

Evliyâ Çelebi'nin İstanbul'u: Metin Madenciliği Temelli Bir Yaklaşım

Evliya Çelebi's Istanbul: A Text Mining-Based Approach

FATMA ALADAĞ

Marmara Üniversitesi.

(fatma.aladag@marmara.edu.tr), ORCID: 0000-0003-3455-6181.

Başvuru/Submitted: 29.08.2025. Kabul/Accepted: 02.10.2025.

“ ” Aladağ, Fatma. “Evliyâ Çelebi'nin İstanbul'u: Metin Madenciliği Temelli Bir Yaklaşım.”

Zemin, s. 10 (2025): 42-63.

DOI: <https://doi.org/10.5281/zenodo.17649961>.

Özet: Bu çalışma, Evliyâ Çelebi'nin (ö. 1687'den sonra), on yedinci yüzyıla ait *Seyâhatnâme*'sinin İstanbul cildinin Doğal Dil İşleme (Natural Language Processing, NLP), çok yöntemli metin madenciliği ve GPT tabanlı yaklaşımlarla incelenmesine yönelik bir metodolojik çerçeve sunmaktadır. Araştırmada, eş-oluşum ve frekans analizi, Latent Dirichlet Allocation (LDA) ile konu modellemesi, BERT tabanlı dönüştürücü modeller aracılığıyla duygu analizi ve adlandırılmış varlık tanıma (NER) uygulamaları kullanılarak *Seyâhatnâme*'deki semantik ve tematik örüntüler belirlenmiştir. Çalışma, Osmanlı Türkçesinin hesaplamalı analiz yöntemlerine entegrasyonunda karşılaşılan dilbilimsel zorlukları ele almakta; tarihsel metinlere uyarlanmış dil modellerinin gerekliliğine ve bu bağlamda GPT'nin sağlayabileceği katkılara işaret etmektedir. Bu yaklaşım, nicel ve tematik analizleri bir araya getirerek, *Seyâhatnâme*'nin sosyo-kültürel ve tarihsel içeriğinin farklı boyutlarının incelenmesine olanak tanımakta ve dijital beşerî bilimler ile dijital tarih alanında Osmanlı çalışmaları literatürüne katkı potansiyeli sunmaktadır. **Anahtar Kelimeler:** Evliyâ Çelebi, Osmanlı Çalışmaları, İstanbul, LLM, NLP, Dijital Beşerî Bilimler, Şehir Tarihi.

Abstract: This study presents a methodological framework for examining Evliyâ Çelebi's 17th-century Istanbul travelogue through Natural Language Processing (NLP), multi-method text mining, and GPT-based approaches. The research applies co-occurrence and frequency analysis, topic modeling using Latent Dirichlet Allocation (LDA), sentiment analysis via BERT-based transformer models, and named entity recognition (NER) to identify semantic and thematic patterns within the text. The study addresses the linguistic challenges of integrating Ottoman Turkish into computational analysis methods, highlighting the need for language models adapted to historical texts and noting the potential contributions of GPT in this context. By combining quantitative and thematic analyses, the approach enables the exploration of multiple dimensions of the travelogue's socio-cultural and historical content and offers the potential to contribute to the literature on Ottoman studies within the fields of digital humanities and digital history.

Keywords: Evliyâ Çelebi, Ottoman Studies, LLM, NLP, Istanbul, Digital Humanities, Urban History.

Dijital beşerî bilimler, tarihsel metinlerin metin madenciliği yöntemleriyle incelenmesine yeni imkânlar açmakta ve disiplinler arası yaklaşımları mümkün kılmaktadır. Bu alan, Evliya Çelebi'nin (ö. 1687'den sonra) *Seyâhatnâmesi* gibi tarihsel ve kültürel bakımdan zengin kaynakların gelişmiş hesaplamalı tekniklerle çok boyutlu değerlendirilmesine olanak tanır. Osmanlı Türkçesinin karmaşık dil yapısı, Doğal Dil İşleme (Natural Language Processing, NLP) yöntemleriyle ele alındığında dönemin sosyal, kültürel ve mekânsal dinamiklerine dair daha önce ortaya konmamış bulgulara ulaşmaya katkı sağlar.¹

Metin madenciliği, veri bilimi, dilbilim ve bilgisayar biliminin kesişiminde konumlanan, büyük hacimli metinlerden istatistiksel modeller ve makine öğrenmesiyle anlamlı örüntüler çıkaran bir yöntemdir.² Ancak Osmanlı Türkçesinin morfolojik yapısı, Arapça-Farsça etkiler ve standartlaşmamış yazım biçimleri, standart NLP araçlarının yetersiz kalmasına yol açar; bu nedenle tokenizasyon, lemmatizasyon ve sözdizimsel etiketleme gibi adımlar tarihsel dile uyarlanmalıdır.³

Bu makale, on yedinci yüzyıl seyyahı Evliya Çelebi'nin *Seyâhatnâmesi*'nin İstanbul bölümünü metin madenciliği teknikleriyle inceleyen bir uygulama ör-

1 Fatma Aladağ, "Dijital Beşerî Bilimler ve Türkiye Araştırmaları: Bir Literatür Değerlendirmesi," *Türkiye Araştırmaları Literatür Dergisi*, s. 18 (2020): 773-796, <https://dergipark.org.tr/tr/pub/talid/issue/58073/800406>; Bu alanda yapılan çalışmaları, kişi ve yayınları bir araya getiren bir platform olarak bkz. Digital Ottoman Studies Platform (<https://www.digitalottomanstudies.com>), (erişim 12 Kasım 2025).

2 Ronen Feldman and James Sanger, *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data* (Cambridge: Cambridge University Press, 2007), <https://doi.org/10.1017/CBO9780511546914>.

3 *Tokenizasyon*, metnin sözcüklere ya da birimlere ayrılması; *lemmatizasyon*, sözcüklerin kök ya da sözlük biçimine indirgenmesi; *sözdizimsel etiketleme (POS tagging)* ise her sözcüğün dilbilgisel türünün (isim, fiil, sıfat vb.) belirlenmesi işlemidir. Ayrıca Osmanlı Türkçesi metinlerin dijitalleştirilmesi için yürütülen çalışma ve projelerin kapsamlı bir literatürü için bkz. Fatma Aladağ ve Elif Derin Can, "Programlamadan Yapay Zekâya," *Dijital Osmanlı Çalışmaları*, ed. Yunus Uğur (İstanbul: Vakıf Bank Yayınları, 2023); Süphan Kırmızıaltın ve David Wrisley, "Automated Transcription of Historical Periodicals with HTR: Challenges and Futures," *Digital Humanities Quarterly* 16, no. 2 (2022); Esmâ F. Bilgin Taşdemir, "Printed Ottoman Text Recognition Using Synthetic Data and Data Augmentation," *International Journal on Document Analysis and Recognition*, no. 26 (2023): 273-287; Osmanlı Türkçesi dijitalleştirme çalışmalarına dair bir platform olarak bkz. Digital Ottoman Corpora (<https://www.digitalottomancorpora.org>), (erişim 12 Kasım 2025).

neği sunmaktadır. Metin, edebî bir eser olmanın ötesinde Osmanlı coğrafyasına dair mimari, gündelik yaşam ve idari yapılar hakkında ayrıntılı bilgiler içerir; bu yönüyle çoklu yöntem metin madenciliği stratejileri için elverişli bir örnektir.

Çalışmada frekans analizi (*frequency analysis*), eş-oluşum ağı (*co-occurrence*), konu modellemesi (*topic modeling*) ve duygu analizi (*sentiment analysis*) yöntemleri hem modern Türkçe versiyon hem de Osmanlı Türkçesinin Latin harfli transkripsiyonu üzerinde uygulanmış; bu iki metin katmanı arasındaki sonuçlar karşılaştırılmıştır. Böylece, tarihsel içerik bakımından tam bir yorum hedeflenmemiş, bunun yerine Osmanlı Türkçesinin hesaplamalı analizi için metodolojik bir çerçeve sunulmuştur. Çalışma, tarih ile veri analizi disiplinleri arasında besleyici bir tartışma zemini oluşturmayı ve gelecekteki araştırmalara yönetsel katkı sağlamayı amaçlamaktadır.

