

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**BİNALARIN MAKİNE ÖĞRENMESİ
ALGORİTMALARIYLA MEKÂNSAL
ÖZELLİKLERİNDEN YARARLANARAK
FONKSİYONEL OLARAK SINIFLANDIRILMASI**

Ayça Nursena BİLGİN

YÜKSEK LİSANS TEZİ

Harita Mühendisliği Anabilim Dalı

Uzaktan Algılama ve Coğrafi Bilgi Sistemleri Programı

Danışman

Prof. Dr. Ali Melih BAŞARANER

Şubat, 2025

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**BİNALARIN MAKİNE ÖĞRENMESİ ALGORİTMALARIYLA
MEKÂNSAL ÖZELLİKLERİNDEN YARARLANARAK
FONKSİYONEL OLARAK SINIFLANDIRILMASI**

Ayça Nursena BİLGİN tarafından hazırlanan tez çalışması 26.02.2025 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Harita Mühendisliği Anabilim Dalı, Uzaktan Algılama ve Coğrafi Bilgi Sistemleri Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Prof. Dr. Ali Melih BAŞARANER
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Prof. Dr. Ali Melih BAŞARANER, Danışman
Yıldız Teknik Üniversitesi

Prof. Dr. Fusun BALIK ŞANLI, Üye
Yıldız Teknik Üniversitesi

Prof. Dr. Necla ULUGTEKİN, Üye
İstanbul Teknik Üniversitesi

Danışmanım Prof. Dr. Ali Melih BAŞARANER sorumluluğunda tarafımca hazırlanan “Binaların Makine Öğrenmesi Algoritmalarıyla Mekânsal Özelliklerinden Yararlanarak Fonksiyonel Olarak Sınıflandırılması” başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Ayça Nursena BİLGİN

İmza



Aileme
ve
eşime

TEŞEKKÜR

Bu tez çalışmamın tamamlanmasında bana yol gösteren, bilgi ve tecrübelerini benimle paylaşan tez danışmanım Prof. Dr. Ali Melih BAŞARANER'e, bu uzun ve zorlu süreçte beni yalnız bırakmayan, her zaman yanımda olan, en büyük destekçilerim olan aileme, annem Mürüvvet KORKMAZ'a, babam Şeref KORKMAZ'a ve sevgili eşim Hasan BİLGİN'e sonsuz sabır ve anlayışları için teşekkür ederim.

Ayça Nursena BİLGİN

İÇİNDEKİLER

SİMGE LİSTESİ	vii
KISALTMA LİSTESİ	viii
ŞEKİL LİSTESİ	x
TABLO LİSTESİ	xii
ÖZET	xiii
ABSTRACT	xv
1 GİRİŞ	1
1.1 Literatür Özeti	2
1.2 Tezin Amacı	5
1.3 Hipotez	6
2 ŞEKİL ANALİZİ VE SINIFLANDIRMA	7
2.1 Şekil Analizi	8
2.2 Mekânsal Özellikler	9
2.2.1 Büyüklük Göstergeleri	10
2.2.2 Şekil Göstergeleri	11
2.2.3 Mekânsal İlişkiler	19
2.2.4 Nicel Öznitelikler	20
2.3 Sınıflandırma	20
2.3.1 Sınıflandırma Algoritmaları	21
2.4 Sınıflandırma Algoritmalarının Performanslarının Değerlendirilmesi	37
3 UYGULAMA	44

3.1 Çalışma Alanı.....	44
3.2 Veri Hazırlığı	45
3.2.1 Veri Toplama ve Manipülasyonu	45
3.2.2 Büyüklük Göstergelerinin Hesaplanması	47
3.2.3 Şekil Göstergelerinin Hesaplanması	50
3.2.4 Mekânsal İlişkilerin Hesaplanması	56
3.2.5 Nicel Özniteliklerin Hesaplanması	58
3.2.6 Fonksiyonel Sınıflandırma	59
3.3 Model Eğitimi ve Değerlendirilmesi.....	62
3.3.1 Uygulama 1	63
3.3.2 Uygulama 2	69
3.3.3 Tartışma	75
4 SONUÇ	77
KAYNAKÇA	78
A SINIFLANDIRMA İÇİN LİTERATÜR ARAŞTIRMASI	82
B İNDİRİLEN VERİ SETİNİN SINIFLANDIRMASI	84
C UYGULAMA 1: CART ALGORİTMASI KODLARI VE OPTİMİZASYON SÜRECİ	87
D UYGULAMA 2: RASTGELE ORMAN ALGORİTMASI KODLARI VE OPTİMİZASYON SÜRECİ	89
E UYGULAMA 1: SINIFLARIN PERFORMANS METRİKLERİ	91
F UYGULAMA 2: SINIFLARIN PERFORMANS METRİKLERİ	92
TEZDEN ÜRETİLMİŞ YAYINLAR	93

SİMGE LİSTESİ

A_{CK}	Çokgenin Alanı
C_{CK}	Çokgenin Çevresi
A_{DB}	Dışbükey Gövdenin Alanı
C_{EAD}	Eşit Alanlı Dikdörtgenin Çevresi
C_{EAK}	Eşit Alanlı Karenin Çevresi
X^2	Ki-Kare
A_{MASD}	Minimum Alan Sınırlayıcı Dikdörtgenin Alanı
C_{MASD}	Minimum Alan Sınırlayıcı Dikdörtgenin Çevresi
$MASD_K$	Minimum Alan Sınırlayıcı Dikdörtgenin Kısa Kenarı
$MASD_U$	Minimum Alan Sınırlayıcı Dikdörtgenin Uzun Kenarı

KISALTMA LİSTESİ

AdaBoost	Adaptive Boosting
CART	Classification Regression Trees
D	Dairesellik
DB	Dışbükeylik
DD	Dikdörtgensellik
DN	Doğru Negatif
DP	Doğru Pozitif
DVM	Destek Vektör Makineleri
EDG	Eşdeğer Dikdörtgen Göstergesi
EPISCOPE	Energy Performance Indicator Tracking Schemes for the Continuous Optimisation of Refurbishment Processes in European Housing Stocks
EPSG	The European Petroleum Survey Group
GENIUS	Generator of Interactive Urban Blocks
ID3	Iterative Dichotomiser – 3
IPCC	Intergovernmental Panel on Climate Change
K	Karesellik
k-NN	K-En Yakın Komşu
MASD	Minimum Alan Sınırlayıcı Dikdörtgen
MSD	Minimum Sınırlayıcı Dikdörtgen
NAD	The North American Datum
P	Pürüzlülük
POI	Point of Interest
RO	Rastgele Orman
ROC	Receiver Operating Characteristic Curve
ROC AUC	Receiver Operating Characteristic Curve Area the Curve
SGB	Stochastic Gradient Boosting
TABULA	Typology Approach for Building Stock Energy Assessment
U	Uzanim

XGBoost	eXtreme Gradient Boosting
YN	Yanlış Negatif
YP	Yanlış Pozitif



ŞEKİL LİSTESİ

Şekil 2.1 Şekil analiz süreçleri	7
Şekil 2.2 Çalışma kapsamında binalara yönelik olarak dikkate alınan mekânsal özellikler	10
Şekil 2.3 Çokgen ve eşit çevreli çemberi	11
Şekil 2.4 Dikdörtgensellik	13
Şekil 2.5 Minimum alanlı sınırlayıcı dikdörtgenin bulunması	13
Şekil 2.6 Minimum sınırlayıcı dikdörtgen	14
Şekil 2.7 Minimum sınırlayıcı çokgenin boyutu çokgenin yönünden etkilenir	14
Şekil 2.8 Karesellik	15
Şekil 2.9 Dışbükeylik	16
Şekil 2.10 Eşdeğer dikdörtgen göstergesi	17
Şekil 2.11 Uzanım	18
Şekil 2.12 Pürüzlülük	19
Şekil 2.13 Makine öğrenmesi paradigmaları	22
Şekil 2.14 Karar ağacı yapısı	23
Şekil 2.15 Karar ağacı düğümleri	23
Şekil 2.16 Minimum nesne budaması tekniği	27
Şekil 2.17 Hata tabanlı budama tekniği	28
Şekil 2.18 Topluluk öğrenmesi algoritması yapısı	29
Şekil 2.19 Torbalama (bagging) ve güçlendirme (boosting)	30
Şekil 2.20 Hiperdüzlem ile sınıfların ayrılması	33
Şekil 2.21 k-NN algoritmasının görsel çizimi	36
Şekil 2.22 ROC eğrisi ve ROC AUC	41
Şekil 3.1 Miami-Dade ilinin konumu	44
Şekil 3.2 Makine öğrenmesi ile model oluşturma	45
Şekil 3.3 Çalışma alan sınırları	46
Şekil 3.4 Bina alanı göstergesinin konut ve konut dışı binalara göre değer dağılımı	47
Şekil 3.5 Parsel alanı göstergesinin konut ve konut dışı binalara göre değer dağılımı	48

Şekil 3.6 Voronoi çokgeni alanı göstergesinin konut ve konut dışı binalara göre değer dağılımı	48
Şekil 3.7 Bina alanı göstergesinin bina fonksiyonlarına göre değer dağılımı	49
Şekil 3.8 Parsel alanı göstergesinin bina fonksiyonlarına göre değer dağılımı	49
Şekil 3.9 Voronoi çokgeni alanı göstergesinin bina fonksiyonlarına göre değer dağılımı	50
Şekil 3.10 Konut ve konut dışı binaların şekil gösterge değerlerinin dağılımları	52
Şekil 3.11 Dairesellik göstergesinin bina fonksiyonlarına göre değer dağılımı	53
Şekil 3.12 Dikdörtgensellik göstergesinin bina fonksiyonlarına göre değer dağılımı	53
Şekil 3.13 Karesellik göstergesinin bina fonksiyonlarına göre değer dağılımları	54
Şekil 3.14 Dışbükeylik göstergesinin bina fonksiyonlarına göre değer dağılımları	54
Şekil 3.15 Eşdeğer dikdörtgen göstergesinin bina fonksiyonlarına göre değer dağılımları	55
Şekil 3.16 Uzanım göstergesinin bina fonksiyonlarına göre değer dağılımları	55
Şekil 3.17 Pürüzlülük göstergesinin bina fonksiyonlarına göre değer dağılımları	56
Şekil 3.18 Bazı bina ayak izlerinin Voronoi çokgenlerine alansal oranları	56
Şekil 3.19 Konut ve konut dışı binaların mekânsal ilişki gösterge değerlerinin dağılımları	57
Şekil 3.20 Bina fonksiyonlarının mekânsal ilişki gösterge değerlerinin dağılımları	57
Şekil 3.21 Kat adedi göstergesinin konut ve konut dışı binalara göre değer dağılımı	58
Şekil 3.22 Bina yüksekliği göstergesinin konut ve konut dışı binalara göre değer dağılımı	58
Şekil 3.23 Kat adedi göstergesinin bina fonksiyonlarına göre değer dağılımı	59
Şekil 3.24 Bina yüksekliği göstergesinin bina fonksiyonlarına göre değer dağılımı	59
Şekil 3.25 Tez kapsamında yapılan sınıflandırma	62
Şekil 3.26 Model eğitimi ve değerlendirmesi akış şeması	63
Şekil 3.27 CART algoritması göstergelerinin önem değerleri	66
Şekil 3.28 Rastgele orman algoritması gösterge önem değerleri	67
Şekil 3.29 Aşırı gradyan artırma algoritması gösterge önem değerleri	68
Şekil 3.30 CART algoritması gösterge önem değerleri	71
Şekil 3.31 Rastgele orman algoritması gösterge önem değerleri	73
Şekil 3.32 Aşırı gradyan artırma algoritması gösterge önem değerleri	74

TABLO LİSTESİ

Tablo 2.1 Sınıflandırma yaklaşımları.....	20
Tablo 2.2 Hata matrisi.....	38
Tablo 2.3 Psikiyatristlerin hastalara koydukları tanılar	42
Tablo 2.4 K istatistiği uyum düzeyleri	42
Tablo 3.1 Uygulama 1'deki konut ve konut dışı binaların örnek bina ayak izleri ve şekil gösterge değerleri	50
Tablo 3.2 Uygulama 2'deki bina fonksiyonlarının örnek bina ayak izleri ve şekil gösterge değerleri.....	52
Tablo 3.3 Fonksiyonel sınıflandırma için literatür araştırması	60
Tablo 3.4 CART algoritmasının hata matrisi	65
Tablo 3.5 DVM algoritmasının hata matrisi	66
Tablo 3.6 Rastgele orman algoritmasının hata matrisi.....	67
Tablo 3.7 Aşırı gradyan artırma algoritmasının hata matrisi	68
Tablo 3.8 k-en yakın komşu algoritmasının hata matrisi	69
Tablo 3.9 CART algoritmasının hata matrisi	71
Tablo 3.10 Rastgele orman algoritmasının hata matrisi.....	72
Tablo 3.11 Aşırı gradyan artırma algoritmasının hata matrisi	73
Tablo 3.12 k-en yakın komşu algoritmasının hata matrisi	74

Binaların Makine Öğrenmesi Algoritmalarıyla Mekânsal Özelliklerinden Yararlanarak Fonksiyonel Olarak Sınıflandırılması

Ayça Nursena BİLGİN

Harita Mühendisliği Anabilim Dalı

Uzaktan Algılama ve Coğrafi Bilgi Sistemleri Programı

Yüksek Lisans Tezi

Danışman: Prof. Dr. Ali Melih BAŞARANER

Dünya genelinde insanlar, kırsal alanlardan kentlere doğru göç etmektedir. Bu göç hareketi karmaşık yapılaşmayı ve kent yönetimi için zorlukları beraberinde getirmektedir. Kentler için oldukça büyük öneme sahip olan acil durum ve kaynakların yönetimi, iklim değişikliği modellerinin oluşturulması gibi kent yönetim planlarının temel veri kaynağını farklı bina fonksiyonları oluşturmaktadır. Binaların kullanım türlerinin belirlenmesi planlama süreçlerinin temel bir adımıdır. Bu çalışmada, bina detaylarının (nesnelerinin) coğrafi bilgi sistemi (CBS) ortamında elde edilmiş çeşitli mekânsal özelliklerinden (büyüklük, şekil, mekânsal ilişki metrikleri ve niceliksel öznitelikler) yararlanarak fonksiyonel (kullanım türü) sınıflandırmasında farklı makine öğrenmesi algoritmalarının performansları değerlendirilmiş ve böylece kentsel planlama ve kent yönetimi süreçlerine katkı sağlanması amaçlanmıştır. Deneysel çalışmada Florida eyaletinin Miami-Dade ilindeki bina detay verileri kullanılarak iki uygulama gerçekleştirilmiştir. Uygulamalarda kullanılan algoritmalar Grid Search yöntemi ile en iyi performansı

gösterecek olan hiper parametreleri ile eğitilmiştir. İlk uygulamada binalar konut ve konut dışı olmak üzere iki farklı fonksiyona göre sınıflandırılmıştır. Sınıflandırmada CART, destek vektör makineleri (DVM), rastgele orman (RO), aşırı gradyan artırma (XGBoost) ve k – En Yakın Komşu (k-NN) algoritmaları kullanılmıştır. Algoritmaların F1-skor değerleri sırasıyla 0.83, 0.81, 0.83, 0.83 ve 0.74'tır. Topluluk öğrenmesi algoritmalarının daha iyi performans gösterdiği görülmüştür. İkinci uygulamada (1) fabrika ve depo, (2) konut, (3) ticaret ve (4) kamu hizmet binaları olmak üzere dört farklı fonksiyona göre sınıflandırılmıştır. Sınıflandırmada CART, rastgele orman, aşırı gradyan artırma ve k-NN algoritmaları kullanılmıştır. Algoritmaların F1-skor değerleri sırasıyla 0.63, 0.68, 0.68 ve 0.60'tır. Topluluk öğrenmesi algoritmalarının daha iyi performans gösterdiği görülmektedir. Hata matrisleri incelendiğinde her iki uygulamada da algoritmaların hepsi konut sınıfında yüksek doğruluk göstermiştir, fabrika ve depo sınıfı bunu takip etmektedir. Kamu hizmet binaları en az doğruluğun gözlemlendiği sınıftır. Algoritmaların hepsi kamu hizmeti sınıfına ait olan bazı binaları ticaret; ticaret sınıfına ait olan bazı binaları ise kamu hizmeti ve konut olarak sınıflandırmıştır. Metin ve konum tabanlı sosyal medya verileri ve sokak görünümü verilerinin çalışmalara dâhil edilmesiyle elde edilen sonuçların desteklenmesi sağlanabilir.

Anahtar Kelimeler: Bina detayları, fonksiyonel sınıflandırma, mekânsal özellikler, makine öğrenmesi

ABSTRACT

Functional Classification of Buildings Based on Their Spatial Characteristics Using Machine Learning Algorithms

Ayça Nursena BİLGİN

Department of Geomatic Engineering
Master of Science Thesis

Supervisor: Prof. Dr. Ali Melih BAŞARANER

People are migrating from rural areas to cities all over the world. This migration movement brings with it complex structures and challenges for urban management. Different building functions constitute the basic data source of urban management plans such as emergency and resource management, which are of great importance for cities, and the creation of climate change models. Determining the types of use of buildings is a fundamental step in planning processes. In this study, the performances of different machine learning algorithms in functional (type of use) classification were evaluated by utilizing various spatial characteristics (size, shape, spatial relationship metrics and quantitative attributes) obtained from the geographic information system (GIS) environment of building features (objects), and thus, it was aimed to contribute to urban planning and urban management processes. In the experimental study, two applications were carried out using building detail data from Miami-Dade County, Florida. The algorithms used in the applications were trained with the Grid Search method and the hyper parameters that would show the best performance. In the first application, buildings were

classified according to two different functions: residential and non-residential. CART, support vector machines (SVM), random forest (RF), extreme gradient boosting (XGBoost) and k-Nearest Neighbors (k-NN) algorithms were used in the classification. The F1-score values of the algorithms are 0.83, 0.81, 0.83, 0.83 and 0.74, respectively. Ensemble learning has been shown to perform better. In the second application, (1) factory and warehouse, (2) residential, (3) commercial and (4) public service buildings were classified according to four different functions. CART, random forest, extreme gradient boosting (XGBoost) and k-NN algorithms were used in the classification. The F1-score values of the algorithms are 0.63, 0.68, 0.68 and 0.60, respectively. Ensemble learning algorithms were found to perform better. When the error matrices are examined, all of the algorithms showed high accuracy in the residential class in both applications, followed by the factory and warehouse class. Public service buildings are the class with the least accuracy. All of the algorithms classified some buildings belonging to the public service class as commercial; and some buildings belonging to the commercial class as public service and residential. The results obtained can be supported by including text and location-based social media data and street view data in the studies.

Keywords: Building features, functional classification, spatial characteristics, machine learning

1 GİRİŞ

Şehirlere doğru olan göçler her geçen yıl artmaktadır. Birleşmiş Milletlerin “Dünya Nüfus Beklentileri” ile ilgili raporunda 2050 yılında dünya nüfusunun %68’inin şehirlerde yaşayacağı tahmin edilmektedir (*World Urbanization Prospects: The 2018 Revision*, 2019). Bu durum beraberinde şehrsel yapılaşmada önemli ölçüde değişikliğe ve gelişime yol açmaktadır. Yapılaşmadaki değişim ve gelişim süreçlerinin yönetilebilmesi için şehirlerin temel yapıtaşları olan binalar ve sahip oldukları fonksiyonlar yüksek öneme sahiptirler.

Yapılaşma sürecinin ilk adımlarında insanlar barınma ihtiyaçlarını karşılamak için konut inşa ederler. Ancak sürecin devamında insanların yaşam, çalışma ve dinlenme gibi aktivitelerinin bir sonucu olarak çeşitli bina fonksiyonları ortaya çıkar ve şehrsel yapılaşma daha da karmaşık bir hale gelir. Bu durum kuşkusuz acil durumların ve kaynakların (enerji, su, elektrik gibi) yönetilmesi, iklim değişimi modellerinin oluşturulması ve sera gazı emisyonlarının etkilerinin değerlendirilmesi gibi şehir yönetiminde büyük öneme sahip süreçleri etkilemektedir.

Plan ve modeller oluşturulurken alınan kararların temelinde doğru ve güvenilir mekânsal veriler yer alır. Bu veriler, geometrik ve semantik nitelikte olup, geometrik veriler genellikle mekânsal veri toplama teknikleri ve semantik veriler ise genellikle ek saha çalışmaları ile elde edilir. Ayrıca, önceden mevcutsa bu tip veriler, yerel yönetimler başta olmak üzere çeşitli kamu kurumları veya özel kuruluşlardan da temin edilebilir. Saha çalışmaları büyük ölçüde insan gücüne dayanır, fazla zaman alır, güçlü finansal kaynaklar gerektirir ve hata payı yüksek veriler ortaya çıkarır (Li vd., 2023). Ayrıca insan gücüne dayanan yöntemlerde verileri sürekli olarak güncellemek mümkün değildir. Yerel yönetimlerin verileri ise nadiren ve kısmi olarak erişilebilirdir, bu da eksiksiz çalışmalar ortaya konulmasında engel oluşturmaktadır. Diğer veri kaynakları arasında ilgi noktası

(POI) verileri, sokak görünümü verileri ve coğrafi referanslı sosyal medya verileri sayılabilir (Lin vd., 2021).

1.1 Literatür Özeti

Binalar farklı disiplinlere sahip birçok araştırmacı tarafından kentlerin problemlerini çözmek ya da sahip oldukları potansiyellerini ortaya çıkarmak için temel veri kaynağı olarak kullanılmaktadır. Bu temel veri kaynağı dünyamızın geleceğine yön verebilecek özellikle sürdürülebilirlik konusunda da birçok çalışmada kullanılmaktadır. Araştırmacılar karbon emisyonlarının ve küresel iklim değişikliklerinin temel sebebi olarak kentleşmeyi göstermektedir. Küresel ölçekte incelendiğinde 2019 yılında karbon emisyonlarının %31'inin binalardan kaynaklandığı görülmektedir. Çalışmalar farklı bina fonksiyonlarının kaynaklar ve enerji tüketimindeki etkilerinin de farklı olduğunu göstermektedir. Hükümetlerarası İklim Değişikliği Paneli (Intergovernmental Panel on Climate Change - IPCC) raporuna göre 2019 yılındaki enerji kullanımının %50'sini konutlar, %32'sini konut dışı binalar ve %18'ini ise yapısal emisyonlar oluşturmaktadır (Bashmakov vd., 2022). Enerji taleplerinin azaltılması ve enerji sistemlerinin geliştirilmesi için özellikle son yıllarda birçok çalışma yapılmaktadır. Bu çalışmalara Avrupa'daki binaların enerji performansını değerlendirmek ve iyileştirmek için geliştirilen 2009 yılında çalışmalarına başlanan "Bina Stok Enerji Değerlendirmesi için Tipoloji Yaklaşım" (Typology Approach for Building Stock Energy Assessment - TABULA) projesi ve bu projeyi takiben 2013 yılında çalışmalarına başlanan "Avrupa Konut Stoklarında Yenileme Süreçlerinin Sürekli Optimizasyonu için Enerji Performans Göstergesi İzleme Şemaları" (Energy Performance Indicator Tracking Schemes for the Continuous Optimisation of Refurbishment Processes in European Housing Stocks - EPISCOPE) projesi örnek olarak verilebilir. Tornay vd. 2017 yılında yaptıkları çalışmada TABULA ve EPISCOPE projelerine eleştirel bir bakış açısı ile bakarak 'Etkileşimli Kentsel Blok Jeneratörü' (GENerator of Interactive Urban blockS - GENIUS) projesini geliştirdiler. GENIUS projesi diğer iki çalışmadan farklı olarak sadece konutların enerji tüketimlerine değinmeyerek konut dışı binaları da çalışma kapsamının içine alarak toplamda 12 farklı bina fonksiyonu çerçevesinde geliştirilmiştir. Çalışmada asıl amaçlanan farklı bina fonksiyonlarının sahip oldukları özelliklerin kentsel enerji dengesini ne ölçüde

etkilediğini simüle etmektir. Xiao vd. (2022) çalışmalarında geri yayılımlı sinir ağı modelini kullanarak Çin'in Chongqing kentindeki konut binalarının enerji talebini oluşturdular. Çalışmalarının mevcuttaki ve kente gelecekte inşa edilecek binalara enerji tasarruf programları ve ölçümleri için kılavuz olabileceğini ifade etmişlerdir. Li vd. (2022), bina yüksekliği ve bina ayak izi verilerinin yerel iklime olan etkilerini incelemişlerdir. Kentsel binalar ve üç boyutlu morfolojilerinin, kentsel iklim ve hava kirleticilerinin yayılmasında önemli rol oynadığını vurgulamışlardır.

Farklı bina fonksiyonları deprem araştırmalarında da önemli bir veri kaynağı olmuştur. Yang vd. Çin'de kentsel yapı dayanıklılığını ölçmek için yapı malzemesi yoğunluğu veri tabanı oluşturmuşlardır (Xiao vd., 2022). Bu veri tabanının, yapı malzemesi stok analizi, işlenmemiş malzeme çıkarma talebi, inşaat yıkım atık tahmini ve geri dönüşüm potansiyeli gibi konularda önemli bir gösterge olduğunu vurgulamışlardır. Jing vd. Çin'deki farklı bina fonksiyonlarının deprem risk analizi ve deprem sigortası araştırmaları için önemli bir veri kaynağı olduğunu vurgulamışlardır (Xiao vd., 2022).

Farklı bir disiplin açısından incelendiğinde bina fonksiyonları kentsel arazi kullanım haritalarının oluşturulmasında da veri kaynağı olarak kullanılmaktadır. Niu vd. (2017) binaların işlevlerine ilişkin bilgilerin şehir planlaması için önemli bir temel olduğunu vurgulayarak arazi kullanım haritalarını çok daha ayrıntılı bir mekânsal ölçekte oluşturmuşlardır. Taksi "Küresel Konumlama Sistemi", WeChat (bir mesajlaşma uygulaması) verilerini ve bina ayak izi verilerini kullanmışlar. Çin'in Tianhe Bölgesi'nde 68.997 bina ayak izi verisi ile yaptıkları çalışmada arazi kullanımını 6 bölgeye (konut, ofis, okul, ticari bina, otel, hastane) ayırmışlar ve %72,22 nihai tespit oranı ve %65'in üzerinde doğruluk oranı elde etmişlerdir. Zhang vd. (2020) arazi kullanım sınıflandırmasını iyileştirmek için arazi örtüsü ve bina fonksiyonlarını kullanmışlardır. Rastgele orman algoritması ile iki veri setini birleştirmişlerdir. İki verinin birlikte kullanılmasının F1 puanını ve duyarlılığı arttırdığını göstermişlerdir. Du vd. (2024) Çin'in Pekin ve Şanghai kentlerinin arazi kullanım haritalarını oluşturmak için denetlenen ve denetlenmeyen algoritmaları birleştiren iki adımlı bir yöntem önermişlerdir. İlgi noktası (POI) ve Open Street Map verilerini kullanmışlardır. Bu iki adımlı birleştirmeye birlikte sınıflandırma için ek manuel örneklerle gerek olmadığını ve ilgi noktası verilerinin sınıflandırma üzerindeki etkisinin azaltılacağını vurgulamışlardır. Yöntemlerinde

ilk olarak denetimsiz algoritmada bina ayak izleri geometrik özelliklerine göre kümelenir ve daha sonra bu kümeler denetimli algoritmada sınıflandırma birimi olarak kullanılır. Denetimli algoritmada ise bina kümelerinin ortalama geometrik özellikleri ve ilgi noktası verilerinin yoğunluk değerleri ile sınıflandırıcı eğitilir. Çalışmada doğruluk değerleri Pekin ve Şanghay için sırasıyla %83,9 ve %81,8 olarak elde edilmiştir. Çalışmanın kentsel gelişim planlamaları için önemli çıkarımlar barındırdığını vurgulamışlardır. Zhong vd. (2017) anketlerden ve akıllı ulaşım kart sistemlerinden elde ettikleri ulaşım verilerinden bina fonksiyonlarını ve bu fonksiyonlardan da kentsel arazi kullanım haritalarını oluşturmuşlardır. Çalışmalarını arazi kullanımının ve ulaşım sistemlerinin çok karışık olmadığı Singapur'da gerçekleştirmişlerdir. Lin vd. (2021) kentsel ortamdaki mekânsal yapıyı keşfetmek için mekânsal benzerlikten çıkardıkları bina fonksiyonlarını ve ilgi noktası verilerinin çekirdek yoğunluğunu kullanmışlardır.

Kartografik genelleştirme de binaların özelliklerinin veri kaynağı olarak kullanıldığı önemli çalışma konularından biridir. Farklı uygulama alanları için farklı ölçek ve türlerde haritalara ihtiyaç duyulur. Kartografik genelleştirme, bu ihtiyaçlara bir çözüm olarak ortaya çıkmıştır. Bu bağlamda, aynı ya da ortak fonksiyonel tanıma sahip binalar birlikte işlem görürler. Ayrıca, fonksiyonel olarak önemli ve etrafındakilerden farklı olanların yönlendirme için referans alınma potansiyellerinden dolayı hedef ölçekte korunması amaçlanır. Genelleştirme süreçlerinde binaların şekilleri genellikle etkilenir. Başaraner ve Selçuk (2004) bina genelleştirmesi için bazı şekil göstergeleri (metrikleri) üretmişlerdir. Kompaktlık, dikdörtgensellik, dışbükeylik, uzanım, köşe sayısı, granülerlik ve doğrultu üretilen şekil göstergeleridir.

Binaların geometrik ve topolojik özelliklerinin temel veri kaynağı olarak kullanıldığı bir diğer disiplin ise kent morfolojisidir. Farklı ölçeklerdeki kent morfolojisini tanımlamak için binaların hangi özelliklerinin en yararlı olduğu bu bilim dalının temel araştırma konularından biridir. Schirmer ve Axhausen (2015), Zürih kentini 4 farklı coğrafi ölçekte ((1) bina, (2) bina etki alanı, (3) mahalle (4) belediye) morfolojik açıdan incelemişlerdir. İlk ölçekte binaların iki ve üç boyutlu öznitelik verilerini kullanmışlardır. Elde ettikleri bina tipolojilerinin konum modellerinde ve hedonik modellerde önemli bir girdi sağlayacağını vurgulamışlardır. Bir üst ölçekte binaların mekânsal yapılanışını inceleyebilmek

için binalar arası mesafeyi 300 metre olarak sınırlandırmışlardır. Kentsel alanlar, avlular ve bina etki alanları bu ölçekten elde edilebilir. İlk iki ölçekteki verilerden belirledikleri merkez noktasından itibaren yürüme mesafesindeki alanları mahalle ölçeği olarak belirtmişlerdir. Bu ölçeğin yerel yoğunluk hakkında bilgi verebileceğini vurgulayan araştırmacılar aynı zamanda sadece kullanım çeşitliliğinin değil yapı türlerinin çeşitliliğinin de mekânın önemli bir karakteristiğini oluşturacağını belirtmiştir. Son ölçekte belediyenin mevcut sınırları dikkate alınmıştır. Bu ölçekten kentlerin tek ya da çok merkezli yapısı, kullanılmayan ve doğal alan rezervleri ve kentlerin büyüme potansiyelleri hakkında bilgi elde edilebileceğini vurgulamışlardır. Yoshida vd. (2004), kentsel morfolojik özelliklerin türetilmesi için binaları mekânsal birim olarak kullanmışlardır. Öngörülen alan başına yüzey alanı (1), öngörülen alan başına hacim (2), bina alanının arazi alanına oranı (3), binaların ortalama yüksekliği (4), binaların birim hacim başına yüzey alanı (5) ve binaların ortalama hacmi (6) türettikleri özelliklerdir. 1 ve 2 numaralı özelliklerin morfoloji ve arazi kullanım arasındaki ilişkileri açıkladığını, kalan değişkenlerinde kentsel morfolojinin önemli yönlerini sergilediğini vurgulamışlardır. Çalışmalarının kentsel alanlardaki değişim süreçlerinin izlenmesine temel oluşturacağını ifade etmişlerdir.

