



TUFUAB -MERSİN 2021

<https://tufuab2021.mersin.edu.tr/>



CBS ve Makine Öğrenmesi yöntemleri ile konut amaçlı taşınmazların toplu değerlendirilmesi

Muhammed Oğuzhan Mete^{*1}, Tahsin Yomralıoğlu¹

¹İstanbul Teknik Üniversitesi, İnşaat Fakültesi, Geomatik Mühendisliği Bölümü, İstanbul, Türkiye

Anahtar Kelimeler

Taşınmaz Değerleme,
Toplu Değerleme,
Nominal Değerleme,
CBS,
Yapay Zeka,
Makine Öğrenmesi

ÖZ

Toplu taşınmaz değerlendirme uygulamalarında Coğrafi Bilgi Sistemleri (CBS) ve Makine Öğrenmesi yöntemleri yaygın olarak kullanılmaktadır. Konut amaçlı değerlendirme çalışmalarda binaların yüzölçümü, oda sayısı gibi fiziksel özelliklerine odaklanılmakta, konumsal kriterler yeterince analiz edilmemektedir. Öte yandan önemli yerlere yakınlık, deniz veya orman manzarası, düz bir topoğrafya gibi kriterler ise taşınmaz değerini önemli derecede etkileyen konumsal faktörlerdendir. Bu çalışmada Türkiye ve Birleşik Krallık'ta bulunan konut amaçlı taşınmazların toplu değerlendirilmesi için CBS ve Makine Öğrenmesi entegre edilerek hibrit bir yaklaşım geliştirilmiştir. Açık lisanslı coğrafi veri kaynakları kullanılarak yakınlık, yüzey ve görünürlük analizleri gerçekleştirilmiş, CBS tabanlı Nominal Değerleme Yöntemi ile taşınmaz değer haritası üretilmiştir. Ayrıca XGBoost, CatBoost, LightGBM ve Rastgele Orman gibi topluluk öğrenmesi regresyon algoritmaları kullanılarak fiyat tahmini için değerlendirme modelleri oluşturulmuştur. CBS analizlerinden elde edilen konumsal kriter puanları öznetelik zenginleştirme amacıyla değerlendirme verilerine dahil edilmiş, tahmin doğruluğu önemli derecede artırılmıştır. Çalışmada Rastgele Orman modeli ile diğer regresyon yöntemlerine kıyasla daha yüksek bir doğruluğa ulaşılmıştır.

GIS and Machine Learning based mass valuation of residential properties

Keywords

Real Estate Valuation,
Mass Valuation,
Nominal Valuation,
GIS,
Artificial Intelligence,
Machine Learning

ABSTRACT

Geographic Information Systems (GIS) and Machine Learning methods are widely used in mass property valuation practices. Focusing on physical attributes of properties, locational criteria are insufficiently used during the price prediction process. Whereas, locational criteria like proximity to important places, sea or forest views, flat topography are some of the spatial factors that extremely affect the property value. In this study, a hybrid approach is developed by integrating GIS and Machine Learning for automated mass valuation of residential properties in Turkey and the United Kingdom. GIS-based Nominal Valuation Method was applied to produce a land value map by carrying out proximity, terrain, and visibility analyses. Besides, ensemble regression methods like XGBoost, CatBoost, LightGBM, and Random Forest are built for price prediction. Spatial criteria scores obtained from GIS analyses were included in the price prediction data for feature enrichment purpose. Results showed that adding locational factors to the property price data increased the prediction accuracy dramatically. It also demonstrated that Random Forest was the most successful regression model compared to other ensemble methods.