***Seyâhatnâme*'nin Çoklu Yöntem Metin Madenciliği ve NLP Uygulamaları ile Analizi**

Çoklu yöntem yaklaşımı, metnin çeşitli boyutlarını farklı analitik tekniklerle inceleyerek, içerikteki tematik yapının ve ilişkisel örüntülerin çok katmanlı bir görünümünü sağlar. Örneğin, frekans analizi aracılığıyla metindeki yüksek tekrar oranına sahip terimler ve kavramlar tespit edilebilirken; duygu analizi ile anlatıların barındırdığı duygusal eğilimler ve değerlendirme tonları hakkında ipuçları elde edilebilir. Eş-oluşum ağı analizi, belirli bir kelime veya kelime öbeğinin diğer terimlerle kurduğu bağların niceliksel ve görsel temsillerini sunar. Konu modellemesi ise kelime dağılımı kalıplarından hareketle metin içindeki tematik alanların olasılıksal olarak tanımlanmasına imkân verir.

Evliya Çelebi'nin on yedinci yüzyıl dil özelliklerini yansıtan Osmanlı Türkçesi ile kaleme aldığı *Seyâhatnâmesi*, çoklu yöntem metin madenciliği uygulamaları açısından yüksek analitik değer taşımaktadır. Bununla birlikte, bu tür bir hesaplamalı yaklaşımın tarihsel dil verisine uygulanması, Osmanlı Türkçesinin ortaya koyduğu dilbilimsel zorlukların ayrıntılı biçimde değerlendirilmesini gerektirir. Bu çalışmada, söz konusu metin üzerinde yürütülen analizlerde hem dilin yapısından hem de mevcut NLP araçlarının sınırlılıklarından kaynaklanan güçlükler, ayrıca bu sürecin sunduğu kolaylaştırıcı unsurlar sistematik biçimde ele alınmaktadır. Tarihi belgelerde NER uygulamalarında karşılaşılan sorunları kapsamlı biçimde tartışan Ehrmann ve arkadaşları, özellikle diakritik işaretlerin normalizasyonu, yazım varyantları, OCR kaynaklı gürültü ve bağlam kaymaları

gibi zorluklara dikkat çekmektedir.⁴ Bu çerçevede, Osmanlı Türkçesine özgü dilsel güçlükler de şu başlıca alanlarda özetlenebilir: (i) izafet yapıları, morfolojik çözümlemede segmentasyon hatalarına yol açmaktadır; (ii) diakritik işaretler, tokenizasyon ve normalizasyon süreçlerinde ek adımlar gerektirmektedir; (iii) yazım varyantları (ör. *İstanbul/İslambol/İstambol*), standartlaştırma ihtiyacını gündeme getirmektedir; (iv) standartlaşmamış imla, lemmatizasyon ve istatistiksel analizlerde hata payını artırmaktadır; (v) toponim çeşitleri, tarihsel gazetteer eşleştirmeleri olmaksızın doğru biçimde sınıflandırılmamaktadır. Bu sorun alanlarının sistematik biçimde tanımlanması, yöntemin yeniden üretilebilirliğini güçlendirmekte ve sonraki çalışmalara aktarılabilirliğini artırmaktadır.

Metinden Veriye Dönüşüm: *Seyâhatnâme*'yi Analize Hazırlama ve Ön İşleme

Seyâhatnâme metninin hesaplamalı yöntemlerle analiz edilebilmesi için, yapılandırılmamış metnin analize uygun hâle getirilerek ön işleme süreçlerinden geçirilmesi gerekmektedir. Çalışmanın yönetsel çerçevesi, bütüncül bir iş akışı üzerine kurulmuştur. Bu iş akışı; ön işleme (temizleme, tokenizasyon, lemmatizasyon, stop-word ayıklama) adımlarını, ardından farklı NLP tabanlı analiz yöntemlerinin (NER, frekans analizi, eş-oluşum ağları, LDA konu modellemesi, duygu analizi) uygulanmasını ve nihayetinde bulguların görselleştirme ile raporlanmasını içermektedir.

Bu çalışmada, *Seyâhatnâme*'nin hem Osmanlı Türkçesi çeviriyazısı⁵ hem de modern Türkçe metni⁶ için paralel ön işleme adımları izlenmiştir. İlk aşamada tipografik hatalar giderilmiş, rakamlar ve analize gereksiz işaretler temizlenmiştir. Ardından metinler tokenizasyon işlemine tabi tutulmuş; cümleler kelimelere ve diğer anlamlı birimlere (token) ayrılmıştır. Sonraki aşamada, lemmatizasyon ve kök ayıklama (stemming) teknikleri uygulanarak kelimeler kök ya da lemma

4 Mathieu Ehrmann et al., “Challenges in Named Entity Recognition for Historical Texts,” *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries (JCDL)* (2023), <https://doi.org/10.1145/3604931>.

5 Evliyâ Çelebi, *Seyâhatnâme: I. Kitap*, haz. Robert Dankoff, Seyit Ali Kahraman ve Yücel Dağlı (İstanbul: YKY, 2008).

6 Evliyâ Çelebi, *Günümüz Türkçesiyle Evliya Çelebi Seyahatnamesi: İstanbul*, haz. Yücel Dağlı ve Seyit Ali Kahraman (İstanbul: YKY, 2003). Burada “modern” ifadesi, metnin günümüz Türkçesine uyarlanmış hâlinde ziyade, tarihsel dil özelliklerini kısmen koruyan sadeleştirilmiş versiyonunu tanımlamakta; dolayısıyla çağdaş Türkçeden belirgin biçimde ayrılmaktadır.

biçimlerine indirgenmiştir. Bu işlem, metin içinde kavramsal bütünlüğün korunması ve istatistiksel analizlerin güvenilirliğinin artırılması açısından önemlidir. Bununla birlikte, lemmatizasyonun başarımları oranı yalnızca Osmanlı Türkçesi için değil, modern Türkçe için de sınırlıdır. Bu nedenle çalışmada, yaygın NLP kütüphaneleri olan NLTK ve spaCy'nin yanı sıra, Türkçeye özgü Zemberek,⁷ Turkish-Lemmatizer⁸ ve TRMOR⁹ gibi araçlardan yararlanılmıştır.

Bu metodolojik süreç, *Seyâhatnâme* metninin farklı metin madenciliği teknikleri ile analizine imkân tanıyacak nitelikli bir veri seti üretmeyi amaçlamaktadır. Ön işleme tamamlandıktan sonra, metin üzerinde frekans analizi, eş-oluşum ağı analizi ve duygu analizi gibi geniş bir yelpazeyi kapsayan NLP tabanlı yöntemler uygulanmıştır.

***Seyâhatnâme*'nin Adlandırılmış Varlık Tanıma (NER) ile Frekans Analizi** Adlandırılmış Varlık Tanıma (NER), Osmanlı Türkçesi metinlerde kişi adları, zaman ifadeleri, mekânlar, olay adları, kurumlar ve nesne adları gibi belirli kategorilere ait öğelerin otomatik olarak tespit edilmesini sağlar. Bu yöntem sayesinde metinde geçen özel adlar sistematik biçimde ayrıştırılabilmekte ve farklı kategoriler altında sınıflandırılabilmektedir.¹⁰ NER, tarihsel metinlerin dijitalleştirilmesi ve içerik zenginleştirilmesi süreçlerinde önemli bir rol oynamakta; özellikle tarihsel metin analizi ve dijital beşerî bilimler çalışmalarında, büyük hacimli metin külliyatlarında özel adların ve terimlerin otomatik olarak belirlenmesi ve kategorize edilmesi için yaygın biçimde kullanılmaktadır.¹¹

7 Ahmet A. Akın, *Zemberek-NLP* (GitHub repository), <https://github.com/ahmetaa/zemberek-nlp> (erişim 12 Haziran 2025).

8 Abdullatif Köksal, *Turkish-Lemmatizer* (GitHub repository), <https://github.com/akoksal/Turkish-Lemmatizer> (erişim 13 Haziran 2025).