1.2 Tezin Amacı

Kent yönetiminde önemli bir paya sahip olan acil durum ve kaynakların yönetimi, iklim değişikliği modellerinin oluşturulması, kentsel morfoloji, kentsel arazi kullanım haritalarının oluşturulması ve kartografik genelleştirme çalışmalarında bina fonksiyonları temel veri kaynağı olarak kullanılmaktadır. Bina fonksiyonlarının saha çalışmalarından elde edilmesi genellikle zaman alan ve güçlü finansal kaynaklar gerektiren süreçlerdir.

Bu tez çalışmasının temel amacı, bina fonksiyonlarını mekânsal özelliklerinden yararlanarak makine öğrenmesi yöntemleri ile daha hızlı, daha doğru ve daha ekonomik bir şekilde tespit etmek ve sınıflandırmaktır.

1.3 Hipotez

Hipotez 1 – Bina fonksiyonları, binaların mekânsal özellikleri (büyüklük, şekil, mekânsal ilişkiler ve nicel özellikler) ile ilişkili olabilir.

Hipotez 2 – Bina fonksiyonları, farklı kategorilerde incelenebilir.

Hipotez 3 – Bina fonksiyonlarının belirlenmesi için mekânsal özellikleri ölçen göstergeler (metrikler/ölçüler) kullanılarak farklı makine öğrenmesi yöntemleriyle farklı sınıflandırma doğrulukları elde edilebilir.

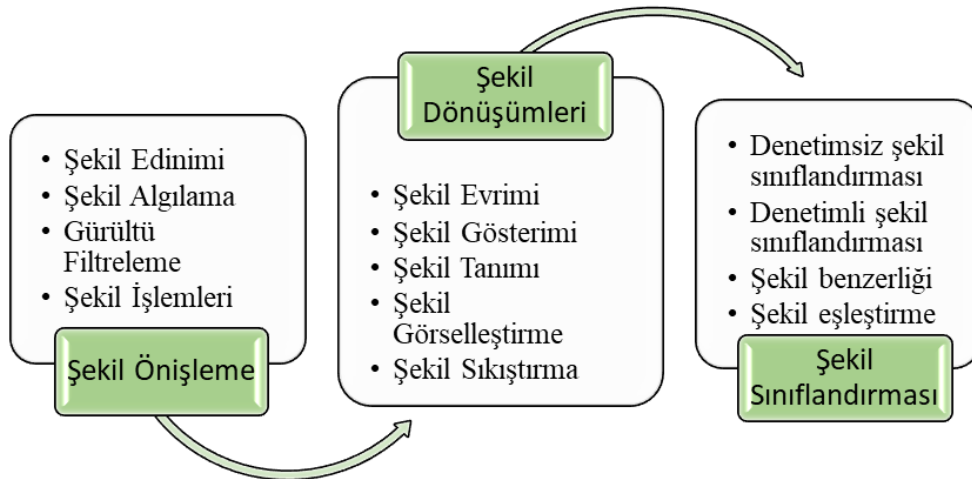


ŞEKİL ANALİZİ VE SINIFLANDIRMA

Yaşadığımız dünyayı daha iyi tanımlayabilmek ve anlayabilmek için mekânsal analizlere ihtiyaç duyulur. Mekânsal analizler, nesnelerin temel özelliklerini ve ilişkilerini araştırmak ve böylece çeşitli bilgiler elde etmek için kullanılır. Bu bilgiler insanların geçmiş deneyimleri, duyguları ve algılama yetenekleri ile yorumlanarak yararlı bilgiler haline gelir. Özellikle kentlerde binalara ilişkin bilgilerin elde edilmesi kent yönetimi için önemlidir. Bu kapsamda şekil analizleri, bina ve bina gruplarının özelliklerinin araştırılmasında kullanılabilir.

Şekil, ana hat parçaları ile gösterilen, uzamda nesne varlığı ile ortaya çıkan yapıdır. Ana hatlarının sorgulanması ile ortaya çıkar. Şekil aynı zamanda bir nesnenin etrafında var olan alanın yapısı olarak da ifade edilebilir (Leymarie, 2006). Şekiller görsel dilin kelimeleri olarak da ifade edilebilir (Da Fona Costa ve Cesar, 2018).

Şekil analizi veya şekil tanıma yöntemleri, nesnelerin şekillerini inceleyerek onlarla ilgili anlamlı bilgiler çıkarmak için kullanılan bir yöntemdir (Loncaric, 1998). Şekil analizi nesnelerin içsel özelliklerini ve ilişkilerini keşfeder (Da Fona Costa ve Cesar, 2018). Şekil analiz süreçleri Şekil 2.1’de görülmektedir.



Şekil 2.1 Şekil analiz süreçleri (Da Fona Costa ve Cesar, 2018)

2.1 Şekil Analizi

Şekil analiz süreçleri şekil önışleme adımı ile başlar. Şekil önışleme; şekil edinimi, şekil algılama, gürültü filtreleme, şekil işlemleri adımlarından oluşmaktadır. (1) Şekil edinimi, bir görüntüye ulaşmak ve onu bilgisayar tarafından işlenebilmesi için dijitalleştirmektir. (2) Şekil algılama, nesnenin görüntüdeki ilgisiz yapılardan ayrıştırılmasıdır. (3) Gürültü filtreleme, bir önceki işlem adımlarından kaynaklanan ya da görüntünün kendisinde var olan gürültülü değışikliklerin azaltılması ya da ortadan kaldırılmasıdır. (4) Şekil işlemleri; normalleştirme, etkileşimli düzenleme, şekil ekleme, kesişim gibi çok sayıda farklı yöntemler (Da Fona Costa ve Cesar, 2018).

Şekil dönüşümleri, şekilden bilgi çıkarmaktır, dolayısıyla şeklin analiz edilebilmesi için en önemli adımdır. Şekil dönüşümleri ile şekil uygun bir şekilde temsil edilir. Şekil dönüşüm yöntemleri; şekil evrimi, şekil gösterimi, şekil tanımı, şekil görselleştirme ve şekil sıkıştırmadır. (1) Şekil evrimi, zaman içerisinde şeklini değıştirmiş nesneleri analiz eder. (2) Şekil gösterimi ile bir nesnenin şekli matematiksel olarak temsil edilir. Böylece şeklin önemli özellikleri korunur ve şekil kolay işlenir. (3) Şekil tanımı birden fazla yöntem içermektedir:

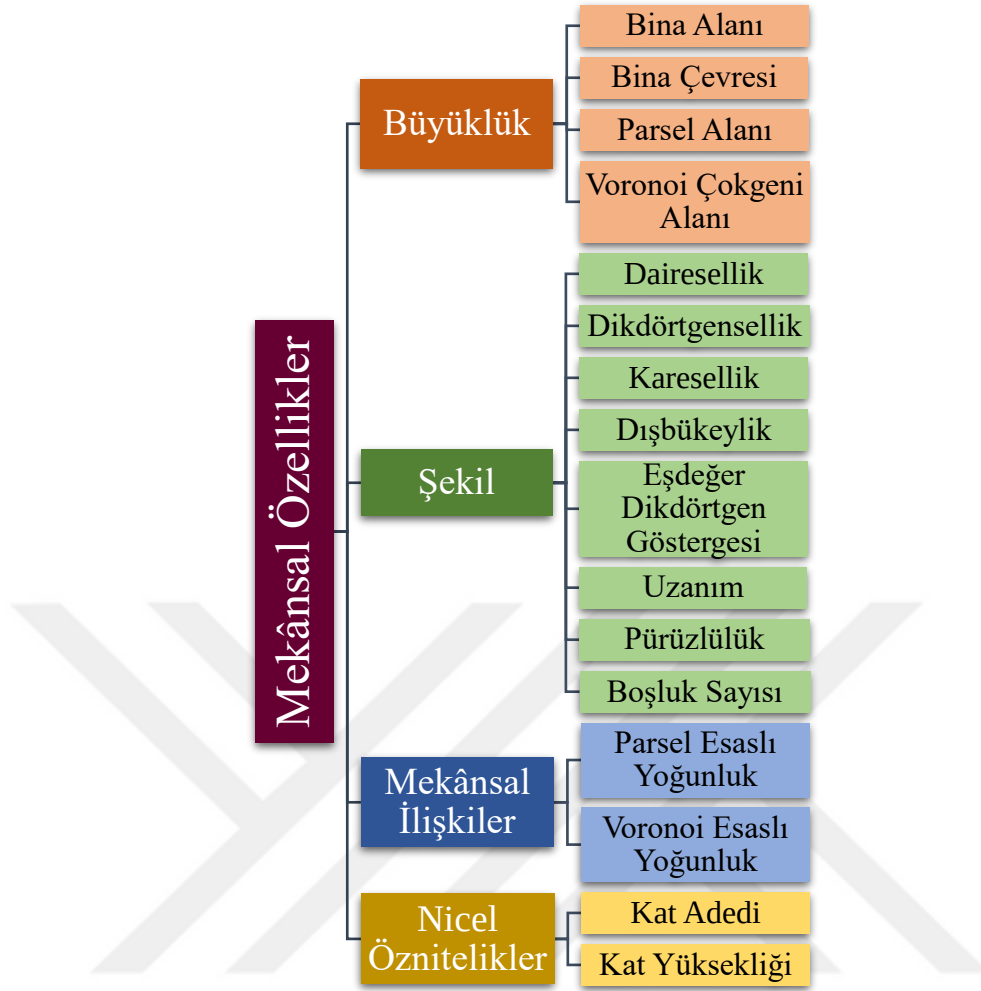
- Şekil ölçüleri (göstergeleri), şekilleri tanımlamada en yaygın kullanılan yöntemlerdendir. Şeklin sayılarla tanımlanarak anlamlı bir şekilde temsil edilmesini sağlar. Bu tez kapsamında şekil analizi sürecinde bu yöntem seçilmiştir. Tezde kullanılan şekil göstergeleri 2.2.2 numaralı başlıkta detaylıca incelenecektir.
- Şekil dönüşümleri, bir şeklin farklı bir temsile dönüştürülmesidir. Fourier en yaygın kullanılan yöntemlerden biridir.
- Şekil ayrıştırma, şekli daha basit parçalara ayırma işlemidir.
- Veri yapıları aracılığıyla şekil tanımı; ağaç yapılar, sinir hücreleri, nehir ağları gibi karmaşık şekillerin matematiksel olarak temsil edilmesidir.

(4) Şekil görselleştirme verinin insanların algılayabileceğı şekilde sunulmasıdır. Böylelikle şekil analiz araçlarının geliştirilmesi ve kontrol edilmesi sağlanır. (5) Şekil sıkıştırma ile büyük miktarda veri işlenebilir. Bu yöntem ile depolama ve hesaplama maliyetleri azaltılır. Örneğin iki boyutlu veriler tek boyutlu yapılarda temsil edilerek saklanır (Da Fona Costa ve Cesar, 2018).

Şekil sınıflandırması, şekillerin farklı kategorilere ayrıştırılması işlemidir. Şekil sınıflandırması yöntemleri; denetimsiz sınıflandırma, denetimli sınıflandırma, şekil benzerliği ve şekil eşleştirme'dir. (1) Denetimsiz şekil sınıflandırması ile önceden sınıf tanımı yapılmamış veriler sınıflandırılır, genellikle kümeleme algoritmaları tarafından gerçekleştirilir. (2) Denetimli şekil sınıflandırması önceden sınıf tanımı yapılmış ve örneklerin mevcut olduğu verilerde gerçekleştirilir. Bu tezde denetimli sınıflandırma yöntemleri kullanılacaktır. Kullanılan sınıflandırma algoritmaları 2.3.1 bölümünde ayrıntılı olarak incelenmektedir. (3) Şekil benzerliği, iki şeklin birbirine ne kadar benzediğini ölçer. Bu ölçüm şekil benzerliği kriterleri ile sağlanır. (4) Şekil eşleştirme, iki veya daha fazla şeklin nokta nokta ilişkilendirildiği işlemdir (Da Fona Costa ve Cesar, 2018).

2.2 Mekânsal Özellikler

Çalışma kapsamında binaların mekânsal özellikleri, dört bileşenle tanımlanmıştır: büyüklük, şekil, mekânsal ilişkiler ve nicel öznel özellikler. Şekil 2.2'de bu bileşenler ve bu bileşenleri ölçmekte kullanılan göstergeler (metrikler) verilmiştir. Binalar fonksiyonel olarak makine öğrenmesi algoritmalarıyla sınıflandırılırken bu göstergelerin değerleri girdi olarak kullanılmıştır.



Şekil 2.2 Çalışma kapsamında binalara yönelik olarak dikkate alınan mekânsal özellikler

2.2.1 Büyüklük Göstergeleri

Büyüklük göstergeleri; bina (çokgen) alanı, bina (çokgen) çevresi, parsel alanı ve Voronoi çokgeni alanı bileşenlerinden oluşmaktadır. Bina çevresi yalnızca şekil göstergeleri hesaplamalarında kullanılan bir göstergedir. Bina alanı ise hem şekil göstergesi hesaplamalarında kullanılmış hem de doğrudan algoritmaya bir girdi olarak sunulmuştur. Bina alanı tez kapsamında CBS programı ile Eşitlik 2.1'deki ve çokgen çevresi Eşitlik 2.2'deki gibi hesaplanmaktadır (Başaraner M. , 2020).

$$A_{CK} = \frac{1}{2} \sum_{i=1}^n (y_i x_{i+1} - x_i y_{i+1}) \quad (2.1)$$

$$C_{CK} = \sum_{i=1}^n \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \quad (2.2)$$

Bu eşitlikte;

A_{CK} : Çokgenin alanını,

C_{CK} : Çokgenin çevresini,

x_i : Çokgenin i'inci köşe noktasının x koordinatını,

y_i : Çokgenin i'inci köşe noktasının y koordinatını,

n : Köşe sayısını ifade etmektedir.

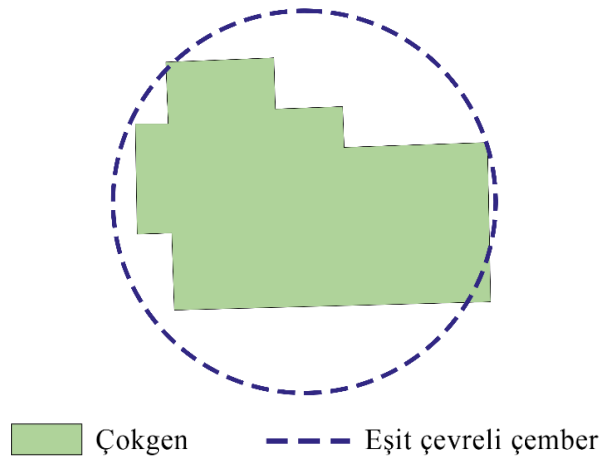
Parsel alanı indirilen veri setinden elde edilmiştir. Voronoi çokgeni alanı ise CBS programı ile oluşturulan Voronoi çokgenlerinin alanlarının hesaplanması ile elde edilmiştir.

2.2.2 Şekil Göstergeleri

Şekil göstergeleri; dairesellik, dikdörtgensellik, karesellik, dışbükeylik, eşdeğer dikdörtgen indeksi, uzanım, pürüzlülük bileşenlerinden oluşmaktadır.

2.2.2.1 Dairesellik

Dairesellik bir çokgenin şekil olarak daireye ne kadar benzediğini açıklar (Şekil 2.3). Çokgen ile eşit çevreli çemberi arasındaki alan sapmasını ölçer. Çokgenin alanı ve çevresine göre Eşitlik 2.3'teki gibi hesaplanır (Başaraner M. , 2020).



Şekil 2.3 Çokgen ve eşit çevreli çemberi

$$D = \frac{4\pi A_{CK}}{C_{CK}^2} \quad (2.3)$$

Bu eşitlikte;

D : Daireselliği,

A_{CK} : Çokgenin alanını,

C_{CK} : Çokgenin çevresini ifade etmektedir.

2.2.2.2 Dikdörtgensellik

Dikdörtgensellik (rectangularity) çokgenin alanının, çokgenin minimum alanlı sınırlayıcı dikdörtgenine oranı olarak hesaplanır (Eşitlik 2.4). Çokgenin içe doğru ne kadar girintili olduğunu hesaplar. Aynı zamanda çokgenin dikdörtgene ne kadar benzediğini ortaya koyar Şekil (2.4) (Başaraner M., 2020; Başaraner ve Çetinkaya, 2017).

$$DD = \frac{A_{CK}}{A_{MASD}} \quad (2.4)$$

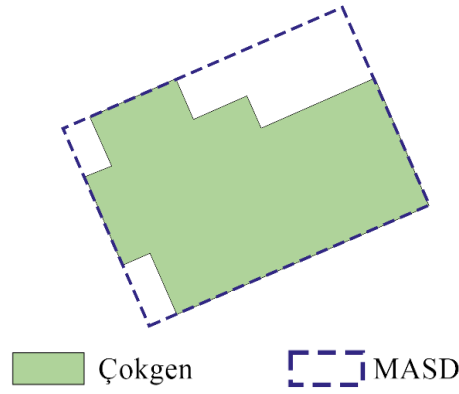
Bu eşitlikte;

DD : Dikdörtgenselliği

A_{CK} : Çokgenin alanını

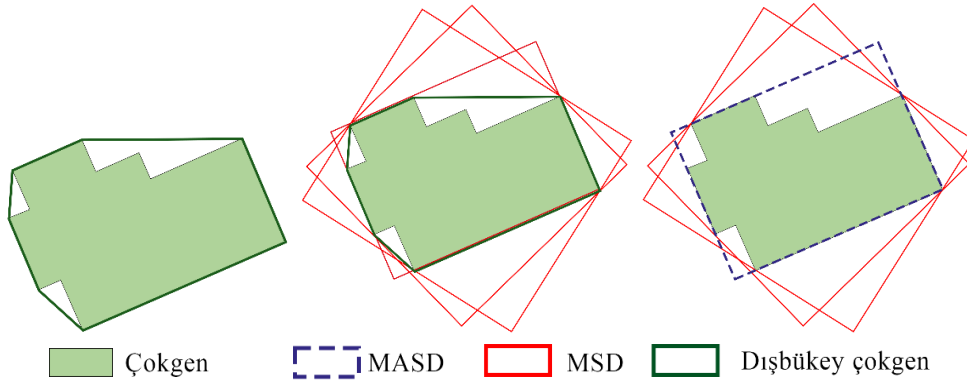
A_{MASD} : Çokgenin minimum alanlı sınırlayıcı dikdörtgeninin alanını ifade etmektedir.

Dikdörtgensellik (0, 1] aralığında değer alır. Değerin 1 olması çokgenin dikdörtgen olduğunu, 1'e yakın olması dikdörtgene benzeyen bir çokgen olduğunu gösterir. 1'den belirgin uzaklaşan değerler ise çokgenin dikdörtgenden farklı olduğunu gösterir.



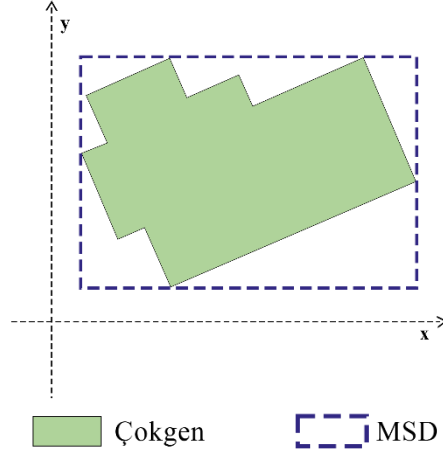
Şekil 2.4 Dikdörtgensellik

Minimum alanlı sınırlayıcı dikdörtgen (MASD) ile minimum sınırlayıcı dikdörtgen (MSD) şekil analizi için oldukça önemli ve birbirlerinden farklı kavramlardır. MASD bir şekli tam olarak kapsayan en az alana sahip olan dikdörtgendir (Şekil 2.4). MASD'nin bulunmasında birçok algoritma kullanılmaktadır. Yaygın olarak kullanılan algoritmanın ilk adımında çokgenin dışbükey çokgeni bulunur (Şekil 2.5). İkinci adımda dışbükey çokgenin minimum sınırlayıcı çokgenleri bulunur (Şekil 2.5). Freeman ve Shapira'nın teoremine göre bir çokgenin MASD'sinin dışbükey çokgeninin kenarlarından biriyle aynı doğrultuda bir kenarı vardır. Dışbükey çokgenin her bir kenarına paralel olacak MSD'ler elde edilir ve aralarından en az alana sahip olan seçilir (Şekil 2.5) (Toussaint, 1983).



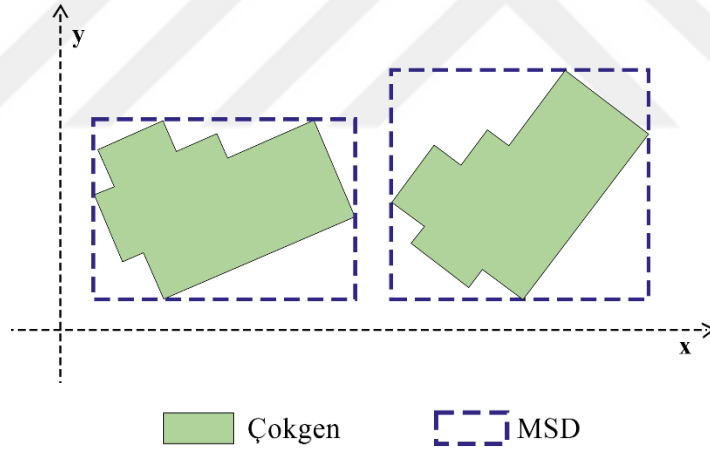
Şekil 2.5 Minimum alanlı sınırlayıcı dikdörtgenin bulunması

MASD'de kenarların koordinat eksenlerine paralel olması şart değildir. Ancak MSD'de kenarlar koordinat eksenlerine paralel olmalıdır (Şekil 2.6) (Başaraner M. , 2020).



Şekil 2.6 Minimum sınırlayıcı dikdörtgen

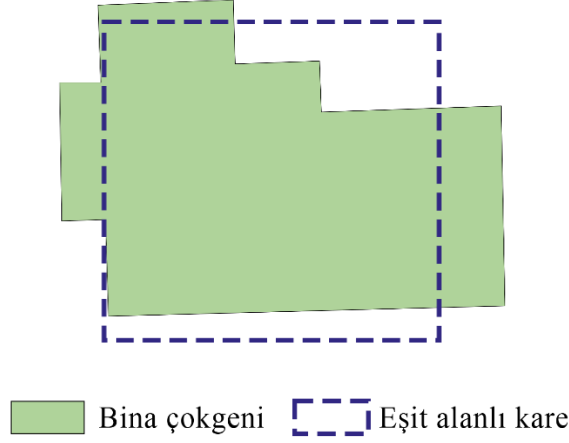
MSD'nin boyutu çokgenin doğrultusundan etkilenir. Aynı çokgen döndürüldüğünde MSD'nin boyutu değişir (Şekil 2.7). Ancak MASD'nin boyutu çokgen döndürülse de değişmez. Bunlar dikkat edilmesi gereken temel farklardır. Bu nedenle dikdörtgensellikte MASD kullanılmalıdır (Başaraner M. , 2020).



Şekil 2.7 Minimum sınırlayıcı çokgenin boyutu çokgenin yönünden etkilenir

2.2.2.3 Karesellik

Karesellik (squareness) çokgenin eşit alanlı karesinin çevresi ile çokgenin çevresinin arasındaki çevre sapmasını ölçer (Eşitlik 2.5) (Şekil 2.8) (Başaraner ve Çetinkaya, 2017).



Şekil 2.8 Karesellik

$$K = \frac{C_{EAK}}{C_{CK}} = \frac{4\sqrt{A_{CK}}}{P_{CK}} \quad (2.5)$$

Bu eşitlikte;

K : Kareselliği,

C_{EAK} : Çokgenin eşit alanlı karesinin çevresini,

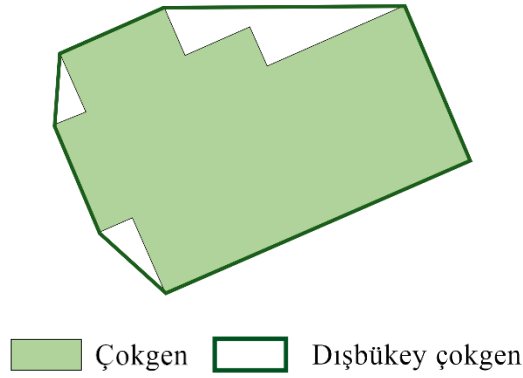
C_{CK} : Çokgenin çevresini

A_{CK} : Çokgenin alanını ifade etmektedir.

Karesellik (0,1.128] aralığında değer alır. Değerin 1 olması çokgenin kare olduğunu, 1'e yakın olması kareye benzeyen bir çokgen olduğunu gösterir. 1'den uzak değerler ise çokgenin kareden farklı olduğunu gösterir.

2.2.2.4 Dışbükeylik

Dışbükeylik (convexity), çokgenin içe veya dışa doğru ne kadar bükümlü olduğunu ölçer. Çokgenin alanı ile çokgenin minimum dışbükey çokgeni (convex hull) arasındaki alan sapmasını ölçer (Eşitlik 2.6) (Şekil 2.9) (Başaraner M. , 2020).



Şekil 2.9 Dışbükeylik

$$DB = \frac{A_{CK}}{A_{DB}} \quad (2.6)$$

Bu eşitlikte;

DB : Dışbükey çokgeni,

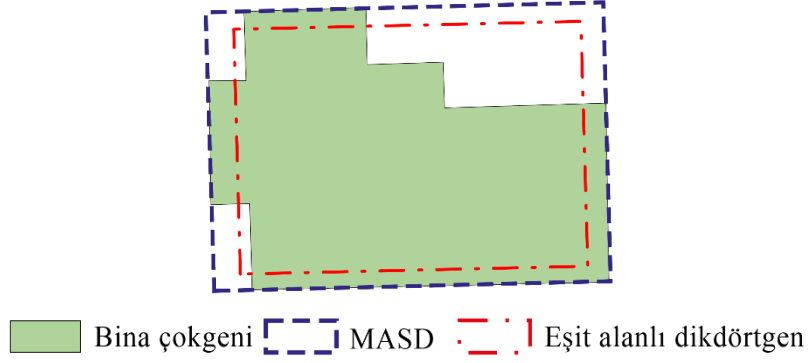
A_{CK} : Çokgenin alanını,

A_{DB} : Çokgenin dışbükey gövdesinin alanını ifade eder.

Dışbükeylik $(0,1]$ aralığında değer alır. Değerin 1 olması çokgenin dışbükey olduğunu, 1'e yakın olması dışbükeyliğe yakın olduğunu gösterir. 1'den uzak değerler ise çokgenin içbükey olduğunu gösterir.

2.2.2.5 Eşdeğer Dikdörtgen Göstergesi

Eşdeğer dikdörtgen göstergesi (equivalent rectangular index - EDG) dikdörtgenselliği iyileştirmeyi hedefler. Dikdörtgensellik göstergesi çokgenin çıkıntılarına karşı çok hassastır. Bu durum MASD'nin boyutunda önemli bir artışa neden olur ve bu da gösterge değerini etkiler. EDG, bu problemi giderir. Teorik olarak, EDG hesabı için, MASD'nin alanı çokgenin alanına eşit oluncaya kadar ölçeklenir (Şekil 2.10). Çokgenin eşit alanlı dikdörtgeninin çevresinin, çokgenin çevresine oranıyla hesaplanır (Eşitlik 2.9).



Şekil 2.10 Eşdeğer dikdörtgen göstergesi

$$k = \sqrt{\frac{C_{CK}}{A_{MASD}}} \quad (k \leq 1) \quad (2.7)$$

$$C_{EAD} = k \times C_{MASD} \quad (2.8)$$

$$EDG = \frac{C_{EAD}}{C_{CK}} = \sqrt{\frac{A_{CK}}{A_{MASD}}} \times \frac{C_{MASD}}{C_{CK}} \quad (2.9)$$

Eşitlik 2.9'u oluşturan elemanları detaylıca incelersek; eşit alanlı dikdörtgen, minimum sınırlayıcı dikdörtgenin genişlik ve uzunluğu dikkate alınarak eşit oranda ölçeklendirilir. EAD ile çokgenin alanının eşit olduğu varsayılarak k hesaplanır (Eşitlik 2.7). Elde edilen k katsayısı ve minimum sınırlayıcı alanın çevresi çarpılarak eşit alanlı dikdörtgenin çevresi hesaplanır (Eşitlik 2.8).

Bu eşitlikte;

EDG : Eşdeğer dikdörtgen göstergesi,

C_{CK} : Çokgenin çevresini,

A_{CK} : Çokgenin alanını,

A_{MASD} : MASD'nin alanını,

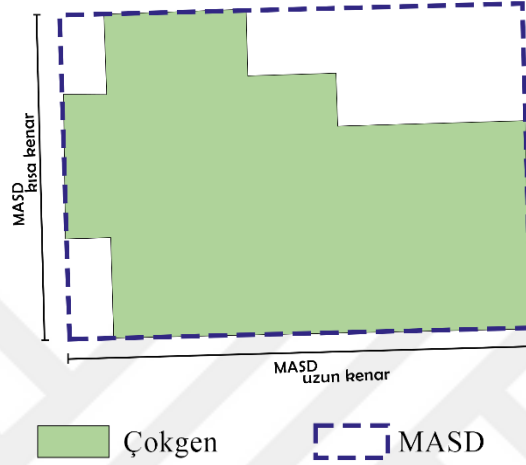
C_{MASD} : MASD'nin çevresini,

C_{EAD} : Eşit alanlı dikdörtgenin çevresini,

k : ölçekleme katsayısını ifade etmektedir.

2.2.2.6 Uzanım

Uzanım (elongation) şekil karmaşıklığı ile doğrudan ilişkili olmayan ancak şekillerin kompaktlığının belirlenmesinde yardımcı olan bir göstergedir. Uzanım göstergesini hesaplamak için çokgenin MASD'si elde edilir ve bu dikdörtgenin kısa kenarının, uzun kenarına oranı hesaplanır. Uzanım (0,1] arasında değer alır (Eşitlik 2.10) (Şekil 2.11) (Başaraner M. , 2020).