* Sorumlu Yazar

^{*}(metemu@itu.edu.tr) ORCID ID 0000 - 0002 - 9312 - 1965
(tahsin@itu.edu.tr) ORCID ID 0000 - 0002 - 8532 - 747X

1. GİRİŞ

Taşınmaz mallar mülkiyet güvencesi ve değerli varlık olma özellikleri sebebiyle sosyal ve ekonomik yönlerden önemli bir çekiciliğe sahiptirler. Öte yandan taşınmaz değeri, sürdürülebilir kalkınma hedeflerine ulaşmak için gerekli olan küresel arazi yönetimi modelinin önemli bir bileşenidir (Enemark, 2004). Etkin bir taşınmaz piyasasının oluşumunu destekleyen taşınmaz değerine alım satım, vergilendirme, kamulaştırma, sermaye piyasası faaliyetleri gibi birçok işlemde ihtiyaç duyulmaktadır (FAO, 2017; McDermott, Myers ve Augustinus, 2018; UN-GGIM, 2019). Bu sebeple taşınmaz değerinin standartlara uygun, objektif bir yaklaşımla değerlendirilmesi oldukça önemlidir.

Taşınmazlara dayalı işlemlerde piyasa değerinin belirlenmesi için her yıl sayısız değerlendirme raporu oluşturulmaktadır. Öte yandan vergilendirme, kamulaştırma, imar uygulamaları, portföy yönetimi gibi çeşitli uygulamalarda hızlı ve otomatik değerlendirme yaklaşımlarına ihtiyaç duyulmaktadır. Literatürde Hedonik Değerleme, Nominal Değerleme, Çoklu Regresyon Analizi, Topluluk Öğrenmesi Yöntemleri, Yapay Sinir Ağları (YSA) gibi toplu değerlendirme yöntemleri yukarıda belirtilen amaçlar için yaygın olarak kullanılmaktadır (Jahanshiri, Buyong, Rashid ve Shariff, 2011; Mohd, Jamil, Johari, Abdullah ve Masrom, 2020; Pagourtzi, Assimakopoulos, Hatzichristos ve French, 2003; Wang ve Li, 2019).

Açık veri kavramı, son dönemde ülkelerin yürürlüğe koyduğu yasal mevzuat altyapısı ile önemli ölçüde artmıştır (Open Data Barometer, 2017). Açık verilerin ekonomik değer, şeffaflık, verimlilik ve kamu hizmetlerini iyileştirme gibi birçok faydası vardır (The European Data Portal, 2021). Devletlerin ve belediyelerin açık veri portalları sayesinde kullanıcılar güncel ve doğru verilere kolaylıkla ulaşabilmektedir. Ülkemizde de son dönemde açık veri portallarının yaygınlaştığı ve coğrafi verilerin vatandaşlar ile paylaşılmasıyla birlikte endüstrileşme yolunda önemli bir katma değer oluşturduğu gözlemlenmektedir.

Birleşik Krallık, 2010 yılından itibaren "data.gov.uk" veri portalı üzerinden çeşitli alanlarda çok sayıda veriye açık erişim sağlamıştır. HM Land Registry - Price Paid Data (PPD), İngiltere ve Galler'deki konutların çeşitli fiziksel özelliklerini ve alım-satım kayıtlarını içeren bir açık devlet verisidir (Price Paid Data, 2021). Öte yandan Energy Performance Certificates (EPC) verileri de İngiltere ve Galler'deki binaların toplam yüzölçümü, enerji verimlilik değerleri, karbondioksit emisyonları, ısıtma maliyetleri gibi birçok faydalı öznitelik bilgisini bünyesinde barındıran bir açık veridir (EPC, 2021). Öte yandan ülkenin harita kurumu Ordnance Survey tarafından da birçok temel coğrafi veri setleri açık lisans ile paylaşılmaktadır.

Bu çalışmanın temel amacı, Coğrafi Bilgi Sistemleri (CBS) ve Makine Öğrenmesi algoritmalarının birlikte kullanılmasıyla yüksek doğruluğa sahip bir taşınmaz değerlendirme modelinin oluşturulmasıdır. Bu kapsamda CBS tabanlı Nominal Değerleme Yöntemi ile bilimsel ve nesnel değerlendirmeye dayalı bir taşınmaz değer haritasının oluşturulması, Türkiye ve İngiltere'de konut

vasıflı taşınmazların toplu değerlemesi için çeşitli Makine Öğrenmesi regresyon modellerinin performanslarının karşılaştırılması hedeflenmektedir. Ayrıca yakınlık, yüzey ve görünürlük gibi coğrafi analizler sonucunda elde edilen nominal kriter puanlarının değerlendirme verilerine eklenmesiyle birlikte öznitelik zenginleştirmesini ve konumsal kriterlerin değer tahmin doğruluğuna katkısını incelemektedir.