9 Ayla Kayabaş ve öte., "TRMOR: A Finite-State-Based Morphological Analyzer for Turkish," *Turkish Journal of Electrical Engineering and Computer Sciences* 27, no. 5 (2019): 3837-3851, <https://doi.org/10.3906/elk-1902-125>.

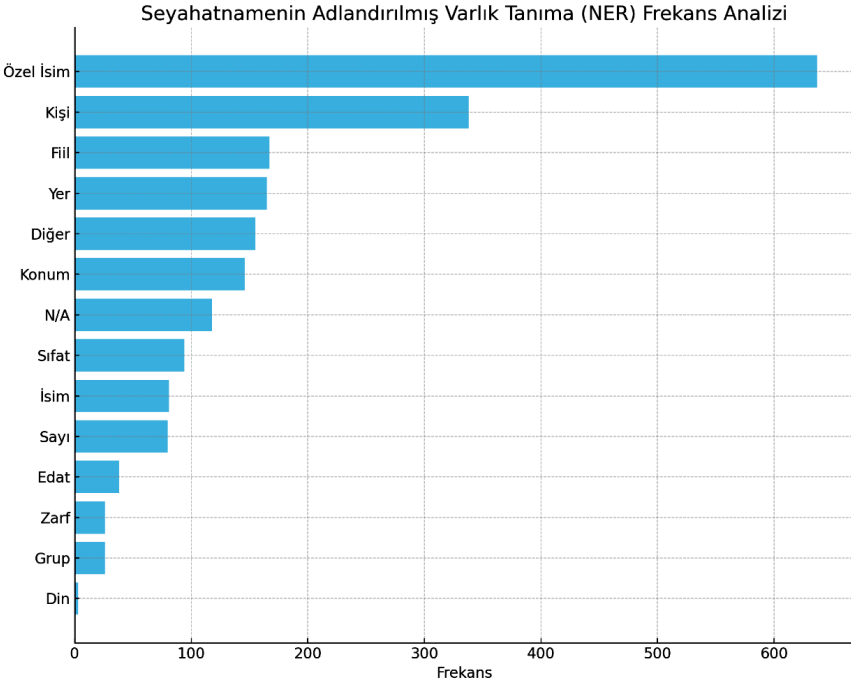
10 Dawn Archer, "Data Mining and Word Frequency Analysis," *Research Methods for Reading Digital Data in the Digital Humanities* (Edinburgh: Edinburgh University Press, 2016), 72-92, <https://doi.org/10.1515/9781474409629-006>.

11 Bruno Martins, Patricia Murrieta-Flores ve Miguel Won, "Ensemble Named Entity Recognition (NER): Evaluating NER Tools in the Identification of Place Names in Historical Corpora," *Frontiers in Digital Humanities*, no. 5 (2018): 333859, <https://doi.org/10.3389/fdigh.2018.00002>.

Frekans analizi ile NER'in bir arada kullanılması, metinlerde geçen isimlerin ve terimlerin dağılımını ortaya çıkarmanın yanı sıra, bu terimlerin sosyo-kültürel, ekonomik ve politik bağlam içindeki öneminin değerlendirilmesine de imkân tanır.

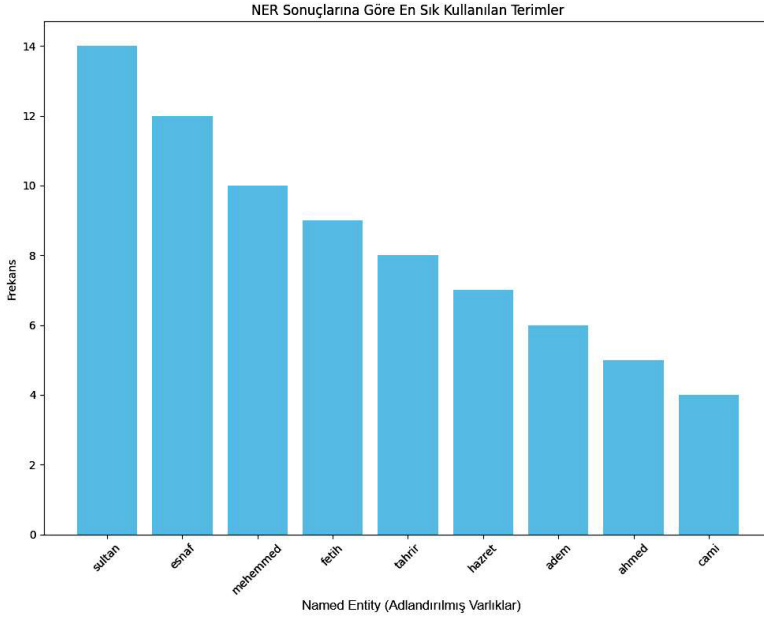
Bu çalışmada, Osmanlı Türkçesi metinler üzerinde NER ve kategorizasyon işlemleri için OpenAI tarafından geliştirilen büyük dil modeli GPT-4 Turbo kullanılmıştır. Metinler Python ortamında NLTK ve spaCy ile ön işleme tabi tutulmuş; durak kelimeler (*stop-words*) ve noktalama işaretleri temizlenmiş, tokenizasyon ve temel normalizasyon gerçekleştirilmiştir. Frekans analizi ve özet istatistikler Pandas ile hesaplanmış, gerektiğinde görselleştirmeler Matplotlib ile üretilmiştir.

Elde edilen sonuçlar, *Seyâhatnâme*'de geçen kişi adları, yer adları, unvanlar, dini ve idari terimler gibi çeşitli adlandırılmış varlık kategorilerinin dağılımını ortaya koymaktadır (*Görsel 1*). Bu analiz, hem metnin içerik yapısındaki odak alanları hem de tematik yoğunlukları belirleme açısından önem taşımaktadır.



Görsel 1: *Seyâhatnâme* NER kategorilerine göre frekans dağılımı.

Bununla birlikte, NER sonuçları modern dil modellerinin Osmanlı Türkçesi gibi morfolojik açıdan karmaşık ve yazım varyantlarının çok olduğu bir dile uygulanmasındaki sınırlılıkları da açıkça göstermektedir. Bazı farklı varlık kategorilerine ait kelimelerin “özel isim” (*proper noun*) olarak etiketlenmesi ve modelin tanımlayamadığı varlıkların yüksek oranda bulunması bu sınırlılıklara örnektir.¹²



Görsel 2: En sık görülen terimler.

Osmanlı Türkçesi üzerinde mevcut yapay zekâ tabanlı NER sistemlerinin performansı, ince ayar (*fine-tuning*) ve dil uyarlamaları yapılmadan optimal seviyeye ulaşmamaktadır. Mevcut model çıktılarının frekans analizi incelendiğinde (Görsel 2), gerçek anlamda adlandırılmış varlık kategorisine giren öğeler arasında “Sultan” ve “Mehmed” öne çıkmaktadır. Bununla birlikte, sistem çıktılarında yer alan “esnaf,” “fetih,” “tahrir,” “hazret” ve “camii” gibi terimler, tek başlarına özel ad niteliği taşımamakta; yalnızca belirli bir özel adla birleştiğinde (örn. *İstanbul’un Fethi*, *Süleymaniye Camii*, *Üsküdar esnafı*) adlandırılmış varlık olarak kabul edilebilmektedir.

¹² Burada ayrıca şu ayrımı vurgulamak gerekir: POS etiketleri dilbilgisel türlere, NER etiketleri ise adlandırılmış varlıklara karşılık gelmektedir. Bu çalışmada odak NER sonuçlarıdır; POS çıktıları yalnızca morfolojik çözümlemenin yan çıktıları olarak değerlendirilmiştir.

