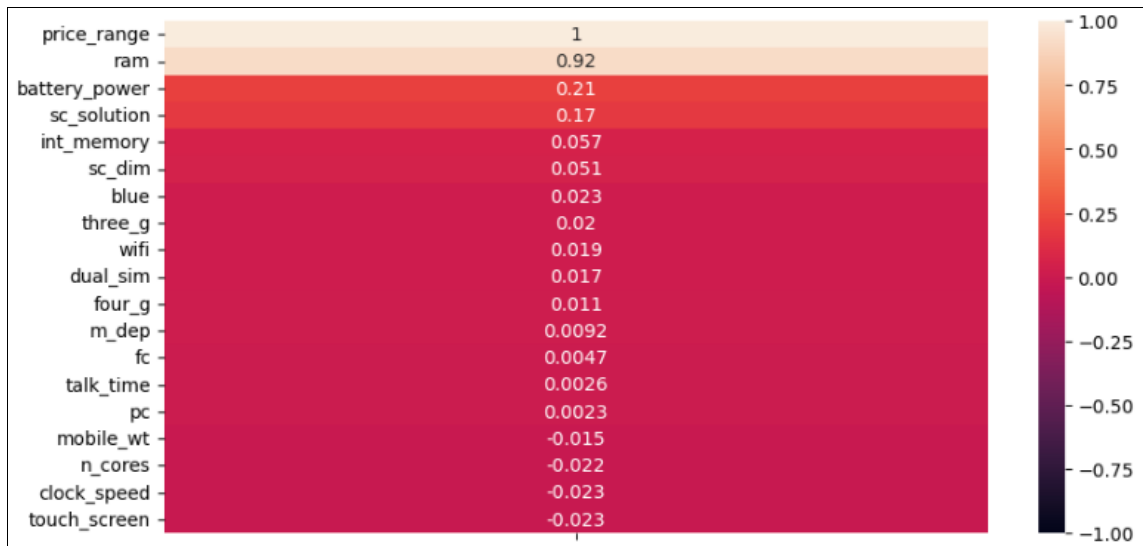
Özellik çıkarımı kapsamında, telefonların dikey ve yatay ekran çözünürlüklerini ifade eden “px_height” ve “px_width” özelliklerinin (öklit yöntemiyle) kareleri toplamının karekökü alınarak, “sc_solution” isimli yeni bir özellik oluşturulmuştur. Benzer işlem telefonların dikey ve yatay ekran boyutlarını temsil eden “sc_h” ve “sc_w” özelliklerinden “sc_dim” isimli yeni bir özelliğin ortaya çıkartılmasında da kullanılmıştır. Aşağıda yer alan Tablo 4.8’de özellik çıkarımı yöntemiyle elde edilen özellikler yer almaktadır.

	battery_power	blue	clock_speed	dual_sim	fc	four_g	int_memory	m_dep	mobile_wt	n_cores	pc	ram	sc_solution	sc_dim	talk_time	three_g	touch_screen	wifi	price_range
0	842	0	2.2	0	1	0	7	0.6	188	2	2	2549	756.0	11.0	19	0	0	1	1
2	563	1	0.5	1	2	1	41	0.9	145	5	6	2603	2131.0	11.0	9	1	1	0	2
4	1821	1	1.2	0	13	1	44	0.6	141	2	14	1411	1711.0	8.0	15	1	1	0	1
5	1859	0	0.5	1	3	0	22	0.7	164	1	7	1067	1935.0	17.0	10	1	0	0	1
6	1821	0	1.7	0	4	1	10	0.8	139	8	10	3220	1087.0	15.0	18	1	0	1	3
...
1992	674	1	2.9	1	1	0	21	0.2	198	3	4	1180	1898.0	7.0	4	1	1	1	0
1994	858	0	2.2	0	1	0	50	0.1	84	1	2	3978	1511.0	23.0	3	1	1	0	3
1997	1911	0	0.9	1	1	1	36	0.7	108	8	3	3057	1848.0	9.0	5	1	1	0	3
1998	1512	0	0.9	0	4	1	46	0.1	145	5	5	869	750.0	21.0	19	1	1	1	0
1999	510	1	2.0	1	5	1	45	0.9	168	6	16	3919	895.0	19.0	2	1	1	1	3

