



Web Kazıma Yöntemlerinin Tüketici Fiyat Endeksi Çalışmalarında Kullanılması Türkiye Örneği

Yeliz YALÇIN

Ankara Hacı Bayram Veli Üniversitesi, Ekonometri Bölümü / Prof. Dr.

yeliz.yalcin@hbv.edu.tr

Orcid No: 0000-0002-7141-3086

Burak BAYSAL

Ankara Hacı Bayram Veli Üniversitesi, Ekonometri Bölümü / Lisansüstü Öğrencisi

burak.baysal@hbv.edu.tr

Orcid No: 0009-0008-5431-3172

Özet

Tüketici Fiyat Endeksleri, hanehalklarının tüketimine konu olan mal ve hizmetlerin fiyatlarının genel düzeyinde zaman içerisinde meydana gelen değişimlerin bir ölçüsü olarak tanımlanmaktadır. Gelişen teknoloji ile birlikte özellikle fiyat istatistikleri bağlamında kullanılacak yeni veri derleme yöntemleri ortaya çıkmış, bu sayede Tüketici Fiyat Endeksi çalışmaları için oldukça hızlı bir biçimde girdi sağlanabilmesi mümkün hale gelmiştir. Bu yöntemlerden biri internet üzerindeki bilgilerin otomatik kazıcılar vasıtası ile yapısal veriye dönüştürülmesini ve veri tabanlarına kaydedilmesini sağlayan web kazıma yöntemidir. Bu sayede geleneksel veri derleme yöntemlerinin kullanımı ile aylık olarak üretilen Tüketici Fiyat Endeksi göstergelerinin günlük veya haftalık olarak daha yüksek frekanslarla hesaplanabilmesi mümkün hale gelmiştir. Çalışmada, web kazıma yöntemleri ile internet üzerinden derlenecek fiyatlar kullanılarak günlük ve aylık Tüketici Fiyat Endeksleri elde edilmesi amaçlanmıştır. Uluslararası anlamda kabul görmüş endeks hesaplama yöntemleri kullanılarak ulaşılan bulgular dikkate alınan dönemde web kazıma yöntemleri ile elde edilen sonuçların resmi istatistiklerle tutarlı sonuçlar sağladığını ortaya koymuştur.

Anahtar sözcükler: Tüketici Fiyat Endeksi, Web kazıma, Metin madenciliği.

Sorumlu Yazar / Corresponding Author: Burak BAYSAL, Ankara Hacı Bayram Veli Üniversitesi, Ekonometri Bölümü.

Atıf / Citation: YALÇIN Y., BAYSAL B. (2024). Web Kazıma Yöntemlerinin Tüketici Fiyat Endeksi Çalışmalarında Kullanılması Türkiye Örneği. *İstatistik Araştırma Dergisi*, 14 (2), 46-58.

Makalede yer alan görüş ve düşünceler yazarlarına ait olup dergiyi, editörlerini ve Türkiye İstatistik Kurumu'nu bağlamamaktadır.

Using Web Scraping Methods in Consumer Price Index Studies, Türkiye Example

Abstract

Consumer Price Indices are defined as a measure of the changes in the prices of goods and services that are subject to household consumption over time. With the developing technology, new data collection methods have emerged that can be used especially in the context of price statistics. In this way, it has become possible to provide a quick input to the Consumer Price Index studies. The web scraping method, which enables the information on the Internet to be converted into structured data by means of automatic scrapers and saved in databases, is one of these methods. In this way, it has become possible to calculate the Consumer Price Index indicators, which are produced monthly with traditional data collection methods, on a daily or weekly basis. In the study, it is aimed to obtain daily and monthly Consumer Price Index by using web scraping methods and prices compiled over the internet. Findings obtained by using internationally accepted index calculation methods revealed that the results obtained by web scraping methods provide results consistent with official statistics.

Keywords: Consumer Price Index, Web scraping, Text mining.

1. Giriş

Endeksler, bir değişkenin zamana veya mekana göre ne şekilde değiştiğinin sayısal bir ölçüsüdür. Bir ekonomide tüketiciler tarafından satın alınan milyonlarca farklı mal ve hizmet yer almaktadır. Bu mal ve hizmet fiyatlarının genel düzeyinde zaman içerisinde meydana gelen değişimler ise Tüketici Fiyat Endeksleri (TÜFE) yardımı ile ölçülmektedir. TÜFE verileri başlı başına uluslararası karşılaştırmalarda kullanılan bir gösterge olmanın yanı sıra ücret ve maaşların güncellenmesi, birçok değişken için reel değerlerin hesaplanması, enflasyon hedeflemesi ve fiyat istikrarının izlenmesi gibi alanlarda kullanılmaktadır.

TÜFE hesaplamalarının gerçekleştirilebilmesi için öncelikle toplumun harcama yapısını temsil edecek bir mal ve hizmet sepeti belirlenmesi, bu sepette yer alan her bir maddenin göreceli ağırlığının hesaplanması, tüketime konu olan mal ve hizmetlerin uygun bir sınıflama sistemine göre tasnif edilmesi ve son olarak baz alınan dönem ve cari dönem fiyatlarının derlenmesi gerekmektedir. Tüm bu girdilerin elde edilmesinin ardından uygun endeks formülü kullanılarak endeks hesaplaması gerçekleştirilmektedir. Bu girdilerin tamamının bir araya getirilmesi ve TÜFE hesaplamalarının yapılmasından ulusal istatistik ofisleri sorumludur. Özellikle geleneksel veri derleme yöntemleri kullanımının ciddi bir zaman ve işgücü maliyeti gerektirmesinden dolayı bireysel olarak bu hesaplamaların yapılması son derece zor görünmektedir. Ancak günümüzde teknolojinin gelişmesi ile alternatif veri derleme yöntemleri ortaya çıkmış, bu sayede daha düşük zaman ve işgücü maliyetleri ile veri derleme uygulamaları gerçekleştirilmesi mümkün hale gelmiştir. Bu çalışmanın da konusunu oluşturan TÜFE çalışmaları açısından başlıca alternatif veri kaynakları olarak internet ve barkod verileri öne çıkmaktadır. Bilindiği üzere son dönemde tüketicilerin satın alma alışkanlıkları önemli ölçüde değişmiş, online alışverişlerin hacmi ciddi oranda artmıştır. Bununla birlikte çok sayıda işyeri geleneksel mağazalarının yanı sıra kendi web siteleri üzerinden de satış yapmaya başlamış, bazı işyerleri ise geleneksel mağazacılık anlayışını terk ederek yalnızca internet üzerinden satış gerçekleştirir hale gelmiştir. İnternet, ekonomi üzerinde son derece etkili olurken sürekli büyüyen bir veri kaynağı haline de gelmiştir. Özellikle TÜFE çalışmaları açısından bu verilerin kullanılması istatistiki birimlerin cevaplayıcı yükünün azaltılması ve manuel veri derlemeden kaynaklanan maliyetlerin azaltılması noktalarında faydalı görülmektedir. Ayrıca daha yüksek frekans ve hacimlerde veri derlemenin mümkün hale gelmesi, istatistik üretim süreçlerinin hızlandırılması açısından da faydalı olacaktır. EUROSTAT (2020) çalışmasına göre; artan gözlem sayıları ve web kazıma teknolojilerinin otomasyonu vasıtasıyla yeni göstergelerin geliştirilmesi, veri kalıplarının daha iyi anlaşılması, fiyatların günlük, saatlik ve daha yüksek frekanslarla izlenmesi mümkün hale gelecektir. Ayrıca fiyat ölçümlerini destekleyecek ilave metaveriler toplanabilecektir.