Bu bulgular, Osmanlı Türkçesi metinlerde adlandırılmış varlık tespitinin tarihsel bağlamı çözümlemede sunduğu potansiyeli ortaya koyarken, aynı zamanda dilsel karmaşıklığın ve mevcut sınırlılıklarının metodolojik olarak dikkate alınması gerektiğini göstermektedir.

“İstanbul” ve “Mahalle”nin Ağ Tabanlı Semantik Profili: Eş-Oluşum (Co-occurrence) Analizi

Metin madenciliğinde eş-oluşum analizi, bir terimin başka terimlerle aynı bağlamda (örneğin aynı cümle veya belirli bir kelime penceresi içinde) birlikte görülme sıklığını inceleyerek bu terimler arasındaki anlamsal yakınlığı ortaya koyar. Bu yöntem, yalnızca kelime frekanslarını hesaplamaktan öte, kavramlar arası ilişki ağını modellemeye imkân verir. Çoğunlukla korelasyon katsayıları veya benzerlik ölçümleri gibi istatistiksel yöntemler aracılığıyla nicel değerlendirmeler yapılır.¹³ Bu çalışmada, *Seyâhatnâme*'nin modern Türkçe versiyonunda “İstanbul” ve “mahalle” kelimelerinin semantik ilişki ağları analiz edilmiştir. Karşılaştırma amacıyla, aynı yöntem Osmanlı Türkçesi versiyonunda “İsambol” terimi üzerinden de uygulanmıştır. Analiz, Doğal Dil İşleme (NLP) teknikleri ile yürütülmüş; seçilen hedef kelimelerin bağlamında yer alan komşu kelimeler belirlenerek bu kelimeler arasındaki ilişkiler ağ yapısında görselleştirilmiştir.

Analiz öncesinde modern Türkçe ve Osmanlı Türkçesi metinleri üzerinde çok aşamalı bir ön işleme süreci uygulanmıştır. İlk olarak metinlerdeki tüm noktalama işaretleri ve rakamsal ifadeler temizlenmiştir. Ardından, modern Türkçe metinler için mevcut *stop-word* listesi kullanılarak bağlaç, zamir ve ek gibi işlev sözcükleri çıkarılmış; Osmanlı Türkçesi metinleri için ise tarihsel bağlamı dikkate alan, genişletilmiş özel bir *stop-word* listesi hazırlanmıştır. Bu çalışmada kullanılan *stop-word* listesi, bağlaçlar, zamirler, ekler ve sık kullanılan işlevsel sözcüklerden oluşmaktadır. Osmanlı Türkçesi versiyonları için ayrıca tamlama ekleri ve bağlaçlara yönelik genişletilmiş bir liste oluşturulmuş, böylece metnin semantik içeriğini taşımayan unsurlar ayıklanmıştır. Bu kapsamda özellikle tamlama ekleri (-t, -i, -u, -ü), ilgi ekleri (-ın, -in, -un, -ün) ve bağlaçlar (ve, ile, dahi,

13 Stefan Bordag, “A Comparison of Co-occurrence and Similarity Measures as Simulations of Context,” *Computational Linguistics and Intelligent Text Processing: CICLing 2008*, ed. Alexander Gelbukh, *Lecture Notes in Computer Science*, vol. 4919 (Berlin-Heidelberg: Springer, 2008), https://doi.org/10.1007/978-3-540-78135-6_5.

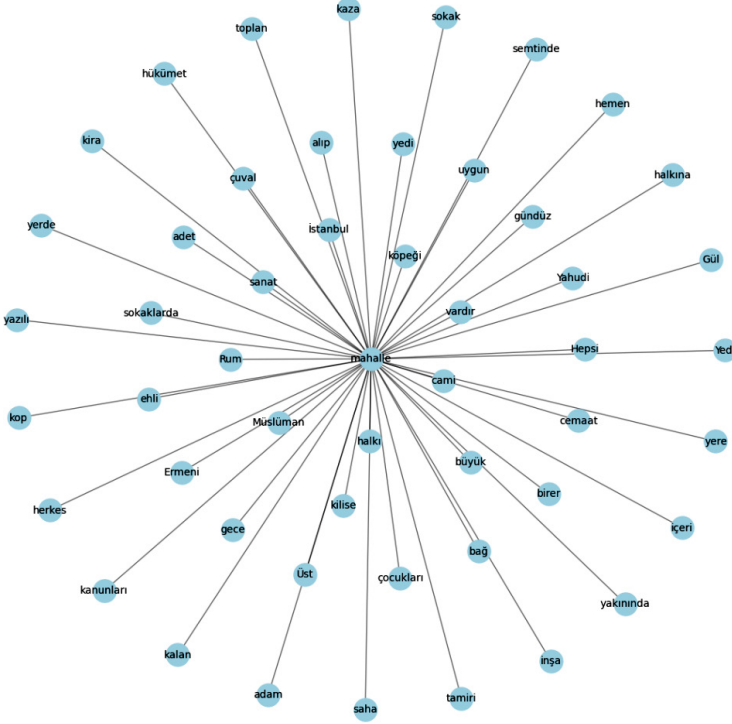
Analiz bulguları, “İstanbul” kelimesinin metin içerisinde farklı tematik kümelerle çok yönlü ilişkiler kurduğunu ortaya koymaktadır. Coğrafi bağlamda “Üsküdar,” “Galata,” “Eyüp,” “boğazi,” “şehir” ve “deniz” gibi terimlerle eş-oluşum ilişkileri, kentin topografik ve mekânsal referans noktalarıyla güçlü bir bağlantıya işaret etmektedir. Zamansal bağlamda “eski” ve “zamanında” gibi ifadeler, İstanbul’un tarihsel sürekliliğini ve geçmişe dönük anlatıların yoğunluğunu yansıtmaktadır. Mimari bağlamda “kale,” “cami,” “Ayasofya,” “han” ve “imar” gibi terimler, kentin anıtsal yapıları ve kentsel dönüşüm süreçleriyle ilişkisini göstermektedir. İdari ve siyasi düzlemde “padişah,” “sultan,” “Mehmed,” “Süleyman” ve “fetih” gibi referanslar, İstanbul’un imparatorluk merkezî kimliğini vurgulamaktadır. Niceliksel ve niteliksel tanımlamalar bağlamında “büyük,” “nice” ve “meşhur” gibi sıfatlar ise kentin hem fiziksel hem de sembolik büyüklüğüne yapılan atıfları pekiştirmektedir. Bununla birlikte, elde edilen ağ grafiği yalnızca kelimeler arasındaki eş-oluşum ilişkilerinin görsel bir temsili olup, bu ilişkilerin bağlamsal ve söylemsel anlamının derinlemesine kavranabilmesi için ayrıntılı bir söylem analizine ihtiyaç duyulmaktadır.¹⁵

Benzer şekilde “mahalle” kelimesi üzerinden gerçekleştirilen eş-oluşum analizi sonuçlarında yine tematik örüntüler takip edilebilmektedir (*Görsel 4*). Çok kültürlü topluma ve sosyal hayata işaret eden “cami,” “kilise,” “Ermeni,” “Rum,” “Müslüman,” “Yahudi” gibi atıfların yanı sıra kentsel mekân göstergeleri olarak “sokak,” “semt,” “İstanbul” gibi ve topluluk ya da kurum referansları olarak ise bağlama göre “hükümet,” “halk,” “herkes,” “cemaat,” “ehli,” “sanat” gibi atıflar görülmektedir.