Şekil 2.11 Uzanım

$$U = \frac{MASD_K}{MASD_U} \quad (2.10)$$

Bu eşitlikte;

U : Uzanımı,

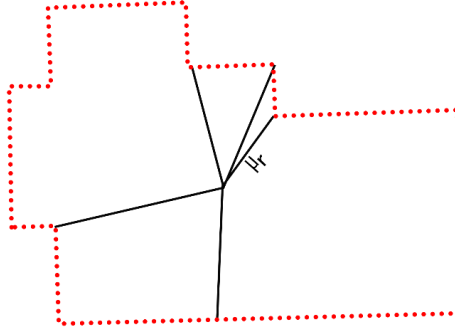
MASD_K : MASD'nin kısa kenar uzunluğunu,

MASD_U : MASD'nin uzun kenar uzunluğunu ifade etmektedir.

2.2.2.7 Pürüzlülük

Pürüzlülük (roughness), bir tür kompaktlık ölçüsüdür. Alanın çevreye oranına dayalı diğer kompaktlık ölçülerine göre uzanıma daha az duyarlı ve girinti-çıkıntılara daha duyarlıdır. Eşitlik 2.11'de görüldüğü üzere radyal çizgilerin ortalama uzunluğu, A_{CK} ve C_{CK} parametreleri kullanılarak Eşitlik 2.11 ile hesaplanır. Bu işlem öncesinde, bina kenar çizgisine sıklaştırma işlemiyle 300 nokta eklenmesi önerilmiştir. Herhangi bir katsayı kullanmadan daire için P değeri

1/42.62'ye eşit olur. Daire için 1 değerini elde etmek için oran 42.62 ile çarpılır (Şekil 2.12) (Başaraner ve Çetinkaya, 2017).



Şekil 2.12 Pürüzlülük

$$P = \frac{\mu_r^2}{A_{CK} + C_{CK}^2} \times (\pi + 4\pi^2) = \frac{\mu_r^2}{A_{CK} + C_{CK}^2} \times 42.62 \quad (2.11)$$

Bu eşitlikte;

P : Pürüzlülüğü,

A_{CK} : Çokgenin alanını,

C_{CK} : Çokgenin çevresini,

μ_r : Çokgenin ağırlık merkezinden sınır boyunca sıklaştırılan noktalarına ulaşan ışınların ortalama uzunluğunu ifade eder.

2.2.3 Mekânsal İlişkiler

Mekânsal ilişkiler; parsel esaslı yoğunluk ve Voronoi esaslı yoğunluk bileşenlerinden oluşmaktadır. Yoğunluk, bir binanın biçimini betimlemek ve fonksiyonlarını belirlemek için kullanılır (Kong vd., 2024). Çalışmada, bina ayak izinin yoğunluğu iki farklı yöntemle hesaplanmıştır. İlk olarak, bina (ayak izi) alanının parsel alanına oranı hesaplanarak yoğunluk değeri elde edilmiştir. İkinci yöntemde Voronoi diyagramlarından yararlanılmıştır. Voronoi diyagramları binaların mekânsal dağılım ve etkileşimlerini analiz etmede kullanılır. Bina (ayak izi) alanının, Voronoi çokgeni alanına oranı hesaplanarak, binaların gerçek kapladığı alan ile etki alanları arasındaki ilişki incelenmiştir.

2.2.4 Nicel Öznitelikler

Nicel öznitelikler; kat adedi ve kat yüksekliği bileşenlerinden oluşmaktadır. Kat adedi ve kat yüksekliği indirilen veri setinden elde edilmiştir.

2.3 Sınıflandırma

Sınıflandırma, girdi özelliklerinin bir veya daha fazla sınıfa atanması işlemidir. Bilgi temelli yaklaşımlar (tümdengelimsel) ve veri temelli yaklaşımlar (tümevarımsal) olarak iki ana başlık altında incelenebilir. Tablo 2.1’de her iki yaklaşım arasındaki farklar görülmektedir. Bilgi temelli yaklaşımların sürekli gelişim ve değişim sürecinde olan alanlarda uygulanması zor olabilir. Veri temelli yaklaşım bu noktada önem kazanmaktadır. Bu yaklaşımda modellemeler makine öğrenmesi algoritmaları tarafından otomatik olarak yapılır (Hecht vd., 2015). Bu tezde sınıflandırma için veri temelli yaklaşımlardan (makine öğrenmesi algoritmaları) yararlanılacaktır.

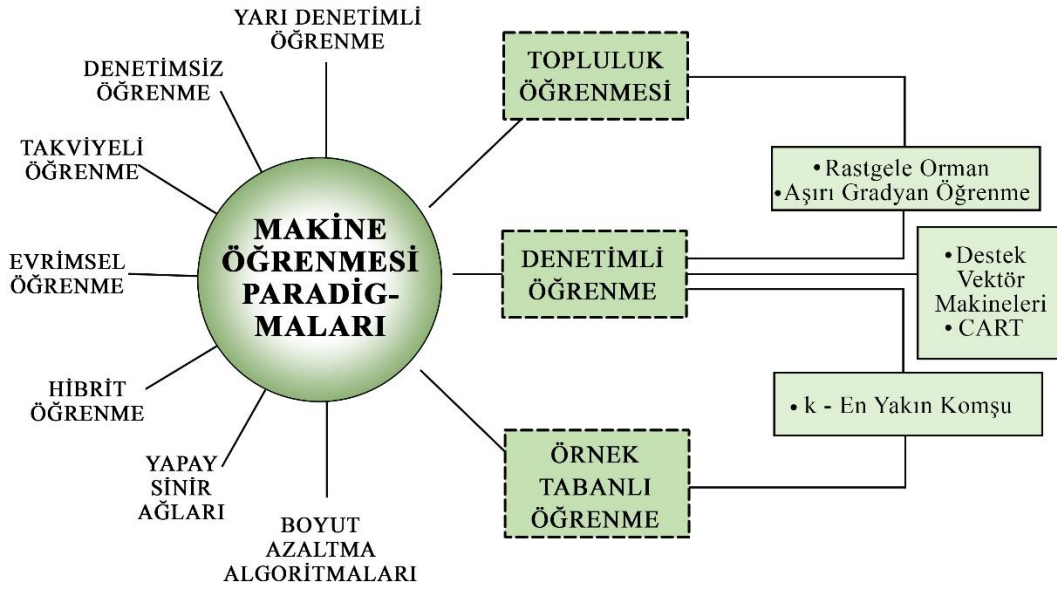
Tablo 2.1 Sınıflandırma yaklaşımları (Hecht vd., 2015)

<i>Özellik</i>	<i>Bilgi Temelli Yaklaşım</i>	<i>Veri Temelli Yaklaşım</i>
Model Oluşturma	Uzmanlar tarafından oluşturulan kılavuzlar, kural kümeleri ve açıklamalar	Otomatik öğrenme algoritmaları ile oluşturulan modeller
Algoritma	Belirli, önceden tanımlanmış algoritmalar	Veriye dayalı olarak öğrenen, esnek algoritmalar
Hata Payı	Düşük hata payı (Eğer kurallar doğruysa)	Veri kalitesine ve algoritma seçimine bağlı olarak değişken hata payı
Veri Boyutu	Az miktarda veri	Büyük miktarda veri
Esneklik	Düşük esneklik, sistemin yapısı sabit	Yüksek esneklik, yeni verilere ve durumlara uyum sağlama
Uygulama Alanları	Net kuralların olduğu, uzman bilgisinin önemli olduğu alanlar	Büyük veri analitiği, tahminleme, sınıflandırma gibi alanlar

2.3.1 Sınıflandırma Algoritmaları

Başlangıçta bilgisayar sistemleri var olan kurallar üzerinden başarılı bir şekilde işlemler gerçekleştirebilmekteydi. Ancak sistemler anlamsal akıl yürütmekten uzaktı. E. A. Feigenbaum önderliğinde bir grup bilim insanı bu problem üzerinde çeşitli çalışmalar gerçekleştirdiler. Akıllı olabilmeleri için makinelerin bilgi edinme yeteneklerinin olması gerektiğini savundular. Böylelikle yapay zeka araştırmaları 1970’li yılların ortalarında başlamış oldu. Aslında makinelerin öğrenme olasılığına Turing 1950’li yıllarda yaptığı çalışmalar ile çoktan değinmişti. Arthur Samuel tarafından 1950’li yıllarda geliştirilen dama oyununun oynanarak zamanla hamle geliştirmeye başlaması gibi adımlarında makine öğrenmesinin temelini oluşturduğu düşünülmektedir. 1960 – 1970 yılları karar teorileri ve desteğe dayalı öğrenme sistemlerinin gelişmeye başladığı dönemlerdir. 1980’li yıllara gelindiğinde mantık tabanlı öğrenme ve karar ağaçları tarafından temsil edilen sembolik öğrenme, örneklerden öğrenmenin ana akımı olmuştur. Karar ağaçları günümüzde araştırmacılar tarafından birçok çalışmada yer verilen ve bu tez kapsamında da değinilecek olan ağaç tabanlı bir yaklaşımdır. İnsanların bilgiyi en aza indirgemesi olarak ifade edilebilir. Karar ağaçları 2.3.1.1’de ayrıntılı olarak ele alınmaktadır (Zhou, 2021).

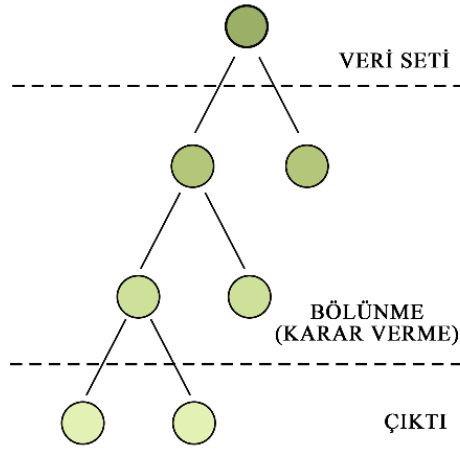
Makine öğrenmesi algoritmalarının sayısı bu alandaki gelişmelerle birlikte sürekli artmaktadır. Algoritmanın nasıl eğitildiği ve eğitim sonundaki çıktının kullanılabilirliğine bağlı olarak, makine öğrenmesi algoritmaları 10 paradigma altında incelenebilir (Şekil 2.13). Bu tez kapsamında topluluk öğrenmesi, denetimli öğrenme ve örnek tabanlı öğrenme paradigmalarının rastgele orman, aşırı gradyan öğrenme, destek vektör makineleri, CART (Classification and Regression Trees) ve k-en yakın komşu sınıflandırma algoritmaları incelenecektir (Alzubi vd., 2018).



Şekil 2.13 Makine öğrenmesi paradigmaları

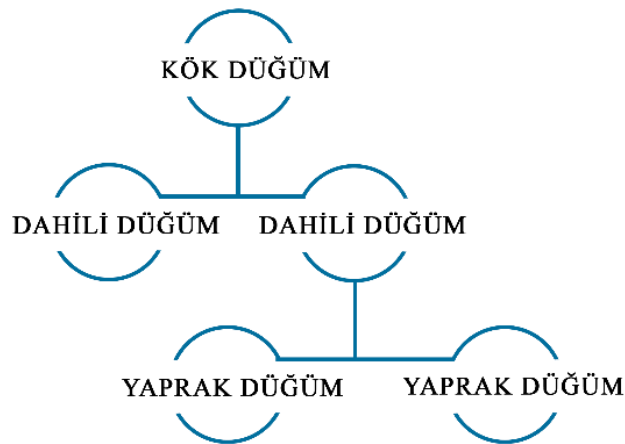
2.3.1.1 Karar Ağaçları

Karar ağaçlarının temeli Clopper ve Pearson'ın 20.yüzyılın ortalarındaki çalışmalarına dayanmaktadır. Araştırmacılar o dönemde ikili karar süreçleri kavramını ortaya atmışlardır. Ancak modern uygulamalar ilk olarak 1984 yılında Breiman tarafından geliştirilmiştir. Breiman veri kümesini ikili bir şekilde bölerek karar ağaçları oluşturan CART algoritmasını geliştirmiştir (Şekil 2.14). Bu algoritmayla birlikte Gini katsayısı ve ikili bölme gibi kavramları da tanıtmıştır. Quinlan 1986 yılında verileri alt kümelere ayırmak için bilgi kazancını kullanan ID3 (Iterative Dichotomiser-3) algoritmasını geliştirmiştir. Quinlan 1993 yılında ID3 algoritmasının geliştirilmiş bir versiyonu olan C4.5'i sunmuştur. Yeni algoritma gürültülü verilere karşı daha dayanıklı ve sayısal verilerle de çalışabilmektedir. Ayrıca aşırı uyumu önlemek için çeşitli budama tekniklerini kullanmaktadır (Mienye ve Jere, 2024).



Şekil 2.14 Karar ağacı yapısı

Temelde karar ağacı algoritmalarının çalışma prensipleri oldukça benzerdir. Karar ağaçları tüm veri kümesini temsil eden kök düğümlere sahiptirler (Şekil 2.15). Kök düğümlerin gelen kenarları yoktur. Diğer tüm düğümlerin gelen kenarları vardır. Gelen ve giden kenarları olan düğümler “dahili” veya “test” düğümü olarak adlandırılır (Şekil 2.15). Kök ve dahili düğümlerde bölünme belirli bir kritere göre gerçekleşir (Şekil 2.15). Bu aşamada en iyi bölünmeye yol açan eşik ve özellik bölme kriterleri tarafından belirlenir. Bölme kriterinin belirlenmesinde veri setinin (düğümün) özellikleri, problemin türü ve modelin amacı önemli rol oynamaktadır. Bölme kriterlerinin doğru seçilmesi modelin nihai performansını önemli ölçüde etkilemektedir. Farklı karar ağaçları farklı bölme kriterlerini kullanabilir. Burada temel amaç kök ya da dahili düğümün en anlamlı şekilde bölünebilmesini sağlamaktır (Rokach ve Maimon, 2013; Mienye ve Jere, 2024).



Şekil 2.15 Karar ağacı düğümleri

Popüler bölünme kriterleri, “Gini Katsayısı (Gini Impurity)”, “Bilgi Kazancı (Information Gain)”, “Bilgi Kazanç Oranı (Information Gain Ratio)” ve “Ki-Kare Testi (Chi-Square Test)” dir.

Gini katsayısı daha önce de değinildiği üzere Breiman tarafından CART algoritması için geliştirilen bir bölünme kriteridir. Kaynaklarda “Gini Kirliliği” ve “Gini İndeksi” olarak da ifade edilmektedir. Gini katsayısı, her düğümün saflık derecesini ya da başka bir ifade ile yanlış sınıflandırılma olasılığını ölçer. Bir düğümdeki tüm veriler tek bir sınıfa ait ise o sınıfın saflık derecesi artar, yanlış sınıflandırılma olasılığı azalır ($Gini(S) = 0$) ve düğüm “saf düğüm” olarak adlandırılır. Saf düğümde tekrar bir bölünme gerçekleştirilmez. Düğümdeki tüm veriler farklı sınıflara ait ise bu düğümün saflık derecesi azalır ve yanlış sınıflandırılma olasılığı artar ($Gini(S) = 1$) ve düğüm “kirli düğüm” olarak adlandırılır. Bu düğümde gini katsayısı en yüksek olan sınıf seçilir ve bölünme devam eder. Gini katsayısı formülü Eşitlik 2.12’de görülmektedir. Değer (0,1] aralığındadır. Değerin 0 çıkması düğümün mükemmel saflık derecesinde olduğunu, 1 çıkması ise maksimum kirlilik derecesinde olduğunu gösterir (Mienye ve Jere, 2024).

$$Gini(S) = 1 - \sum_{i=1}^n p_i^2 \quad (2.12)$$

Bu eşitlikte;

S : Örnek düğümü,

n : Benzersiz sınıf sayısını,

p_i : i sınıfına ait olan örneklerin oranını ifade eder.

Bilgi kazancı Quinlan tarafından ID3 ve C4.5 algoritmaları için geliştirilen bölünme kriteridir. Temeli bilgi teorisindeki entropi kavramına dayanmaktadır. Veri kümesindeki (düğüm) belirsizlik, öngörülemezlik ya da rastgelelik entropi ile ölçülür. Entropi değeri (0-1] aralığındadır. Düğümdeki veriler ne kadar homojen ise ya da başka bir ifade ile veriler tek bir sınıfa ait ise entropi düşüktür. Düğümdeki veriler heterojen bir yapıda yani farklı sınıflara ait ise entropi yüksektir. Bir düğümün entropisi şu şekilde hesaplanır (Eşitlik 2.13) (Mienye ve Jere, 2024).

$$Entropi(S) = - \sum_{i=1}^n p_i \log_2(p_i) \quad (2.13)$$

Bu eşitlikte;

S : Örnek düğümü,

n : Benzersiz sınıf sayısını,

p_i : Örneklerin i sınıfına oranını ifade eder.

Bilgi kazancı ise bölünme sonucunda elde edilen alt düğümlerin ana düğümden ne kadar farklı olduğunu ölçer. Eşitlik 2.14'te de görüldüğü üzere her bir özellik için alt düğümler oluşturulur ve bölünme sonucunda oluşan bu düğümlerin ne kadar bilgi sağladığı bilgi kazancı ile ölçülür. Bilgi kazancı (0,1] aralığındadır. Bilgi kazancının 0 olması durumunda ana düğümden herhangi bir özelliğe göre oluşturulan alt düğümler bilgi kazandırmazlar. Bilgi kazancının 1 olduğu durumda ise bir özelliğe göre yapılan bölünme ile tüm alt düğümlerin homojen hale geldiğini ifade etmektedir.

$$Bilgi\ Kazancı(S) = \sum_{v \in Değerler(A)} \frac{|S_v|}{|S|} E(S_v) \quad (2.14)$$

Bu eşitlikte;

S : Örnek düğümü,

A : Düğümün özelliklerini,

$|S|$: Düğümdeki örnek sayısını,

$|S_v|$: Özellik A'ya göre bölünme sonucu oluşan v. alt düğümdeki örnek sayısını ifade eder.

Bilgi kazanç oranı temelleri bilgi kazancı ve entropiye dayanan bir bölme kriteridir (Eşitlik 2.15). Quinlan'ın C4.5 algoritması için geliştirdiği kriterdir. Ana kümenin birden fazla özelliğe göre alt kümeler ayrıldığı durumlarda bilgi kazanç oranı kriteri en anlamlı bölünmeyi sağlayan özelliği seçer. Bu seçim sonucu bölünme bir özellik üzerinden gerçekleşir (Mienye ve Jere, 2024).

$$Bilgi\ Kazanç\ Oranı(S, A) = \frac{Bilgi\ Kazancı\ (S, A)}{Bölünmüş\ Bilgi(S, A)} \quad (2.15)$$

Ki-Kare (X^2) bölme kriteri veri setindeki sınıf etiketlerinin aynı veri setindeki mevcut öznitelikler ile olan ilişkisini inceler (Eşitlik 2.16). Örneğin, cinsiyetin (öznitelik) öğrencilerin başarısı (sınıf etiketi) üzerindeki etkisinin araştırıldığı bir çalışmada ki-kare testi kız öğrencilerin başarı oranının erkek öğrencilerin başarı oranından anlamlı derecede farklı olup olmadığını söyler (Mienye ve Jere, 2024).

$$X^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (2.16)$$

Bu eşitlikte;

r : A özniteliğinin kategori sayısını,

k : Düğüm sayısını,

O_{ij} : Hücre (i,j) de gözlenen frekansı,

E_{ij} : Hücre (i,j)'deki bağımsızlık sıfır hipotezi altında beklenen frekansı ifade eder.

E_{ij} Eşitlik 2.17'de görüldüğü gibi hesaplanmaktadır.

$$E_{ij} = \frac{(satur_toplami \times sutun_toplami)}{toplami \text{ örneklem}} \quad (2.17)$$

Ki-kare değerinin 0 çıkması iki değişken arasında hiçbir ilişki olmadığını, tamamen bağımsız olduklarını gösterir. Yüksek bir değer çıkması ise iki değişkenin arasında bir ilişkinin olduğunu, bağımlı olduklarını gösterir.

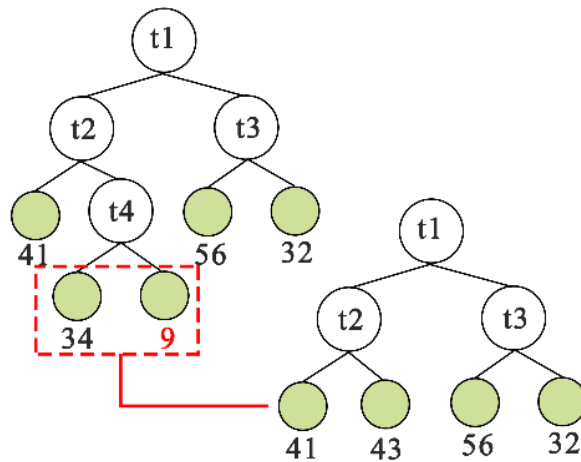
Karar ağaçlarında herhangi bir durdurma kriteri olmaz ise bölünme işlemleri sonsuza kadar devam edebilir. Bu durum aşırı öğrenmeye (overfitting) yol açar ve modelin performansını oldukça düşürür. Bu durumun önüne geçebilmek için çeşitli budama (pruning criteria) ve durdurma kriterleri (stopping criteria) geliştirilmiştir. Budama veya durdurma ile giden kenarları olmayan ve sadece gelen kenarı olan düğümler “yapraklar” ve “karar” düğümleri olarak adlandırılmaktadır (Şekil 2.15). Bu iki kriter (durdurma ve budama) temelde birbirinden farklı olsa da birbirini tamamlayan yapılar oldukları için birlikte incelenecektir.

Bratko ve Bohanec budamayı “doğruluktan basitliğe geçiş” olarak ifade etmişlerdir. Çeşitli çalışmalar budamayla birlikte özellikle gürültülü alanlarda genel performansın iyileştirilebileceğini göstermiştir (Rokach ve Maimon, 2013). Budama kriterleri 2 ana başlık altında incelenebilir; ön budama (pre-pruning) ve son budama (post-pruning).

Ön budama farklı kaynaklarda ileri budama, çevrimiçi budama ya da erken durdurma olarak da ifade edilebilmektedir (Mienye ve Jere, 2024; Patel ve Upadhyay, 2012). Ön budamada ağaç oluşturulurken bazı dallar erkenden sonlandırılır, önemsiz dalların oluşumu engellenir. Ön budama kriterleri çoğu zaman durdurma kriterleri olarak da kullanılmaktadır. Her iki kriterde de ağacın büyümesi şu durumlarda durdurulur:

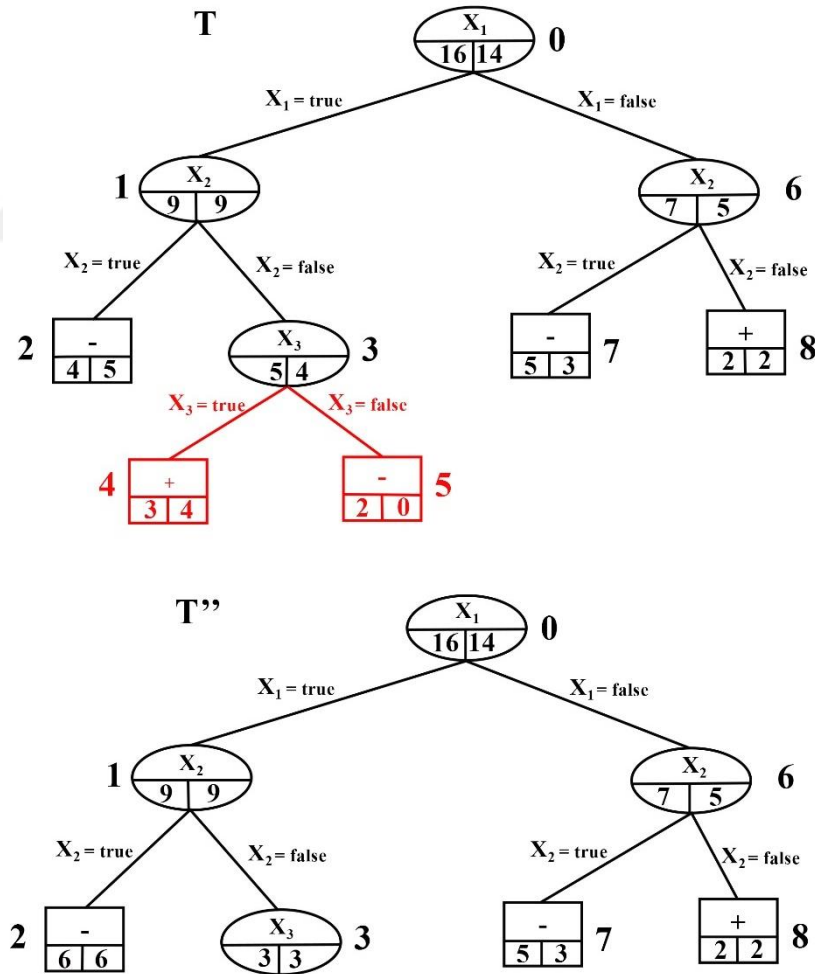
- Önceden belirlenmiş maksimum derinliğe ulaşıldığında,
- Yapraklardaki örnek sayısı belirli bir eşiğin altına düştüğünde (minimum nesne budaması tekniği ile)
- Yeni bölünme ile elde edilen bilgi kazancının belirli bir eşik değerinin altında olduğunda (Mienye ve Jere, 2024).

Şekil 2.16’da minimum nesne budaması tekniği gösterilmektedir. Karar ağacında minimum nesne sayısı 30 olarak belirlenmiştir. t4 düğümü bölündüğünde oluşan yeni düğümler 34 ve 9 nesneye sahiptir. 9 nesne önceden belirlenen eşiğin altında ($9 < 30$) olduğu için minimum nesne budaması tekniği ile t4 düğümünün bölünmesini durdurmuştur (Patel ve Upadhyay, 2012).



Şekil 2.16 Minimum nesne budaması tekniği

Son budama, bazı kaynaklarda geriye doğru budama olarak da geçmektedir. Bu teknikte ağacın büyümesini durduran ön budamanın aksine ağacın tamamen büyümesi beklenir, sonra geriye dönülerek budama işlemleri gerçekleştirilir. Popüler son budama teknikleri: azaltılmış hata budaması, hata tabanlı budama, minimum hata budaması, maliyet tabanlı budamadır (Esposito vd.,1997; Patel ve Upadhyay, 2012). Şekil 2.17’de hata tabanlı budama tekniği son budamaya örnek olarak gösterilmektedir. Hata tabanlı budama alt dallardan kök düğüme geçiş stratejisi ile düğümleri ziyaret eder. Bu tekniğin budamanın yanı sıra aşılama yaptığı söylenebilir. İlk adımda budama işlemi gerçekleştirilir (T’). Kırmızı ile işaretlenen 4. ve 5. dallar kaldırılır. Sonrasında ana düğüme aşılama yapılır (T’'). Her bir düğümün hata oranını ve ağacın genel hata oranını hesaplar. Kaldırıldığı zaman ağacın genel hata oranını büyük oranda arttırmayacak olan düğüm budanır (Esposito vd., 1997).

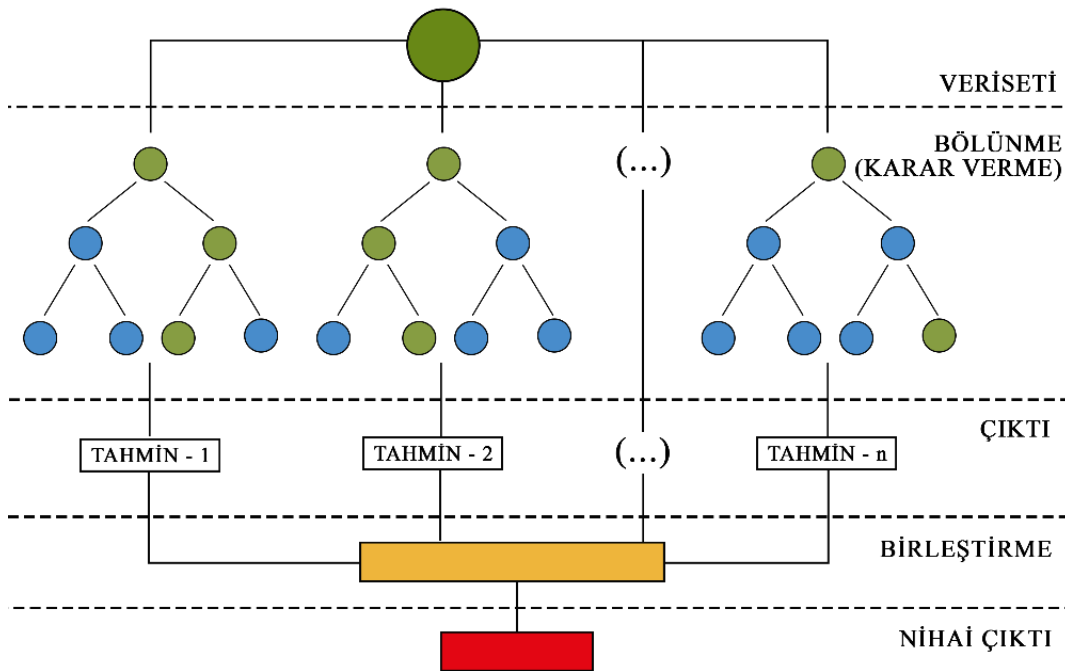


Şekil 2.17 Hata tabanlı budama tekniği (Esposito vd., 1997)

Budama veya durdurma işlemleri bireysel öğrenme algoritmalarını nihai sonuca ulaştırmaktadır. Ancak toplu öğrenme algoritmalarında nihai sonuca birden fazla karar ağacının birleştirilmesi ile ulaşılmaktadır. 2.3.1.2 bölümünde birleştirme işlemlerine detaylıca değinilecektir.