Web kazıma ile elde edilen verilerin fiyat çalışmalarında kullanılmasına yönelik ilk büyük çalışma MIT (Massachusetts Institute of Technology) tarafından yürütülen "Billion Price Project" adlı projedir. Cavallo ve Rigobon (2016) yaptıkları çalışmada çeşitli ülkeler için internet üzerinden fiyat derlemesi gerçekleştirmiş, günlük ve aylık frekanslarla endeks değerleri hesaplamış ve kapsadıkları ürünler özelinde resmi istatistiklerle karşılaştırmalar yapmışlardır. Örneğe seçilen iş yerleri hem çevrimiçi hem de çevrimdışı satış yapan iş yerlerinin

arasından seçilmiştir. Fiyatı derlenen ürünler ise ağırlıkların mevcut olduğu, resmi istatistikler kapsamında fiyat derlemesi yapılan maddeler arasında yer almaktadır. Üçüncü parti uygulamalar yerine iş yerinin kendi web sitesinden veri derlemesi yapılmaktadır. Yaklaşık 25 ülke için TÜFE ağırlığının %70'inden fazlası kapsamaktadır. Çalışma özellikle Arjantin'in enflasyonu üzerine odaklanmış, açıklanan resmi rakamlar ile çalışma kapsamında hesaplanan rakamlar arasında üç kata varan farklar ortaya çıkmıştır.

Billion Price Project, internet üzerinden derlenen fiyatların TÜFE çalışmalarında kullanılabilmesine yönelik önemli bir emsal teşkil etmekle birlikte ulusal istatistik ofisleri de internetten fiyat kazıma yöntemlerini resmi istatistik alanında kullanmaktadır. Bu bağlamda 34 farklı ülkenin veri derleme yöntemleri incelenmiş, bunların 19'unun web kazıma yöntemlerini çeşitli düzeylerde kullandığı görülmüştür. Ülkemizde ise Gençler ve Ulusoy (2021) çalışmasında TÜİK tarafından yayınlanan madde sepeti ve ağırlıkları kullanarak günlük endeks hesaplamaları gerçekleştirmişlerdir. Çalışma sonucunda elde edilen bulgular Enflasyon Araştırma Grubu (ENAG) adı ile yayınlanmaktadır. Bu çalışmanın ENAG enflasyon hesaplamasından farklılıkları şu şekildedir;

- Bu çalışmada sabit sepet yaklaşımına sadık kalınmış, fiyatı derlenen ürünlerin kalitesinin sabit tutulması sağlanmıştır. ENAG çalışmasında ise hesaplamada kullanılan ürünlerin devamlı olarak değiştiği ve herhangi bir kalite düzeltmesi uygulanmadığı anlaşılmaktadır.
- Bu çalışmada literatürde önerilen endeks formülleri olan Jevons endeks formülü, üst düzey endekslerde ise Laspeyres endeks formülü kullanılmaktadır. ENAG çalışmasında ise özellikle günlük zincirlemeden kaynaklı bir zincir sapması olup olmadığına yönelik bulguların paylaşılmasının önemli olduğu düşünülmektedir. EUROSTAT (2022: 9) çalışmasına göre zincirleme sıklığının artırılması endeksin sistematik olarak yanlışlaşmasına, özellikle uzun dönemde makul olmayan sonuçların ortaya çıkmasına neden olmaktadır.
- Bu çalışmada Laspeyres endeks formülünün gereği olarak baz dönemi, ağırlıkların da referans dönemi olan aralık ayı olarak belirlenmiştir. ENAG endeks hesaplamasında ise baz dönemi 26 ağustos olarak belirlenmiş, bunun gerekçesi açıklanmamıştır. Ayrıca ağırlıklarda fiyat güncelleme (price update) işleminin gerçekleştirilip gerçekleştirilmediği de belirtilmemiştir.
- Bu çalışmada fiyatı sürekli olarak derlenebilen ürünlerin (Eşleşen modeller yöntemi) fiyatları kullanılmaktadır. Bu nedenle baz döneminde fiyatının derlenmesi mümkün olmayan ürünler endeks hesaplamasının dışında bırakılmaktadır. ENAG endeksinde özellikle güçlü mevsimsel özellik gösteren ürünlere ilişkin uygulamadan bahsedilmemiştir.
- Son farklılık ise kullanılan ağırlıklar ve sınıflama ile ilgilidir. Bu çalışmada COICOP sınıflaması kamuoyu ile paylaşılan en alt düzeyde kullanılmış olup derlenen tüm fiyatlara metin madenciliği yöntemleri de kullanılarak 7'li düzeyde kod ataması gerçekleştirilmiştir. Ayrıca TÜFE sepetinde yer alan maddelerin tamamının fiyatı derlenemediği için ağırlıklarda yanlışlık oluşmaktadır. Bu nedenle ağırlıklar ana gruplardan başlanarak kapsanan maddelere yeniden dağıtılmıştır. ENAG çalışmasında COICOP sınıflamasının hangi düzeyde kullanıldığı anlaşılamamış olup yalnızca ana gruplar düzeyinde kullanıldığı değerlendirilmiştir. Ağırlıkların doğru şekilde kullanılabilmesi için fiyatların ağırlıklarla en uyumlu şekilde sınıflandırılması gerekmektedir.

İnternet fiyatları kullanılarak günlük TÜFE hesaplanmasına yönelik bir diğer çalışma İsviçre için gerçekleştirilmiştir. İnternet üzerinden derlenen fiyatlar ve banka kartı harcama verileri kullanılarak oluşturulan ağırlıklar vasıtası ile günlük fiyat endeksleri elde edilmiştir. Bu endeksler resmi istatistikler ile karşılaştırılmıştır (Santiago ve Sarah, 2020).

Bununla birlikte ulusal istatistik ofisleri de web kazıma verilerini farklı kapsamlarda endeks hesaplamalarına dahil etmektedir. TÜİK, derlemekte olduğu fiyatların yaklaşık %5'ini web kazıma yöntemi ile derlerken beyaz eşyalar, ev elektroniği, mobilya, otomobil ve otobüs biletleri bu kapsamda yer almaktadır (TÜİK, 2024). Benzer şekilde BLS derlediği fiyatların yaklaşık %8'ini web üzerinden elde etmektedir (U.S Bureau of Labor Statistics, 2020). Avrupa ülkelerindeki durum ele alındığında ise birçok ülkede yine web kazıma yöntemlerinin farklı kapsamlarda kullanıldığı göze çarpmaktadır.

Bu farklılıklar bağlamında çalışmanın literatüre katkılarını aşağıdaki gibi sıralamak mümkün olacaktır:

- i. Çalışma uluslararası TÜFE metodolojisi ile tamamen uyumlu bir hesaplama ortaya koymaktadır.
- ii. TÜFE çalışmalarında web kazıma yöntemi ile derlenmesi uygun olmayan maddeler ortaya konulmuştur.
- iii. TÜFE çalışmaları bakımından web kazıma yöntemi kullanılmasının ortaya çıkaracağı olası zorluklara yer verilmiştir.
- iv. Eşleşen modeller yöntemiyle fiyat derlenmesi sağlanarak kalite faktörünün zaman içerisinde sabit tutulması garanti edilmiştir.
- v. COICOP sınıflaması, alt düzeyde kod atanarak ve metin madenciliğiyle yapılmış; ağırlıklar daha doğru sınıflandırılmıştır.