1393 rows × 19 columns

Tablo 4.8. Özellik Çıkarımı Sonrası: Veri Seti

Özellik çıkarımı sonrası, veri setinin tüm boyutlarını kapsayan ve Şekil 4.3’te gösterilen korelasyon matrisi incelenerek telefon fiyatları üzerinde etkili olan üç özellik tespit edilmiştir: ram, batarya gücü (“battery_power”), dikey ve yatay ekran piksel yoğunluğundan türetilen çözünürlük (“sc_solution”). Dikey ve yatay ekran boyutlarını temsil eden “sc_h” ve “sc_w” özelliklerinden türetilen “sc_dim” isimli özelliğin sınıf etiketi ile korelasyonu 0,051, yani çok düşük olduğu için modele dahil edilmemiştir. Benzer gerekçeyle sınıf etiketiyle korelasyonu düşük diğer boyutlar da modele dahil edilmemiştir. Boyut indirgeme kapsamında, problemin boyutu 20’den 3’e düşürülmüştür. Böylece accuracy metriğinde bir artış olsun ya da olmasın, problemin karmaşıklığının azaltılması ve modelin genelleme (generalization) yapabilme kabiliyetinin artırılması amaçlanmaktadır.



Şekil 4.3. Özellik Çıkarımı Sonrası: Tüm Boyutların Özellik Sınıfı “price_range” ile Korelasyonu

Boyut indirgeme ve özellik çıkarımı öncesi sürece benzer şekilde, 1393 örnekten oluşan veri setinin yüzde 30’u test, yüzde 70’ise eğitim için kullanılmıştır. Oluşturulan KNN modelinde, deneme-yanılma yöntemiyle

yapılan arařtırmalar sonucunda, uzaklık ölçüsü olarak “manhattan”, ağırlık (weights) yöntemi olarak ise uzaklıklara göre veri noktalarını eşit bir şekilde ağırlıklandıran “uniform” parametresi kullanılmıştır.

Tablo 4.9’da temsil edilen boyut indirgeme ve özellik çıkarımı sonrası elde edilen veri seti, KNN algoritmasına girdi olarak verilmiştir.

	ram	battery_power	sc_solution	price_range
0	2549	842	756.0	1
2	2603	563	2131.0	2
4	1411	1821	1711.0	1
5	1067	1859	1935.0	1
6	3220	1821	1087.0	3
...
1992	1180	674	1898.0	0
1994	3978	858	1511.0	3
1997	3057	1911	1848.0	3
1998	869	1512	750.0	0
1999	3919	510	895.0	3

1393 rows × 4 columns

Tablo 4.9. Özellik Çıkarımı Sonrası: Boyut İndirgeme

En uygun k değerini belirlemek amacıyla maksimum 20 iterasyondan oluşan bir döngüden ve “accuracy” metriğinden yararlanılmıştır.

k	test	train
1	0.89	1.0
2	0.871	0.938
3	0.904	0.952
4	0.885	0.938
5	0.904	0.945
6	0.907	0.93
7	0.902	0.934
8	0.904	0.926
9	0.897	0.932
10	0.9	0.922
11	0.9	0.927
12	0.9	0.928
13	0.888	0.926
14	0.88	0.925
15	0.885	0.925
16	0.895	0.922
17	0.892	0.918
18	0.89	0.919
19	0.89	0.916
20	0.888	0.92

Tablo 4.10. Boyut İndirgeme ve Özellik Çıkarımı Sonrası: Accuracy Metriğiyle En Uygun K Değerinin Tespit Edilmesi

İterasyon sonuçlarının yer aldığı Tablo 4.10'a göre boyut indirgeme ve özellik çıkarımı sonrası, 0,91 (0,907) "accuracy" değeriyle en uygun k değeri 6 olarak tespit edilmiştir. Test ve eğitim setlerine ilişkin "accuracy" metriklerinin birbirine çok yakın olması (0,907'e 0,930) modelde aşırı öğrenme/ezberleme (overfitting/memorization) probleminin olmadığına işaret etmektedir. Diğer taraftan test ve eğitim veri setlerine ilişkin performans metriklerinin (somut probleme göre değişmekle birlikte) yüzde 90 üzerinde gerçekleşmesi, modelin yetersiz öğrenme (underfitting) problemine maruz kalmadığını ifade etmektedir.