2.3.1.2 Topluluk Öğrenmesi Algoritmaları

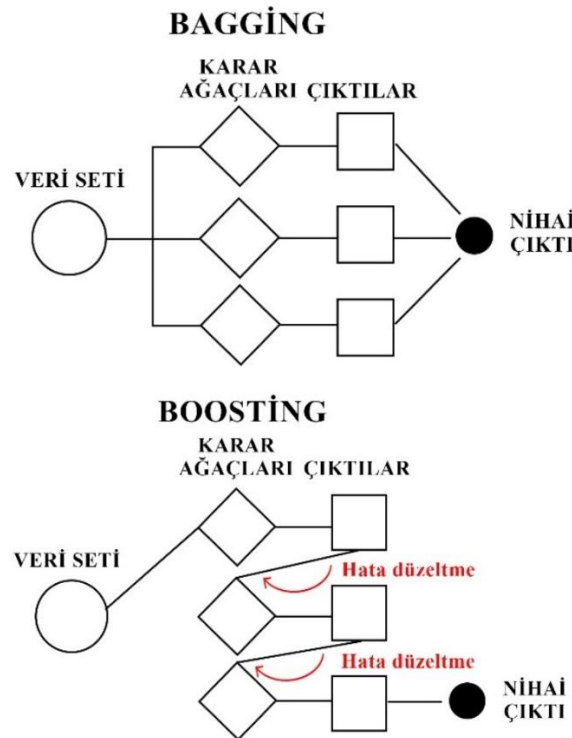
CART, ID3 ve C4.5 algoritmaları karar ağaçlarının temelini oluşturan bireysel öğrenici (single learners) algoritmalarıdır. Veri kümeleri yüksek boyutlu ve dengesiz olduğunda öğrenme bu algoritmalar için bazen zorlayıcı olabilmektedir. Araştırmalar algoritmaların bu durumda düşük performans gösterme eğiliminde olduğunu göstermiştir. Bu nedenle yeni ve geliştirilmiş karar ağacı modellerine ihtiyaç duyulmuştur. Bu sorunlara çözüm olarak “topluluk öğrenmesi algoritmaları” (ensemble learning) geliştirilmiştir. Topluluk öğrenmesi algoritmaları birden fazla karar ağacının birleştirilmesi ile oluşur (Şekil 2.18). Algoritmada tek bir modelden gelen tahminler yerine birden fazla modelden gelen tahminler birleştirilerek değerlendirilir. Maliyet açısından incelendiğinde her iki öğrenme algoritması arasında önemli bir fark yoktur. Bireysel öğrenmede parametrelerin ve modelin belirlenmesi süreci toplu öğrenme algoritmalarında da bulunmaktadır. Toplu öğrenme farklı olarak birleştirme adımını içerir. Bunun maliyeti genellikle çok küçüktür (Zhou, 2012).



Şekil 2.18 Topluluk öğrenmesi algoritması yapısı

Toplu öğrenmenin temelini hangi araştırmacının hangi çalışması ile ortaya attığı tam olarak bilinmemektedir. Ancak çoğu çalışma Dasarathy ve Sheela'nın 1979 yılında yayınladıkları makalelerinde toplu öğrenmeye atıfta bulunduklarını ve temellerinin bu çalışmaya dayandığını savunmaktadır (Mienye ve Sun, 2022). 1990'lı yıllara gelindiğinde araştırmalar hız kazanmıştır. Hanson ve Salamon, sınıflandırıcıların birleştirilmesi ile yapılan tahminlerin tek sınıflandırıcı tahminlerinden (en başarılı olan tek sınıflandırıcı olsa da) genellikle daha üstün performans sağladığını göstermişlerdir (Zhou, 2012). Aynı yıllarda Schapire, zayıf sınıflandırıcıların birleşerek güçlü sınıflandırıcıları oluşturmasını sağlayan “güçlendirme tekniğini” (boosting method) önermiştir. Bu teknik günümüzde çalışmalarda oldukça sık kullanılan ve bu tezde de yer verilecek olan aşırı gradyan artırma (XGBoost) gibi tekniklerin gelişmesinde önemli rol oynamıştır.

Birleştirmenin temel hedefi karar ağacının varyansını azaltmaktır. Varyans modelin farklı eğitim kümelerinde ne kadar farklı sonuçlar verdiğinin bir ölçüsüdür. Varyansın yüksek olması modelin farklı eğitim kümelerinde eğitildiğinde çok farklı tahminler üreteceğini göstermektedir. Aynı zamanda yüksek varyans modelin aşırı uyum (overfitting) sorunu ile karşı karşıya olduğunu gösterir. Bu aşamada karar ağacı tarafından üretilen farklı modeller “torbalama (bagging)” ya da “güçlendirme (boosting)” yöntemleri ile birleştirilir ve nihai karara ulaşılır (Şekil 2.19).



Şekil 2.19 Torbalama (bagging) ve güçlendirme (boosting)

Torbalama 1996 yılında Breiman tarafından geliştirilen bir birleştirme yöntemidir. Önyükleme toplama olarak da bilinmektedir. Torbalama İngilizce’de önyükleme anlamına gelen “bootstrapping” ve toplama anlamına gelen “aggregating” kelimelerinin birleşiminin kısaltmasıdır. Torbalamada ilk olarak veri setinden rastgele örnekler alınır (eğitim veri seti). Her bir örneklem üzerinden çeşitli tahmin modelleri (karar ağaçları) oluşturulur. Nihai sonuca elde edilen tahmin modellerinin çoğunluğunun oyu ile varılır (Şekil 2.18) (Eşitlik 2.18). Bu yöntem yüksek boyutlu verilerde iyi performans gösterir (Mohammed ve Kora, 2023).

$$f(x) = \frac{1}{B} \sum_{B=1}^B f_b(x) \quad (2.18)$$

Bu eşitlikte;

$F(x)$: Nihai sonucu,

B : Toplam model sayısını,

$f_b(x)$: b. modelin tahminini ifade eder. Bu tezde de kullanılacak olan rastgele orman algoritması (random forest algorithm) torbalama yönteminin kullanıldığı en popüler algoritmadır.

Güçlendirme yöntemi Schapire tarafından 1997 yılında geliştirilmiştir. Bu yöntemde her bir model bir önceki modelin hatalarını düzeltmeye çalışır. Modeller ardışık bir şekilde oluşturulur ve her model bir önceki modelin hatasının düzeltilmiş halidir (Şekil 2.19). Her adımda hataların düzeltilmesi ile güçlendirme oluşturulması hesaplamayı maliyetli hale getirmektedir. Güçlendirme sayısı arttıkça model aşırı uyuma karşı savunmasız hale gelmektedir. Güçlendirme hesabı Eşitlik 2.19’da görülmektedir (Mohammed ve Kora, 2023).

$$f(x) = \sum_t \alpha_t h_t(x) \quad (2.19)$$

Bu eşitlikte;

$f(x)$: Nihai sonucu,

h_t : h. modelin tahminini,

α_t : t. modelin ağırlığını ifade eder. α_t Eşitlik 2.20’de görüldüğü gibi hesaplanmaktadır. ϵ_t t. modelin hata oranını ifade etmektedir.

$$\alpha_t = 0.5 * \ln \left(\frac{(1 - \epsilon_t)}{\epsilon_t} \right) \quad (2.20)$$

Yanlış sınıflandırılmış olan modellerin hata oranı artar. Böylece bir sonraki modelin hatalı örneklerle odaklanması sağlanır. Modelin hata oranı (ϵ_t) ne kadar yüksekse ağırlığı da o kadar düşüktür (Kulkarni ve Kelkar, 2014).

Güçlendirmenin sık kullanıldığı algoritmalar şunlardır: “Uyarlanabilir Güçlendirme (Adaptive Boosting - **AdaBoost**)”, “Rastgele Gradyan Güçlendirme (Stochastic Gradient Boosting - **SGB**)” ve “Aşırı Gradyan Güçlendirme (Extreme Gradient Boosting – **XGBoost**)”.

2.3.1.3 Destek Vektör Makineleri

Destek Vektör Makineleri (Support Vector Machine – DVM) Vapnik tarafından geliştirilmiştir. DVM ile doğrusal olmayan problemler ile seyrek örnekleme ve küresel optimum problemleri ve yüksek boyutlu veri problemleri çözülebilir. Başlangıçta yalnızca ikili sınıflandırma problemlerinde kullanılan algoritma çeşitli geliştirmelerle birlikte çoklu sınıflandırma problemlerinde de kullanılabilir. Genellikle grafikte görselleştirilen bu algoritmada her veri ögesi n boyutlu uzayda bir nokta olarak ifade edilir. n, veri setinde bulunan özelliklerin sayısını ifade etmektedir. Her özelliğin değeri, karşılık gelen koordinatın değeridir. Algoritmada sınıflar bir hiper düzlem ile ayrılır. Hedef, sınıflar arasındaki marjı maksimum hale getirmektir (Şekil 2.20) (Alzubi vd., 2018; Haitao vd., 2007).

$$y_i[(w \cdot x_i) + b] - 1 \geq 0 \quad \forall_i \quad (2.24)$$

Optimum hiper düzlem bulma işlem adımları ise Eşitlik 2.25'te görülmektedir. Bu eşitlikte, C, ceza parametresini ifade eder. Hata ξ 'nin kenar dengesini kontrol etmek için kullanılır.

$$\begin{aligned} \min & \frac{1}{2} ||w||^2 + C \sum_{i=1}^l \xi_i \\ \text{s. t. } & y_i[w \cdot \phi(x_i) + b] \geq 1 - \xi_i \\ & \xi_i \geq 0, \quad i = 1, 2, \dots, l \end{aligned} \quad (2.25)$$

Primal problemin adımları yukarıda belirtildiği gibidir. Ancak bazı durumlarda dual probleme geçiş yapılması gerekebilir (Eşitlik 2.26). Bu adımda Lagrange Çarpanları (α_i ve α_j) tekniğinden yararlanılır. Bu teknik kısıtlı optimizasyon problemlerinde kullanılır. Bir fonksiyonun maksimum veya minimum değerlerini bulurken aynı zamanda belirli koşullara (kısıtlara) da uyulması gerektiğinde kullanılır. DVM'de bu kısıtlar, destek vektörlerinin marj üzerinde olması gerektiği gibi durumlardır.

$$\begin{aligned} \min & \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j K(x_i, y_j) - \sum_{i=1}^l \alpha_i \\ \text{s. t. } & \sum_{i=1}^l y_i \alpha_i = 0 \\ & 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, l \end{aligned} \quad (2.26)$$

Dual problemlerde kullanılan çekirdek fonksiyonları ($K(x_i, y_j)$) doğrusal olmayan ayırım problemlerinin çözülmesini sağlar. Çekirdek fonksiyonlar verileri daha yüksek boyutlu bir uzaya dönüştürür ve orada doğrusal olarak ayırır. Gaussian Radius Basis Function, Polynominal ve Sigmoid fonksiyonları sıklıkla kullanılan çekirdek fonksiyonlarıdır.

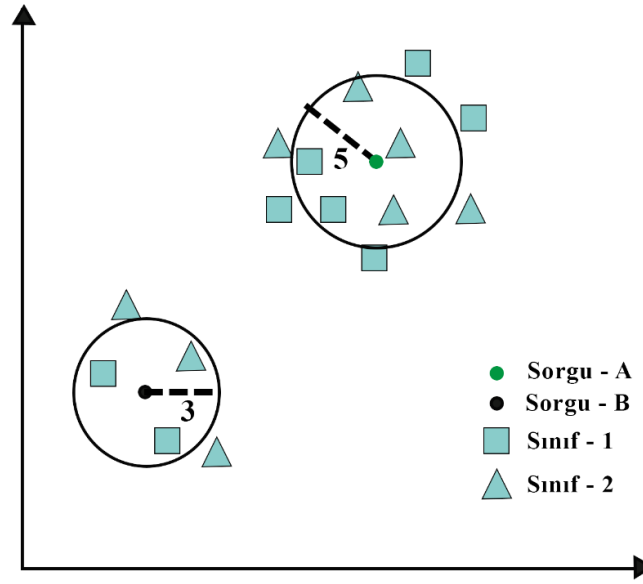
Son karar fonksiyonu ise Eşitlik 2.27'deki gibidir. Yeni bir veri noktasının hangi sınıfa ait olduğu bu eşitlik ile belirlenir.

$$f(x) = \text{sgn}\left(\sum_{sv} y_i \alpha_i K(x, x_i) + b\right) \quad (2.27)$$

Çoklu sınıflandırmada Bir-Herkese-Karşı (One-vs-All), Bir-Bire-Karşı (One-vs-One) ve Yönlendirilmiş-Döngüsel-Grafik (Directed Acyclic Graph) yöntemleri kullanılır.

2.3.1.4 K-En Yakın Komşu

k-En Yakın Komşu algoritması (k-Nearest Neighbour – k-NN) genellikle sınıflandırma amaçları için kullanılır ve denetlenen öğrenme algoritmasıdır. k-NN algoritması eğitim verilerinin özelliklerini ve etiketlerini hesaba katarak etiketlenmemiş verilerin sınıflandırmasını amaçlar. Algoritma ilk adımda test verilerine en çok benzeyen eğitim verilerini arar. Test verilerine en yakın olan “k” sayıda eğitim verisi bulur. Yakınlık genellikle Öklid mesafesi gibi bir ölçü birimi ile hesaplanır. Manhattan mesafesi de kullanılan yakınlık ölçülerinden biridir. Test verisini “k” sayıda ulaştığı eğitim verisinin çoğunluğunun oyuyla bir sınıfa atar. Şekil 2.21’de k-NN algoritmasının çalışma prensibi görülmektedir. “k” sayısının 5 olarak belirlendiği Sorgu-A incelendiğinde, test verisine en yakın 5 nokta (veri) bulunur. Bu 5 noktadan 2’si Sınıf-1, 3’ü ise Sınıf-2’ye aittir. Çoğunluk Sınıf-2’ye ait olduğu için A verisi bu sınıfa atanır. Benzer şekilde “k” sayısının 3 olarak belirlendiği Sorgu-B’de test verisi B’ye en yakın 3 nokta (veri) bulunur. Bu 3 noktadan 2’si Sınıf-1’e, 1’i Sınıf-2’ye ait olduğu için çoğunluğun oyu ile B verisi Sınıf-1’e atanır (Uddin vd., 2022).



Şekil 2.21 k-NN algoritmasının görsel çizimi (Uddin vd., 2022)

k-NN algoritması farklı boyutlardaki (özellik sayısına sahip) ve farklı sınıf sayısına sahip veri setleri için sınıflandırmada kullanılmaktadır. Algoritma sınıflandırma yaparken “k” sayıda en yakın veriye odaklandığı için gürültü verilerin olduğu (hatalı veya anlamsız veriler) veri setlerinde de sınıflandırma için kullanılmaktadır. k-NN algoritması, verileri ölçeklendirerek farklı değer aralıklarındaki verilerle başa çıkabilmektedir. Ayrıca, verileri normalleştirerek de daha iyi sonuçlar elde edebilmektedir (Uddin vd., 2022).

k-NN makine öğrenmesi algoritmaları arasında en basit, uyarlanabilir ve anlaşılır bir algoritma olsa da bazı sınırlamaları bulunmaktadır. Bu sınırlar geliştirilen farklı k-NN varyantları ile aşılmaktadır. Bu varyantlar, “k” değerini en iyi hale getirme, mesafe hesaplamalarını iyileştirme, farklı veri noktalarına önem verme (ağırlıklandırma) ve eğitim verilerini düzenleme gibi farklı algoritmik yaklaşımlar kullanılmaktadır. k-NN algoritmasının farklı varyantlarını öneren çok sayıda çalışma bulunmaktadır. Varyantların çoğu “k” değerini bulmaya odaklanır. Araştırmacıların bazıları geliştirdikleri “uyarlanabilir k-NN” algoritmaları ile her veri seti için en uygun “k” değerini bulmayı amaçlamışlardır. Diğer araştırmacılar ise daha gelişmiş bir varyant olan “yerel olarak uyarlanabilir k-NN” varyantını önermişlerdir. Bu varyantta farklı sınıflandırma yöntemleri bir araya getirilerek ve azınlık sınıfları da dikkate alınarak performansın artırılması amaçlanmıştır. “k” değerini bulma ihtiyacını ortadan kaldıran veya farklı algoritmalarla çalışan k-NN varyantları da bulunmaktadır (Uddin vd., 2022).

k-NN algoritmasının farklı yönlerini iyileştirmeye odaklanan çeşitli varyantlarda bulunmaktadır. Bu varyantlar k-NN'nin doğruluğunu ve verimliliğini arttırmayı amaçlamaktadır. Ağırlık atfına odaklanan varyantlar bunlardan biridir. Bu varyantlar komşuların sınıflandırmaya olan katkılarını, mesafe ve sınıf sıklıklarına göre hesaplayarak belirlerler. Böylece tüm komşular eşit önemde sınıflandırmaya katılmaz. Diğer varyant “ağırlıklı karşılıklı k-NN”, bu varyantta eğitim veri seti benzer özelliklere sahip veri noktalarının bir araya geldiği karşılıklı kümeler ayrılır. Bu kümeleme işlemi ile veri setindeki gürültü azaltılır ve daha anlamlı gruplar elde edilir. Kümeler içindeki her noktanın sınıflandırmaya olan katkısını hesaplamak için küme içindeki diğer noktalar ile olan ilişkisine göre ağırlık değeri hesaplanır. Test verisinin sınıflandırıldığı son adımda ise test veri noktasına en yakın olan karşılıklı komşu bulunur. Geliştirilen diğer varyant ise “bulanık k-NN” varyantıdır. Bu varyantta bir veri noktasının her bir sınıfa ait olma olasılığı hesaplanır ve veri noktası ait olma olasılığı en yüksek olan sınıfa atanır. Ayrıca sıklıkla kullanılan Öklid ve Manhattan mesafelerine alternatif olarak kullanılan ve mevcut mesafe ölçümlerinin sonuçlarını iyileştiren algoritmalar da bulunmaktadır (Uddin vd., 2022).

2.4 Sınıflandırma Algoritmalarının Performanslarının Değerlendirilmesi

Karar sınıflarının değerlendirilmesi karar ağacı projelerinde kritik bir adımdır. Bu adım modelin tahminleri ne kadar doğru yaptığına dair bir ölçüt sunar, modele olan güvenilirliği belirler ve modelin geliştirilmesi gereken yönlerini tespit eder. Farklı modellerin karşılaştırılmasına olanak tanır. Sınıflandırma problemlerinde performansın değerlendirilmesi için sık kullanılan metrikler: doğruluk, özgüllük, duyarlılık ve kesinliktir. Tüm bu metriklerin temeli hata (karışıklık) matrisine (confusion matrix) dayanmaktadır.

Hata matrisi modelin tüm tahminlerini görsel bir özetle sunar. Matrisin boyutunu ($m \times m$) modelin sınıf sayısı (m) belirler. Tablo 2.2’de görüldüğü üzere gerçek değerler pozitif (positive) ve negatif (negative) olarak, tahmini değerler ise doğru (true) ve yanlış (false) olarak ifade edilmektedir (*Machine Learning*, 2011).

Tablo 2.2 Hata matrisi

		GERÇEK DEĞER	
		POZİTİF	NEGATİF
TAHMINİ DEĞER	POZİTİF	DOĞRU POZİTİF (DP)	YANLIŞ POZİTİF (YP)
	NEGATİF	YANLIŞ NEGATİF (YN)	DOĞRU NEGATİF (DN)

Gerçek ve tahmini değerlerin kesişimi ile dört farklı sonuç elde edilmektedir:

- Doğru Pozitif (DP) (True Positives), bir sınıflandırma modelinin pozitif olarak sınıflandırdığı ve gerçekte de pozitif olan örnekleri ifade eder. Yani hem modelin tahmini hem de gerçek durum pozitiftir.
- Doğru Negatif (DN) (True Negatives), bir sınıflandırma modelinin negatif olarak sınıflandırdığı ve gerçekte de negatif olan örnekleri ifade eder. Yani hem modelin tahmini hem de gerçek durum negatiftir.
- Yanlış Pozitif (YP) (False Positives), bir sınıflandırma modelinin negatif olması gereken bir durumu pozitif olarak sınıflandırması durumudur. Yani, modelin bir durumun var olduğunu yanlışlıkla belirtmesidir.
- Yanlış Negatif (YN) (False Negatives), bir sınıflandırma modelinin negatif olarak sınıflandırdığı ancak gerçekte pozitif olan örnekleri ifade eder. Yani, modelin bir durumun varlığını yanlışlıkla reddetmesi anlamına gelir.

DP ve DN sınıflandırmanın doğru yapıldığını, YP ve YN ise sınıflandırmanın yanlış yapıldığını ifade eder. Matrisin sol üst ile sağ alt köşesini birleştiren köşegen değerlerinin yüksek, diğer değerlerin ise 0 veya 0'a yakın olması modelin performansının yüksek olduğunu gösterir (*Machine Learning*, 2011).

Doğruluk (accuracy), modelin performansını ölçen popüler ve basit ölçütlerden biridir. Modelin ne kadar doğru tahminler yaptığını gösterir. Modelin verdiği tüm cevapların kaç tanesinin doğru olduğunu ölçer (Eşitlik 2.28). En iyi doğruluk oranı 1.00, en kötü oran ise 0.00'dır (Mohammed ve Kora, 2023).

$$\text{Doğruluk} = \frac{\text{doğru tahmin sayısı}}{\text{toplam tahmin sayısı}} = \frac{TP + TN}{TP + TN + FN + FP} \quad (2.28)$$

Doğruluk testi bazı durumlarda tek başına yetersizdir. Özellikle dengesiz veri setlerinde yanıltıcı olabilmektedir. Dengesiz veri setinde bir sınıf diğerine göre daha çok örnek içermektedir. Bu da modelin daha fazla örnek içeren sınıfı daha iyi öğrenmesine yol açar. Model çoğunluk sınıfına yönelir ve azınlık sınıfını göz ardı eder. Yüksek bir doğruluk oranına rağmen azınlık sınıfındaki örneklerin yanlış sınıflandırılması, özellikle tıbbi teşhis gibi alanlarda, ciddi sonuçlara yol açabilir (Mohammed ve Kora, 2023).

Özgüllük (specificity) gerçekte negatif olan örneklerin doğru bir şekilde negatif olarak sınıflandırılma oranıdır (Eşitlik 2.29). Duyarlılık ile ters orantılıdır. Özgüllük (0,1] aralığında değer alır. En iyi, en kötü özgüllük oranı yoktur. Çalışma özelinde değişmektedir (Mohammed ve Kora, 2023).

$$\begin{aligned} \text{Özgüllük} &= \frac{\text{negatif sınıflandırılan, gerçekte negatiflerin sayısı}}{\text{toplam gerçekte negatiflerin sayısı}} \\ &= \frac{TN}{TN + FP} \end{aligned} \quad (2.29)$$

Duyarlılık (sensitivity), hatırlama (recall) olarak da bilinmektedir. Bir sınıflandırma modelinin pozitif olarak sınıflandırdığı ve gerçekte de pozitif olan örneklerin, toplam pozitif örneklerle oranıdır (Eşitlik 2.30). En iyi duyarlılık oranı 1.00, en kötü duyarlılık oranı ise 0.00'dır (Mohammed ve Kora, 2023).

$$\begin{aligned} \text{Duyarlılık} &= \frac{\text{pozitif sınıflandırılan, gerçekte pozitiflerin sayısı}}{\text{toplam gerçekte pozitiflerin sayısı}} \\ &= \frac{TP}{TP + FN} \end{aligned} \quad (2.30)$$

Kesinlik (precision), bir sınıflandırma modelinin pozitif olarak sınıflandırdığı ve gerçekte pozitif olan örneklerin, sınıflandırıcı tarafından pozitif olarak tahmin edilen toplam örnek sayısına oranıdır (Eşitlik 2.31). En iyi kesinlik oranı 1.00, en kötü kesinlik oranı ise 0.00'dır (Mohammed ve Kora, 2023).

$$\begin{aligned}
\text{Kesinlik} &= \frac{\text{pozitif sınıflandırılan, gerçek pozitiflerin sayısı}}{\text{pozitif tahmin edilenlerin sayısı}} \\
&= \frac{TP}{TP + FP}
\end{aligned} \tag{2.31}$$

F1 Skoru (F1 Score) duyarlılık ve kesinliğin harmonik ortalaması alınarak hesaplanır (Eşitlik 2.32). Bir modelin gerçek pozitifleri ne kadar iyi tespit ettiğini ve pozitif olarak sınıflandırdığı örneklerin ne kadarının gerçekten pozitif olduğunu ölçer. F1 skoru (0,1] aralığında bir değer alır. Değer 1'e yaklaştıkça modelin performansının iyi olduğu kabul edilir (Mohammed ve Kora, 2023).

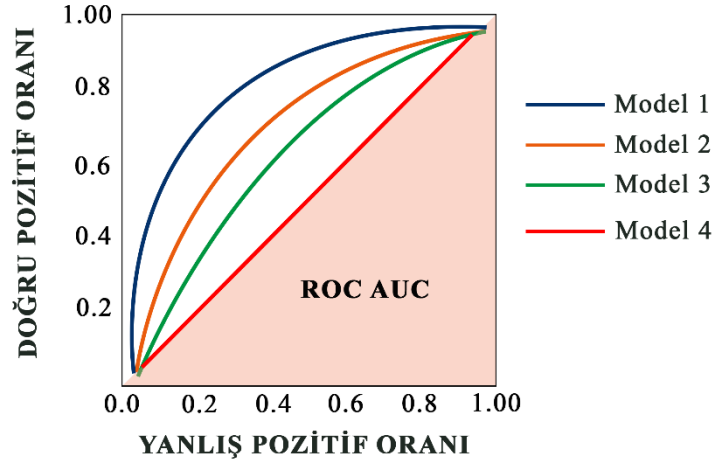
$$F1 \text{ Skor} = \frac{2 \times \text{Kesinlik} \times \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \tag{2.32}$$

F1 skorunda duyarlılık ve kesinlik arasındaki dengeye önem verilmektedir. Bu nedenle aritmetik ortalama yerine harmonik ortalama tercih edilmiştir. Aritmetik ortalama ile hesaplansaydı; kesinlik değerinin 1, duyarlılık değerinin 0 ya da kesinlik değerinin 0, duyarlılık değerinin 1 olduğu durumlarda 0,5 sonucuna ulaşıldı. Bu da modelin performansının ortalama olduğunu göstermektedir. Ancak her iki değerden birinin 0 olması bu durumun tam aksini belirtmektedir.

ROC Eğrisi (Receiver Operating Characteristic Curve – Alıcı İşletim Karakteristik Eğrisi), gerçek pozitif oran ile yanlış pozitif oran arasındaki dengeyi görselleştiren bir şekildir (Şekil 2.22). Bu metriktan çeşitli bilgiler elde edilebilmektedir:

- Birden fazla modelin performansı karşılaştırılabilir,
- Bir ya da birden fazla modelde karar verme sınırı (eşik değerini) belirlenebilir,
- Bir modelin performansı ölçülebilir.

ROC AUC (ROC Area Under the Curve – ROC Eğrisi Altındaki Alan), model performansını değerlendirmek için kullanılır (Şekil 2.22). Değerleri (0-1) aralığındadır, 0 kötü bir sınıflandırma, 1 ise iyi bir sınıflandırma yaptığını göstermektedir (Vujovic, 2021).



Şekil 2.22 ROC eğrisi ve ROC AUC

Kappa istatistiği, iki puanlayıcı arasındaki uyumun derecesini belirlemek için Cohen (1960) tarafından geliştirilmiştir. Fleiss (1971) ikiden fazla puanlayıcıda kullanılabilmesi için Kappa istatistiğini genelleştirmiştir. Brennan ve Prediger (1981) Kappa'nın puanlama sürecinde nesne ve bireylerin, puanlayıcıların puanlamalarının ve puanlamada kullanılan kategorilerin birbirinden bağımsız olduğunu ifade etmişlerdir. Kappa istatistiği puanlayıcılar arası gözlenen uyumun içinden şansa dayalı uyumun çıkarılmasını sağlar (Eşitlik 2.33) (Bilgen ve Doğan, 2017).

$$Kappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e} \quad (2.33)$$

Bu eşitlikte;

$\bar{P}$: gözlenen uyumluluk oranını,

$\bar{P}_e$: şansa uyumluluk oranını ifade eder.

Kılıç (2015), Kappa istatistiğini bir klinikte görev yapan iki psikiyatristin 50 hastaya koydukları tanımlar üzerinden açıklamıştır (Tanı A, Tanı B). Psikiyatristlerin tanımları Tablo 2.3'te görülmektedir.

Tablo 2.3 Psikiyatristlerin hastalara koydukları tanıları

	<i>Gözlemci 2 Tanı A</i>	<i>Gözlemci 2 Tanı B</i>	<i>TOPLAM</i>
<i>Gözlemci 1 Tanı A</i>	26	7	33 (%66)
<i>Gözlemci 1 Tanı B</i>	9	8	17 (%34)
<i>TOPLAM</i>	35 (%70)	15 (%30)	50 (%100)

Örnek üzerinden $\bar{P}$ ve $\bar{P}_e$ ifadeleri anlatılacak olursa, $\bar{P}$ her iki gözlemcinin de aynı tanıyı koydukları hastaların oranlarının toplamıdır (Eşitlik 2.34).

$$\bar{P} = \frac{(\text{ortak gözlemlenen sınıf 1}) + (\text{ortak gözlemlenen sınıf 2})}{\text{toplam örnek sayısı}} \quad (2.34)$$

$$\bar{P} = \frac{26 + 8}{50} \cong 0,70 \quad (\text{Tablo 2.3'teki değerler})$$

$\bar{P}_e$, gözlemcilerin sınıflara göre gözlemledikleri örnek sayıları ayrı ayrı çarpılır ve sonrasında çarpımlar toplanır (Eşitlik 2.35).

$$\begin{aligned} \bar{P}_e = & \left(\begin{array}{l} 1. \text{gözlemci Tanı A gözlemleri} \\ \times 2. \text{gözlemci Tanı A gözlemleri} \end{array} \right) \\ & + \left(\begin{array}{l} 1. \text{gözlemci Tanı B gözlemleri} \\ \times 2. \text{gözlemci Tanı B gözlemleri} \end{array} \right) \end{aligned} \quad (2.35)$$

$$\bar{P}_e = (0.66 * 0.70) + (0.34 * 0.30) = 0.564$$

Kappa (K) istatistiği -1 ile +1 arasında değer almaktadır. K değerinin 0'dan küçük olması iki puanlayıcı arasındaki uyumun şansın ötesinde düşük olduğunu göstermektedir ve dikkate alınmamaktadır. K istatistiğinin yorumlanması için Landis ve Koch (1977) tarafından önerilen uyum düzeyleri kullanılmaktadır (Tablo 2.4) (Bilgen ve Doğan, 2017).

Tablo 2.4 K istatistiği uyum düzeyleri

<i>K istatistiği</i>	<i>Uyumun Gücü</i>
< 0.00	Zayıf
0.00 - 0.20	Önemsiz
0.21 - 0.40	Düşük
0.41 - 0.60	Orta
0.61 - 0.80	Önemli
0.81 - 1.00	Çok Yüksek

K deęerinin 1 olması iki puanlayıcı arasında mükemmel bir uyum olduęu, řans faktörünün etkisinin olmadığını göstermektedir. Örnekteki k deęeri Eřitlik 2.36'daki gibi hesaplanmaktadır.

$$K = \frac{0.700 - 0.564}{1 - 0.564} = \frac{0.136}{0.436} = 0.312 \quad (2.36)$$

İki psikiyatristin tanısı zayıf düzeyde uyum göstermektedir, řans oranının yüksek olduęu söylenebilir.