Çalışmanın ikinci bölümünde hesaplamada kullanılan endeks formüllerinden, üçüncü bölümde veri toplama süreci ve endeks hesaplamasından, dördüncü bölümde çalışmada elde edilen bulgulardan bahsedilecek, tartışma ve sonuç bölümü ile çalışma sonlandırılacaktır.

2. Yöntem

2.1. Tüketici Fiyat Endeksi

TÜFE hesaplamaları iki aşamalı olarak gerçekleştirilmektedir. İlk aşamada ağırlık verilebilen en alt düzeyde endeksler hesaplanmakta ikinci aşamada ise sınıflamanın hiyerarşik olarak üst basamaklarına çıkılmaktadır.

İlk aşamada ağırlık verilemeyen fiyatlar toplulaştırılarak endekslenmektedir. Literatürde bu işlemde kullanılacak üç temel endeks formülü yer almaktadır. Bunlar Dutot, Carli ve Jevons endeks formülleridir.

Bir endeks formülünün geçerli kabul edilebilmesi için arzu edilen bazı istatistiksel özellikleri ihtiva etmesi gerekmektedir. Bu bağlamda endeks rakamlarının bazı aksiyomları sağlayıp sağlamadığının test edilmesi gerekmektedir. Bu testler literatürde aksiyomatik testler olarak anılmakta olup endeks formüllerinin birbirlerine karşı üstünlüklerinin ve zayıflıklarının anlaşılmasında kullanılmaktadır (Armknrecht ,1996). Bu testlerden en yaygın olarak kullanılanları zaman çevrimi testi, eş ölçülebilirlik testi ve geçişlilik testidir.

Zaman çevrimi testine göre, iki farklı periyotta gözlenen fiyatlar yer değiştirdiğinde endeks de aynı şekilde yer değiştirmelidir. Geçişlilik testi, iki periyot arasındaki zincirlenmiş endeks ile doğrudan endeksin eşit olup olmadığını sınımlamaktadır. Eş ölçülebilirlik testi ise fiyatı ölçülen ürünlerin ölçüm birimleri ile ilgilidir. Bu test endeksin ölçü biriminden etkilenip etkilenmediğini ölçer. Jevons endeksi bu testlerin tamamında başarılı sonuçlar ortaya koyarken Dutot endeksi eş ölçülebilirlik testinde Carli endeksi ise zaman çevrimi ve geçişlilik testlerinde başarısız sonuçlar üretmektedir (Fenwick, 2009: 154,155). Aksiyomatik testlerde en başarılı sonuçları Jevons endeksi sağladığından literatürde alt düzey endekslerde Jevons önerilmektedir. Jevons endeksi iki farklı şekilde hesaplanabilir. Temel toplamlar düzeyinde karşılaştırılan iki dönemde de eşleşen fiyatların geometrik ortalamalarının oranı şeklinde hesaplanabileceği gibi göreceli fiyatların geometrik ortalaması biçiminde de hesaplanabilir (Graf, 2020: 124).

$$I_j^{t/0} = \prod \left(\frac{P_i^t}{P_i^0} \right)^{1/n} = \frac{\prod (P_i^t)^{1/n}}{\prod (P_i^0)^{1/n}} \quad (2.1)$$

Yukarıdaki denklemde $I_j^{t/0}$ t dönemi endeks rakamını, P_i^t i. ürünün t dönemindeki fiyatını, P_i^0 i. ürünün baz dönemi fiyatını, n ise ürün sayısını ifade etmektedir. Buna göre hesaplanan endekslerden sınıflamanın üst basamaklarına çıkılırken kullanılacak Laspeyres, Paashe, Fisher, Törnqvist gibi çok sayıda endeks formülü bulunmaktadır. Ancak elde edilebilen en güncel harcama verisi bir önceki yıla ait olduğu için yalnızca Laspeyres formülü kullanılabilir.

$$I_{LA}^{t/0} = \frac{\sum P_i^t Q_i^0}{\sum P_i^0 Q_i^0} \quad (2.2)$$

Yukarıdaki denklemde P_i^0 i. ürünün baz dönem fiyatını, P_i^t i. ürünün t dönemindeki fiyatını, Q_i^0 i. ürünün baz dönem miktarını, Q_i^t i. ürünün t dönemindeki miktarını, $I_{LA}^{t/0}$ ise Laspeyres fiyat endeksini ifade etmektedir (EUROSTAT, 2018: 170). Buna göre ilk aşamada hesaplanan endeks değerlerinin ağırlıklı ortalaması alınarak sınıflamanın üst basamaklarına çıkılmaktadır. Bu işlem yapılırken baz dönemi için hesaplanan ağırlık seti kullanılmaktadır.

2.2. Web kazıma

Web kazıma; web sitelerinde yer alan bilgilerin hızlı, verimli ve otomatik bir biçimde çıkarılarak bu bilgilerin kullanılabilir bir formatta sunulmasıdır (Castrillo-Fernández, 2015). Daha açık bir ifade ile web sitelerinde yer alan bilgiler alınarak tanınabilir hale getirilir ve bir veri tabanına kaydedilir. Web kazıma işlemleri Python ve R başta olmak üzere Java, Visual Basic gibi birçok programlama dili kullanılarak yapılabilmektedir. Bu paketlerin yanı sıra programlama bilgisi gerektirmeyen üzerine tıklama (point and click) şeklinde çalışan çeşitli programlar da yer almaktadır.

Web kazıma amacı ile en yaygın olarak kullanılan ve tavsiye edilen programlama dilleri Python ve R dilleridir. Hem Python hem de R web kazıma amacıyla geliştirilmiş kütüphaneler içermektedir. Bu kütüphaneler aracılığı ile HTML sayfasının indirilerek hedef bilgilerin seçilmesi sağlanmaktadır (EUROSTAT, 2020). Web kazıma işlemlerinde Python programlama dili sıklıkla kullanılmakta olup temelde üç paket üzerinden gerçekleştirilmektedir. Bunlar BeautifulSoup, Selenium ve Scrapy'dir. Bu paketlerden BeautifulSoup, HTML ve XML dosyalarından veri çekmek üzere oluşturulmuş bir Python kütüphanesidir (Richardson, 2019). HTML yapılarını anlaşılır hale getirmenin yanı sıra XML yapılarını temsil eden, kolay geçişli Python nesnelere sunarak dağınık web'i biçimlendirmeye ve düzenlemeye yardımcı olur (Mitchell, 2015: 6). Kullanımı kolay olmakla birlikte sunucuya talep yollama işlemi gerçekleştirilememektedir. Ayrıca yalnızca statik web sitelerinde kullanılabilir. Elementleri kullanıcı hareketlerine göre güncellenebilen, dinamik yapıları web sitelerinde kullanılamamaktadır.