Modern Türkçe versiyon, tarihsel bir anlatı niteliği taşımasının yanı sıra çeviri metin olması nedeniyle, gerek kelime dağarcığı ve dil bilgisi gerekse söz dizimi bakımından günümüz Türkçesinden belirgin farklılıklar göstermektedir. Bu durum, yürütülen analizlerin sonuçlarını doğrudan etkilemekte ve mevcut dil işleme sistemleri açısından, Osmanlı Türkçesi metinlerinin analizini çok daha güç ve karmaşık hâle getirmektedir. Nitekim, eş-oluşum analizi Osmanlı

¹⁵ Ağ grafiğinde bazı varyantların (*cami/camii, fetih/feth, Ayasof/Ayasofya, şehir/şehirinde*) ayrı düğümler hâlinde görünmesi dikkat çekmektedir. Bu tür farklılıklar büyük ölçüde ön işleme ve transliterasyon sürecindeki sınırlılıklardan kaynaklanmakta olup, görselleştirmede fazladan düğümler yaratabilmektedir. Yorumlama aşamasında ise bu biçimler tekil karşılıkları altında değerlendirilmiştir.

'mahalle' için Eş-oluşum (co-occurrence) Kelimeleri (50)



Görsel 4: "Mahalle" için eş-oluşuma sahip 50 kelime (Modern Türkçe versiyonu).

Türkçesi modeli kullanılarak da gerçekleştirilmiş; ancak dil modelinin özellikle tokenizasyon ve lemmatizasyon aşamalarında istenen başarıyı sağlayamadığı gözlemlenmiştir (Görsel 5). Bunun temel nedeni, mevcut Doğal Dil İşleme (NLP) modellerinin modern Türkçeyi işlemek üzere tasarlanmış olması ve metni çağdaş kelime dağarcığına göre tokenize etmesidir. Bu bağlamda, çalışmada Osmanlı Türkçesine özgü daha anlamlı sonuçlar elde edebilmek amacıyla tamlama ekleri ve bağlaçları kapsayan genişletilmiş *stop-word* listesi ön işleme sürecine dâhil edilmiştir.

tokenize etse dahi, bağlamsal anlamın eksik ya da hatalı yorumlanmasına yol açabilmektedir. Ayrıca, dönemin yazımsal çeşitliliği ve standart bir imla sisteminin bulunmaması, modelin eğitimi sırasında bu varyasyonların sistematik biçimde dikkate alınmasını zorunlu kılmaktadır. Bu bağlamda, özellikle Osmanlı Türkçesi için kapsamlı, etiketlenmiş ve yüksek kaliteli veri setlerinin hazırlanması kritik önemdedir. Bununla birlikte, Osmanlı Türkçesinin modern Türkçeye aktarımında kullanılacak transkripsiyon modelinin belirlenmesi, söz konusu veri setlerinin yapısının tutarlılığı açısından belirleyici olacaktır. Dolayısıyla, dijital ortamda üretilecek Osmanlı Türkçesinin Latin harflerine aktarımında kullanılacak çeviriyazı standardının belirlenmesi hem model eğitimi hem de sonraki dil işleme süreçleri açısından önemli bir kazanım sağlayacaktır. Böyle bir veri temeli üzerinde geliştirilecek dil modelleri, Osmanlı Türkçesinin morfolojik yapısını, dil bilgisi kurallarını ve yazımsal varyasyonlarını daha etkin biçimde öğrenerek, doğru ve anlamlı tokenizasyon ile lemmatizasyon işlemlerini mümkün kılacaktır.

***Seyâhatnâme*'nin Latent Dirichlet Allocation (LDA) ile Konu Modellemesi**

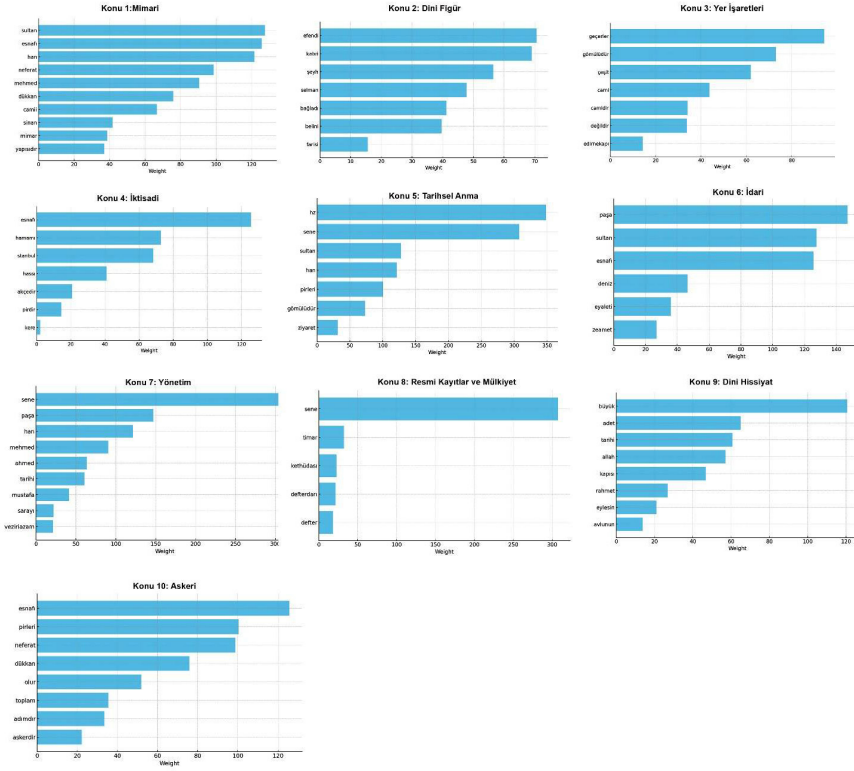
Seyâhatnâme'nin İstanbul'a odaklanan anlatısında makro ölçekte hangi tematik konuların öne çıktığını sistematik olarak ortaya koymak, hesaplamalı yöntemlerle mümkündür. Metin madenciliği ve doğal dil işleme (NLP) alanında sıklıkla kullanılan Latent Dirichlet Allocation (LDA), metinler içerisindeki gizli (*latent*) temaları ortaya çıkarmak amacıyla geliştirilmiş istatistiksel bir yöntemdir.¹⁶ LDA, gözetimsiz (*unsupervised*) makine öğrenmesi yaklaşımıyla, bir metin külliyatındaki tekrarlayan ve eş-oluşum sergileyen kelime gruplarını kümeleyerek bu kümeleri “konu” (*topic*) olarak adlandırır. Böylece, farklı metinler arasındaki benzerlikler tespit edilebilir, özetleme yapılabilir ve metinsel örüntüler sistematik biçimde analiz edilebilir.¹⁷

Bu çalışmada, LDA yöntemi *Seyâhatnâme*'nin modern Türkçe versiyonu üzerinde uygulanmış; İstanbul anlatısındaki tematik yapının niceliksel olarak incelenmesi hedeflenmiştir. Ön işleme sürecinde *stop-word*, sayılar ve anlamsal

¹⁶ David M. Blei, Andrew Y. Ng and Michael I. Jordan, “Latent Dirichlet Allocation,” *Journal of Machine Learning Research*, no. 3 (2003): 993-1022.

¹⁷ Galip Aydın ve İbrahim R. Hallaç, “Türkçe Metinlerde Otomatik Konu Tespiti,” *Fırat Üniversitesi Mühendislik Bilimleri Dergisi* 33, s. 2 (2021): 599-606.

değeri düşük ifadeler metinden ayrılmıştır. Elde edilen konu kümelerine ait başlıklar, bağlam analiziyle ve GPT-4 tabanlı yardımcı sınıflandırma desteğiyle tanımlanmıştır (Görsel 6).



Görsel 6: LDA ile Seyâhatnâme'nin konu modellemesi.