2. YÖNTEM

Konumsal kriterlerin taşınmaz değeri üzerindeki etkisini ortaya koymak için CBS'den yararlanmak mümkündür. Toplu değerlendirme çalışmalarında CBS'nin yönetim ve görselleştirme işlevlerinin yanısıra, güçlü konumsal analiz kabiliyeti de bulunmaktadır. Coğrafi Ağırlıklı Regresyon, Mekansal Analiz ve Nominal Değerleme gibi CBS tabanlı toplu değerlendirme yöntemleri, geniş alanlarda bulunan taşınmazların hızlı, doğru ve objektif bir şekilde değerlendirilmesini sağlamaktadır.

Nominal Değerleme, farklı etki düzeylerine sahip kriterlerin puanlama usulüne dayalı ağırlıklı toplamını ifade eden, Ağırlıklı Doğrusal Kombinasyon yaklaşımını baz alan stokastik bir yöntemdir (Mete ve Yomralıoğlu, 2019; Yomralıoğlu, 1993). Değeri etkileyen her bir kriter belli bir ölçekte (örneğin 0-100 aralığında) puan verilerek hem arazi hem de binaların değerleri nesnel bir yaklaşımla, parametrik olarak belirlenebilmektedir. Bir taşınmazın toplam nominal değeri, parsel veya piksel alanı ile çarpılan kriter puanlarının ağırlıklı toplamı olarak hesaplanmaktadır (1).

$$V_i = S_i * \sum_{j=1}^k (f_{ji} * w_j) \quad (1)$$

öyle ki;