Selenium doğrudan web kazıma amacıyla geliştirilmiş bir araç değildir. Asıl amacı web tarayıcılarının otomasyonudur (Selenium, 2022). Selenium, amacı her ne kadar web kazıma olmasa da bu işlemler için de kullanılabilir. Bir web tarayıcısı aracılığı ile insan benzeri bir davranış sergilediğinden tüm web sitelerine erişilebilmekte ve buralardaki bilgiler alınabilmektedir. Selenium doğrudan bir web tarayıcısı üzerinden web sayfalarını taramakta ve buradaki verileri kaydetmektedir. Bu çalışma biçimi nedeniyle hız açısından dezavantajlıdır. Selenium sunucu bazlı olarak çalışan, çoğunlukla JavaScript kodlarının tetiklenmesine olanak sağlayan dinamik web sitelerinden veri alınması için uygun bir çözümdür. Bir sayfa üzerinde tıklama, sayfayı hareket ettirme, form doldurma gibi birçok işlem de gerçekleştirilebildiği için web kazıma kullanıma mükemmel biçimde uyumlu olarak görülmektedir (EUROSTAT, 2020).

Scrapy, yüksek seviyeli bir web tarama ve web kazıma çerçevesidir. Web sayfalarının taranması ve bu sayfalarda yer alan yapısal verilerin çıkarılmasında kullanılır. Bunun yanı sıra otomatik testlerden veri madenciliğine kadar çok geniş bir alanda farklı görevlerde kullanılabilir (Scrapy developers, 2022). Scrapy, diğer araçlara kıyasla çok daha hızlı olmasına karşılık Javascript kullanan web sitelerinde kullanılamamaktadır. İlave bir araç kullanılarak bu sorunun üstesinden gelinebilmektedir.

Yeni bir veri derleme yöntemi olarak web kazıma uygulamasının kalite ve verimlilik bakımlarından TÜFE çalışmalarında kullanılmaya son derece uygundur. Web kazıma yöntemlerinin TÜFE çalışmalarına entegre edilmesi ile birlikte yüksek frekanslı veriler daha düşük maliyetle derlenebilecektir (Griffioen ve Ten Bosch, 2016).

Bu çalışmada; kullanımının kolay olması, bir web tarayıcı aracılığı ile insan benzeri bir davranış göstermesi nedeniyle her türden web sitesinin kazınmasına olanak sağlaması ve yasaklanma olasılığının daha düşük olduğunun değerlendirilmesi nedenleri ile Selenium ve BeautifulSoup paketleri kullanılmıştır.

3. Veri Toplama Süreci ve Endeks Hesabı

TÜFE hesaplamasının gerçekleştirilebilmesi için hanchalkı harcamalarını temsil edici bir mal ve hizmet sepetine, tüketime konu mal ve hizmetlerin tasnif edilmesinde kullanılacak bir sınıflamaya, sepette yer alan mal ve hizmetlerin ağırlıkları ile baz ve cari dönem fiyatlarına ihtiyaç duyulmaktadır.

Tablo 3.1. Ana gruplara göre fiyat derlenen işyeri sayıları.

Ana Grup	İşyeri Sayısı*
Genel	53
Gıda ve alkolsüz içecekler	4
Giyim ve ayakkabı	11
Konut	8
Ev eşyası	11
Ulaştırma	20
Haberleşme	6
Eğlence ve kültür	6
Lokanta ve oteller	4
Çeşitli mal ve hizmetler	7

* Web sitelerinde birden fazla ana grup verisi derlenebildiğinden ana gruplar toplamı geneli vermemektedir.

Çalışmada kullanılan işyerlerinin COICOP ana gruplarına göre sayıları Tablo 3.1’de yer almaktadır. Madde sepeti ve ağırlıklar TÜİK web sitesinden alınmış, sınıflama olarak COICOP sınıflaması kullanılmıştır. Fiyatların derleneceği internet siteleri ulusal düzeyde faaliyet gösteren, satış düzeyi itibariyle temsil edici olduğu düşünülen web siteleri arasından karar örnekleme yönteminde seçilmiştir.

Tablo 3.2. Kapsanan maddelerin ana grup içerisindeki ağırlıkları.

Ana Grup	Kapsama Oranı
Gıda ve alkolsüz içecekler	89.02
Alkollü içecekler ve tütün	93.74
Giyim ve ayakkabı	82.4
Konut	77.97
Ev eşyası	72.43
Sağlık	0
Ulaştırma	78.69
Haberleşme	80.16
Eğlence ve kültür	52.94
Eğitim	0
Lokanta ve oteller	69.48
Çeşitli mal ve hizmetler	63.51
Toplam	75.25

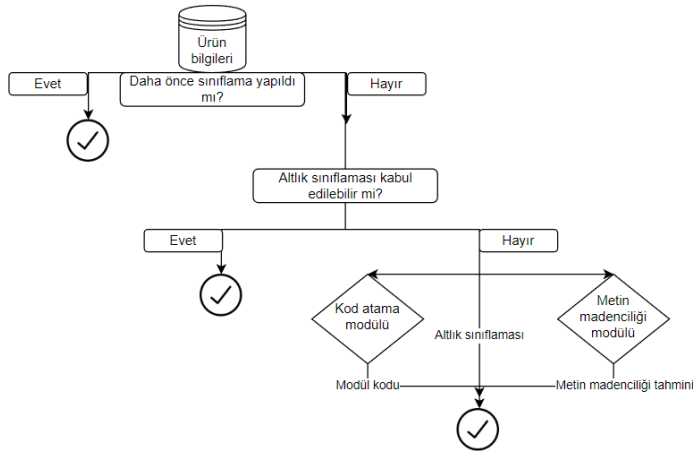
2022 TÜFE madde sepetinde yer alan 409 maddenin 233 tanesinin fiyatı hesaplamaya dahil edilmiştir. Kapsanan maddelerin ilgili ana grup içerisindeki ağırlıkları Tablo 3.2’de yer almaktadır. Burada yer alan değerler ana grup toplamının yüze eşit olduğu duruma göre hesaplanmıştır. Ana grup içerisindeki ağırlığı yüksek maddelerin fiyatlarının web üzerinden derlenmesi mümkün olmayan sağlık ve eğitim ana grupları çalışma kapsamı dışında bırakılmıştır. Bu nedenle ağırlıkları da sıfır olarak ele alınmıştır.

Tablo 3.3. Ana grup bazında ağırlıklar.

Ana Grup	Resmi Ağırlık	Yeni Ağırlık
Gıda ve alkolsüz içecekler	25.32	26.73
Alkollü içecekler ve tütün	4.31	4.55
Giyim ve ayakkabı	6.42	6.78
Konut	14.12	14.91
Ev eşyası	8.85	9.34
Sağlık	3.24	0
Ulaştırma	16.8	17.73
Haberleşme	3.78	3.99
Eğlence ve kültür	3.06	3.23
Eğitim	2.03	0
Lokanta ve oteller	7.11	7.51
Çeşitli mal ve hizmetler	4.96	5.24

Çalışmada kullanılan ağırlık seti üzerinde ise birtakım işlemler gerçekleştirilmiştir. TÜİK’in TÜFE madde sepetinde yer alan tüm ürünlerin fiyatlarının derlenememesi nedeniyle ağırlık yapısı bozulmakta ve yeniden dağıtılma ihtiyacı doğmaktadır. Ağırlıklar ana gruplardan başlanarak her bir düzeyde yeniden dağıtılmış, bu sayede ağırlık yapısı korunmuştur. Ayrıca fiyat derlenemeyen ana grupların ağırlıkları diğer gruplara kendi ağırlıkları oranında dağıtılmıştır. Ana gruplar düzeyindeki TÜİK resmi ağırlıkları ve hesaplanan yeni ağırlıklar Tablo 3.3’te yer almaktadır.