LDA, her metni belirli oranlarda birden fazla konunun bileşimi olarak tanımlar; her konu ise kelimeler üzerindeki olasılık dağılımıyla temsil edilir. Konular, “Konu 1”den “Konu 10”a kadar numaralandırılmış, her birinde öne çıkan kelimelerin göreceli ağırlıkları hesaplanmıştır. Örneğin, Konu 1: Mimari başlığı, sıklıkla mimari unsurlarla ilişkilendirilen kelimelerden oluşmaktadır. Bu kümeleneşte Mimar Sinan’a (ö. 1588) yönelik yoğun referanslar, Evliya Çelebi *Seyâhatnâmesi*’nin “Mimar Sinan eserleri rehberi”¹⁸ olarak değerlendirilmesini

18 Uğur Demir, “Seyyâh-ı Âlem Evliyâ Çelebi’nin Kaleminden Cihan Mimarî Sinan,” *Mimar Sinan ve Su*, haz. Coşkun Yılmaz (İstanbul: Sultangazi Belediyesi, 2017), 11-49; Emin Selçuk

desteklemektedir. Konu 2: Dinî Figür ve Konu 9: Dinî Hissiyat, dinî söylem ve imgelerin güçlü bir şekilde metinde yer aldığını göstermektedir. Konu 4: İktisadi ve Konu 6: İdari kümeleri sosyoekonomik ve yönetsel terimlere işaret ederken, Konu 7: Yönetim ve Konu 8: Resmî Kayıtlar ve Mülkiyet bürokratik yapı ve mülkiyet ilişkilerini yansıtmaktadır.

Bu yaklaşım, Evliya Çelebi'nin metninin niteliksel analizine niceliksel bir boyut eklemekte; metni tematik bölümlere ayırarak her bir bölümün etrafında şekillenen kavramları görünür kılmaktadır. Bununla birlikte, *Seyâhatnâme* mikro ölçekte çok daha zengin ve ayrıntılı temalar barındırmaktadır. Mevcut LDA uygulaması, özellikle daha gelişmiş dil modelleriyle desteklendiğinde, bu ince tematik ayrışmaları daha net ortaya çıkarabilecektir. Bu tür bir modelleme, *Seyâhatnâme*'nin tüm ciltlerine uygulanarak makro ve mikro ölçekteki konu dağılımlarının karşılaştırmalı olarak incelenmesine olanak sağlayacaktır.

Türkçe tarihsel metinlerde LDA uygulaması belirli dilsel sınırlılıklara sahipken, Osmanlı Türkçesi için bu süreç çok daha karmaşık hâle gelmektedir. Türkçenin dilbilgisel özelliklerinin hesaplamalı analizlerde sağlıklı biçimde temsil edilebilmesi için kapsamlı ön işleme adımları ve model iyileştirmeleri gereklidir. Özellikle doğru tokenizasyon ve lemmatizasyon, konu modellemesinin başarısında kritik rol oynamaktadır. Bu nedenle, standart bir LDA algoritmasının, Osmanlı Türkçesine özgü özel uyarlamalar yapılmadan yüksek başarı göstermesi beklenmemektedir.

İstanbul, Galata ve Üsküdar: Mekânların Duygu (Sentiment) Analizi

Duygu analizi, doğal dil işleme (NLP) alanında metinlerde ifade edilen duygusal eğilimleri veya tutumları belirlemeye yönelik bir yöntemdir. Temel olarak sözcüklerin ve ifadelerin taşıdığı olumlu, olumsuz ya da tarafsız anlamların sistematik biçimde sınıflandırılmasını sağlar. Bu yöntem günümüzde sosyal medya içerikleri, kullanıcı yorumları ve haber metinleri gibi geniş yelpazede metinlerin duygusal tonunu analiz etmek için yaygın olarak kullanılmaktadır.

Gelişmiş NLP teknikleri, tarihsel metinlerde de duygu analizinin uygulanmasına imkân tanımaktadır. Duygu analizinin tarihsel metinlere uygulanması, bu metinlerin barındırdığı sosyo-kültürel bağlamların ve duygusal alt örüntülerin

Taşar, "Tahrir, Tasnif ve Tavsif: Evliyâ Çelebi Seyahatnâmesinde Şehir ve Mimari" (Doktora tezi, Mardin Artuklu Üniversitesi, 2020).

anlaşılması açısından önemli bir analitik çerçeve sunar. Yöntem, yazarın belirli bir konuya veya temaya karşı tutumunu olumlu, olumsuz ya da tarafsız biçimde belirlemek ve bu tutumları kategorize etmek amacıyla geliştirilmiştir. Bu nedenle, tarihsel anlatılara uygulandığında, metnin çok katmanlı yorumlanmasına ve yazarın bakış açısının daha incelikli biçimde çözümlenmesine katkı sağlar.¹⁹

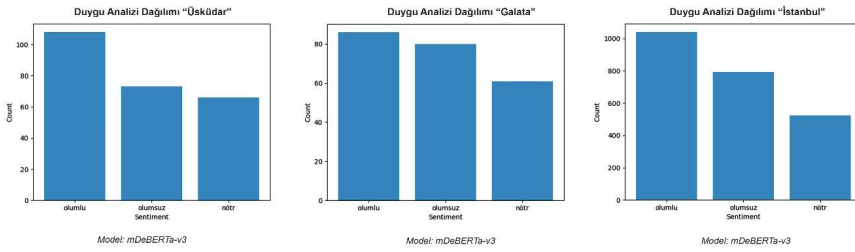
Evliya Çelebi'nin tarihî ve kültürel açıdan oldukça zengin bilgiler içeren *Seyâhatnâmesi* bu yöndeki bir duygu analizi için önemli bir potansiyel taşımadır. Bu çalışmada *Seyâhatnâme*'nin hem modern hem de Osmanlı Türkçesi çeviri yazısı üzerinde duygu analizi gerçekleştirilmiştir. Analizler karşılaştırmalı olarak Türkçe tabanlı ve diller arası geliştirilen BERT *transformer* modelleri uygulanarak yapılmıştır ve başarı sonuçları test edilmiştir. Analizler *zero-shot* çıkarım (eğitim verisi olmaksızın sınıflandırma) yaklaşımıyla yürütülmüştür.

Analiz sürecinde, *Seyâhatnâme* metinleri ön işleme adımlarından geçirilmiş ve İstanbul/İslambol (Suriçi), Galata ve Üsküdar anlatıları ayrı veri setleri hâline getirilmiştir. Türkçe tabanlı Turkish-BERT²⁰ modeli, dilsel uyumu optimize etme yeteneği nedeniyle; çok dilli mDeBERTa-v3²¹ modeli ise bağlamsal anlam çıkarımı kapasitesi nedeniyle tercih edilmiştir. Her iki model de metinleri işleyerek, ilgili mekânlar hakkında olumlu, olumsuz veya nötr duygu etiketleri üretmiştir. Modeller, her cümle için bu etiketlere karşılık gelen olasılık tabanlı skorlar hesaplamış; en yüksek olasılığa sahip etiket, cümlelerin duygu kategorisi olarak atanmıştır. Skor değerleri, cümlelerin sınıflandırma güven düzeyini yansıtan ek bir ölçüt olarak değerlendirilmiştir. Üç mekânsal bölge için gerçekleştirilen analizlerde cümleler duygu etiketleriyle işaretlenmiş; model doğruluğu rastgele seçilmiş 10'ar cümlelik kontrol veri setleri üzerinden test edilmiştir.

¹⁹ Pansy Nandwani and Rupali Verma, "A Review on Sentiment Analysis and Emotion Detection from Text," *Social Network Analysis and Mining*, no. 11 (2021): 81, <https://doi.org/10.1007/s13278-021-00776-0>; Mayur Wankhade, Annavarapu Chandra Sekhara Rao and Chaitanya Kulkarni, "A Survey on Sentiment Analysis Methods, Applications, and Challenges," *Artificial Intelligence Review*, no. 55 (2022): 5731-5780, <https://doi.org/10.1007/s10462-022-10144-1>.

²⁰ BERT-base Turkish Sentiment Model (Hugging Face), <https://huggingface.co/savasy/bert-base-turkish-sentiment-cased> (erişim 22 Temmuz 2025).

²¹ BERT-base mDeBERTa-v3-base-xnli-multilingual-nli-2mil7 (Hugging Face), <https://huggingface.co/MoritzLaurer/mDeBERTa-v3-base-xnli-multilingual-nli-2mil7> (erişim 24 Temmuz 2025).