Sonuçlar boyut indirgeme ve özellik çıkarımı öncesi sonuçlarıyla büyük ölçüde benzer gerçekleşmiştir. Diğer bir ifadeyle, daha az sayıda özellik kullanılarak performanstan ödün verilmeden modelin karmaşıklığı düşürülmüştür.

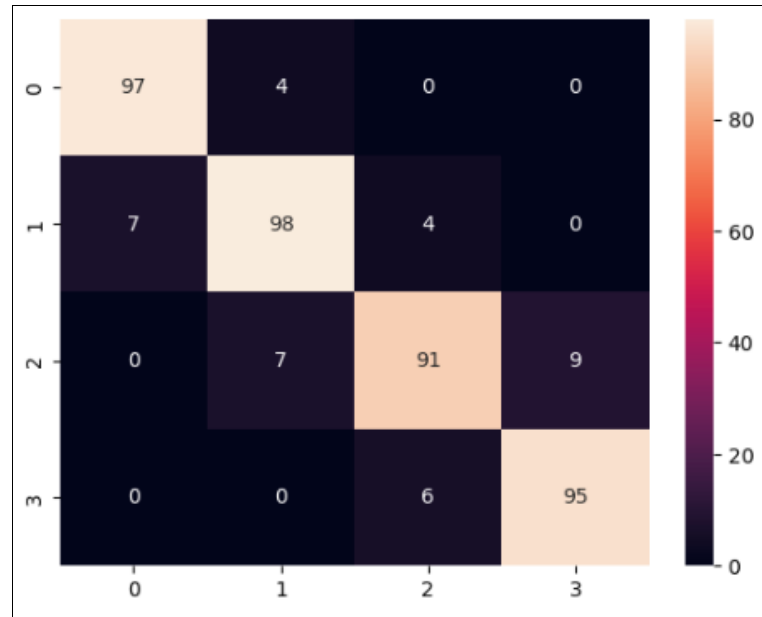
5. DENEYSEL ÇALIŞMA SONUÇLARI

Boyut indirgeme ve özellik çıkarımı öncesinde, 20 özellik ve bir sınıf özelliğiyle oluşturulan KNN modelinin performans metrikleri Tablo 5.1de sunulmuştur.

	precision	recall	f1-score
0	0.93	0.96	0.95
1	0.90	0.90	0.90
2	0.90	0.85	0.88
3	0.91	0.94	0.93
accuracy			0.91
macro avg	0.91	0.91	0.91
weighted avg	0.91	0.91	0.91

Tablo 5.1. Boyut İndirgeme ve Özellik Çıkarımı Öncesi: Performans Metrikleri

Boyut indirgeme ve özellik çıkarımı öncesinde oluşturulan KNN modeli, Şekil 5.1'de gösterilen karışıklık matrisine göre düşük (0 no.lu) fiyat sınıfından sadece 4 örneği, orta (1 no.lu) fiyat sınıfından 11 örneği, yüksek (2 no.lu) fiyat sınıfından 16 örneği ve çok yüksek (3 no.lu) fiyat sınıfından ise 6 örneği hatalı tahmin etmiştir.



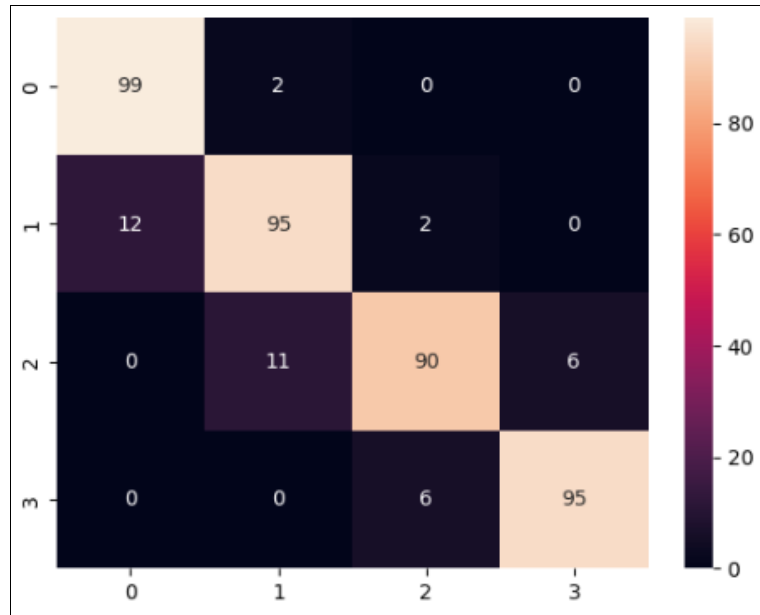
Şekil 5.1. Boyut İndirgeme ve Özellik Çıkarımı Öncesi: Karışıklık Matrisi