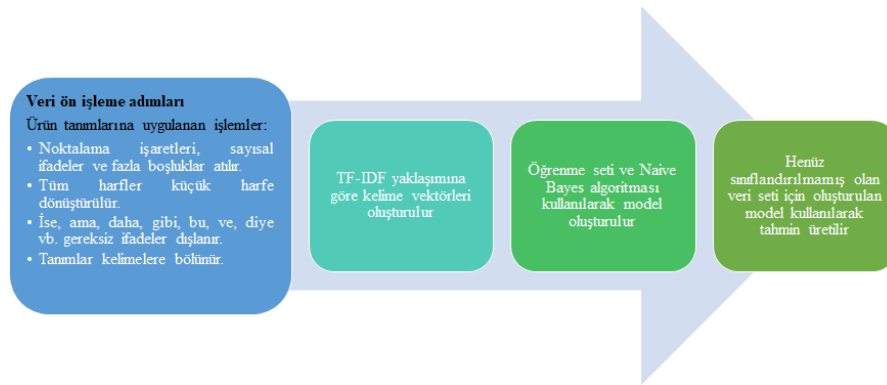
Fiyat derleme uygulamasına başlamadan önce web sayfalarının linkleri ve bu linklerde yer alan ürünlere atanacak COICOP kodları bir tabloda kayıtlanmıştır. Oluşturulan bu tablo çalışmaya altlık teşkil etmektedir ve izleyen bölümde fiyatı derlenen ürünlerin sınıflandırılmasında birincil olarak bu tabloda yer alan kodlamalar kullanılmıştır. Bu yöntem otomobiller, akaryakıtlar, cep telefonu görüşme ücretleri, giyim ürünleri ve ayakkabılar gibi birçok madde için uygun sonuçlar sağlamakla birlikte her ürün için aynı durum geçerli değildir. Özellikle market ürünlerinde web sayfalarının COICOP sınıflamasına uygun biçimde bölünmesi mümkün olmamaktadır. Örneğin nohut, mercimek, pirinç, bulgur gibi bakliyat ürünlerine bazı web sitelerinde tek bir link üzerinden bağlantı yapılmaktadır. Bu ve benzeri durumlar için ayrıca bir kod atama çalışması yapılması gerekmektedir. Çalışmada fiyatı kazınan her ürüne COICOP 7’li düzeyde kod ataması yapılmıştır. Bu düzey kamuoyu ile paylaşılan en detaylı sınıflama düzeyidir.



Şekil 3.1 Ürünlerin sınıflanma süreci.

Ürünlerin kod atama süreci Şekil 3.1’de özetlenmiştir. Kod atama uygulaması geliştirilmeden önce manuel olarak yaklaşık yetmiş bin ürüne kod atanması işlemi gerçekleştirilmiş, “if else” bloklarından oluşan kod atama modülü oluşturulmuştur. Bu modülde yalnızca market ve teknoloji market ürünleri kapsama alınmıştır. Ayrıca altlık tablosunda yer alan linklere karşılık tanımlanmış olan kodlardan güvenilir olanlar tespit edilerek ilgili kodların doğrudan atanması sağlanmıştır. Ürünler sınıflama işlemine sokulmadan önce hali hazırda kodlanmış ürünler arasında yer alıp almadığına bakılmaktadır. Daha önce sınıflama işlemi yapılmışsa aynı sınıf etiketi atanarak sınıflandırma işlemi tamamlanmaktadır. Daha önce sınıflama işlemi yapılmamış ise altlık tablosunda ürün linkine karşılık gelen sınıf etiketinin doğru kabul edilen sınıflar arasında yer alıp alınmadığı incelenmektedir. Ürünün kabul sınıflar arasında yer alması halinde bu sınıf etiketi doğrudan atanmakta aksi halde yazar tarafından oluşturulan kod atama modülü ve metin madenciliği uygulamasının yer aldığı metin madenciliği modülüne sokularak birer sınıf etiketi atanmaktadır. Son olarak altlık tablosunda yer alan sınıf etiketi de dikkate alınarak çoğunluk oyu yöntemi ile sınıflama işlemi tamamlanmaktadır. Çoğunluk oyu yöntemi kod ataması için yeterli değil ise yani atanan üç sınıf etiketi de birbirinden farklı ise kod atama işlemi manuel olarak gerçekleştirilmektedir.

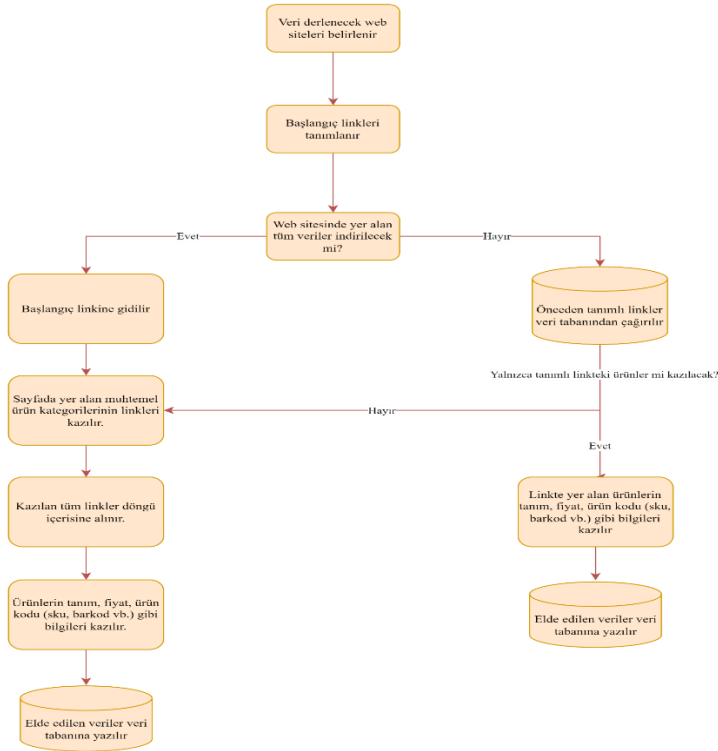
Metin madenciliği uygulamasına ilişkin yürütülen süreç ise Şekil 3.2’de yer almaktadır.



Şekil 3.2 Metin madenciliği uygulaması.

Buna göre önce metinlere ön işleme yapılarak metinler standart bir formata dönüştürülmüş devamında kelime vektörleri oluşturulmuştur. Tahminde Naive Bayes algoritması kullanılmıştır.

Çalışmada kullanılan fiyatlar ise web kazıma yöntemleri kullanılarak internet üzerinden derlenmiştir. Uygulamada Python dili kullanılmış Selenium ve BeautifulSoup paketlerinden yararlanılmıştır.



Şekil 3.3 Metin madenciliği uygulaması.