V: Toplam nominal değer

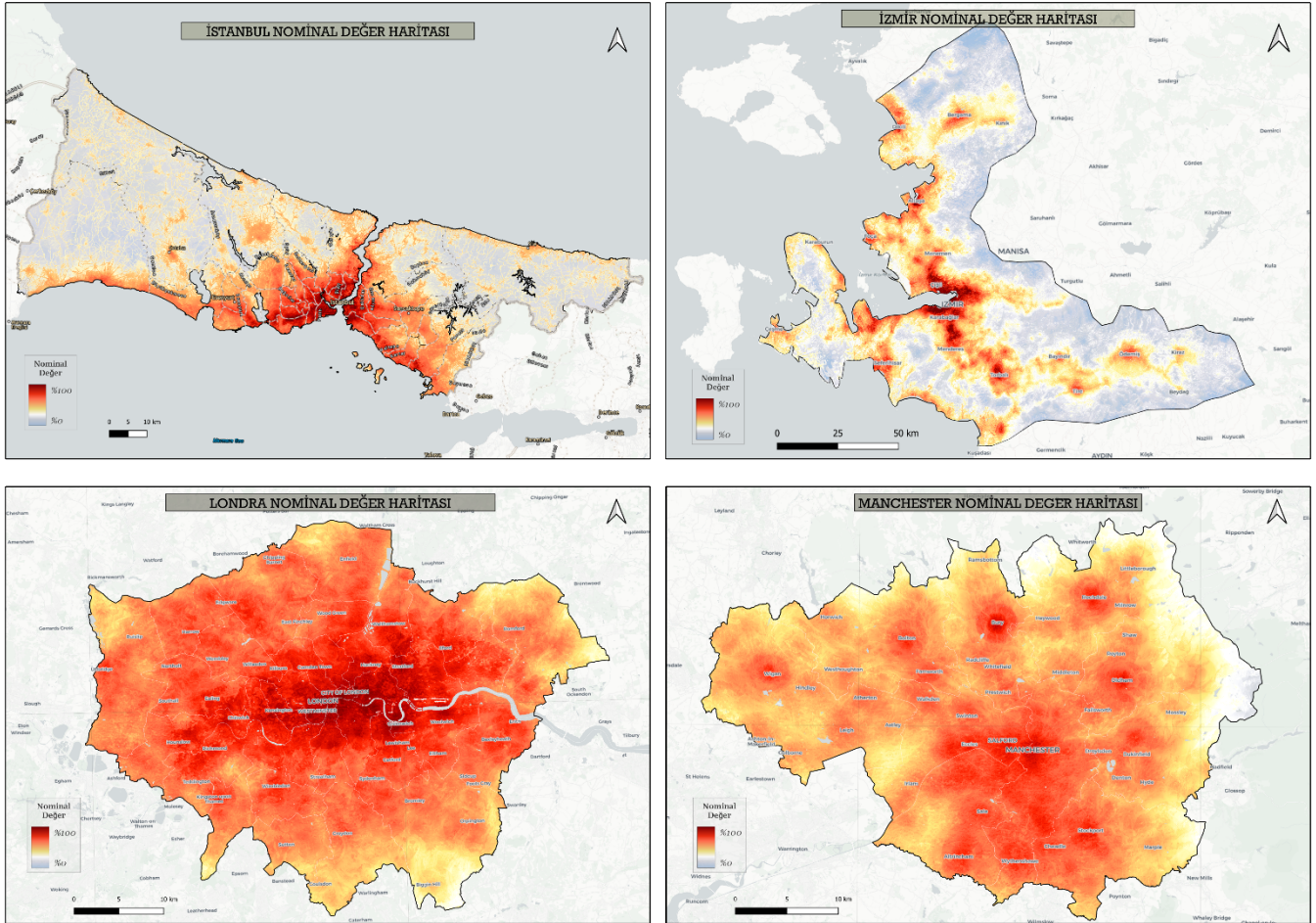
S: Parsel veya piksel alanı

f: Faktör puanı

w: Faktör ağırlığı

Makine Öğrenmesi yöntemleri finans, yatırım, gayrimenkul değerlendirme gibi çeşitli ekonomik alanlarda yaygın olarak kullanılmaktadır. Toplu taşınmaz değerlendirme uygulamalarında, objektif ve doğru fiyat tahmini için çoğunlukla regresyon analizleri ve YSA'dan yararlanılmaktadır. Regresyon analizi, bir bağımlı değişken ile bir veya daha fazla bağımsız değişken arasındaki ilişkileri kestirmek için kullanılır. Taşınmaz değerlemede Doğrusal, Ridge, Lasso, Polinomal ve Bayesçi Regresyon gibi çeşitli yöntemlerin yanı sıra Topluluk Öğrenmesi regresyon modelleri de yaygın olarak kullanılmaktadır.

CBS ve Makine Öğrenimi yöntemleri, toplu taşınmaz değerlendirme amaçlı matematiksel model oluşturmak için yaygın olarak kullanılmaktadır. Bu çalışmada CBS'nin konumsal analiz yeteneği ile Makine Öğrenmesinin tahminleme algoritmaları bir araya getirilmiş, öznitelik zenginleştirme yaklaşımı ile yüksek doğruluğa sahip bir değerlendirme modeli geliştirilmiştir.



Şekil 1. Türkiye ve İngiltere çalışma bölgeleri için üretilen nominal değer haritaları

Taşınmaz değer haritaları, bir bölgedeki fiyat değişimlerinin ortaya konmasını sağlamakta ve yönetim, planlama, vergilendirme gibi faaliyetleri kolaylaştırmaktadır. Çalışmada İstanbul ve İzmir şehirleri için belediyelerin kullanıma sunduğu açık veriler ile OpenStreetMap verileri; Londra ve Manchester şehirleri için ise Ordnance Survey ve OpenStreetMap açık veri kaynakları kullanılmış, CBS tabanlı Nominal Değerleme yöntemi ile yakınlık, yüzey ve görünürlük analizleri gerçekleştirilerek arazi değer haritaları üretilmiştir (Şekil 1).