Boyut indirgeme ve özellik çıkarımı sonrasında, 3 özellik ve bir sınıf özelliğiyle oluşturulan KNN modelinin performans metrikleri Tablo 5.2’de sunulmuştur.

	precision	recall	f1-score
0	0.89	0.98	0.93
1	0.88	0.87	0.88
2	0.92	0.84	0.88
3	0.94	0.94	0.94
accuracy			0.91
macro avg	0.91	0.91	0.91
weighted avg	0.91	0.91	0.91

Tablo 5.2. Boyut İndirgeme ve Özellik Çıkarımı Sonrası: Performans Metrikleri

Boyut indirgeme ve özellik çıkarımı sonrasında oluşturulan KNN modeli, Şekil 5.2’de gösterilen karışıklık matrisine göre düşük (0 no.lu) fiyat sınıfından 2 örneği, orta (1 no.lu) fiyat sınıfından ise 14 örneği hatalı tahmin etmiştir. Benzer şekilde söz konusu model yüksek (2 no.lu) ve çok yüksek (3 no.lu) fiyat sınıfından, sırasıyla, 17 ve 6 örneği hatalı tahmin etmiştir .



Şekil 5.2. Boyut İndirgeme ve Özellik Çıkarımı Sonrası: Karışıklık Matrisi

Aşağıda yer alan Tablo 5.3, özellik çıkarımı ve boyut indirgeme öncesi ile sonrası karışıklık matrislerinden türetilmiş olup modellerin sınıf bazında (test veri setine göre) tahmin ettikleri doğru ve yanlış örnek sayılarını göstermektedir.

Fiyat Sınıfları	Özellik Çıkarımı ve Boyut İndirgeme	Doğru	Yanlış	Toplam
Düşük (0)	Öncesi	97	4	101
	Sonrası	99	2	
Orta (1)	Öncesi	98	11	109
	Sonrası	95	14	
Yüksek (2)	Öncesi	91	16	107
	Sonrası	90	17	
Çok Yüksek (3)	Öncesi	95	6	101
	Sonrası	95	6	
Tüm Sınıflar	Öncesi	381	37	418
	Sonrası	379	39	

Tablo 5.3. Boyut İndirgeme ve Özellik Çıkarımı Öncesi ve Sonrası: Doğru ve Yanlış Sınıflandırılan Örnek Sayılarının Karşılaştırılması

Tablo 5.3'e göre düşük (0 no.lu) fiyat sınıfından sonraki model, önceki modelden 2 eksik (yani 2 örneği), orta (1 no.lu) fiyat sınıfından ise 3 fazla (yani 14 örneği), yüksek (2 no.lu) fiyat sınıfından ise 1 fazla (yani 17 örneği) hatalı tahmin etmiştir. Söz konusu model çok yüksek (3 no.lu) fiyat sınıfından, önceki model ile aynı sayıda (yani 6 örnek için) hatalı tahminde bulunmuştur.

Yeni durumda (sonraki) modelin genel performansında kayda değer herhangi bir kötüleşme olmadığı, aksine bazı fiyat sınıflarında küçük de olsa iyileşme olduğu, dolayısıyla daha az sayıda özellik ile benzer performansın elde edildiği görülmektedir. Veri madenciliğinin temel amaçlarından birinin de faydalı bilginin (knowledge) genelleme (generalization) yapabilme kabiliyetini barındırması gerektiği göz önüne alınırsa, benzer performansa sahip daha az karmaşık bir modelin, genelleme yapabilme kabiliyetinin de daha fazla olduğu öne sürülebilir.