Çalışmada öncelikle karar örneklemesine göre fiyatı derlenecek işyerleri seçilmiştir. Fiyatların derlenmesi noktasında iki tip uygulama yapılabilmektedir. İlk olarak web sitesinde yer alan tüm ürünlerin fiyatları derlenebilir (Toplu kazıma). Bu yöntemde TÜFE sepetinde yer almayan ürünlerin fiyatları da derleneceğinden bunların ayıklanması ve COICOP kodu atanması zorlaşacaktır. Diğer yöntem ise hedeflenmiş kazımadır. Bu yöntemde ise önceden tanımlanmış başlangıç linkleri ve altında yer alan linklerdeki veriler kazılmaktadır. İkinci yöntemin dezavantajı ise takip edilecek link sayısı çoğalacağından bunların sağlıklı olup olmadığının takip edilmesinden kaynaklanan zaman maliyetidir. Bu çalışmada hibrit bir yaklaşım kullanılmış olup hem toplu kazıma hem de hedeflenmiş kazıma yöntemleri birlikte kullanılmıştır. Çalışmada kullanılan fiyat derleme süreci Şekil 3.3'te yer almaktadır. Buna göre hangi web sitelerinden hangi yöntemle veri derleneceği kararlaştırılır. Toplu kazıma yapılacak web sitesinin başlangıç linkine gidilir ve sitede yer alan sayfalar gezilerek fiyat kazılacak linkler taranır. Bu linkler döngü içerisine alınarak endeks hesaplamasında ihtiyaç duyulan ürün kodu, barkod, ürün tanımı, fiyat gibi veriler derlenir. Hedeflenmiş kazıma yapılacak ise zaten başlangıç linkleri önceden tanımlanmış olduğu için bu linklere gidilerek varsa alt linkler de toplu kazıma yöntemine benzer şekilde taranır yoksa doğrudan tanımlı linklerdeki veriler derlenerek süreç tamamlanır.

TÜFE çalışmalarında en çok satış yapan işyerlerinden en çok satılan ürünlerin fiyatlarının derlenmesi ve mümkün olduğu ölçüde aynı ürünün fiyatının izlenmesi esastır. Fiyat değişimlerinin kalite faktöründen bağımsız olması gerekmektedir. Bu nedenle aynı ürünün fiyatının herhangi bir nedenle izlenememesi durumunda benzer kalitede yeni ürünlerin ikame alınması, fiyat değişiminin ikame alınan ürünün kalitesinden kaynaklanması durumunda ise uygun kalite düzeltme yöntemleri kullanılarak fiyatların kalite etkisinden arındırılması gerekmektedir.

Çalışmada; ulusal düzeyde faaliyet gösteren ve online satış gerçekleştiren işyerlerinden yirmi milyonun üzerinde fiyat derlenmiş, endeks hesaplaması fiyatı sürekli olarak derlenebilen ürünler üzerinden madde düzeyinde Jevons endeks formülü, sınıflamanın hiyerarşik olarak daha üst düzeylerinde ise Laspeyres endeks formülü kullanılarak gerçekleştirilmiştir. Ürün takibi mümkün olduğu ölçüde işyerlerince sağlanan barkod, sku (stock keeping unit) ve ürün kodları ile yapılmıştır. Barkod, sku veya ürün kodlarına ulaşamadığı hallerde ürün tanımlarından faydalanılmıştır. Endeks hesaplamasında fiyatları devamlı olarak takip edilebilen ürünler kullanıldığından herhangi bir kalite düzeltmesi ihtiyacı ortaya çıkmamıştır.

Genel uygulama fiyatları sürekli derlenen ürünlerin hesaplamada kullanılması olmakla birlikte bazı istisnai uygulamalar da kullanılmıştır. Bu uygulamalar şunlardır:

- Gerçek kiraların derlenmesinde sürekli aynı konutun fiyatının izlenmesi mümkün olmamakla birlikte değişen konutların aynı kalitede olup olmadığı da bilinmemektedir zira kiraları etkileyen çok sayıda etken bulunmaktadır. Bunun yanı sıra gerçek kiraların takip edilmesinde stoktaki kiracıların başka bir deyişle

mevcut kontratı devam eden kiraların fiyat gelişmelerinin de izlenmesi gerekmektedir. İnternet üzerinden bu fiyatlar derlenememektedir. Bu nedenlerle gerçek kiraların metodolojiye uygun bir biçimde tamamen internet üzerinden derlenmesi uygun değildir. Ancak gerçek kiraların ağırlığının çok yüksek olması nedeniyle çalışmada konutların yaş aralıkları, oda sayıları, lokasyonları ve metre kareleri sınırlanarak bu değişkenlerin belirlenen aralıklarda tutulması sağlanmış ve endeks hesaplamasında kullanılmıştır.

- Paket tur fiyatlarının derlenmesinde ulaşım araçları ve konaklama koşullarının sabit tutulması mümkün olmamıştır.
- Otel fiyatlarının derlenmesinde tek kişi için gecelik fiyatlar derlenmiştir. Ancak standart oda, aile odası, süit oda vb. ayırımına gidilmemiştir.
- TÜFE metodolojisi gereği hanehalkı tüketimine konu mal ve hizmet fiyatları endeks hesaplamasında kullanılmalıdır. Bu bağlamda altın fiyatlarının derlenmesinde gram altın fiyatları yatırım harcaması olarak değerlendirilmiş altın bileklik fiyatları derlenerek kullanılmıştır.
- İnternet üzerinden satışı yasaklı olan bütün ürünlerinin fiyatları manuel olarak girilmiştir.

Çalışmada fiyat derleme periyodu 1 Aralık 2021- 25 Nisan 2022 tarihleridir. Fiyat derleme sürecinin başlangıç tarihinin 1 Aralık olarak belirlenme nedeni resmi istatistik kapsamında kullanılan ağırlık setinin referans döneminin 2021 Aralık ayı olmasıdır. Yeni bir ağırlık setinin hesaplanmasına olanak sağlayacak daha zamanlı bir veri bulunmamaktadır. Gerek günlük endekslerin gerekse aylık endekslerin hesaplanmasında aynı ağırlık seti ve aynı fiyatlar kullanılmaktadır. Endeks hesaplamasının günlük olması bakımından veri derlenen zaman periyodunun yeterli olduğu değerlendirilmektedir.

4. Bulgular

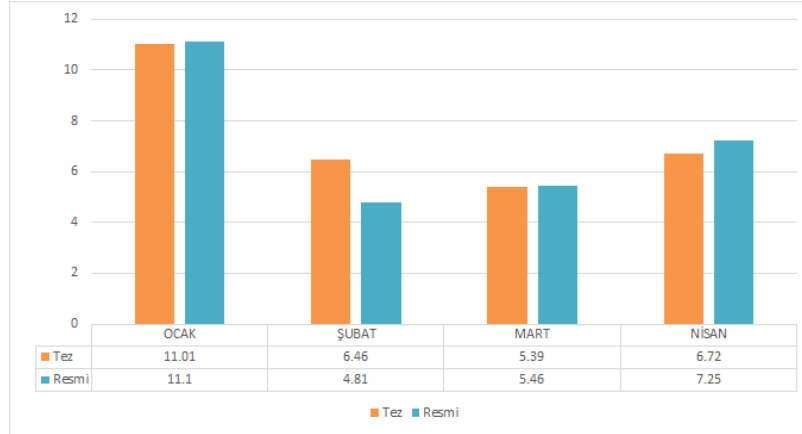
Bu bölümde internet üzerinden derlenen fiyatlar kullanılarak hesaplanan aylık endeks değişim oranları ile resmi istatistik kapsamında hesaplanan TÜFE verilerine karşılaştırmalı olarak yer verilmiştir. Her ne kadar bu bölümde iki verinin karşılaştırması yapılsa da aşağıdaki nedenlerle veriler tam anlamıyla karşılaştırılabilir değildir (TÜİK, 2022):

- Kapsam farklılıkları,
- İşyeri ve ürün örneklemelerinin farklı olması,
- Veri derleme sıklıklarının farklı olması,
- Yerel etkilerin göz ardı edilmesi.