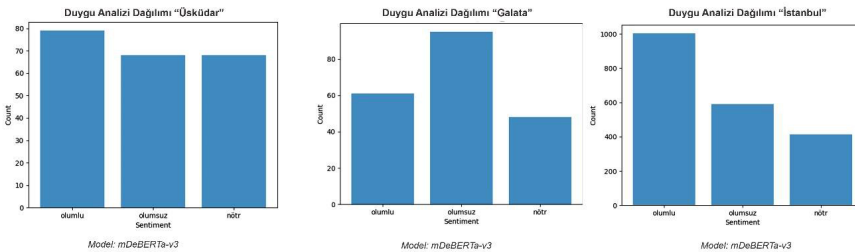


Görsel 7: İstanbul, Galata ve Üsküdar için mDeBERTa-v3 ile duygu analizi (Modern Türkçe).

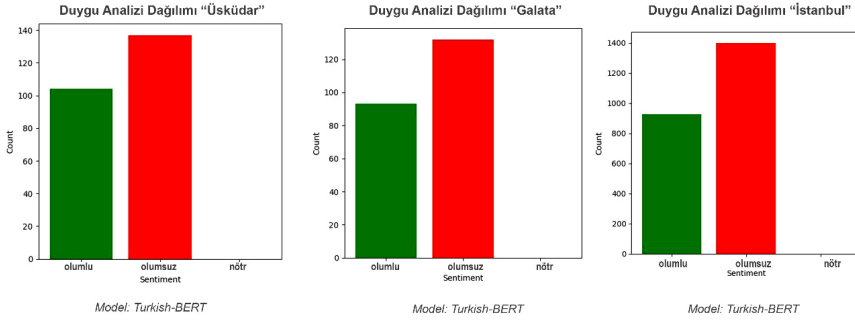
Modern Türkçe metinlerde mDeBERTa-v3 modelinin genel doğruluk oranı %66,67 olarak ölçülmüştür. Üç mekânsal bölge için de 'olumlu' etiketli cümlelerin oranı yüksektir (Görsel 7). Doğru sınıflandırılan cümlelerde ortalama skor 0,589, yanlış sınıflandırılanlarda ise 0,484 olarak bulunmuştur. Bu durum, daha yüksek skorların doğru sınıflandırma ile pozitif korelasyona sahip olduğunu göstermektedir. Kategori bazında, nötr cümlelerin %72,73'ü, olumlu cümlelerin %61,54'ü ve olumsuz cümlelerin %66,67'si doğru etiketlenmiştir. Bulgular, tarihsel anlatıya sahip Türkçe metinlerde duygu analizi performansının artırılması için modelin daha fazla geliştirilmesi gerektiğini ortaya koymaktadır.

Aynı model Osmanlı Türkçesi metinlerde uygulandığında, doğruluk oranının %40'a düştüğü gözlenmiştir. Ortalama duygu skoru 0,525, standart sapma ise 0,145'tir. Bu sonuç, modelin sınırlı düzeyde bir ayrıştırma yeteneğine sahip olduğunu, ancak anlamlı bir iyileştirme (fine-tuning) olmadan yüksek başarıya ulaşamayacağını göstermektedir (Görsel 8).

Turkish-BERT modeli modern Türkçe metinlerde uygulandığında ise genel doğruluk oranı yalnızca %16,67'dir (Görsel 9). Bu, modelin cümlelerin



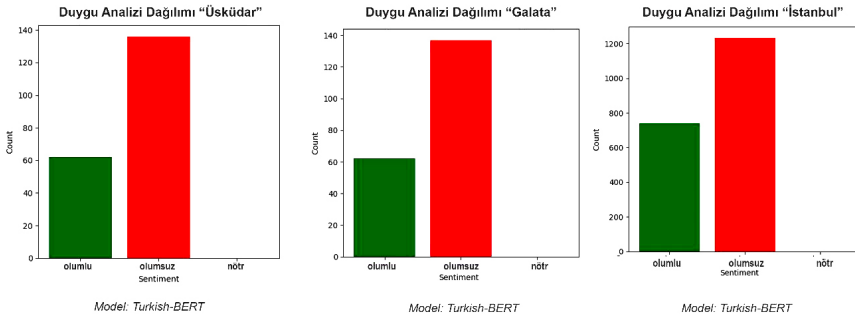
Görsel 8: İstanbul (İslambol), Galata ve Üsküdar için mDeBERTa-v3 ile duygu analizi (Osmanlı Türkçesi).



Görsel 9: İstanbul, Galata ve Üsküdar için *Turkish-BERT* ile duygu analizi (Modern Türkçe).

duygularını yalnızca altıda bir oranında doğru tespit edebildiğini ve oldukça düşük bir başarı gösterdiğini ortaya koymaktadır.

Model Osmanlı Türkçesi versiyona uygulandığında da benzer şekilde başarısız olduğu ölçülmüştür. Veri kümesindeki duygu etiketlerinin dağılımı büyük ölçüde yanlıştır ve cümlelerin %73,33'ü olumsuz ve yalnızca %26,67'si olumlu olarak sınıflandırılmıştır (*Görsel 10*).



Görsel 10: İstanbul (İslambol), Galata ve Üsküdar için *Turkish-BERT* ile duygu analizi (Osmanlı Türkçesi).

Bu karşılaştırma, çok dilli modelin (mDeBERTa-v3) başarı oranının, Türkçe tabanlı modele kıyasla görece daha iyi olduğunu, ancak her iki modelin de tarihsel metinlerde duygu ayrımını yeterli düzeyde yapamadığını göstermektedir. Bulgular, Osmanlı Türkçesi gibi dilsel açıdan karmaşık ve yazımsal çeşitliliği yüksek tarihî metinlerde, duygu analizine yönelik dil modellerinin daha geniş, etiketli veri setleriyle yeniden eğitilmesi gerektiğini açıkça ortaya koymaktadır.

Sonuç

Bu çalışma, Evliya Çelebi *Seyâhatnâmesi*'nin İstanbul cildinin modern Doğal Dil İşleme (NLP) ve metin madenciliği teknikleri ile, GPT tabanlı ileri yapay zekâ yöntemleri kullanılarak analizinin ve dijitalleştirilmesinin sunduğu olanakları ortaya koymaktadır. Eserin edebî değeri kadar, Osmanlı İmparatorluğu'nun sosyal, kültürel ve idari yapısına dair içerdiği zengin veriler, hesaplamalı yöntemler aracılığıyla yeniden değerlendirildiğinde, yalnızca bu metne değil, benzer tarihî kaynaklara yönelik de yeni araştırma perspektifleri geliştirmeye imkân tanımaktadır.

Dijital tekniklerin uygulanması, tarihî metinlerin daha geniş bir araştırmacı kitlesi tarafından erişilebilir, anlaşılabilir ve yeniden kullanılabilir hâle gelmesini sağlamaktadır. Bununla birlikte, Osmanlı Türkçesinin yapısal karmaşıklığı ve yazımsal çeşitliliği, hesaplamalı çalışmalarda dikkate alınması gereken önemli metodolojik zorluklar yaratmaktadır. Bu zorluklar, özellikle çok katmanlı morfolojik yapı, Arapça ve Farsça kökenli bileşik ifadelerin çözümlemesi ve Latin harfli transkripsiyonlardaki çeşitlilik gibi dilsel boyutlarda kendini göstermektedir. Ayrıca, tarihî metinlerdeki ortografik belirsizlikler ve dönemsel söz varlığına özgü kavramların güncel dil modellerince yeterince temsil edilememesi, mevcut sistemlerin sınırlarını ortaya koymaktadır. Bu nedenle, Osmanlı Türkçesine özgü NLP modelleri ve yapay zekâ tabanlı araçların geliştirilmesi, hem yazılı kültürel mirasla etkileşimi derinleştirecek hem de araştırma kapasitesini artıracaktır.