Makine Öğrenmesinin temel algoritma türlerinden biri olan Regresyon analizi, taşınmazların fiyat tahmini için yaygın olarak kullanılmaktadır. Araştırma kapsamında çalışma alanı olarak belirlenen iki farklı ülkede konut vasıflı taşınmazların değerlemesi için çeşitli Makine Öğrenmesi regresyon modelleri geliştirilmiştir. İstanbul ve İzmir şehirleri için bir emlak listeleme şirketinin uygulama geliştirme arayüzü (API) kullanılarak konut fiyatları elde edilmiş, Londra ve Manchester şehirleri için ise PPD-EPC veri seti kullanılmıştır. Veri ön işleme adımlarından sonra veriler %80 eğitim, %10 doğrulama, %10 test olarak bölünmüş, değer kestirimi için XGBoost, CatBoost, LightGBM, Rastgele Orman regresyon modelleri oluşturulmuştur. Modellerin tahmin başarısını ölçmek için R^2 , düzeltilmiş R^2 , Ortalama Mutlak Hata (Mean Absolute Error - MAE), Ortalama Mutlak Yüzde Hatası (Mean Absolute Percentage Error - MAPE) ve Karesel Ortalama Hata

(Root Mean Square Error - RMSE) gibi çeşitli doğruluk metrikleri kullanılmıştır.

Taşınmaz değerleme verileri Makine Öğrenmesi tabanlı değerleme çalışmalarında fiyat tahmini için gereken birçok fiziksel özneliği içermektedir. Ancak daha yüksek doğruluklu regresyon modelleri oluşturmak için taşınmaz değeri ile yüksek korelasyona sahip konumsal faktörlerin de sürece dahil edilmesine ihtiyaç vardır (Kiel ve Zabel, 2008; Mete ve Yomralıoğlu, 2021; Wyatt, 1997). CBS, konumsal kriterlerin taşınmaz değeri üzerindeki etkilerini ortaya koyabilmek için kullanılabilecek çok sayıda analiz aracı sunmaktadır. Bu kapsamda regresyon modelinin öznelilik zenginleştirilmesi için yakınlık, yüzey ve görünürlük analizlerinin nominal puanları değerleme verilerine entegre edilmiştir. Zenginleştirilmiş öznelilikler elde edildikten sonra konumsal faktörlerin değerleme doğruluğuna etkisini incelemek amacıyla Makine Öğrenmesi yöntemleri ile regresyon analizi tekrar gerçekleştirilmiştir. Tablo 1’de modellerin doğruluk metrikleri görülmektedir. Özellik seçiminin bir parçası olan özneliliklerin önem skorları, bir değişkenin tahmin sürecini ne kadar etkilediğini yorumlamak için kullanılmaktadır. Çalışmada regresyon modellerine dair permütasyon önem skorları hesaplanmıştır.

3. BULGULAR

XGBoost, LightGBM, Rastgele Orman gibi Topluluk Öğrenmesi yöntemleri, konut fiyat tahminleri için yaygın olarak kullanılmaktadır.

Tablo 1. Regresyon analizleri sonucu elde edilen doğruluk metrikleri

#	Model	R ²	Düzeltilmiş R ²	MAE	MAPE	RMSE
1	XGBoost	0.4341	0.4340	98505.9861	0.4689	204625.5031
1	CatBoost	0.4551	0.4550	98811.1141	0.4737	200791.2485
1	LightGBM	0.4366	0.4365	101291.7355	0.4937	204170.2015
1	Rastgele Orman	0.4690	0.4689	95934.5759	0.4614	198211.9713
2	XGBoost	0.7907	0.7906	65367.1736	0.3107	123585.9848
2	CatBoost	0.8380	0.8379	56668.9762	0.2660	108740.2277
2	LightGBM	0.8058	0.8058	65567.8885	0.3188	119042.0076
2	Rastgele Orman	0.8579	0.8578	44888.7285	0.1941	101847.7448