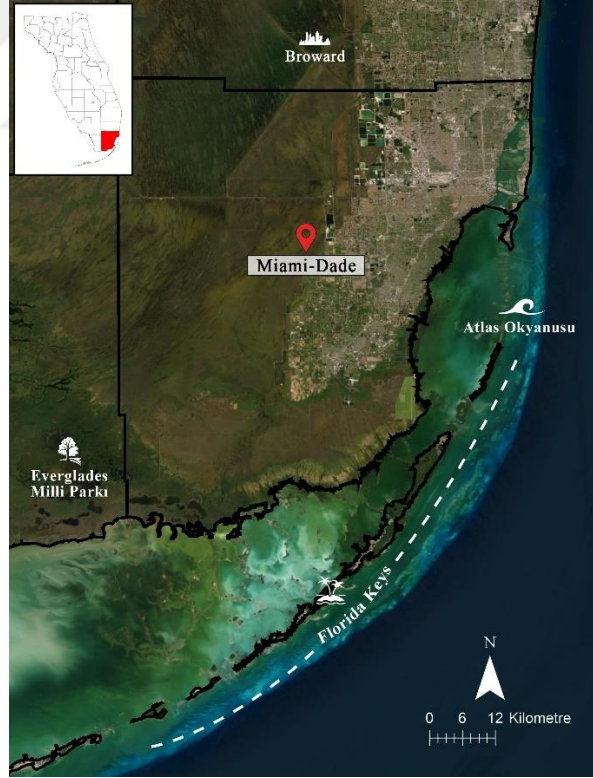


3

UYGULAMA

3.1 Çalışma Alanı

Bina fonksiyonlarının belirlenmesi için çalışma alanı olarak Florida eyaletinin Miami-Dade ili (county) belirlenmiştir. 2020 yılı nüfus verilerine göre Florida Amerika'nın en kalabalık üçüncü eyaleti ve Miami-Dade ise Florida'nın en kalabalık birinci ilidir (United States Census Bureau, 2024). Miami-Dade Florida'nın güneydoğusunda yer almaktadır. Doğusunda Atlas Okyanusu, batısında Everglades Milli Parkı, güneyinde Florida Keys adaları ve kuzeyinde Broward ili (county) ile çevrilidir (Şekil 3.1) (Data USA, 2024).



Şekil 3.1 Miami-Dade ilinin konumu

Miami-Dade ili (county) 2018 yılında PortMiami limanında dünya yolcu rekorunu kırarak Dünya'nın kruvaziyer başkenti olarak anılmıştır. İçerisinde yer alan

uluslararası havalimanı yolcu sayısı açısından Amerika'nın en büyük üçüncü havalimanı iken yük açısından en büyük birinci havalimanıdır. Aynı zamanda Amerika'nın dördüncü büyük kamu okulu bölgesidir (Miami-Dade County, 2024). Çalışma alanında en fazla endüstriye sahip üç işletme sırası ile (1) profesyonel, bilimsel ve teknik hizmetler, (2) sağlık bakımı ve sosyal yardım ve (3) perakende ticarettir (Data USA, 2024). İlin 2010 – 2020 yıllarındaki nüfus artış oranı %7.5'dur (Bu oran Amerika'da %7.2) (USAFacts, 2024). Florida'da en çok büyüyen beş il arasında üçüncü sırada yer almaktadır ve büyüme oranı %13'ün üzerindedir (Florida State Office of Economic and Demographic Research, 2024). Verilerinin erişilebilir olmasının yanında bulunduğu önemli konum ve sahip olduğu kalabalık nüfus nedeniyle Miami-Dade çalışma alanı olarak seçilmiştir. Bu iki faktör beraberinde bina fonksiyonlarında çeşitliliğe yol açmaktadır.

3.2 Veri Hazırlığı

Bu başlık altında verilerin elde edilmesinden model tahmin aşamasına kadar olan süreçler incelenecektir (Şekil 3.2).



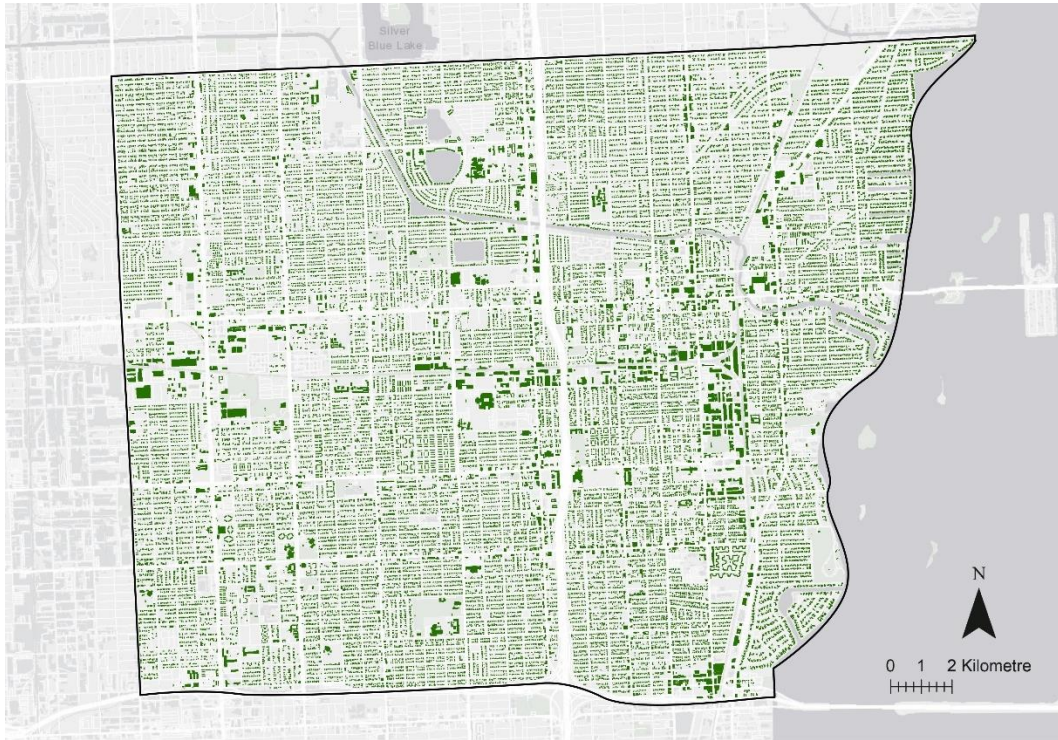
Şekil 3.2 Makine öğrenmesi ile model oluşturma

3.2.1 Veri Toplama ve Manipülasyonu

Uygulamada ArcGIS'in açık veri platformu üzerinde Miami-Dade ili (county) tarafından sunulan veri setleri kullanılmıştır. Bina ayak izi ve ada sınırları olmak üzere iki veri seti kullanılmıştır. Bina ayak izi veri seti ilgili kurum tarafından 12.03.2024 tarihinde güncellenmiş ve 20.03.2024 tarihinde indirilmiştir (ArcGIS Hub, 2024a). Ada sınırları olarak kurum tarafından paylaşılan bir veri setine ulaşamamıştır. Bu nedenle kurum tarafından 7.11.2023 tarihinde güncellenen

kaldırım sınır çizgileri veri seti 26.05.2024 tarihinde indirilmiş ve düzenlenerek ada sınırları olarak kullanılmıştır (ArcGIS Hub, 2024b).

Miami-Dade iline dair elde edilen veri setinde 520.494 adet bina verisi bulunmaktadır. Modelin eğitim süresi ve yönetimi veri boyutundan doğrudan etkilenmektedir. Modelin performansı arttırmak amacıyla veri setinden daha küçük bir alt küme çalışma alanı olarak belirlenmiştir. Çalışma alanı Miami-Dade ilinin en kalabalık şehirleri Miami ve Hialeah olup ve 35084 adet bina ayak izi verisi bulunmaktadır (Şekil 3.3).



Şekil 3.3 Çalışma alan sınırları

Verileri daha anlamlı hale getirebilmek ve analiz edebilmek için veri manipülasyonu adımı sürecin önemli bir parçasıdır. Çalışma alanı için kullanılan koordinat sistemi, orijinal olarak Geographic (NAD 83) EPSG: 4269 sisteminde olup, bu çalışma için Florida bölgesine özgü olan State Plane (TM 81° W) (NAD83_HARN) EPSG: 2881 projeksiyon koordinat sistemine dönüştürülmüştür. Veri setinde yer alan golf sahası ve defin alanı 3.2.4 başlığı altında değinilecek olan sınıflandırmalara dahil olmadığı için silinmiştir.

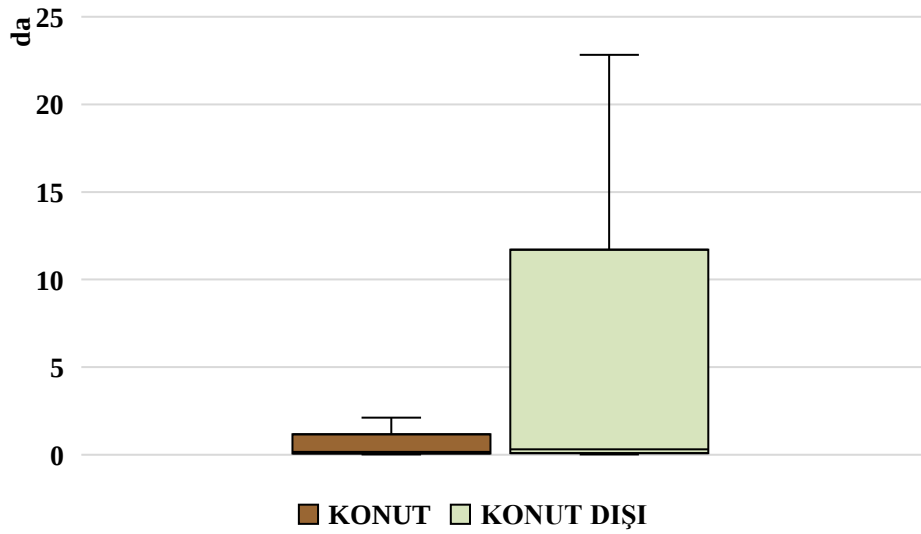
Ada sınırları doğrudan ulaşılabilir bir veri olmadığı için kaldırım çizgisi verileri kullanılmıştır. Shapefile formatında indirilen veriler, ArcMap 10.8’de

düzenlenmiştir. İlk olarak koordinat sistemi bina ayakizi verilerinde olduğu gibi EPSG: 2881 sistemine dönüştürülmüştür. Sonrasında kapalı bir şekil oluşturan ilgili çizgisel veriler birleştirilerek çokgen verisine dönüştürülmüş ve ada verisi olarak kullanılmıştır. Çalışma kapsamında ada verisi olarak kullanılmayacak refüjler silinmiştir ve bazı ada sınırları, yazılımda açılan altlık görüntü üzerinden yeniden çizilmiştir. Tüm bu adımlardan sonra iki veri seti çalışmada bir araya getirilmiştir.

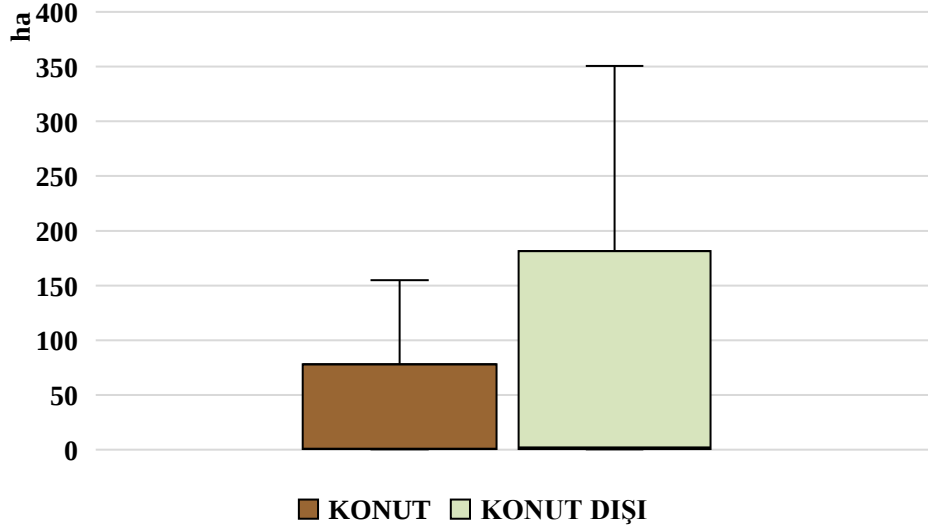
Tez kapsamında iki uygulama gerçekleştirilmiştir. Her iki uygulamada da algoritmalarda en iyi performansı elde etmek için gösterge seçimi ve çıkarımı önemli bir adımdır. Bu adım 3.2.2, 3.2.3 ve 3.2.4 bölümlerinde açıklanmaktadır.

3.2.2 Büyüklük Göstergelerinin Hesaplanması

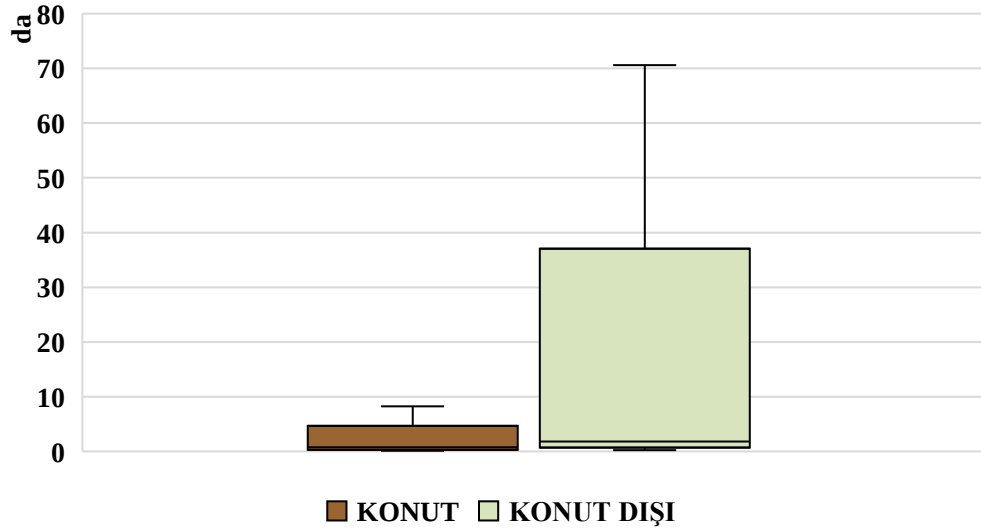
Büyüklük göstergeleri; bina alanı, parsel alanı ve Voronoi çokgeni alanıdır. Uygulama 1 için konut ve konut dışı binaların büyüklük göstergelerinin dağılımlarını gösteren kutu grafikleri Şekil 3.4 – 3.6’da görülmektedir.



Şekil 3.4 Bina alanı göstergesinin konut ve konut dışı binalara göre değer dağılımı

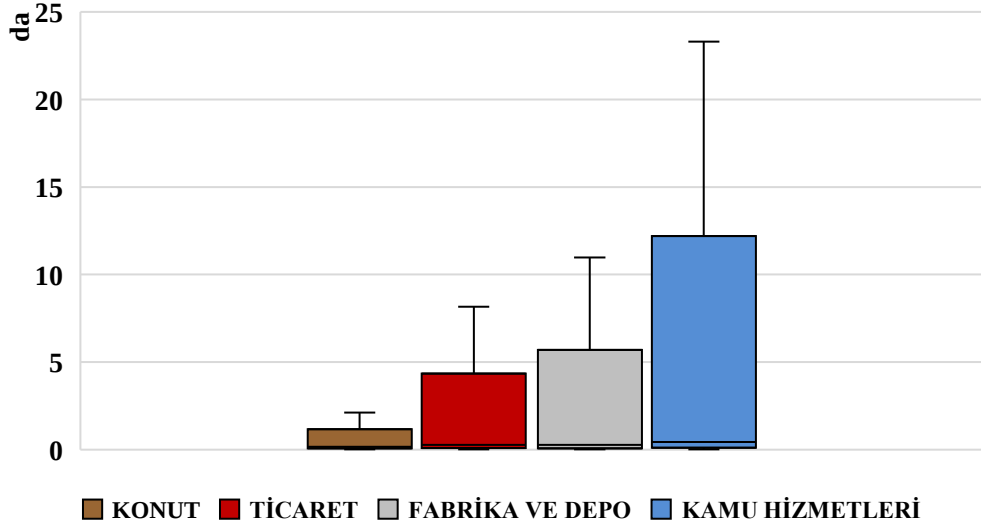


Őekil 3.5 Parsel alanı göstergesinin konut ve konut dıőı binalara göre deęer daęılımı

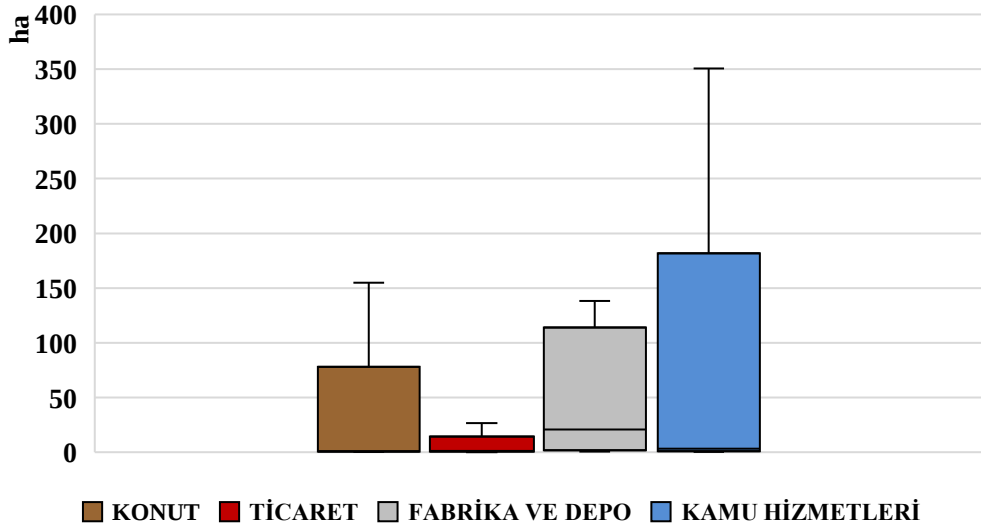


Őekil 3.6 Voronoi çokgeni alanı göstergesinin konut ve konut dıőı binalara göre deęer daęılımı

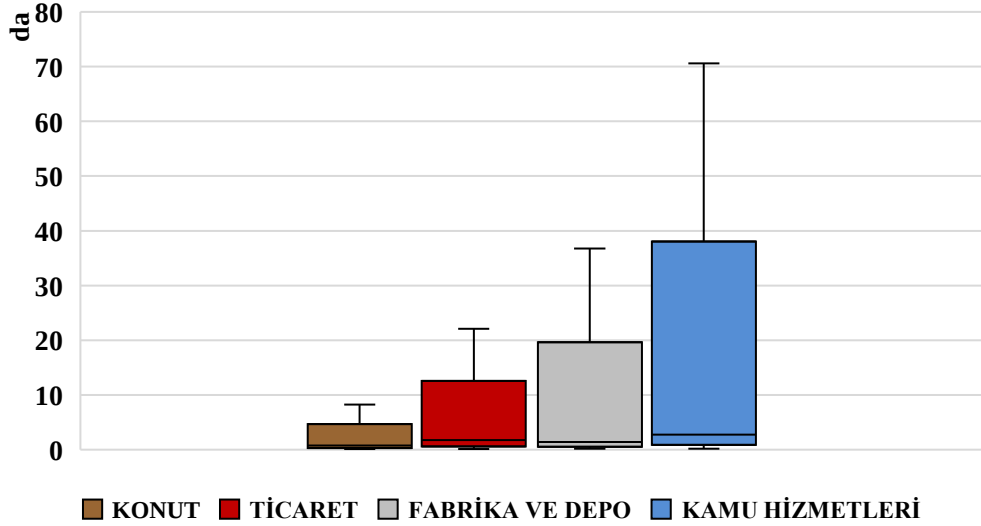
Uygulama 2 için konut, ticaret, kamu hizmetleri ve fabrika ve depo bina fonksiyonlarının büyüklük göstergelerinin daęılımlarını gösteren kutu grafikleri Őekil 3.7 – 3.9’da görölmektedir.



Şekil 3.7 Bina alanı göstergesinin bina fonksiyonlarına göre değer dağılımı



Şekil 3.8 Parsel alanı göstergesinin bina fonksiyonlarına göre değer dağılımı



Şekil 3.9 Voronoi çokgeni alanı göstergesinin bina fonksiyonlarına göre değer dağılımı

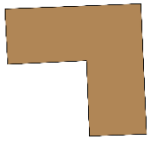
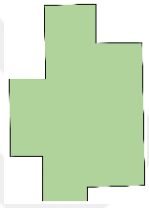
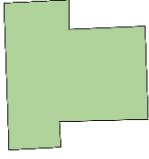
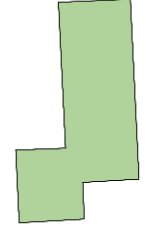
3.2.3 Şekil Göstergelerinin Hesaplanması

Şekil göstergeleri; dairesellik, dışbükeylik, dikdörtgensellik, pürüzlülük, karesellik, eşdeğer dikdörtgen indeksi, uzanım şekil göstergeleri ve bina alanıdır. Bu göstergeler, tez danışmanınca geliştirilen bir CBS eklentisiyle her bina için hesaplanmıştır. Tablo 3.1’de Uygulama 1 için konut ve konut dışı örnek bina fonksiyonlarının şekil gösterge değerleri görülmektedir.

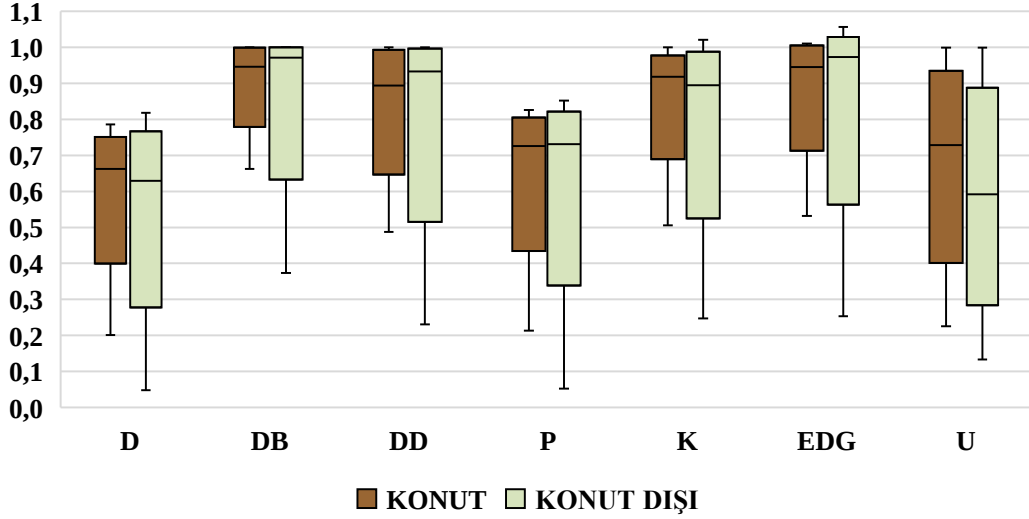
Tablo 3.1 Uygulama 1’deki konut ve konut dışı binaların örnek bina ayak izleri ve şekil gösterge değerleri, D: Dairesellik, DB: Dışbükeylik, DD: Dikdörtgensellik, P: Pürüzlülük, K:Karesellik, EDG: Eşdeğer Dikdörtgen Göstergesi, U: Uzanım

FONKSİYON	BİNA AYAK İZİ	GÖSTERGELER						
		D	DB	DD	P	K	EDG	U
KONUT - 1		0,732	0,967	0,929	0,770	0,965	0,967	0,888
KONUT - 2		0,656	0,963	0,927	0,742	0,914	0,963	0,518

Tablo 3.1 Uygulama 1’deki konut ve konut dışı binaların örnek bina ayak izleri ve şekil gösterge değerleri, D: Dairesellik, DB: Dışbükeylik, DD: Dikdörtgensellik, P: Pürüzlülük, K:Karesellik, EDG: Eşdeğer Dikdörtgen Göstergesi, U: Uzanım (devamı)

FONKSİYON	BİNA AYAK İZİ	GÖSTERGELER						
		D	DB	DD	P	K	EDG	U
KONUT - 3		0,523	0,799	0,665	0,574	0,816	0,816	0,960
KONUT DIŞI - 1		0,568	0,856	0,746	0,607	0,850	0,865	0,689
KONUT DIŞI - 2		0,615	0,881	0,782	0,646	0,885	0,885	0,955
KONUT DIŞI - 3		0,473	0,790	0,651	0,572	0,776	0,808	0,562

Konut ve konut dışı binaların şekil göstergelerinin dağılımlarını gösteren kutu grafikleri Şekil 3.10’da görülmektedir.



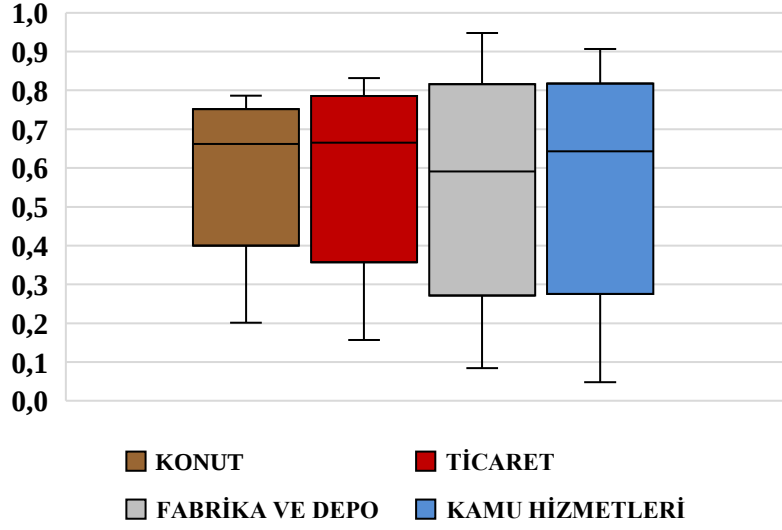
Şekil 3.10 Konut ve konut dışı binaların şekil gösterge değerlerinin dağılımları,
D: Dairesellik, DB: Dışbükeylik, DD: Dikdörtgensellik, P: Pürüzlülük,
K:Karesellik, EDG: Eşdeğer Dikdörtgen Göstergesi, U: Uzanım

Tablo 3.2’de Uygulama 2 için örnek bina fonksiyonlarının şekil gösterge değerleri görülmektedir.

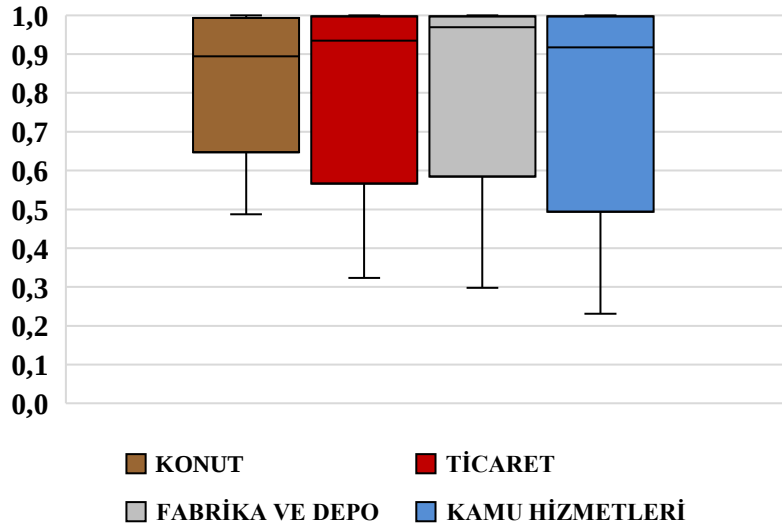
Tablo 3.2 Uygulama 2’deki bina fonksiyonlarının örnek bina ayak izleri ve şekil gösterge değerleri, D: Dairesellik, DD: Dikdörtgensellik, K:Karesellik, DB: Dışbükeylik, EDG: Eşdeğer Dikdörtgen Göstergesi, U: Uzanım, P: Pürüzlülük

GÖSTERGELER	BİNA FONKSİYONLARI			
	KONUT	TİCARET	KAMU HİZMETLERİ	FABRİKA VE DEPO
				
D	0.646	0.637	0.602	0.564
DD	0.854	0.875	0.884	0.872
K	0.907	0.901	0.876	0.847
DB	0.922	0.935	0.901	0.920
EDG	0.924	0.937	0.893	0.879
U	0.681	0.568	0.672	0.578
P	0.69	0.707	0.664	0.620

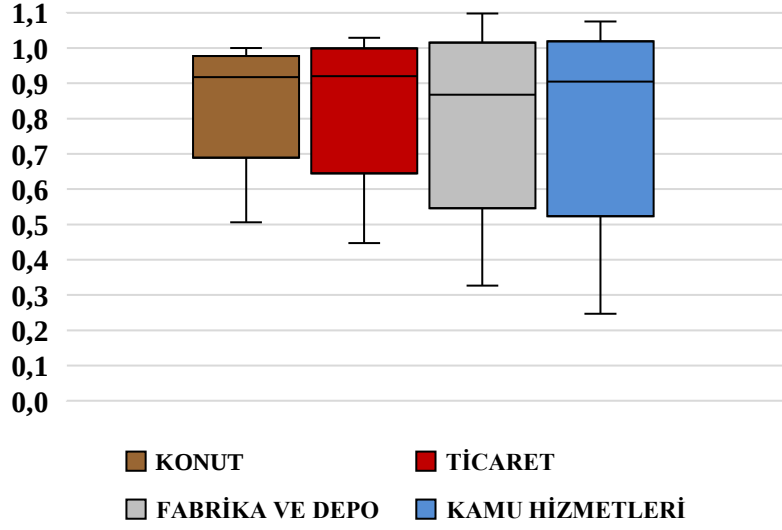
Şekil göstergelerinin fonksiyonlara göre değer dağılımlarını gösteren kutu grafikleri Şekil 3.11 – 3.17’de görülmektedir.