Süreç, Osmanlı İmparatorluğu'nun sosyal ve kültürel dinamiklerini daha kapsamlı biçimde anlamak için disiplinler arası yaklaşımları gerekli kılmaktadır. Ayrıca, bu tür çalışmaların sürdürülebilirliği için dijital arşivlerin erişim, yüksek çözünürlüklü veri sağlama ve uluslararası üst veri (*metadata*) standartları konusunda güçlendirilmesi önem taşımaktadır. Bu bağlamda, özellikle Dublin Core, TEI (Text Encoding Initiative)²², METS/MODS ve yüksek çözünürlüklü görsel veri için IIIF gibi standartların kullanımı hem veri paylaşımını kolaylaştıracak hem de farklı araştırma ortamları arasında birlikte çalışabilirliği artıracaktır.

Sonuç olarak, bu araştırma, Evliya Çelebi'nin *Seyâhatnâmesi*'nden hareketle tarihî metinlerin dijital beşerî bilimler alanında daha etkin temsil edilmesine yönelik yeniden üretilebilir bir iş akışı ortaya koymakta; Osmanlı Türkçesi

22 TEI'nin şematik uyarlaması ve uygulama örnekleri için ayrıca bkz. Fatma Aladağ, "Osmanlı Çalışmalarında Dijital Edisyon ve Metin Kodlama: Text Encoding Initiative (TEI)," *Zemin*, s. 6 (2023): 190-211, <https://doi.org/10.5281/zenodo.10435921>.

üzerine gelecekte yürütülecek çalışmalara hem teknik hem de kavramsal açıdan sağlam bir metodolojik temel sunmaktadır.

Kaynaklar

- Aladağ, Fatma. “Dijital Beşerî Bilimler ve Türkiye Araştırmaları: Bir Literatür Değerlendirmesi.” *Türkiye Araştırmaları Literatür Dergisi*, s. 18 (2020): 773-796. <https://dergipark.org.tr/tr/pub/talid/issue/58073/800406>.
- _____. “Osmanlı Çalışmalarında Dijital Edisyon ve Metin Kodlama: Text Encoding Initiative (TEI).” *Zemin*, s. 6 (2023): 190-211. <https://doi.org/10.5281/zenodo.10435921>
- _____ ve Elif Derin Can. “Programlamadan Yapay Zekâya Osmanlı Türkçesi İçin Metin Tanıma ve Otomatik Transkripsiyon.” *Dijital Osmanlı Çalışmaları*, hazırlayan Yunus Uğur, 1-20. İstanbul: Vakıf Bank Yayınları, 2023.
- Archer, Dawn. “Data Mining and Word Frequency Analysis.” *Research Methods for Reading Digital Data in the Digital Humanities*, 72-92. Edinburgh: Edinburgh University Press, 2016. <https://doi.org/10.1515/9781474409629-006>.
- Aydın, Galip ve İbrahim R. Hallaç. “Türkçe Metinlerde Otomatik Konu Tespiti.” *Fırat Üniversitesi Mühendislik Bilimleri Dergisi* 33, s. 2 (2021): 599-606.
- Bilgin Taşdemir, Esmâ F. “Printed Ottoman Text Recognition Using Synthetic Data and Data Augmentation.” *International Journal on Document Analysis and Recognition*, no. 26 (2023): 273-287.
- Blei, David M., Andrew Y. Ng and Michael I. Jordan. “Latent Dirichlet Allocation.” *Journal of Machine Learning Research*, no. 3 (2003): 993-1022.
- Bordag, Stefan. “A Comparison of Co-occurrence and Similarity Measures as Simulations of Context.” *Computational Linguistics and Intelligent Text Processing: CICLing 2008*, edited by Alexander Gelbukh, *Lecture Notes in Computer Science*, vol. 4919. Berlin-Heidelberg: Springer, 2008. https://doi.org/10.1007/978-3-540-78135-6_5.
- Demir, Uğur. “Seyyâh-ı Âlem Evliyâ Çelebi’nin Kaleminden Cihan Mîmarı Sinan.” *Mimar Sinan ve Su*, hazırlayan Coşkun Yılmaz, 11-49. İstanbul: Sultangazi Belediyesi, 2017.
- Digital Ottoman Corpora* (website). <https://www.digitalottomancorpora.org> (erişim 12 Kasım 2025).
- Digital Ottoman Studies Platform* (website). <https://www.digitalottomanstudies.com> (erişim 10 Kasım 2025).
- Duran, Zeynep. “Evliya Çelebi Seyahatnâmesi’ne Göre İstanbul’da Esnaf, Zanaat ve Ticaret.” Yüksek Lisans Tezi. Kırıkkale Üniversitesi, 2006.
- Ehrmann, Mathieu, Matteo Romanello, Matthias Flückiger and Simon Clematide. “Challenges in Named Entity Recognition for Historical Texts.” *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries (JCDL)* (2023). <https://doi.org/10.1145/3604931>.

- Evliyâ Çelebi. *Günümüz Türkçesiyle Evliyâ Çelebi Seyahatnâmesi: İstanbul*. Hazırlayan Yücel Dağlı ve Seyit Ali Kahraman. İstanbul: YKY, 2003.
- Evliyâ Çelebi. *Seyahatnâme: I. Kitap*. Hazırlayan Robert Dankoff, Seyit Ali Kahraman ve Yücel Dağlı. İstanbul: YKY, 2008.
- Feldman, Ronen ve James Sanger. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge: Cambridge University Press, 2007. <https://doi.org/10.1017/CBO9780511546914>.
- Kayabaş, Ayla, Helmut Schmid, Ahmet Ercan Topçu ve Özkan Kılıç. "TRMOR: A Finite-State-Based Morphological Analyzer for Turkish." *Turkish Journal of Electrical Engineering and Computer Sciences* 27, no. 5 (2019): 3837-3851. <https://doi.org/10.3906/elk-1902-125>.
- Kırmızıaltın, Süphan ve David Wrisley. "Automated Transcription of Historical Periodicals with HTR: Challenges and Futures." *Digital Humanities Quarterly* 16, no. 2 (2022).
- Köksal, Abdullatif. *Turkish-Lemmatizer* (GitHub repository). <https://github.com/akoksal/Turkish-Lemmatizer> (erişim 13 Haziran 2025).
- Martins, Bruno, Patricia Murrieta-Flores ve Miguel Won. "Ensemble Named Entity Recognition (NER): Evaluating NER Tools in the Identification of Place Names in Historical Corpora." *Frontiers in Digital Humanities*, no. 5 (2018): 333859. <https://doi.org/10.3389/fdigh.2018.00002>.
- Nandwani, Pansy and Rupali Verma. "A Review on Sentiment Analysis and Emotion Detection from Text." *Social Network Analysis and Mining*, no. 11 (2021): 81. <https://doi.org/10.1007/s13278-021-00776-0>.
- Taşar, Emin Selçuk. "Tahrir, Tasnif ve Tavsif: Evliyâ Çelebi Seyahatnâmesinde Şehir ve Mimari." Doktora Tezi. Mardin Artuklu Üniversitesi, 2020.
- Wankhade, Mayur, Annavarapu Chandra Sekhara Rao and Chaitanya Kulkarni. "A Survey on Sentiment Analysis Methods, Applications, and Challenges." *Artificial Intelligence Review*, no. 55 (2022): 5731-5780. <https://doi.org/10.1007/s10462-022-10144-1>.
- Zemberek-NLP (GitHub repository). Ahmet A. Akın. <https://github.com/ahmetaa/zemberek-nlp>. (erişim 12 Haziran 2025).
- BERTurk* (Hugging Face). <https://huggingface.co/dbmdz/bert-base-turkish-cased>. (erişim 22 Temmuz 2025).
- BERT-base Turkish Sentiment Model* (Hugging Face). <https://huggingface.co/savasy/bert-base-turkish-sentiment-cased>. (erişim 22 Temmuz 2025).
- mDeBERTa-v3-base-xnli-multilingual-nli-2mil7* (Hugging Face). <https://huggingface.co/MoritzLaurer/mDeBERTa-v3-base-xnli-multilingual-nli-2mil7>. (erişim 24 Temmuz 2025).