6. SONUÇ VE DEĞERLENDİRME

KNN modeliyle yapılan çalışmada akıllı telefon fiyatları üzerinde ram, batarya gücü ("battery_power"), dikey ve yatay ekran piksel yoğunluğundan türetilen çözünürlük ("sc_solution") özelliklerinin büyük ölçüde belirleyici olduğu tespit edilmiştir. Diğer bir ifadeyle söz konusu üç özellik kullanılarak herhangi bir akıllı telefonun düşük, orta, yüksek ve çok yüksek şeklinde ait olduğu fiyat sınıfı, yüzde 91 doğrulukta öngörülmüştür.

Sınıf özelliği hariç, 20 boyutlu veri setiyle yapılan öngörü ile özellik çıkarımı ve boyut indirgeme yöntemleriyle sadece 3 boyuta indirgenen veri setiyle yapılan öngörü çok yakın performanslara sahiptir. Korelasyon matrisi analiz edilerek dikey ve yatay çözünürlükleri ifade eden 2 farklı boyut, özellik çıkarımı yöntemiyle "çözünürlük" ("sc_solution") isminde tek boyuta indirgenmiştir. Benzer şekilde diğer boyutların da hedeflenen sınıf etiketi üzerindeki etkilerinin sınırlı olması sebebiyle, içlerinden sadece en önemli 2 boyutun modele dahil edilmesi suretiyle boyut indirgeme süreci tamamlanmıştır.

Bir modelin mümkün olan en basit karmaşıklık seviyesine sahip olması, veri madenciliğinde arzulanan bir hedeftir. Faydalı bilgiyi (knowledge) ortaya çıkarmayı amaçlayan veri madenciliğinde, modelin genelleme (generalization) kabiliyeti kritik bir unsurdur. Genelleme kabiliyetini arttırmak ise problemin çözümüne ilişkin mümkün olan en az karmaşıklıkta bir model gerektirir. Ayrıca problemin karmaşıklığının azaltılması, KNN gibi

bellek tabanlı çalışan algoritmaların özellikle büyük ölçekli verilerde kullanımını önemli ölçüde olumlu etkileyeceği de söylenebilir.

Sonuç olarak KNN modelinin kullanıldığı bu çalışmada 20 boyutlu bir problem, veri madenciliği teknikleriyle sadece 3 boyuta indirgenmek suretiyle daha az karmaşıktaki modellenmiştir. Böylece veri ön işleme teknikleriyle problemin çözümünün anlaşılabilirliği ve sunumu kolaylaştırılmıştır.

KAYNAKLAR

- [1] Aggarwal, A., et al.(2018). Evaluation of Smartphones in Indian Market Using EDAS. Procedia Computer Science, 132(2018), 236-243.
- [2] Bhattacharjee, D., et al.(2023). An Integrated Machine Learning and DEMATEL Approach for Feature Preference and Purchase Intention Modelling. Decision Analytics Journal, 6(2023), 1-13.
- [3] Singh, R., et al.(2021). Exploratory Data Analysis and Customer Segmentation for Smartphones. International Journal of Scientific Research in Engineering and Management (IJSREM), 5(5), 1-5.
- [4] Anshori, M.Y., et al.(2021). Comparison Backpropagation (BP) and Learning Vector Quantification (LVQ) on Classifying Price Range of Smartphone in Market. Journal of Physics: Conference Series, 1836(2021), 1-10.
- [5] Mehrotra, D., et al.(2021). Multiclass Classification of Mobile Applications as per Eenergy Consumption. Journal of King Saud University–Computer and Information Sciences, 33(2021), 719-727.
- [6] <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>, Eriřim Tarihi: 28.04.2023.
- [7] <https://www.kaggle.com/code/panini92/mobile-price-classification-eda-models>, Eriřim Tarihi: 28.04.2023.
- [8] <https://www.kaggle.com/code/yousefelbaroudy/mobile-price-classification-98>, Eriřim Tarihi: 28.04.2023.
- [9]<https://www.kaggle.com/code/the314arham/99-accuracy-ensemble-learning-xgboost-knn>, Eriřim Tarihi: 28.04.2023.
- [10] <https://www.kaggle.com/datasets/iabhishekoofficial/mobile-price-classification>, Eriřim Tarihi: 28.04.2023.