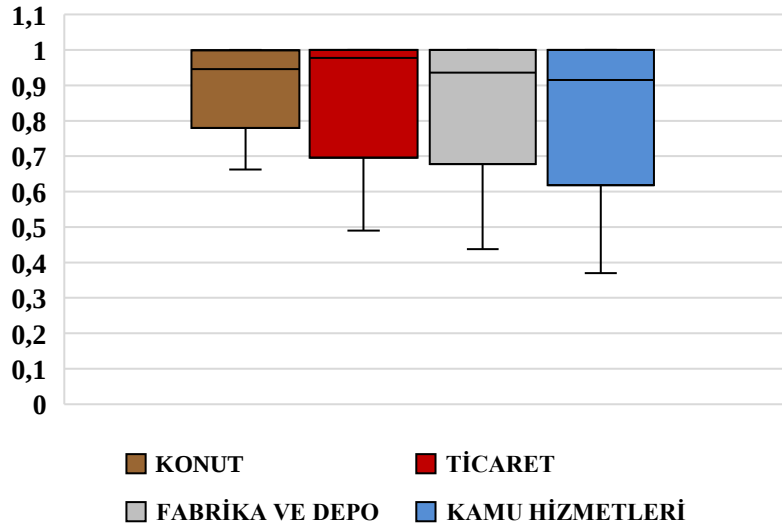
Şekil 3.11 Dairesellik göstergesinin bina fonksiyonlarına göre değer dağılımı



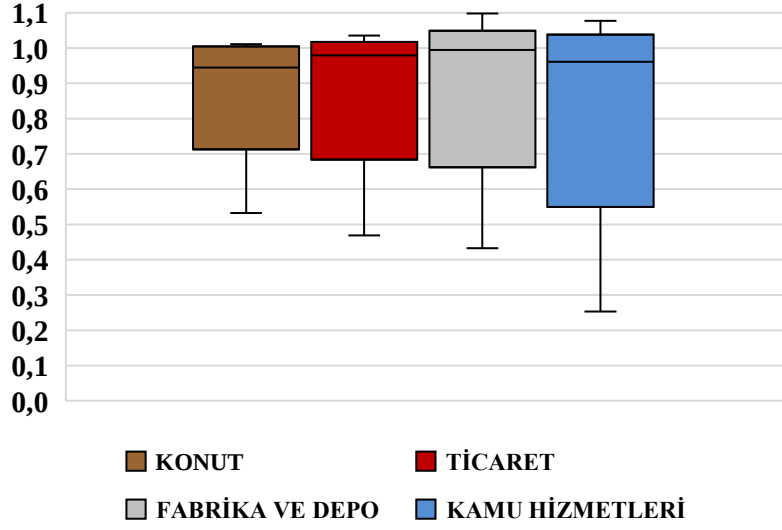
Şekil 3.12 Dikdörtgensellik göstergesinin bina fonksiyonlarına göre değer dağılımı



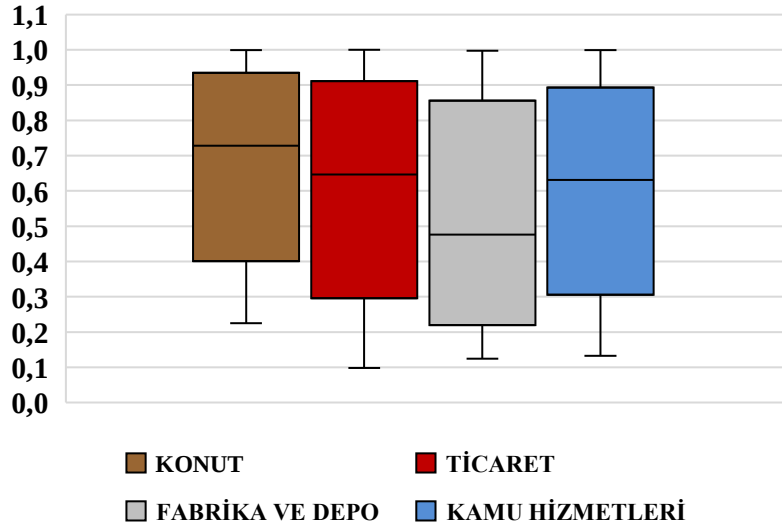
Şekil 3.13 Karesellik göstergesinin bina fonksiyonlarına göre değer dağılımları



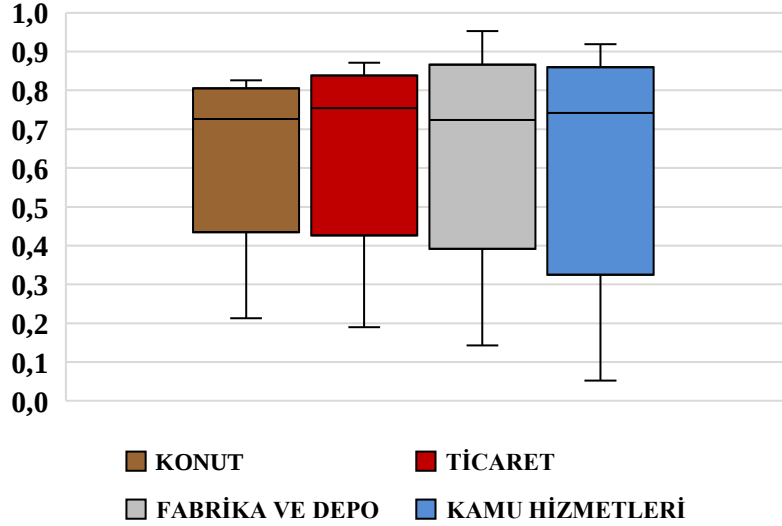
Şekil 3.14 Dışbükeylik göstergesinin bina fonksiyonlarına göre değer dağılımları



Şekil 3.15 Eşdeğer dikdörtgen göstergesinin bina fonksiyonlarına göre değer dağılımları



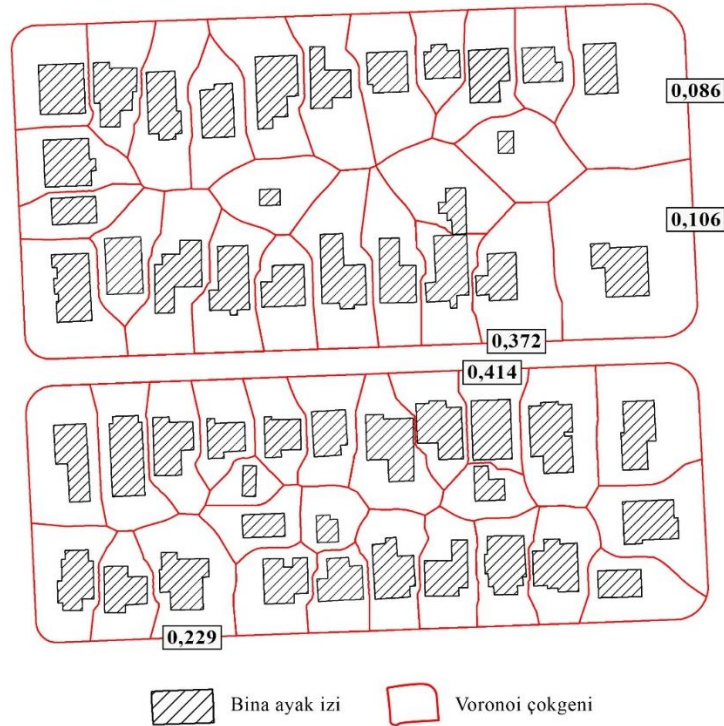
Şekil 3.16 Uzunım göstergesinin bina fonksiyonlarına göre değer dağılımları



Şekil 3.17 Pürüzlülük göstergesinin bina fonksiyonlarına göre değer dağılımları

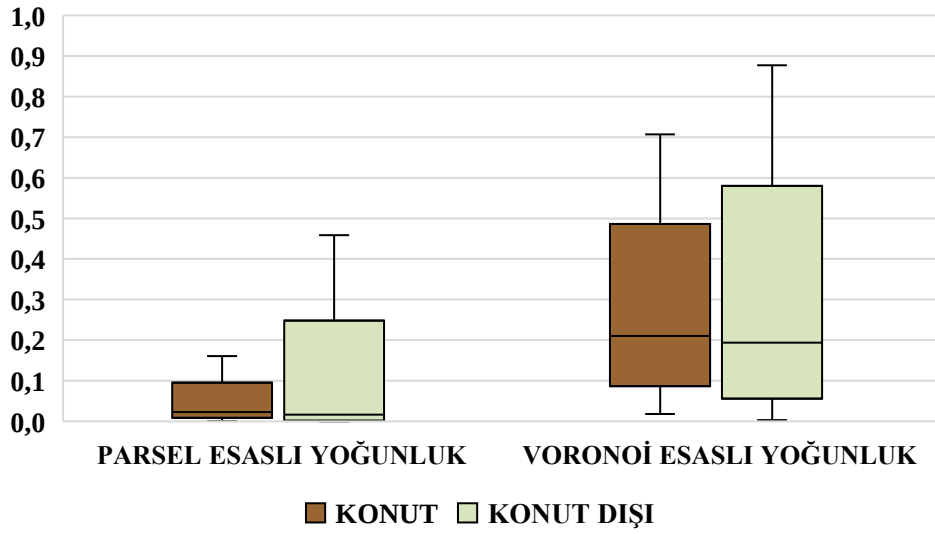
3.2.4 Mekânsal İlişkilerin Hesaplanması

Mekânsal ilişkiler; parsel esaslı yoğunluk ve Voronoi esaslı yoğunluk metrikleriyle ölçülmüştür. Parsel alan bilgileri sadece sayısal değerler olarak elde edilmiştir. Elde edilen veri setinde parsel sınırlarını gösteren çizimler bulunmamaktadır. Voronoi çokgenleri tez kapsamında üretilmiştir. Şekil 3.18’de bazı binaların Voronoi esaslı yoğunluk değerleri görülmektedir.



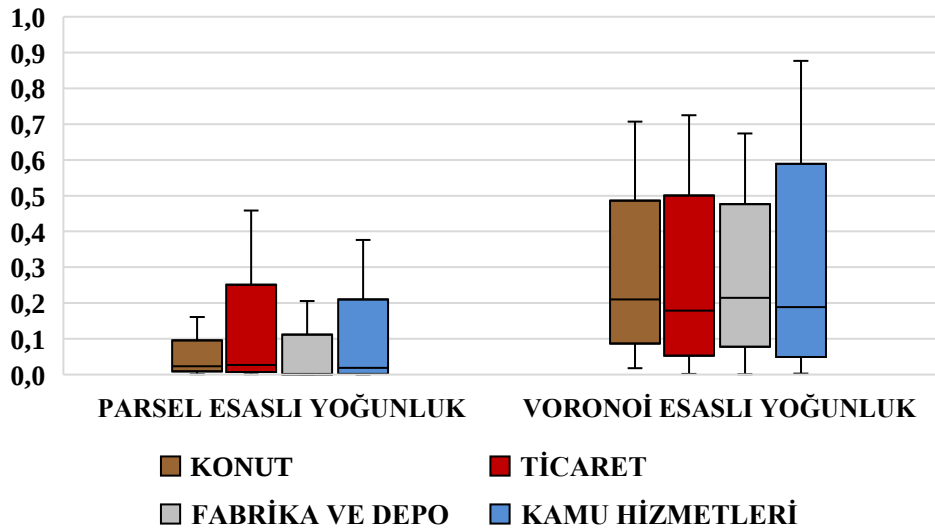
Şekil 3.18 Bazı bina ayak izlerinin Voronoi çokgenlerine alansal oranları

Uygulama 1 için konut ve konut dışı binaların mekânsal ilişki göstergelerinin dağılımlarını gösteren kutu grafikleri Şekil 3.19’da görülmektedir.



Şekil 3.19 Konut ve konut dışı binaların mekânsal ilişki gösterge değerlerinin dağılımları

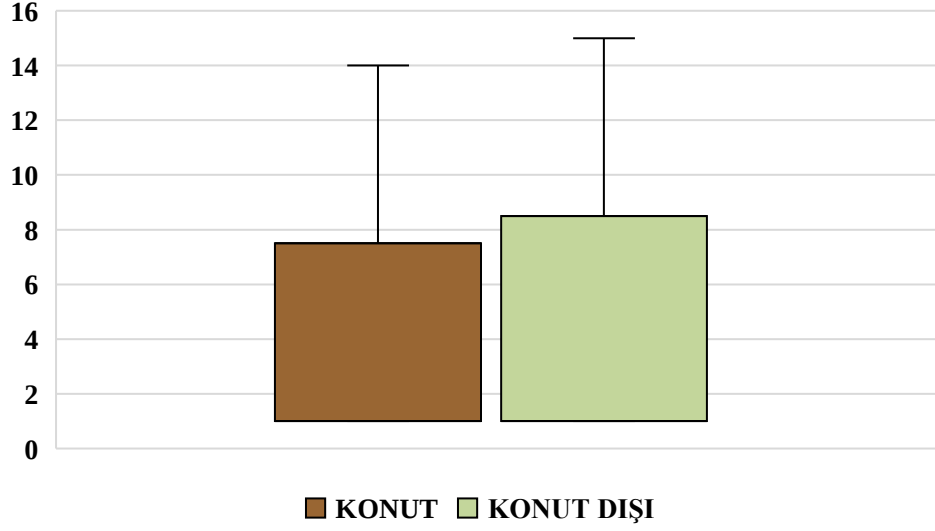
Uygulama 2 için konut, ticaret, kamu hizmetleri ve fabrika ve depo binalarının mekânsal ilişki göstergelerinin dağılımlarını gösteren kutu grafikleri Şekil 3.20’de görülmektedir.



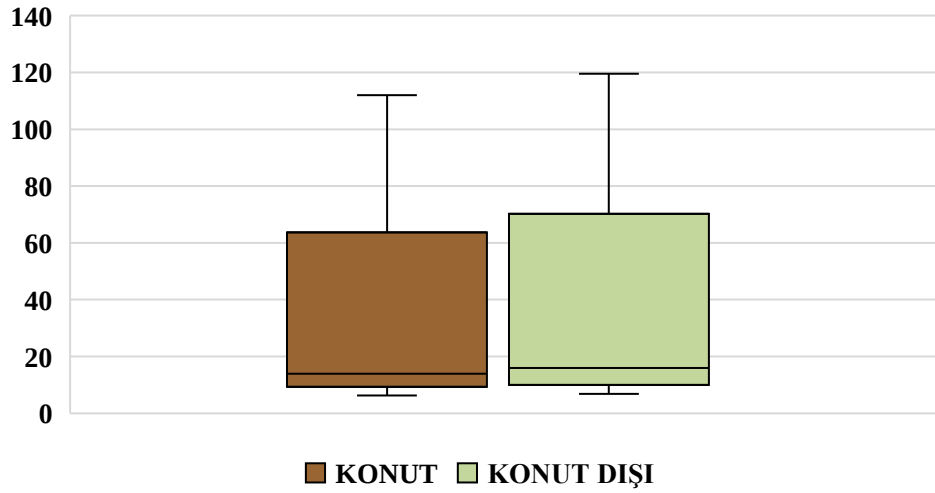
Şekil 3.20 Bina fonksiyonlarının mekânsal ilişki gösterge değerlerinin dağılımları

3.2.5 Nicel Özniteliklerin Hesaplanması

Nicel öznitelik göstergeleri; kat adedi ve kat yüksekliğidir. Uygulama 1 için konut ve konut dışı binaların nicel öznitelik göstergelerinin dağılımlarını gösteren kutu grafikleri Şekil 3.21 – 3.22’de görülmektedir.

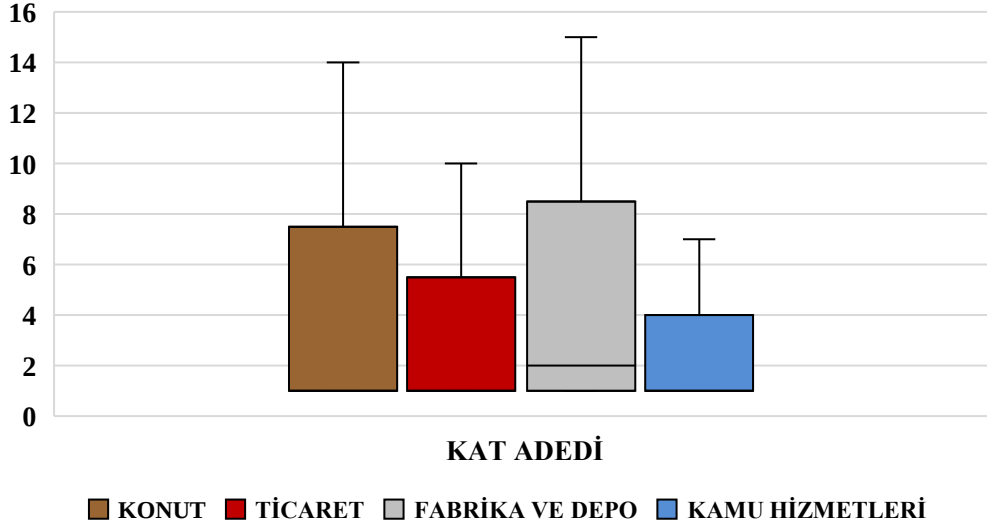


Şekil 3.21 Kat adedi göstergesinin konut ve konut dışı binalara göre değer dağılımı

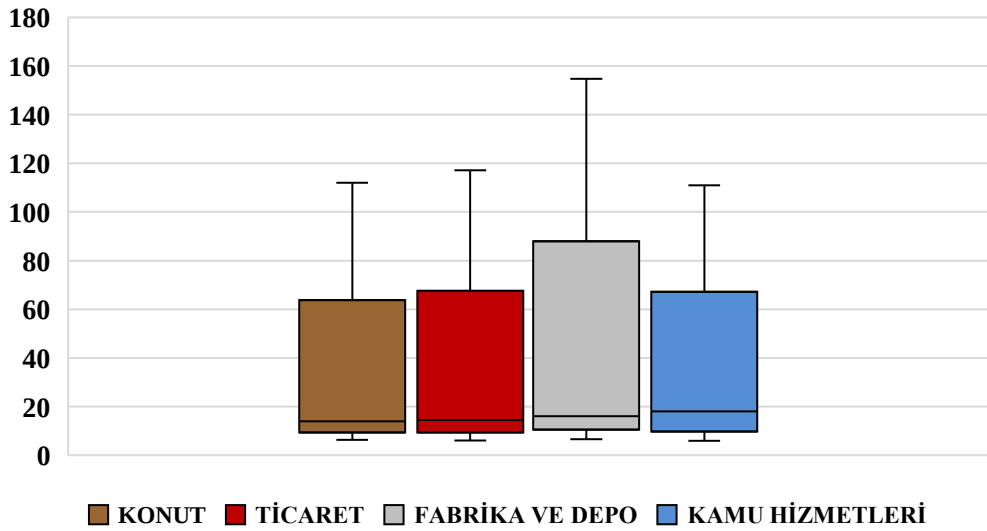


Şekil 3.22 Bina yüksekliği göstergesinin konut ve konut dışı binalara göre değer dağılımı

Uygulama 2 için konut, ticaret, kamu hizmetleri ve fabrika ve depo bina fonksiyonlarının nicel öznitelik göstergelerinin dağılımlarını gösteren kutu grafikleri Şekil 3.23 – 3.24’te görülmektedir.



Şekil 3.23 Kat adedi göstergesinin bina fonksiyonlarına göre değer dağılımı



Şekil 3.24 Bina yüksekliği göstergesinin bina fonksiyonlarına göre değer dağılımı

3.2.6 Fonksiyonel Sınıflandırma

Binaların sahip olduğu fonksiyonlar insanların ihtiyaçları doğrultusunda şekillenir. Binalar ilk olarak barınma ihtiyacını karşılamak için inşa edilmiştir. Artan nüfus, gelişmekte ve değişmekte olan yaşam biçimleri bina fonksiyonlarında çeşitliliğe yol açmıştır. Özellikle kentsel alanlarda bina fonksiyonlarının çeşitliliği giderek artmaktadır. Sınıflandırma ile bina fonksiyonlarının karmaşık çeşitliliği daha basit kategorilere ayrılır. Fonksiyonel sınıflandırma araştırmanın konusu ve kapsamına göre değişmektedir. Tablo 3.3'te araştırmacıların çalıştıkları konu kapsamında

yaptıkları fonksiyonel sınıflandırma görülmektedir. Tablonun devamı ek A’da yer almaktadır.

Tablo 3.3 Fonksiyonel sınıflandırma için literatür araştırması (Du vd., 2020; Du vd., 2024)

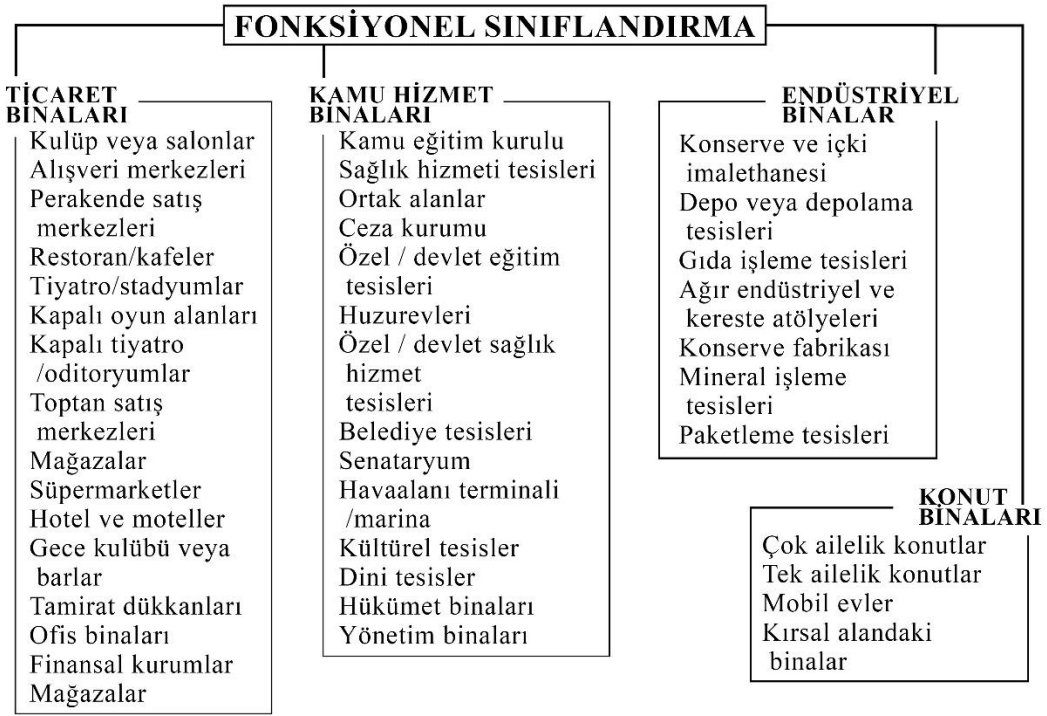
Araştırmacılar	Çalışmanın Amacı	Fonksiyonel Sınıflandırma	
Shouji Du; Shihong Du; Bo Liu; Xiuyuan Zhang; Zhijia Zheng	Kentsel arazi kullanım haritası oluşturma	TİCARET	Ofis, kiralama mağazası, restoran, otel ve finansal hizmetler
		KONUT	Kentsel konutlar ve apartmanlar
		KAMU HİZMETLERİ	Eğitim, sağlık, yönetim binaları ve kamu hizmetleri
		ENDÜSTRİYEL	Üretim tesisleri ve fabrikalar
		GECEKONDU	Kırsal alan ve şehirlerdeki köyler
		AÇIK ALANLAR	Park, yeşil alan, tarım alanları, kurak ve sulak alanlar
Shouhang Du; Meiyun Zheng; Liyuan Guo; Yuhui Wu; Zijuan Li; Peiyi Liu	Fonksiyonel sınıflandırma ile kentsel gelişim planlarına katkı sağlamak	KONUT	Konutlar ve hizmetli apartmanlar
		TİCARET	Şirket, fabrika, alışveriş merkezleri, konaklama tesisleri, finansal ve endüstriyel hizmetler, yiyecek-içecek hizmetleri, spor ve serbest zaman hizmetleri, yaşam hizmetleri ve diğer ticari hizmetler
		YÖNETİM VE KAMU HİZMETLERİ	Yönetim ofisleri, yol yardım tesisleri, kamu hizmetleri, sağlık hizmetleri, kültür ve spor hizmetleri ve diğer kamu hizmetleri
		EĞİTİM HİZMETLERİ	Üniversite ve lise, ilk ve ortaokul ve diğer eğitim hizmetleri

Tablo 3.3 Fonksiyonel sınıflandırma için literatür araştırması (Kong vd., 2024)

(devamı)

Bo Kong; Tinghua Ai; Xinyan Zou; Xiongfeng Yan; Min Yang	Mekânsal veritabanı oluşturmak	KONUT	Yüksek katlı konutlar, apartmanlar ve yurtlar
		TİCARET	Kiralama ofisleri, yeme-içme tesisleri, eğlence tesisleri ve finansal hizmetler
		KIRSAL YAPILAR	Kentsel bloklarla çevrili, yüksek yoğunlukta ve aşırı kalabalık kırsal konutlar
		ORTAK TESİSLER	Sağlık tesisleri, ulaşım tesisleri, spor tesisleri, yönetim birimleri, parklar, huzurevleri
		EĞİTİM	Okul binaları
		FABRİKA VE DEPO	Fabrika, depo ve endüstriyel parklar
		KARMA FONKSİYON	Karma kullanımlar

Ek B’de çalışma kapsamında indirilen veri setinin fonksiyonel sınıflandırılması görülmektedir. Veri setinin fonksiyonel sınıflandırması 52 ana başlık, 97 alt başlıktan oluşmaktadır. Bu tez kapsamında fonksiyonel sınıflandırma dört ana başlık altında incelenmiştir: konut binaları, kamu hizmet binaları, endüstriyel binalar ve ticaret binaları. Veri setindeki mevcut sınıflandırma Şekil 3.25’te görüldüğü gibi Tablo 3.3 ve ek A’daki sınıflandırmalar da dikkate alınarak yeniden gruplandırılmıştır.

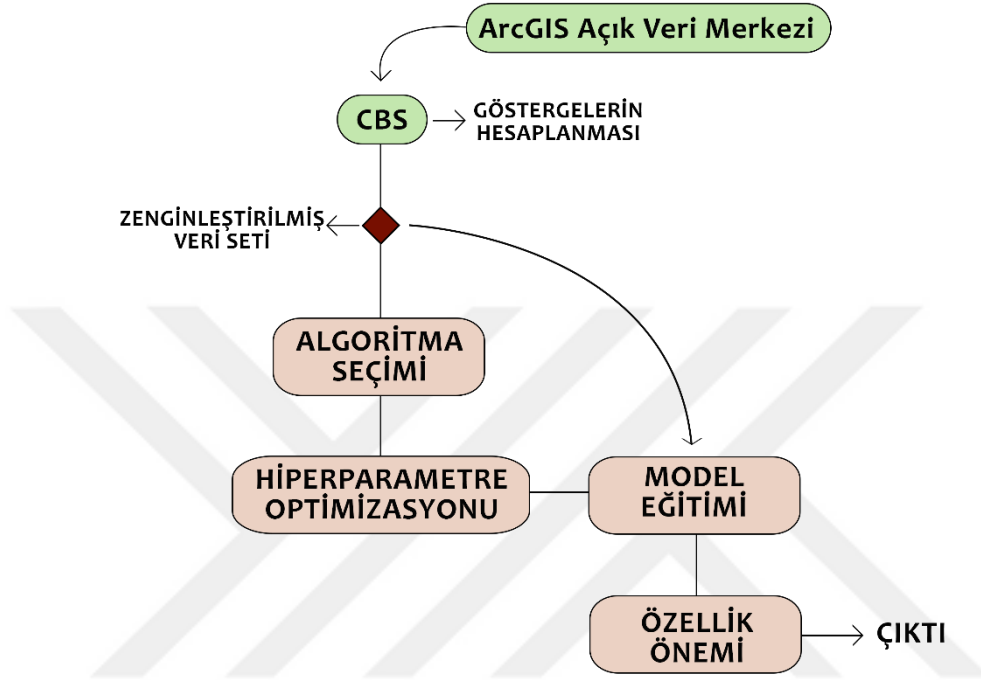


Şekil 3.25 Tez kapsamında yapılan sınıflandırma

3.3 Model Eğitimi ve Değerlendirilmesi

Tez çalışması kapsamında iki uygulama gerçekleştirilmiştir. Uygulama 1’de fonksiyonlar (1) konut ve (2) konut dışı olmak üzere iki sınıfta; Uygulama 2’de ise (1) konut, (2) ticaret, (3) kamu hizmet, (4) fabrika ve depo binaları olmak üzere dört sınıfta incelenmiştir (Şekil 3.25). Şekil 3.26’da her iki uygulamanın da akış şeması görülmektedir. ArcGIS’in açık kaynaklı veri setinden elde edilen ve CBS programı ile çeşitli göstergeler üretilen iki veri seti bir araya getirilerek zenginleştirilmiş veri seti elde edilir. Elde edilen veri seti her iki uygulama için seçilen algoritmalara girdi olarak sunulur. Hiper parametre optimizasyonu adımıyla algoritma için en iyi çalışan parametreler seçilir ve sonraki adımda seçilen parametreler ile algoritma yeniden çalıştırılır (model eğitimi). Göstergelerin performansa katkısının ölçüldüğü özellik önemi adımı yalnızca karar ağacı algoritmalarında (CART, rastgele orman ve aşırı gradyan arttırma) hesaplanmıştır. k-en yakın komşu ve destek vektör makinesi algoritmalarında özellik önemi doğrudan hesaplanabilir değildir. k-en yakın komşu algoritması tüm göstergeleri eşit oranda değerlendirir. Destek vektör makinesi algoritması ise özellik önemini çekirdek fonksiyonlara bağlı olarak hesaplar.

Bu çalışmada kullanılan makine öğrenmesi modellerinin eğitim ve değerlendirme süreçlerinde Python programlama dili ve Jupyter Notebook etkileşimli çalışma ortamı kullanılmıştır. Jupyter Notebook ile veri analizi, model geliştirme ve sonuçların görselleştirilmesi için esnek ve tekrarlanabilir bir çalışma alanı oluşturulmuştur.



Şekil 3.26 Model eğitimi ve değerlendirmesi akış şeması

3.3.1 Uygulama 1

Uygulamada 1700 bina ayak izi verisi kullanılmıştır. Dengeli ve gerçekçi bir veri seti için 1000 adet konut ve 700 adet konut dışı fonksiyonuna sahip bina ayak izi verisi olacak şekilde düzenlenmiştir. Bu uygulamada denetimli öğrenme algoritmalarından CART ve destek vektör makineleri (DVM), çoklu karar ağacı algoritmalarından rastgele orman (RO) ve aşırı gradyan artırma ve örnek tabanlı öğrenme algoritmalarından k-en yakın komşu algoritmaları seçilmiştir. Gösterge seçimi algoritmaların performansında önemli rol oynamaktadır. Gösterge değerlerinin dağılım grafikleri anlamlı ve ayırt edici göstergelerin belirlenmesinde yol gösterici olmuştur. 2.2.2 numaralı “Şekil Göstergeleri” başlığı altındaki dairesellik, dışbükeylik, dikdörtgensellik, karesellik göstergeleri ve bina alanı, kat adedi ve parsel alanı göstergeleri algoritmalara girdi olarak atanmıştır. En iyi

performansı gösteren hiper parametrenin bulunması için Grid Search yöntemi tüm algoritmalara uygulanmıştır.

CART algoritması için yapılan hiper parametre optimizasyonunda;

- max_depth : [3, 5, 7, 9] (maksimum derinlik)
- min_samples_split : [2, 5, 10] (bölünme için minimum örnek sayısı)
- min_samples_leaf : [1, 2, 4] (yapraktaki minimum örnek sayısı)

değerleri denenmiştir. 10-katlı çapraz doğrulama yöntemiyle yapılan deneylerde maksimum derinlik 5, bölünme için minimum örnek sayısı 2, yapraktaki minimum örnek sayısı 2 değerleri için F1 skoru en yüksek değere ulaşmıştır. CART algoritmasının veri ön işleme, hiper parametre optimizasyonu, model eğitim ve değerlendirme aşamalarının kod blokları Ek C’de yer almaktadır.