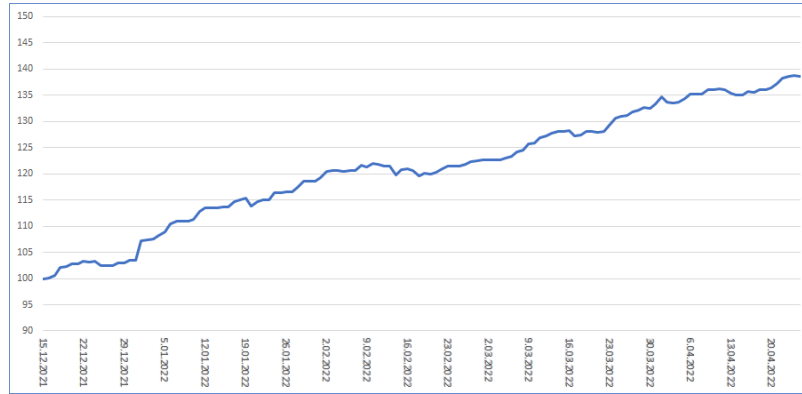
Çalışmada TÜFE genel tahminlerine yönelik hesaplamada kullanılan fiyatlar toplam ağırlığın %75,25'inden oluşmaktadır.

Tablo 4.1'de TÜFE genel endeksine ilişkin aylık değişim oranları yer almaktadır. Buna göre resmi istatistiklerde bir önceki aya göre değişim oranları ocak ayında %11,1, şubat ayında %4,81, mart ayında %5,46 ve nisan ayında %7,25'tir. Bu çalışma bağlamında internet üzerinden derlenen fiyatlarla oluşturulan endekse göre ise ocak ayında %11,01 olan endeks değişim oranı şubat ayında %6,46, mart ayında %5,39 ve nisan ayında %6,72 olmuştur. Buna göre iki hesaplamadan elde edilen sonuçlar ciddi bir farklılık göstermemektedir. Dört aylık kümülatif değişim oranları ise resmi istatistiklere göre %31,71 iken çalışma bulgularında bu oran %32,91 olarak tahmin edilmektedir. Buradan hareketle genel endeks noktasında çok benzer sonuçlar elde edildiği görülmektedir. TÜİK tarafından gerçekleştirilen TÜFE çalışması ile örneklem ve kapsam farklılıkları olmakla birlikte kullanılan endeks formülleri aynıdır. Bunun yanı sıra gerek ağırlıkların gerekse fiyatların referans dönemleri de TÜFE çalışmasına göre belirlenmiştir. İki çalışma her ne kadar tamamen karşılaştırılabilir olmasa da elde edilen sonuçların benzer olması tahmin güvenilirliği açısından olumlu olarak değerlendirilmektedir.

Tablo 4.1. Genel endeks aylık değişim oranları, %.



Şekil 4.1'de ise genel endeksin günlük değerleri yer almaktadır. Buna göre endekste dönem dönem düşüşler olmakla birlikte genel trend yukarı yönlüdür. Ocak ayı başında özellikle yönetilen ve yönlendirilen fiyatlara gelen zamları etkisi belirgin biçimde görülmektedir.



Şekil 4.1. Genel endeks günlük değerleri.

Çalışma sonucunda elde edilen diğer önemli bulgular ise gelecek çalışmalara da altlık teşkil edecek olan karşılaşılan sorunlar ve dikkat edilmesi gereken hususlardır. Bunlar aşağıda özetlenmiştir:

- Her web sitesinin kendine özgü bir sayfa yapısı bulunması nedeniyle her biri özel bir çalışma gerektirmekte bu web sitelerindeki değişikliklerin izlenerek internet gezgininin güncellenmesi gerekmektedir.
- İnsan tanıma uygulamaları internet gezgininin çalışmasını aksatmaktadır.
- İnternet gezgini, web siteleri tarafından tespit edilmesi durumunda kalıcı veya geçici olarak yasaklanabilmektedir. Bu nedenle tasarım olabildiğince insan davranışını yansıtacak biçimde olmalıdır. Bununla birlikte "crawl delay" benzeri yapılarla kazıma işlemi yavaşlatılabilmektedir.
- Bölgesel fiyatlandırma davranışları endekste yansıtılamamaktadır.
- Hanehalkı tüketiminde önemli paya sahip bazı işyerleri (Pazar, manav veya çevrimiçi satış yapmayan marketler) kapsanamamaktadır.
- Güçlü mevsimsel özellik gösteren ürünler sabit ağırlık sisteminde kapsanamamaktadır. Bir yıllık döngünün tamamlanması durumunda imputasyon yöntemleri uygulanabilir.
- Miktar veya adet değişiminden kaynaklanan kalite değişimleri izlenememektedir (Örn. 10 adet olarak satılan bir paket bisküvinin 8 adete düşürülmesi). Bu sorunun çözümü için kayıtlara miktar ve ölçü birimi (Adet, paket, litre, gram, kg vb.) ataması yapılması gerekmektedir.
- Hizmetler ve çevrimiçi satışı yasak olan ürünlerin fiyatları derlenememektedir.
- Ayrı bir HTML etiketi ile ifade edilen indirimlerle ilgili dikkatli olunmalıdır. Birçok web sitesinde indirimli fiyatlar ayrı olarak gösterilmektedir. Genele yaygın indirimlerin endeks hesabında dikkate alınması gerekmektedir.
- Sepet indirimi olarak tanımlanan ve tüm tüketicilere eşit olarak uygulanan indirimlerin endeks hesabına yansıtılması gerekmektedir. Ancak indirimli fiyat sepete gidildiğinde görülebildiğinden bu fiyatlar derlenememiştir.

4. Sonuç ve Tartışma

TÜFE; hanehalklarının tüketimine konu olan mal ve hizmetlerin fiyatlarının genel düzeyinde zaman içerisinde meydana gelen değişimlerin bir ölçüsü olmakla birlikte maaş ve ücretlerin belirlenmesinden kira artışlarına, makroekonomik göstergelerin deflate edilmesinden parasal istikrarın ölçülmesine kadar çok geniş kullanım alanı bulunan bir göstergedir. Bu bağlamda TÜFE'nin uluslararası standartlara uygun, güvenilir ve zamanlı olması gerekmektedir. Günümüzde teknolojinin gelişmesiyle birlikte bu ihtiyaca cevap verecek yeni veri derleme yöntemleri ortaya çıkmış ve bu yöntemler daha yüksek frekanslarda veri derlenebilmesinin önünü açarak daha yüksek frekanslı TÜFE hesaplamalarının yapılabilmesini sağlamıştır. Web kazıma yöntemi bu yöntemlerin arasında yer almaktadır. Web kazıma yöntemi ile internet üzerinde yer alan bilgilerin otomatik bir şekilde alınması, yapısal veriye dönüştürülerek veri tabanlarına kaydedilmesi sağlanmaktadır. TÜFE çalışmaları bağlamında web kazıma yöntemi ile internet üzerinden fiyatlar, fiyatı derlenen ürünün tanımı ve özellikleri otomatik olarak derlenebilmektedir.