1: Öznitelik zenginleştirilmesi öncesi, 2: Öznitelik zenginleştirilmesi sonrası

Bu çalışmada Türkiye'nin İstanbul ve İzmir şehirleri ile İngiltere'nin Londra ve Manchester şehirlerinde toplu taşınmaz değerleme çalışması gerçekleştirmek amacıyla çeşitli topluluk öğrenmesi regresyon yöntemleri kullanılmıştır. Regresyon modellerinin performans metrikleri karşılaştırıldığında, Rastgele Orman algoritmasının diğerlerinden daha yüksek doğruluğa sahip olduğu görülmüştür. XGBoost, CatBoost ve LightGBM regresyon modelleri iyi sonuçlar verse de, Rastgele Orman algoritmasının doğruluğuna erişememişlerdir. Öte yandan eğitim için geçen süreler karşılaştırıldığında, Rastgele Orman modeli en yavaş, LightGBM ise en hızlı model olarak göze çarpmıştır. Elde edilen sonuçlara göre; Rastgele Orman modeli 0.8579 R², %0.1941 MAPE ve 101847.74 RMSE ile konut fiyat tahmininde en başarılı algoritma olmuştur.

Taşınmaz değerleme veri setleri değeri etkileyen fiziksel faktörlerle ilgili birçok özniteliğe sahip olmasına karşın, taşınmazın konumsal veya çevresel özelliklerine dair öznitelikler bulunmamaktadır. Öznitelik verilerini zenginleştirmek ve tahmin doğruluğunu artırmak için CBS analizleri kullanılmış, yakınlık, yüzey ve görünürlük kriterlerinin nominal puanları veri setine dahil edilmiştir. Veri zenginleştirme işleminden sonra R² %38,9 artmış, RMSE ise yaklaşık %48,6 azalmıştır.

4. SONUÇLAR

Taşınmaz değerini esas alan faaliyetlerde toplu değerleme yöntemlerinin kullanımı son yıllarda yaygınlaşmıştır. Hedonik Değerleme, Nominal Değerleme, Regresyon Analizi ve Yapay Sinir Ağları toplu değerleme uygulamalarının öne çıkan yöntemlerinden bazılarıdır. Literatürde konut amaçlı taşınmazların toplu değerlemesinde Makine Öğrenmesi algoritmalarını kullanan çok sayıda çalışma olmasına karşın taşınmaz değerini önemli ölçüde etkileyen konumsal faktörlerden çok binaların fiziksel özelliklerine yoğunlaşıldığı görülmektedir. Bu çalışmada, Türkiye ve İngiltere'deki konutların toplu değerlemesi için CBS tabanlı Nominal Değerleme yöntemi ve çeşitli Makine Öğrenmesi regresyon algoritmaları birlikte kullanılmıştır. Elde edilen sonuçlara göre Rastgele Orman algoritmasının fiyat tahmininde en başarılı model olduğu görülmüştür. Önerilen hibrit yaklaşımın dünya genelinde herhangi bir şehir için taşınmaz değerleme verilerine kolaylıkla uyarlanabileceği, ülkelerin ve değerleme kuruluşlarının

dinamik esaslı toplu taşınmaz değerleme çalışmalarında yüksek doğruluk elde edebileceği öngörülmektedir.

TEŞEKKÜR

Bu çalışma İstanbul Teknik Üniversitesi (İTÜ) Ulusal Coğrafi Bilgi Teknolojileri İnovasyon Uygulama ve Araştırma Merkezi bünyesinde gerçekleştirilmiştir. Ayrıca İstanbul Teknik Üniversitesi Bilimsel Araştırma Projeleri Koordinasyon Birimi tarafından desteklenmiştir. Proje Numarası: MDK-2021-43080.

Yazar Katkı Oranı

Muhammed Oğuzhan Mete: Kavramsallaştırma, Metodoloji, Yazılım, Yazma.

Tahsin Yomralıoğlu: Metodoloji, Yazma, İnceleme ve Düzenleme.

Çıkar çatışmaları

Yazarlar arasında herhangi bir çıkar çatışması bulunmamaktadır.