Destek vektör makineleri algoritması için yapılan hiper parametre optimizasyonunda;

- C : [0.1, 1] (ceza parametresi)
- kernel : ['poly', 'rbf', 'linear'] (çekirdek fonksiyonlar)

değerleri denenmiştir. 10-katlı çapraz doğrulama yöntemiyle yapılan deneylerde ceza parametresi 1, çekirdek fonksiyon 'linear' değerleri için F1 skoru en yüksek değere ulaşmıştır.

Rastgele orman algoritması için yapılan hiper parametre optimizasyonunda;

- n_estimators : [100, 200, 300] (karar ağacı sayısı)
- max_depth : [5, 10, 15] (maksimum derinlik)
- min_samples_split : [2, 5, 10] (bölünme için minimum örnek sayısı)
- min_samples_leaf : [1, 2, 4] (yapraktaki minimum örnek sayısı)

değerleri denenmiştir. 10-katlı çapraz doğrulama yöntemiyle yapılan deneylerde karar ağacı sayısı 300, maksimum derinlik 5, bölünme için minimum örnek sayısı 2 ve yapraktaki minimum örnek sayısı 2 değerleri için F1 skoru en yüksek değere ulaşmıştır.

Aşırı gradyan arttırma algoritması için yapılan hiper parametre optimizasyonunda;

- n_estimators : [100, 200, 300] (karar ağacı sayısı)
- max_depth : [3, 5, 7] (maksimum derinlik)

- learning_rate: [0.1, 0.01, 0.001] (öğrenme oranı)
- subsample : [0.8, 1] (örnek örnekleme oranı)
- colsample_bytree : [0.8, 1] (özellik örnekleme oranı)

değerleri denenmiştir. 10-katlı çapraz doğrulama yöntemiyle yapılan deneylerde karar ağacı sayısı 200, maksimum derinlik 3, öğrenme oranı 0.1, örnek örnekleme oranı 0.8 ve özellik örnekleme oranı 1 değerleri için F1 skoru en yüksek değere ulaşmıştır.

k-en yakın komşu algoritması için yapılan hiper parametre optimizasyonunda;

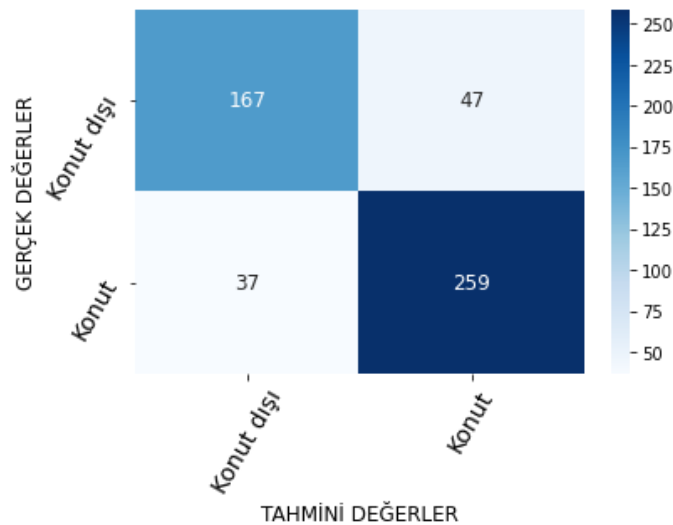
- n_neighbors: [3, 5, 7, 9] (en yakın komşu sayısı)
- metric : ['euclidean', 'manhattan'] (metrik)
- weights : ['uniform', 'distance'] (ağırlık)

değerleri denenmiştir. 10-katlı çapraz doğrulama yöntemiyle yapılan deneylerde en yakın komşu sayısı 7, metrik 'euclidean' ve ağırlık 'uniform' değerleri için F1 skoru en yüksek değere ulaşmıştır.

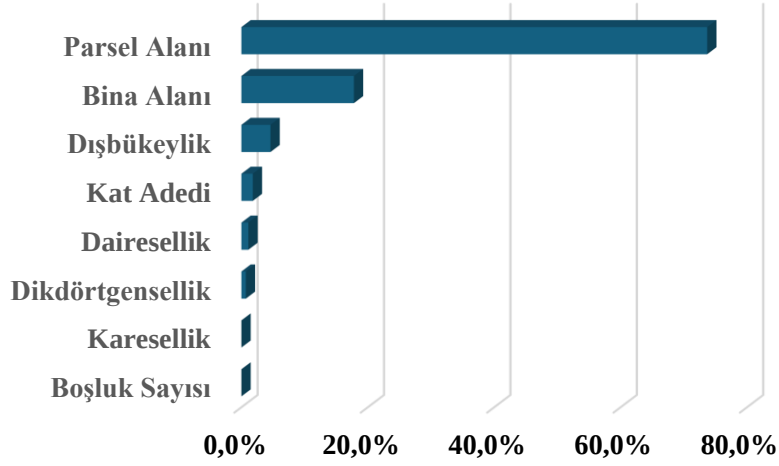
3.3.1.1 Bulgular

CART algoritmasının F1 skor değeri %83.46 ($\kappa = 0.659$) olarak bulunmuştur. Tablo 3.4'te CART algoritmasının hata matrisi görülmektedir. CART algoritması DVM algoritmasına göre daha başarılı performans göstermiştir. Ayrıca en başarılı performans gösteren algoritmadır.

Tablo 3.4 CART algoritmasının hata matrisi



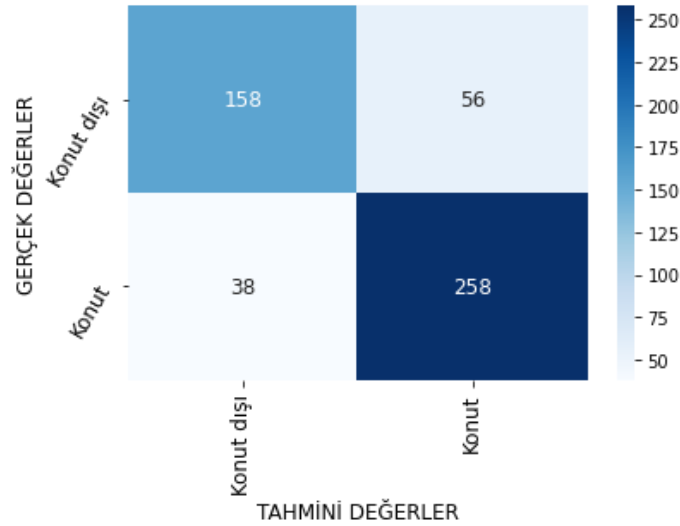
Göstergelerin CART algoritmasındaki önemi, Şekil 3.27’de görülmektedir. Bu şekil, her bir göstergenin modelin tahminlerini ne kadar etkilediğini gösteren önem değerlerini sunmaktadır.



Şekil 3.27 CART algoritması göstergelerinin önem değerleri

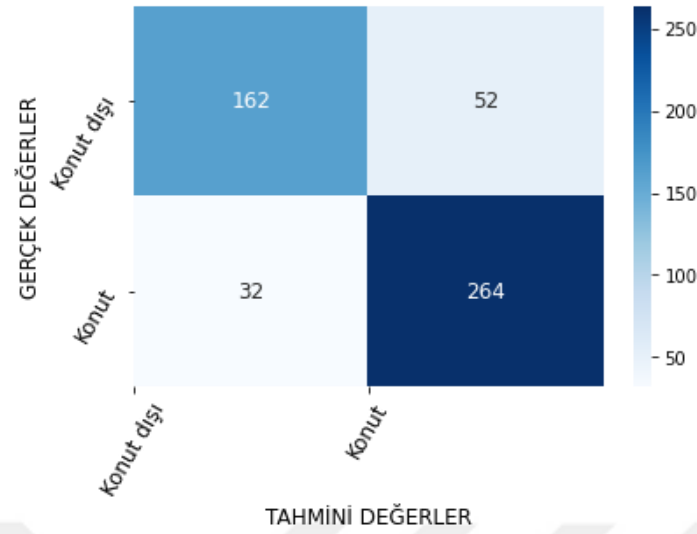
Destek vektör makinesi algoritmasının F1 skor değeri %81.43 ($\kappa = 0.617$) olarak bulunmuştur. Tablo 3.5’te algoritmanın hata matrisi görülmektedir.

Tablo 3.5 DVM algoritmasının hata matrisi

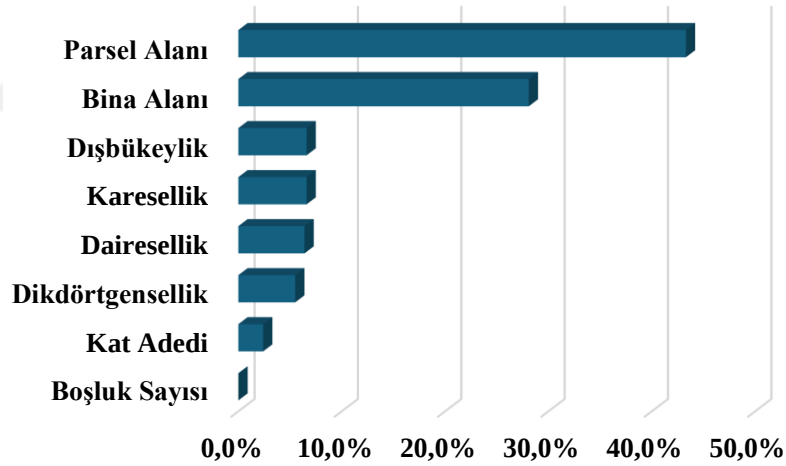


Rastgele orman algoritmasının F1 skor değeri %83.39 ($\kappa = 0.657$) olarak bulunmuştur. Tablo 3.6’da rastgele orman algoritmasının hata matrisi görülmektedir.

Tablo 3.6 Rastgele orman algoritmasının hata matrisi



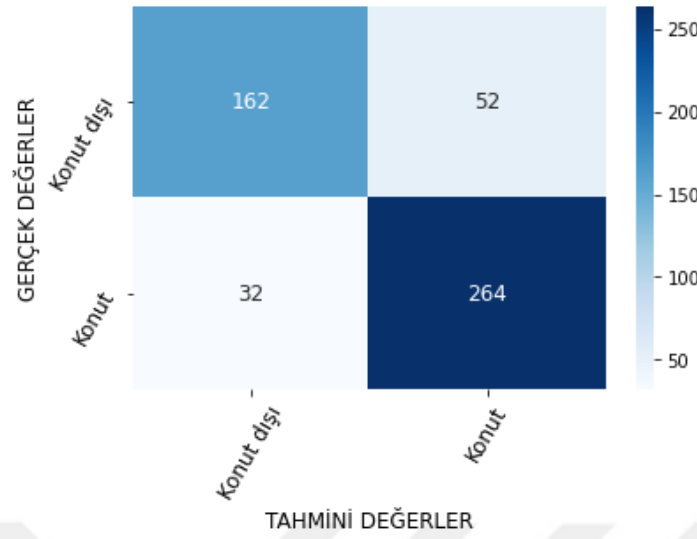
Göstergelerin rastgele orman algoritmasındaki önemi, Şekil 3.28’de görülmektedir. Bu şekil, her bir göstergenin modelin tahminlerini ne kadar etkilediğini gösteren önem değerlerini sunmaktadır.



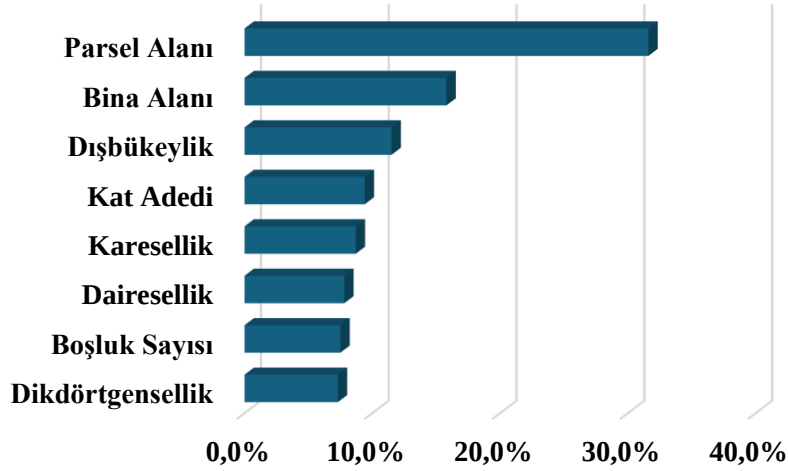
Şekil 3.28 Rastgele orman algoritması gösterge önem değerleri

Aşırı gradyan artırma algoritmasının F1 skor değeri %83.39 ($\kappa = 0.657$) olarak bulunmuştur. Tablo 3.7’de algoritmanın hata matrisi görülmektedir. Topluluk öğrenme algoritmaları aynı performansı göstermiştir. Ayrıca konut binalarının tespit edilmesinde CART algoritmasına göre bu iki algoritma daha başarılı performans göstermiştir.

Tablo 3.7 Aşırı gradyan arttırma algoritmasının hata matrisi



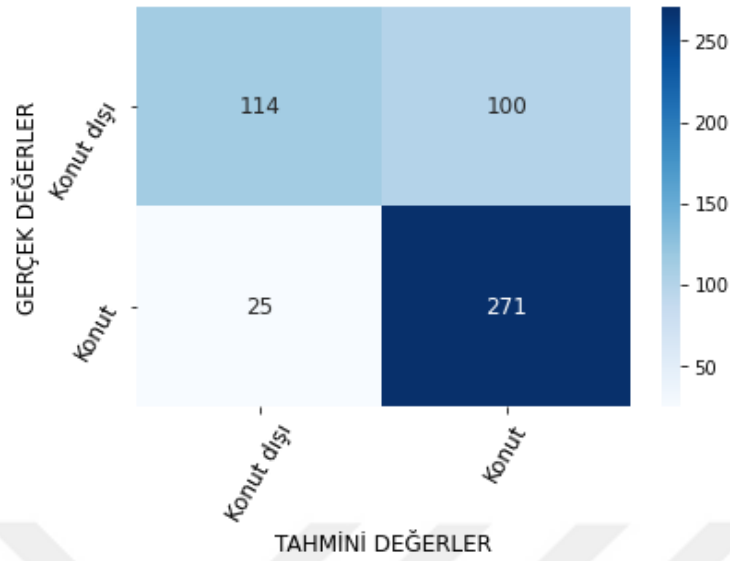
Göstergelerin aşırı gradyan arttırma algoritmasındaki önemi, Şekil 3.29’da görülmektedir. Bu şekil, her bir özelliğin modelin tahminlerini ne kadar etkilediğini gösteren önem değerlerini sunmaktadır.



Şekil 3.29 Aşırı gradyan arttırma algoritması gösterge önem değerleri

k-en yakın komşu algoritmasının F1 skor değeri %74.26 ($\kappa = 0.471$) olarak bulunmuştur. En başarısız performans bu algorithmada görülmüştür. Tablo 3.8’de algoritmanın hata matrisi görülmektedir.

Tablo 3.8 k-en yakın komşu algoritmasının hata matrisi



Tezde incelenen konut ve konut dışı bina sınıflarına ait doğruluk, kesinlik, hassasiyet, F1 skoru ve Kappa değerleri gibi performans metrikleri ek E bölümünde detaylı olarak sunulmuştur.

3.3.2 Uygulama 2

Uygulamada 3509 bina ayak izi verisi kullanılmıştır. Dengeli ve gerçekçi bir veri seti için 1000 adet konut, 1000 adet ticaret, 822 adet kamu hizmetleri ve 687 adet depo ve fabrika fonksiyonuna sahip bina ayak izi verisi olacak şekilde düzenlenmiştir. Bu uygulamada denetimli öğrenme algoritmalarından CART, çoklu karar ağacı algoritmalarından rastgele orman ve aşırı gradyan artırma ve örnek tabanlı öğrenme algoritmalarından k-en yakın komşu algoritmaları seçilmiştir. 2.2.2 numaralı “Şekil Göstergeleri” başlığı altındaki göstergeler, kat adedi, kat yüksekliği, parsel alanı, bina alanı ve yoğunluk hesapları algoritmaya değişken olarak atanmıştır. En iyi performansı gösteren hiper parametrenin bulunması için Grid Search yöntemi tüm algoritmalara uygulanmıştır.

CART algoritması için yapılan hiper parametre optimizasyonunda;

- max_depth : [3, 5, 7, 9] (maksimum derinlik)
- min_samples_split : [2, 5, 10] (bölünme için minimum örnek sayısı)
- min_samples_leaf : [1, 2, 4] (yapraktaki minimum örnek sayısı)

değerleri denenmiştir. 10-katlı çapraz doğrulama yöntemiyle yapılan deneylerde maksimum derinlik 7, bölünme için minimum örnek sayısı 2, yapraktaki minimum örnek sayısı 1 değerleri için F1 skoru en yüksek değere ulaşmıştır.

Rastgele orman algoritması için yapılan hiper parametre optimizasyonunda;

- n_estimators : [100, 200, 300] (karar ağacı sayısı)
- max_depth : [5, 10, 15] (maksimum derinlik)
- min_samples_split : [2, 5, 10] (bölünme için minimum örnek sayısı)
- min_samples_leaf : [1, 2, 4] (yapraktaki minimum örnek sayısı)

değerleri denenmiştir. 10-katlı çapraz doğrulama yöntemiyle yapılan deneylerde karar ağacı sayısı 300, maksimum derinlik 15, bölünme için minimum örnek sayısı 10 ve yapraktaki minimum örnek sayısı 1 değerleri için F1 skoru en yüksek değere ulaşmıştır. Rastgele orman algoritmasının veri ön işleme, hiper parametre optimizasyonu, model eğitim ve değerlendirme aşamalarının kod blokları Ek D’de yer almaktadır.

Aşırı gradyan arttırma algoritması için yapılan hiper parametre optimizasyonunda;

- n_estimators : [100, 200, 300] (karar ağacı sayısı)
- max_depth : [3, 5, 7] (maksimum derinlik)
- learning_rate: [0.1, 0.01, 0.001] (öğrenme oranı)
- subsample : [0.8, 1] (örnek örnekleme oranı)
- colsample_bytree : [0.8, 1] (özellik örnekleme oranı)

değerleri denenmiştir. 10-katlı çapraz doğrulama yöntemiyle yapılan deneylerde karar ağacı sayısı 100, maksimum derinlik 7, öğrenme oranı 0.1, örnek örnekleme oranı 1 ve özellik örnekleme oranı 0.8 değerleri için F1 skoru en yüksek değere ulaşmıştır.

k-en yakın komşu algoritması için yapılan hiper parametre optimizasyonunda;

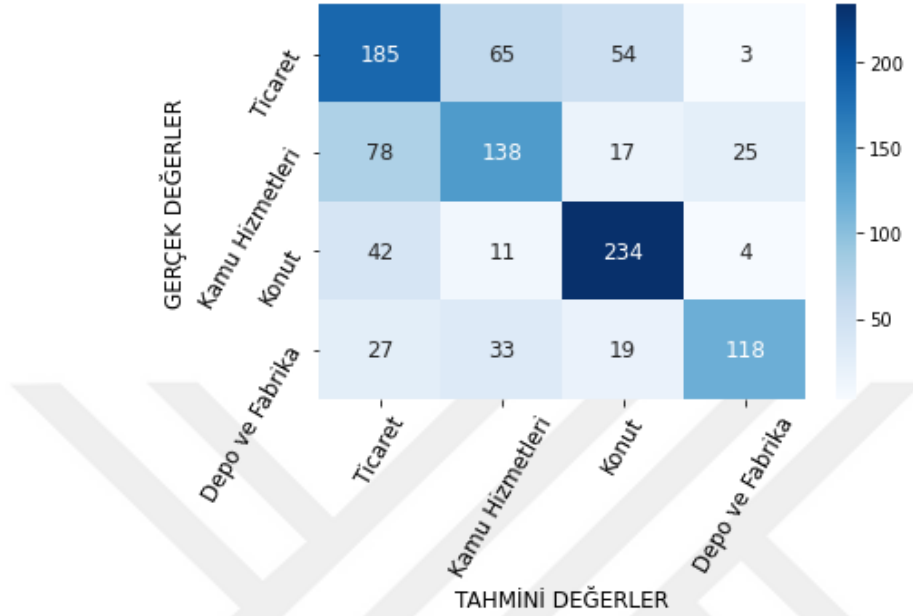
- n_neighbors: [3, 5, 7, 9] (en yakın komşu sayısı)
- metric : ['euclidean', 'manhattan'] (metrik)
- weights : ['uniform', 'distance'] (ağırlık)

değerleri denenmiştir. 10-katlı çapraz doğrulama yöntemiyle yapılan deneylerde en yakın komşu sayısı 9, metrik ‘euclidean’ ve ağırlık ‘uniform’ değerleri için F1 skoru en yüksek değere ulaşmıştır.

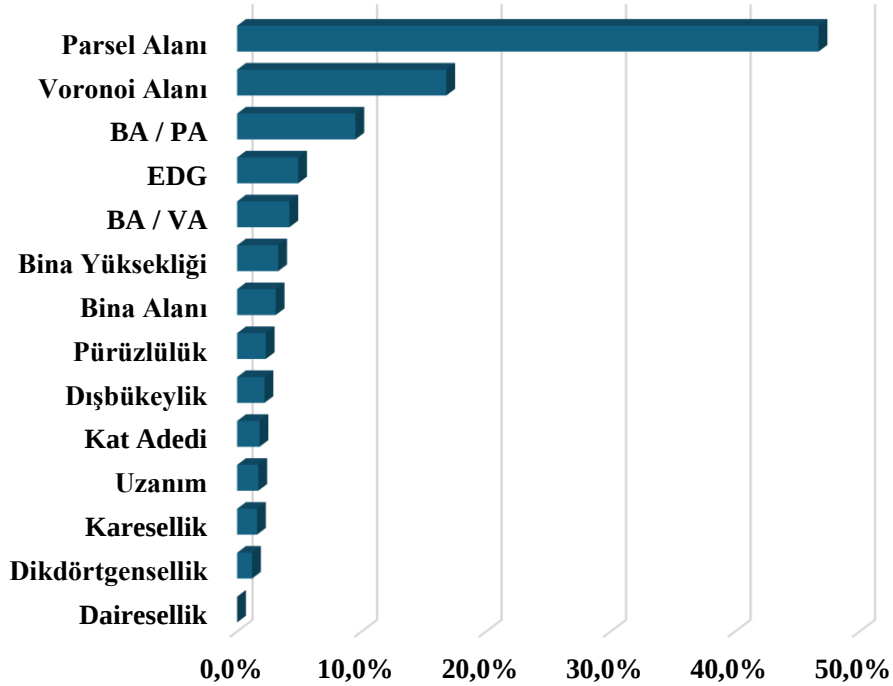
3.3.2.1 Bulgular

CART algoritmasının F1 skor değeri %63.02 ($\kappa = 0.514$) olarak bulunmuştur. Tablo 3.9’da algoritmanın hata matrisi görülmektedir.

Tablo 3.9 CART algoritmasının hata matrisi



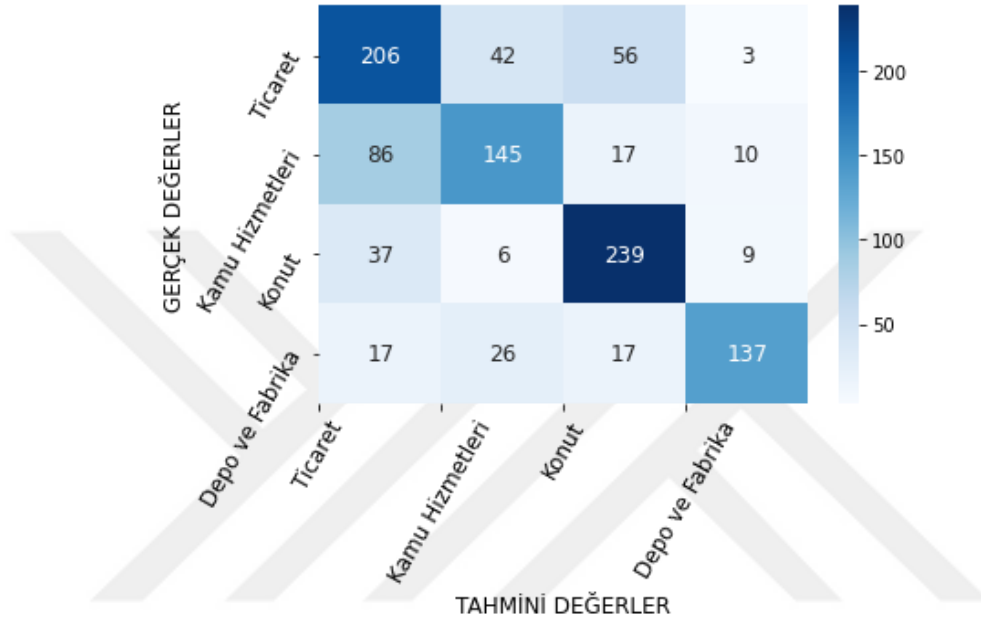
Göstergelerin CART algoritmasındaki önemi, Şekil 3.30’da görülmektedir. Bu şekil, her bir göstergenin modelin tahminlerini ne kadar etkilediğini gösteren önem değerlerini sunmaktadır.



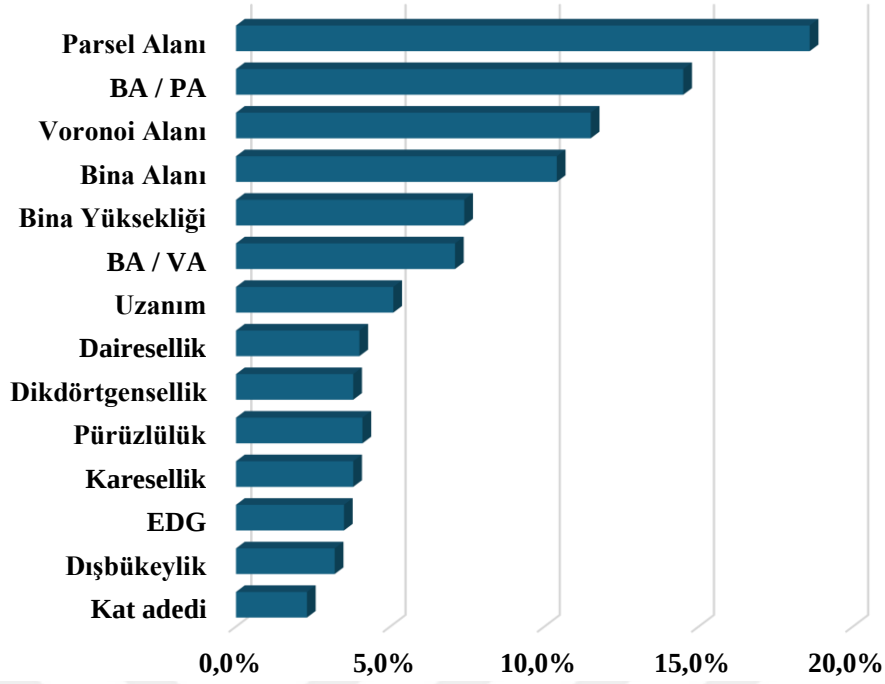
Şekil 3.30 CART algoritması gösterge önem değerleri, BA/PA: Bina Alanı/Parsel Alanı, BA/VA: Bina Alanı/Voronoi Alanı

Rastgele orman algoritmasının F1 skor değeri %68.99 ($\kappa = 0.580$) olarak bulunmuştur. Tablo 3.10’da algoritmanın hata matrisi görülmektedir. Bir diğer topluluk öğrenmesi algoritması olan aşırı gradyan artırma algoritmasına göre daha başarılı performans göstermiştir. Ayrıca en başarılı performans bu algorithmada görülmüştür.

Tablo 3.10 Rastgele orman algoritmasının hata matrisi



Göstergelerin rastgele orman algoritmasındaki önemi, Şekil 3.31’de görülmektedir. Bu şekil, her bir göstergenin modelin tahminlerini ne kadar etkilediğini gösteren önem değerlerini sunmaktadır.



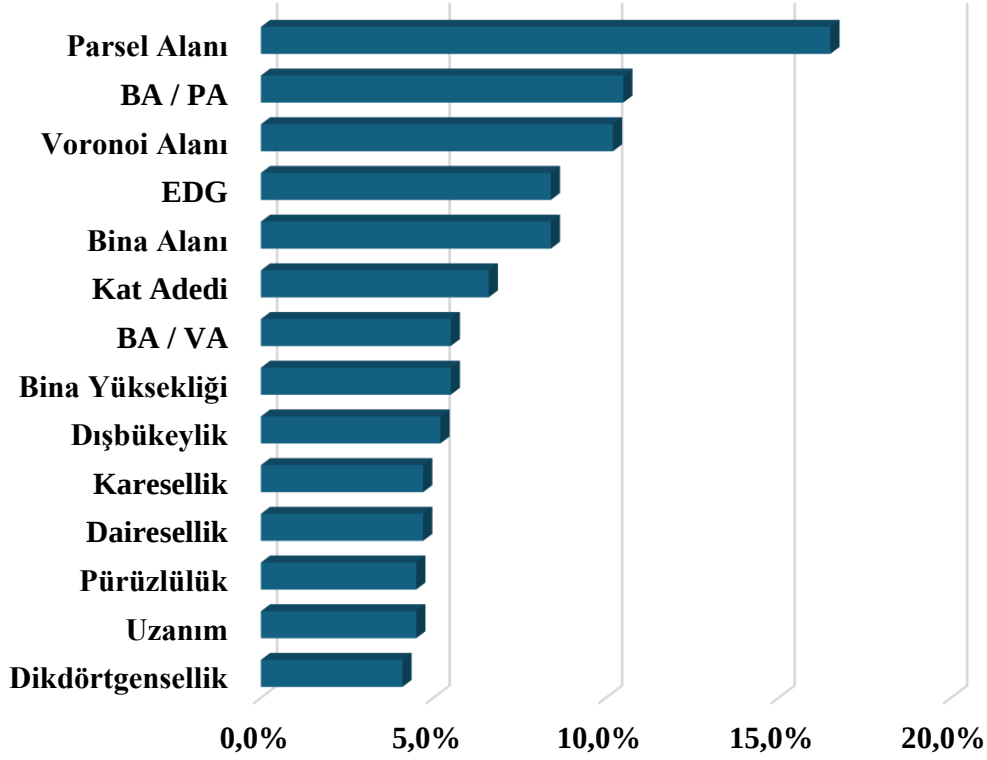
Şekil 3.31 Rastgele orman algoritması gösterge önem değerleri, BA/PA: Bina Alanı/Parsel Alanı, BA/VA: Bina Alanı/Voronoi Alanı

Aşırı gradyan arttırma algoritmasının F1 skor değeri %68.40 ($\kappa = 0.574$) olarak bulunmuştur. Tablo 3.11’de algoritmanın hata matrisi görülmektedir.