Bu çalışmada uluslararası TÜFE hesaplama yöntemlerine mümkün olan en detaylı biçimde uyulmuş bu yöntemlerden sapmalar objektif biçimde ortaya konulmuştur. Ayrıca web kazıma yöntemi ile fiyatının TÜFE metodolojisine uygun olarak derlenmesi mümkün olmayan ürünlere yer verilmektedir. Çalışma bu özellikleri bakımından gelecek çalışmalara güçlü bir temel oluşturma amacına hizmet etmektedir. Çalışma neticesinde ulaşılan bir diğer sonuç TÜFE bakımından web kazıma ile elde edilen verilerin geleneksel yöntemler kullanılarak derlenen verilere göre üstünlük ve zayıflıklarının ortaya konulmuş olmasıdır. Buna göre yüksek frekans ve düşük maliyetle veri derlenebilmesi, veri giriş hatalarının düşük olması, çok yüksek örnek hacimlerine ulaşılabilmesi web kazımının üstünlükleri iken ikame ve kalite düzeltilmesi süreçlerinin otomasyonundaki zorluklar, bölgesel hareketlerin kapsanamaması, satış verilerinin elde edilememesi nedeniyle ürünlerin hanehalkı tüketimindeki öneminin göz ardı edilmesi, verisi kazınan fiyatların sınıflandırılmasındaki zorluklar ve önceki bölümlerde belirtilen bazı ürünlerin fiyatlarının metodolojiye uygun olarak derlenememesi web kazımının zayıflıkları olarak karşımıza çıkmaktadır. Ayrıca web kazıma işlemi için kullanılan botun web siteleri tarafından tespit edilerek kısıtlanması, sunucu isteklerinin geciktirilmesi uygulamaları, insan tanıma uygulamaları, bazı verilerin PDF formatındaki dosyalar biçiminde sunulması, her bir web sayfasının yapısal özelliklerinin farklı olması nedeniyle özel olarak ele alınması ve bu web sitelerinde meydana gelen değişikliklere göre programın güncellenmesinin gerekmesi gibi başka zorluklar da bulunmaktadır.

Veri derleme süreçlerinde yaşanan zorluklar, gerçek kira başta olmak üzere TÜFE metodolojisine uygun şekilde fiyatı derlenemeyecek ürünlerin varlığı gibi sorunlar olmasına karşılık elde edilen bulguların resmi istatistiklerle tutarlı olması, web kazıma yönteminin TÜFE çalışmasında kullanılabilecek anlamlı bir alternatif veri derleme metodu olduğunu göstermiştir.

Gelecekte yapılacak çalışmalarda özellikle devir hızı yüksek ürünler başta olmak üzere ürün ikamesi ve kalite düzeltilmesi süreçlerinin otomasyonu, gerçek kira fiyatlarının metodolojiye uygun biçimde hesaplanabilmesi için stok kiracı oranına bağlı bir hesaplama yöntemi geliştirilmesi ve kapsama oranının artırılması hususlarının ele alınması gerekmektedir. Özellikle yüksek frekans ve hacimli verilerin TÜFE çalışmalarında en etkin şekilde kullanılmasını amaçlayan çok boyutlu endeks hesaplama yöntemlerinin kullanımına yönelik çalışmalar yapılması gerekmektedir.

Çalışmanın sonuçları resmi istatistiklerde kullanılan hesaplama yöntemi ile uyumlu bir biçimde kullanılacak web kazıma verilerinin bir hesaplama yanlılığına neden olmadığını göstermektedir. Bu bakımdan istatistik kurumlarının aylık veya yıllık zincirleme şeklinde eşleşen modeller yaklaşımını kullanarak bu verilerin endeks hesaplamalarında kullanılması veya kapsamının genişletilmesinin değerlendirilmesi tavsiye edilmektedir.

Citation / Atıf: YALÇIN Y., BAYSAL B. (2024). Web Kazıma Yöntemlerinin Tüketici Fiyat Endeksi Çalışmalarında Kullanılması Türkiye Örneği. *İstatistik Araştırma Dergisi*, 14 (2), 46-58.

Kaynaklar

- Armknacht, P. (1996). Improving the Efficiency of the U.S. CPI. *IMF Working Papers*, 96(103).
- Castriello-Fernández, O. (2015). *Web Scraping: Applications and Tools*. European Public Sector Information Platform. https://data.europa.eu/sites/default/files/report/2015_web_scraping_applications_and_tools.pdf, Erişim Tarihi:10.12.2021).
- Cavallo, A., and Rigobon, R. (2016). The Billion Prices Project: Using Online Prices for Measurement and Research. *Journal of Economic Perspectives*, 30(2), 151-178.
- EUROSTAT. (2018). *Harmonized Index of Consumer Prices Methodological Manual*. Luxembourg: Publications office of the European Union.
- EUROSTAT. (2020). Practical Guidelines on Web Scraping for the HICP. (<https://ec.europa.eu/eurostat/web/hicp/methodology>, Erişim Tarihi: 10.12.2021)
- EUROSTAT. (2022). *Guide on Multilateral Methods in the Harmonised Index on Consumer Prices*. Luxembourg: Publications Office of the European Union.
- Fenwick, D. (2009). *Practical Guide to Producing Consumer Price Indices*. New York: UN. (<http://digitallibrary.un.org/record/670125>, Erişim Tarihi: 01.12.2021).
- Gençer, S., ve Ulusoy, V. (2021). Dinamik Ve Yüksek Frekanslı Fiyat Endeksi: Türkiye'nin Günlük Ve Saatlik Bazda Enflasyon Hesaplaması. *Maliye ve Finans Yazıları*, (Özel Sayı 2), s. 173-190.
- Graf, B. (2020). *Consumer Price Index Manual*. USA: International Monetary Fund. (<https://www.elibrary.imf.org/view/book/9781484354841/9781484354841.xml>, Erişim Tarihi: 12.12.2021).
- Griffioen, A. R., & Ten Bosch, O. (2016). On the use of Internet data for the Dutch CPI. CPI. In UNECE Meeting of the Group of Experts on Consumer Price Indices, Geneva.
- Mitchell, R. (2015). *Web Scraping with Python: Collecting Data from Modern Web*. Sebastopol: O'Reilly.
- Richardson, L. (2019). *Beautiful Soup Documentation*. (<https://beautiful-soup-4.readthedocs.io/en/latest/pdf/>, Erişim Tarihi: 04.15.2022).
- Santiago, E. A., and Sarah, M. L. (2020). Tracking inflation on a daily basis. *Swiss Journal of Economics and Statistics*, 156, 1-13.
- Selenium. (2022). Selenium Documentation. (<https://www.selenium.dev/documentation>, Erişim Tarihi: 03.01.2022).
- Scrapy developers. (2022). Scrapy Documentation. (<https://docs.scrapy.org/en/latest/>, Erişim Tarihi: 09.04.2022).
- TÜİK. (2022). Tüketici Fiyat Endeksi Haber Bülteni. (<https://data.tuik.gov.tr/Bulten/Index?p=Tuketici-Fiyat-Endeksi-Ocak-2022-45790>, Erişim Tarihi: 05.04.2022).
- TÜİK. (2024). TÜFE Metodoloji Dokümanı. (<https://data.tuik.gov.tr/Bulten/Index?p=Tuketici-Fiyat-Endeksi-Ekim-2024-5361>, Erişim Tarihi: 12.11.2024).
- U.S Bureau of Labor Statistics. (2020). Handbook of methods. (<https://www.bls.gov/opub/hom/cpi/home.htm>, Erişim Tarihi: 10.10.2022).