Araştırma ve Yayın Etiği Beyanı

Yapılan çalışmada araştırma ve yayın etiğine uyulmuştur.

KAYNAKÇA

- Enemark, S. (2004). Building Land Information Policies. *UN, FIG, PC IDEA Inter-regional Special Forum on The Building of Land Information Policies in the Americas* içinde . Aguascalientes, Mexico: FIG.
- EPC. (2021). Energy Performance of Buildings Data. *UK Department for Levelling Up, Housing and Communities*. 27 Şubat 2022 tarihinde <https://epc.opendatacommunities.org/> adresinden erişildi.
- FAO. (2017). Valuing land tenure rights. *Food and Agriculture Organization*. 27 Şubat 2022 tarihinde www.fao.org/tenure adresinden erişildi.
- Jahanshahi, E., Buyong, T., Rashid, A. ve Shariff, M. (2011). *A Review of Property Mass Valuation Models*. *Pertanika J. Sci. & Technol* (C. 19).
- Kiel, K. A. ve Zabel, J. E. (2008). Location, location, location: The 3L Approach to house price determination. *Journal of Housing Economics*, 17(2), 175–190. doi:10.1016/j.jhe.2007.12.002

- McDermott, M., Myers, M. ve Augustinus, C. (2018). *VALUATION OF UNREGISTERED LANDS: A Policy Guide*.
- Mete, M. O. ve Yomralıoğlu, T. (2019). CBS ile Nominal Taşınmaz Değer Haritası Üretiminde Çözünürlük Araştırması. *Türkiye Arazi Yönetimi Dergisi*, 1(1), 16-23.
- Mete, M. O. ve Yomralıoğlu, T. (2021). Açık Kaynaklı Bulut CBS Yardımıyla Kitleli Taşınmaz Değerleme Uygulaması. *Harita Dergisi*, 165, 28-42.
- Mohd, T., Jamil, N. S., Johari, N., Abdullah, L. ve Masrom, S. (2020). An Overview of Real Estate Modelling Techniques for House Price Prediction. N. Kaur ve M. Ahmad (Ed.), *Charting a Sustainable Future of ASEAN in Business and Social Sciences* içinde (ss. 321-338). Springer Singapore. doi:10.1007/978-981-15-3859-9_28
- Open Data Barometer. (2017). Open Data Barometer 4th Edition. *World Wide Web Foundation*. 20 Aralık 2021 tarihinde <https://opendatabarometer.org/4thedition/> adresinden erişildi.
- Pagourtzi, E., Assimakopoulos, V., Hatzichristos, T. ve French, N. (2003). Real estate appraisal: a review of valuation methods. *Journal of Property Investment & Finance*, 21(4), 383-401. doi:<https://doi.org/10.1108/14635780310483656>
- Price Paid Data. (2021). Price Paid Data. *HM Land Registry*. 27 Şubat 2022 tarihinde <https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads> adresinden erişildi.
- The European Data Portal. (2021). *The Open Data Maturity (ODM) Report 2021*. Luxembourg. https://data.europa.eu/sites/default/files/landscaping_insight_report_n7_2021.pdf adresinden erişildi.
- UN-GGIM. (2019). *Framework for Effective Land Administration (FELA)*. http://ggim.un.org/meetings/2018-Deqing-Expert-Group/documents/Concept_Note.pdf adresinden erişildi.
- Wang, D. ve Li, V. J. (2019). Mass Appraisal Models of Real Estate in the 21st Century: A Systematic Literature Review. *Sustainability*, 11(24), 7006. doi:10.3390/su11247006
- Wyatt, P. J. (1997). The development of a GIS-based property information system for real estate valuation. *International Journal of Geographical Information Science*, 11(5), 435-450. doi:10.1080/136588197242248
- Yomralıoğlu, T. (1993). *A Nominal Asset Value-Based Approach For Land Readjustment And Its Implementation Using Geographical Information Systems*. Doktora Tezi, University of Newcastle upon Tyne, İngiltere.



© Author(s) 2021.

This work is distributed under <https://creativecommons.org/licenses/by-sa/4.0/>