Tablo 3.11 Aşırı gradyan arttırma algoritmasının hata matrisi

GERÇEK DEĞER	Ticaret	Kamu Hizmetleri	Konut	Depo ve Fabrika
	202	40	62	3
	76	142	21	19
	38	7	239	7
TAHMİNİ DEĞER	Ticaret	Kamu Hizmetleri	Konut	Depo ve Fabrika
	19	24	15	139

Göstergelerin aşırı gradyan arttırma algoritmasındaki önemi, Şekil 3.32’de görülmektedir. Bu şekil, her bir göstergenin modelin tahminlerini ne kadar etkilediğini gösteren önem değerlerini sunmaktadır.



Şekil 3.32 Aşırı gradyan artırma algoritması gösterge önem değerleri, BA/PA:

Bina Alanı/Parsel Alanı, BA/VA: Bina Alanı/Voronoi Alanı

k-en yakın komşu algoritmasının F1 skor değeri %60.24 ($\kappa = 0.466$) olarak bulunmuştur. En başarısız performans bu algorithmada görülmüştür. Tablo 3.12’de algoritmanın hata matrisi görülmektedir.

Tablo 3.12 k-en yakın komşu algoritmasının hata matrisi

GERÇEK DEĞERLER	Ticaret	Kamu Hizmetleri	Konut	Depo ve Fabrika
	163	36	101	7
	82	115	50	11
	39	8	237	7
TAHMİNİ DEĞERLER	Ticaret	Kamu Hizmetleri	Konut	Depo ve Fabrika
	24	27	22	124

Tezde incelenen konut, ticaret, kamu hizmetleri, fabrika ve depo bina sınıflarına ait doğruluk, kesinlik, hassasiyet, F1 skoru ve Kappa değerleri gibi performans metrikleri ek F bölümünde ayrıntılı olarak sunulmuştur.

3.3.3 Tartışma

Her iki uygulama değerlendirildiğinde binalar iki fonksiyonel sınıfa ayrıldığında tüm algoritmaların daha başarılı performans gösterdiği görülmektedir. Ancak sınıf sayısı arttıkça algoritmaların performanslarının düştüğü görülmektedir. Her iki uygulamada da topluluk öğrenmesi algoritmaları daha başarılı performans gösterirken, örnek tabanlı bir algoritma olan k-en yakın komşu algoritması en başarısız performansı göstermiştir.

Konut binalarının tespit edilmesi tüm algoritmalarda en yüksek doğruluk oranını göstermiştir. Konut binalarını fabrika ve depo, ticaret ve kamu hizmet binaları takip etmektedir. Hata matrisleri incelendiğinde CART, rastgele orman ve aşırı gradyan artırma algoritmaları gerçekte kamu hizmet sınıfına ait olan bazı binaları ticaret sınıfı; ticaret sınıfına ait olan bazı binaları ise kamu hizmet ve konut sınıfı olarak tahmin etmiştir. k-en yakın komşu algoritması ise gerçekte kamu hizmet sınıfına ait olan bazı binaları ticaret ve konut sınıfı; ticaret sınıfına ait olan bazı binaları ise konut sınıfı olarak tahmin etmiştir.

Şekil göstergelerinin algoritmaların performanslarına olan katkıları incelendiğinde; CART algoritmasının her iki uygulamada da tek bir özelliğe (parsel alanı) daha yüksek oranda odaklandığı görülmüştür. Bu durum veri setinde karmaşıklık (fonksiyonel sınıf sayısı) arttığında performans düşüklüğüne yol açmıştır. Uygulama 1’de CART algoritması en başarılı algoritma iken Uygulama 2’de en başarısız algoritmalarından biridir. RO ve XGBoost algoritmalarının daha fazla sayıda özelliği yakın oranda dikkate aldığı görülmüştür. Bu durum karmaşık veri setlerinde daha başarılı olmalarını sağlamıştır. Her iki algoritmada Uygulama 2’de diğer algoritmalarla göre daha başarılı performans göstermiştir.

Uygulama 1’de algoritmaların göstergelerinin önem değerleri farklılık gösterse de üç algoritma içinde en önemli göstergeler parsel alanı, bina alanı ve dışbükeylik olmuştur. Boşluk sayısı CART ve RO algoritmalarında performansa hiçbir katkı sağlamamıştır. Uygulama 2’de de farklı gösterge önem değerleri olmasına rağmen en önemli göstergeler parsel alanı, Voronoi alanı (bina etki alanı) ve BA/PA oranı

olmuştur. Dairesellik CART algoritmasının performansına hiçbir katkı sağlamamıştır. RO ve XGBoost algoritmalarına %5'in altında katkı sağlayan göstergeler; dikdörtgensellik, karesellik, dairesellik ve pürüzlülüktür.



4 SONUÇ

Bu çalışmada bina fonksiyonlarının sınıflandırılmasında makine öğrenmesi yöntemlerinin sınıflandırma performansına olan etkileri araştırılmıştır. Bu amaçla, binaların mekânsal karakteristikleri (büyüklük, şekil, mekânsal ilişkiler ve nicel öznelilikler) kullanılmıştır. Florida'nın Miami-Dade ili çalışma alanı olarak belirlenmiştir. Veri seti ArcGIS'in açık kaynaklı platformundan elde edilmiştir. Sadece konut ve konut dışı binalarının ve ayrıca dört farklı bina fonksiyonun (konut, ticaret, kamu hizmetleri, fabrika ve depo) tespit edilerek sınıflandırıldığı iki uygulama gerçekleştirilmiştir. Bu kapsamda şu makine öğrenmesi algoritmaları kullanılmıştır: Uygulama 1 (CART, destek vektör makineleri, rastgele orman, aşırı gradyan arttırma ve k - En Yakın Komşu), uygulama 2 (CART, rastgele orman, aşırı gradyan arttırma ve k – En Yakın Komşu).

Sınıf ve gösterge sayısının daha az olduğu Uygulama 1'de tüm algoritmalar Uygulama 2'ye göre daha başarılı performans göstermiştir. Her iki uygulamada da topluluk öğrenmesi algoritmalarıyla daha iyi sonuçlar elde edilmiştir. k-en yakın komşu algoritması ise her iki uygulamada en başarısız performansı göstermiştir.

Gösterge sayısının artması topluluk öğrenmesi algoritmalarının performansını arttırmıştır. Daha fazla göstergede CART algoritması sadece bir göstergeye daha fazla önem verirken rastgele orman ve aşırı gradyan arttırma algoritmaları tüm göstergeleri yakın oranda performansa katmıştır.

Her iki uygulamada da tüm algoritmalar konut sınıfını tespit etmede başarılı iken, ticaret ve kamu hizmetleri sınıflarını tespit etmede başarısız olmuşlardır. Bu sınıfların tespit edilmesinde coğrafi referanslı sosyal medya, sokak görünümü ve taksi yörüngesi verileri gibi verilerin çalışmalara dahil edilmesiyle elde edilen sonuçların desteklenmesi sağlanabilir.

KAYNAKÇA

- Alzubi, J., Nayyar, A. ve Kumar, A. (2018). Machine Learning from Theory to Algorithms: An Overview. *Journal of Physics Conference Series*, 1142, 012012. <https://doi.org/10.1088/1742-6596/1142/1/012012>
- ArcGIS Hub. (2024a, Mart 20). *Building Footprint Ubid Folio*. https://hub.arcgis.com/datasets/7abe7f01391b4113b05030414669fef5_0/explore
- ArcGIS Hub. (2024b, Mayıs 26). *Edge of Pavement 2019*. https://hub.arcgis.com/datasets/1e648aef7989413a9cccea7333b782f4_0/explore
- Basaraner, M. (2020). Geometric and semantic quality assessments of building features in OpenStreetMap for some areas of Istanbul. *Polish Cartographical Review*, 52(3), 94–107. <https://doi.org/10.2478/pcr-2020-0010>
- Basaraner, M. ve Cetinkaya, S. (2017). Performance of shape indices and classification schemes for characterising perceptual shape complexity of building footprints in GIS. *International Journal of Geographical Information Science*, 31(10), 1952-1977. <https://doi.org/10.1080/13658816.2017.1346257>
- Bashmakov, I.A., Nilsson, L.J., Acquaye, A., Bataille, C., Cullen, J.M., de la Rue du Can, S., Fishedick, M., Geng, Y., Tanaka, K. (2022). Industry. In P.R. Shukla, J. Skea, R. Slade, A. Al Khouradje, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz, J. Malley, (eds.) IPCC, 2022: Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK and New York, NY, USA. <https://doi.org/10.1017/9781009157926.013>
- Başaraner, M. ve Selçuk, M. (2004). An Attempt to Automated Generalization of Buildings and Settlement Areas in Topographic Maps. *XXth ISPRS Congress*, 12-23 July 2004, Istanbul. <https://www.isprs.org/proceedings/xxxv/congress/comm4/papers/340.pdf>
- Bilgen, O. B. ve Doğan, N. (2017). Puanlayıcılar arası güvenilirlik belirleme tekniklerinin karşılaştırılması. *Eğitimde Ve Psikolojide Ölçme Ve Değerlendirme Dergisi*, 63. <https://doi.org/10.21031/epod.294847>
- Da Fona Costa, L. ve Cesar, R. M., Jr. (2018). Shape Classification and analysis. In *CRC Press eBooks*. <https://doi.org/10.1201/9781315222325>

- Data USA. (2024, Kasım 19). *Miami-Dade County*. <https://datausa.io/profile/geo/miami-dade-county-fl>
- Du, S., Zheng, M., Guo, L., Wu, Y., Li, Z. ve Liu, P. (2024). Urban building function classification based on multisource geospatial data: a two-stage method combining unsupervised and supervised algorithms. *Earth Science Informatics*, 17(2), 1179–1201. <https://doi.org/10.1007/s12145-024-01250-5>
- Esposito, F., Malerba, D., Semeraro, G. ve Kay, J. (1997). A comparative analysis of methods for pruning decision trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5), 476–493. <https://doi.org/10.1109/34.589207>
- Florida State Office of Economic and Demographic Research. (2024). *The Florida Legislature Econographic News*. Florida State Office of Economic and Demographic Research.
- Haitao, L., Haiyan, G., Yanshun, H., ve Jinghui, Y. (2007, Ocak). SVM Tabanlı Nesne Yönelimli Kentsel Arazi Örtüsü Sınıflandırması için Yüksek Çözünürlüklü Hava Görüntüleri ve LIDAR Verilerinin Füzyonu. *Coğrafi Mekansal Veritabanlarını Güncelleme ISPRS Çalıştayı*, 178-184.
- Hecht, R., Meinel, G. ve Buchroithner, M. (2015). Automatic identification of building types based on topographic databases – a comparison of different data sources. *International Journal of Cartography*, 1(1), 18–31. <https://doi.org/10.1080/23729333.2015.1055644>
- Kilic, S. (2015). Kappa test. *Journal of Mood Disorders*, 5(3), 142. <https://doi.org/10.5455/jmood.20150920115439>
- Kong, B., Ai, T., Zou, X., Yan, X. ve Yang, M. (2024). A graph-based neural network approach to integrate multi-source data for urban building function classification. *Computers Environment and Urban Systems*, 110, 102094. <https://doi.org/10.1016/j.compenvurbsys.2024.102094>
- Kulkarni, S. ve Kelkar, V. (2014). Classification of Multispectral Satellite Images Using Ensemble Techniques of Bagging, Boosting and AdaBoosting. *International Conference on Circuits, Systems, Communication and Information Technology Applications*, 253-258. <https://doi.org/10.1109/CSCITA.2014.6839268>
- Leymarie, F. F. (2006). Thoughts on shape. In *Advances in consciousness research* (pp. 303–350). <https://doi.org/10.1075/aicr.67.19ley>
- Li, K., Li, Y., Yang, X., Liu, X. ve Huang, Q. (2022). Comparing the three-dimensional morphologies of urban buildings along the urban-rural gradients of 91 cities in China. *Cities*, 133, 104123. <https://doi.org/10.1016/j.cities.2022.104123>
- Lin, A., Sun, X., Wu, H., Luo, W., Wang, D., Zhong, D., Wang, Z., Zhao, L. ve Zhu, J. (2021). Identifying urban building function by integrating remote sensing imagery and POI data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 8864–8875. <https://doi.org/10.1109/jstars.2021.3107543>
- Loncaric, S. (1998). A survey of shape analysis techniques. *Pattern Recognition*, 31(8), 983–1001. [https://doi.org/10.1016/s0031-2023\(97\)00122-2](https://doi.org/10.1016/s0031-2023(97)00122-2)

- Machine learning*. (2011). <https://doi.org/10.4018/978-1-60960-818-7>
- Miami-Dade County. (2024, Kasım 19). *About Miami-Dade County*. <https://www.miamidade.gov/global/disclaimer/about-miami-dade-county.page>
- Mienye, I. D. ve Jere, N. (2024). A survey of Decision trees: Concepts, algorithms, and applications. *IEEE Access*, 12, 86716-86727. <https://doi.org/10.1109/access.2024.3416838>
- Mienye, I. D. ve Sun, Y. (2022). A survey of ensemble learning: Concepts, algorithms, applications, and prospects. *IEEE Access*, 10, 99129–99149. <https://doi.org/10.1109/access.2022.3207287>
- Mohammed, A. ve Kora, R. (2023). A comprehensive review on ensemble deep learning: Opportunities and challenges. *Journal of King Saud University - Computer and Information Sciences*, 35(2), 757–774. <https://doi.org/10.1016/j.jksuci.2023.01.014>
- Niu, N., Liu, X., Jin, H., Ye, X., Liu, Y., Li, X., Chen, Y. ve Li, S. (2017). Integrating multi-source big data to infer building functions. *International Journal of Geographical Information Science*, 31(9), 1871–1890. <https://doi.org/10.1080/13658816.2017.1325489>
- Patel, N. ve Upadhyay, S. (2012). Study of Various Decision Tree Pruning Methods with their Empirical Comparison in WEKA. *International Journal of Computer Applications*, 60(12), 20–25. <https://doi.org/10.5120/9744-4304>
- Rokach, L. ve Maimon, O. (2013). Data Mining with Decision Trees. In *Series in machine perception and artificial intelligence*. <https://doi.org/10.1142/9097>
- Schirmer, P. M. ve Axhausen, K. W. (2015). A multiscale classification of urban morphology. *Journal of Transport and Land Use*. <https://doi.org/10.5198/jtlu.2015.667>
- Tornay, N., Schoetter, R., Bonhomme, M., Faraut, S. ve Masson, V. (2017). GENIUS: A methodology to define a detailed description of buildings for urban climate and building energy consumption simulations. *Urban Climate*, 20, 75–93. <https://doi.org/10.1016/j.uclim.2017.03.002>
- Toussaint, G. (1983). Solving geometric problems with the rotating calipers. In *Proceedings of IEEE MELECON'83*. Athens, Greece
- Uddin, S., Haque, I., Lu, H., Moni, M. A. ve Gide, E. (2022). Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. *Scientific Reports*, 12(1). <https://doi.org/10.1038/s41598-022-10358-x>
- United States Census Bureau. (2024). *QuickFacts United States*. <https://data.census.gov>
- USAFACTS. (2024, Kasım 19). *Our Changing Population: Miami-Dade County, Florida*. USAFACTS: <https://usafacts.org/data/topics/peoplesociety/population-and-demographics/ourchangingpopulation/state/florida/county/miami-dadecounty/?endDate=2020-01-01&startDate=2010-01-01>

- Vujovic, Ž. Đ. (2021). Classification model evaluation metrics. *International Journal of Advanced Computer Science and Applications*, 12(6). <https://doi.org/10.14569/ijacsa.2021.0120670>
- World Urbanization Prospects: The 2018 Revision. (2019). <https://doi.org/10.18356/b9e995fe-en>
- Xiao, B., Jia, X., Yang, D., Sun, L., Shi, F., Wang, Q. ve Jia, Y. (2022). Research on Classification Method of Building Function Oriented to Urban Building Stock Management. *Sustainability*, 1-3.
- Yoshida, H. ve Omae, M. (2004). An approach for analysis of urban morphology: methods to derive morphological properties of city blocks by using an urban landscape model and their interpretations. *Computers Environment and Urban Systems*, 29(2), 223-247. <https://doi.org/10.1016/j.compenvurbsys.2004.05.008>
- Zhang, Y., Qin, K., Bi, Q., Cui, W. ve Li, G. (2020). Landscape Patterns and Building Functions for Urban Land-Use Classification from Remote Sensing Images at the Block Level: A Case Study of Wuchang District, Wuhan, China. *Remote Sensing*, 12(11), 1831. <https://doi.org/10.3390/rs12111831>
- Zhou, Z.-H. (2012) Ensemble Methods: Foundations and Algorithms. Chapman and Hall/CRC Press.
- Zhou, Z. (2021). *Machine learning*. Springer.

A

SINIFLANDIRMA İÇİN LİTERATÜR ARAŞTIRMASI

Araştırmacılar	Çalışmanın Amacı	Fonksiyonel Sınıflandırma	
		KONUT	Konut alanları, villa, yurt ve gecekondur
Bing Xiao; Xuexiu Jia; Dong Yang; Lingwen Sun; Feng Shi; Qitong Wang; Yongfei Jia	Binaların kaynak tüketimlerinin çevre üzerindeki etkilerini incelemek	TİCARET	Merkezi iş bölgeleri, restoran, bar, otel, alışveriş merkezler, kuaför, güzellik merkezleri, spor tesisleri, araç satın alma ve kiralama alanları, internet kafe, sinema, konser salonları, tiyatro, ofis, ticari binalar, seyahat acenteleri, bilim ve teknoloji parkları ve endüstriyel parklar
		KAMU	Hükümet binaları, organizasyonlar, sağlık kurumları, kültür ve spor tesisleri, radyo ve televizyon kurumları, sergi salonları, sanat galerileri, hastane, hayvan hastaneleri, sağlık ve eğitim kurumları, yüksek eğitim kurumları, ilkokul, ortaokul, bilimsel araştırma enstitüleri, eğlence parkları, park, müze, galeri, kilise vb.

Araştırmacılar	Çalışmanın Amacı	Fonksiyonel Sınıflandırma	
		KONUT	Konut, hizmet tesisleri
Anqi Lin; Xiaomeng Sun; Hao Wu; Wenting Luo; Danyang Wang; Dantong Zhong; Zhongming Wang; Lanting Zhao; Jiang Zhu	Arazi kullanım haritası oluşturmak	YÖNETİM VE KAMU HİZMETLERİ	Yönetim ofisleri, kitap ve sergi salonları, kültür tesisleri, üniversite ve liseler, ilkokul ve ortaokullar, stadyum ve spor salonları, hastaneler, dini tesisler
		TİCARET VE İŞ TESİSLERİ	Perakende ve toptan işletmeler, yeme-içme tesisleri, otel, finans ve sigorta tesisleri, eğlence tesisleri, rekreasyon ve spor tesisleri, benzin istasyonları vb.
		ENDÜSTRİ VE ÜRETİM	Fabrikalar
		LOJİSTİK VE DEPO	Lojistik depolama tesisleri
		YOL, SOKAK VE ULAŞIM	Kentsel yollar, kentsel tren istasyonları, ulaşım merkezleri, kamusal ulaşım istasyonları, park alanları vb.
		BELEDİYE TESİSLERİ	Su, enerji, gaz, ısı kaynakları, iletişim tesisleri, radyo ve televizyon tesisleri, drenaj tesisleri, temizlik işleri tesisleri, çevre koruma tesisleri, itfaiyeler,

B

İNDİRİLEN VERİ SETİNİN SINIFLANDIRMASI

FONKSİYONLAR	ALT FONKSİYONLAR
10 BİRİMDEN FAZLA ÇOK AİLELİK KONUTLAR	
2-9 BİRİM ARASI ÇOK AİLELİK KONUTLAR	
AÇIK DEPO TESİSLERİ	
AĞIR ENDÜSTRİYEL ALANLAR	Ticari tesisler Ağır endüstriyel veya kereste atölyesi
BELEDİYEYE AİT TESİSLER	Klüp veya salon / özel tesisler Kültürel tesisler Karışık kullanım - hükümet tesisleri Çok ailelik konutlar Belediye tesisleri Perakende satış merkezleri
BÖLGESEL ALIŞVERİŞ MERKEZLERİ	
BÜYÜK MAĞAZALAR	
DEPOLAMA TERMİNALLERİ	
DEVLET HASTANELERİ	
DEVLETE AİT BİNALAR	
DİNİ TESİSLER	
EĞİTİM TESİSLERİ	
FEDERAL DEVLETE AİT BİNALAR	Hükümet binaları Ofis binaları Kamu yönetim binaları Tek ailelik konutlar
FİNANSAL KURUMLAR	
GECE KLUBÜ YADA BARLAR	
GIDA İŞLEME TESİSLERİ	
HAFİF İMALAT TESİSLERİ	Gıda işleme tesisleri
HAVALANI, TERMİNAL YADA MARİNA	
HİZMET İSTASYONLARI	
HOTEL YADA MOTELLER	
HUZUREVLERİ	
İLÇEDEKİ BİNALAR	Yaşam alanları Kulüp veya salon / özel tesisler Toplu konutlar Ticari tesisler Çok ailelik konutlar Ofis binaları Ceza kurumu Tek ailelik konutlar Kamu hizmet binaları Depolama binaları

KAMU EĞİTİM KURULU	
KAMU HİZMET BİNALARI	Ofis binası Kamu hizmet binaları
KAMU YARARI İÇİN OLAN KURULUŞLAR	Sağlık hizmetleri Çok ailelik konutlar Ofis binaları
KAPALI OYUN ALANLARI	
KAPALI TİYATRO VE ODİTORYUM	
KARIŞIK KULLANIM - KONUT ALANLARI	
KARIŞIK KULLANIM - MAĞAZA VE KONUT BİNALARI	Karışık kullanım - ticaret Karışık kullanım - konut Perakende satış merkezleri
KIRSAL ALANDAKİ BİNALAR	
KONSERVE YADA DAMITIM FABRİKASI	Yaşam alanları Otomotiv yada marina tesisleri Konserve fabrikaları Klüp veya salon / özel tesisler Ağır endüstriyel veya kereste atölyeleri Gıda işleme tesisleri Karışık kullanım - ticari alanlar Karışık kullanım - konut alanları Mobil evler Çok ailelik konutlar Ofis binaları Perakende satış alanları Kamu hizmet binaları Depo veya depolama tesisleri Toptan satış merkezleri
KONUT BİNALARI	
MAĞAZALAR	
MİNERAL İŞLEME TESİSLERİ	Mineral işleme Karışık kullanım - endüstri
MOBİL EVLER	
OFİS BİNALARI	
ORTAK ALANLAR	
ORTAK ALIŞVERİŞ MERKEZLERİ	
ÖZEL HASTANELER	
PAKETLEME TESİSLERİ	Gıda işleme tesisleri
PROFESYONEL HİZMET BİNALARI	Sağlık hizmeti tesisleri Karışık kullanım - ticari binalar Ofis binaları
RESTORAN YADA KAFELER	
SENATARYUMLAR	
SİGORTA ŞİRKETLERİ	Ofis binaları
SÜPERMARKETLER	
TAMİRAT DÜKKANLARI	

TARIM ALANINDAKİ BİNALAR	Tek ailelik konutlar
TEK AİLELİK KONUTLAR	1 birimli tek ailelik konutlar
	Tek ailelik konut (ek yaşam alanlı)
	Tek ailelik küme evler
	Tek ailelik konut
TİYATRO VE STADYUMLAR	
TOPTAN SATIŞ TESİSLERİ	
TURİST CAZİBE MERKEZLERİ	
ÜNİVERSİTE YADA KOLEJLER	Özel eğitim
	Devlet okulları



UYGULAMA 1: CART ALGORİTMASI KODLARI VE OPTİMİZASYON SÜRECİ

```
# Kütüphaneleri yükleme
from sklearn.model_selection import GridSearchCV
from sklearn.tree import DecisionTreeClassifier
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, precision_score,
recall_score, f1_score, confusion_matrix, cohen_kappa_score
from sklearn.preprocessing import LabelEncoder

# Veri yükleme
data = pd.read_excel("VERI_DOSYASI_YOLU")

# Veri setini özellik ve sınıf verisi olarak ayırma
X = data.drop(["FID", "Objectid", "Fonk_Sinif", "Bina_alani",
               "Voro_alani", "Bina_yüksek", "BA_Parsela", "BA_VoroA", "RI",
               "ERI", "ELO"], axis = 1)
y = data["Fonk_Sinif"]

# Veri kümesini eğitim ve test verisi olarak ayırma
X_train, X_test, y_train, y_test = train_test_split(X, y,
                                                    test_size = 0.3, random_state = 42)

# Hiper parametre için parametre aralıklarını tanımlama
param_grid = {
    'max_depth': [3, 5, 7, 9],
    'min_samples_split': [2, 5, 10],
    'min_samples_leaf': [1, 2, 4]
}

# Model oluşturma
clf = DecisionTreeClassifier()

# Grid Search ile en iyi model parametrelerini belirleme
grid_search = GridSearchCV(clf, param_grid, cv = 10)
grid_search.fit(X_train, y_train)
```



```
# En iyi parametreleri eğitme
best_model = grid_search.best_estimator_
best_model.fit(X_train,y_train)

# Tahmin ve Değerlendirme
y_pred = best_model.predict(X_test)
y_true = y_test

# Metrik hesaplama
accuracy = accuracy_score(y_true, y_pred)
precision = precision_score(y_true, y_pred, average =
'weighted')
recall = recall_score(y_true, y_pred, average = 'weighted')
f1 = f1_score(y_true, y_pred, average = 'weighted')
kappa = cohen_kappa_score(y_true, y_pred)
```

D

UYGULAMA 2: RASTGELE ORMAN ALGORİTMASI KODLARI VE OPTİMİZASYON SÜRECİ

```
# Kütüphaneleri yükleme
from sklearn.model_selection import train_test_split
import pandas as pd
from sklearn.model_selection import GridSearchCV
from sklearn.ensemble import RandomForestClassifier
from sklearn.preprocessing import LabelEncoder

# Veri yükleme
data = pd.read_excel("VERI_DOSYASI_YOLU")

# Veri setini özellik ve sınıf verisi olarak ayırma
X = data.drop(["FID", "Objectid", "Fonk_Sinif", "Bina_alani",
               "Voro_alani", "Boşluk_Sayi", axis = 1])
y = data["Fonk_Sinif"]

# Veri kümesini eğitim ve test verisi olarak ayırma
X_train, X_test, y_train, y_test = train_test_split(X, y,
                                                    test_size = 0.3, random_state = 42)

# Hiper parametre için parametre aralıklarını tanımlama
param_grid = {
    'n_estimators': [100, 200, 300],
    'max_depth': [5, 10, 15],
    'min_samples_split': [2, 5, 10],
    'min_samples_leaf': [1, 2, 4]
}

# Model oluşturma
rf = RandomForestClassifier()

# Grid Search ile en iyi model parametrelerini belirleme
grid = GridSearchCV(estimator = rf, param_grid = param_grid,
                    cv=10)
```

```
# En iyi parametreleri eğitme
best_model_fit=
RandomForestClassifier(**best_params)best_model.fit(X_train,
y_train)
best_model_fit.fit(X_train, y_train)

# Tahmin ve Değerlendirme
y_pred = best_model.predict(X_test)
y_true = y_test

# Metrik hesaplama
accuracy = accuracy_score(y_true, y_pred)
precision = precision_score(y_true, y_pred, average =
'weighted')
recall = recall_score(y_true, y_pred, average = 'weighted')
f1 = f1_score(y_true, y_pred, average = 'weighted')
kappa = cohen_kappa_score(y_true, y_pred)
```

UYGULAMA 1: SINIFLARIN PERFORMANS METRİKLERİ

CART				
	<i>Doğruluk</i>	<i>Kesinlik</i>	<i>Duyarlılık</i>	<i>F1-Skor</i>
KONUT	0,835	0,846	0,875	0,860
KONUT DIŞI	0,835	0,819	0,780	0,799

Destek Vektör Makineleri				
	<i>Doğruluk</i>	<i>Kesinlik</i>	<i>Duyarlılık</i>	<i>F1-Skor</i>
KONUT	0,816	0,822	0,872	0,846
KONUT DIŞI	0,816	0,806	0,738	0,771

Rastgele Orman				
	<i>Doğruluk</i>	<i>Kesinlik</i>	<i>Duyarlılık</i>	<i>F1-Skor</i>
KONUT	0,835	0,835	0,892	0,863
KONUT DIŞI	0,835	0,835	0,757	0,794

XGBoost				
	<i>Doğruluk</i>	<i>Kesinlik</i>	<i>Duyarlılık</i>	<i>F1-Skor</i>
KONUT	0,835	0,835	0,892	0,863
KONUT DIŞI	0,835	0,835	0,757	0,794

k-En Yakın Komşu				
	<i>Doğruluk</i>	<i>Kesinlik</i>	<i>Duyarlılık</i>	<i>F1-Skor</i>
KONUT	0,755	0,730	0,916	0,813
KONUT DIŞI	0,755	0,820	0,533	0,646

UYGULAMA 2: SINIFLARIN PERFORMANS METRİKLERİ

CART				
	<i>Doğruluk</i>	<i>Kesinlik</i>	<i>Duyarlılık</i>	<i>F1-Skor</i>
<i>TİCARET</i>	0,641	0,557	0,603	0,579
<i>KAMU HİZMETLERİ</i>	0,641	0,559	0,535	0,547
<i>KONUT</i>	0,641	0,722	0,804	0,761
<i>DEPO VE FABRİKA</i>	0,641	0,787	0,599	0,680

Rastgele Orman				
	<i>Doğruluk</i>	<i>Kesinlik</i>	<i>Duyarlılık</i>	<i>F1-Skor</i>
<i>TİCARET</i>	0,904	0,595	0,671	0,631
<i>KAMU HİZMETLERİ</i>	0,690	0,662	0,562	0,608
<i>KONUT</i>	0,690	0,726	0,821	0,771
<i>DEPO VE FABRİKA</i>	0,690	0,862	0,695	0,770

XGBoost				
	<i>Doğruluk</i>	<i>Kesinlik</i>	<i>Duyarlılık</i>	<i>F1-Skor</i>
<i>TİCARET</i>	0,857	0,603	0,658	0,629
<i>KAMU HİZMETLERİ</i>	0,686	0,667	0,550	0,603
<i>KONUT</i>	0,686	0,709	0,821	0,761
<i>DEPO VE FABRİKA</i>	0,686	0,827	0,706	0,762

k-En Yakın Komşu				
	<i>Doğruluk</i>	<i>Kesinlik</i>	<i>Duyarlılık</i>	<i>F1-Skor</i>
<i>TİCARET</i>	0,607	0,529	0,531	0,530
<i>KAMU HİZMETLERİ</i>	0,607	0,618	0,446	0,518
<i>KONUT</i>	0,607	0,578	0,814	0,676
<i>DEPO VE FABRİKA</i>	0,607	0,832	0,629	0,717

TEZDEN ÜRETİLMİŞ YAYINLAR

Konferans Bildirileri

1. Bilgin, A. N. ve Başaraner, A. M. (2024). Binaların mekânsal özelliklerinden yararlanarak makine öğrenmesi algoritmaları ile fonksiyonel olarak sınıflandırılması. *14. Uluslararası Mühendislik, Mimarlık ve Tasarım Kongresi* (s. 973). Güven Plus Grup Danışmanlık.