



Tüketici Fiyat Endeksi (TÜFE) Hesaplamasında Yapay Zekâ Kullanan Çalışmalarının İncelenmesi

Investigation of Studies Using Artificial Intelligence in Calculating Consumer Price Index (CPI)

Abdulcebar On^{*1}, Necaattin Barışçı¹

¹Gazi Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği Bölümü, 06560, Ankara, TÜRKİYE

Başyuru/Received: 11/11/2021

Kabul / Accepted: 30/01/2022

Çevrimiçi Basım / Published Online: 31/01/2022

Son Versiyon/Final Version: 31/01/2022

Öz

Tüketici Fiyat Endeksi (TÜFE), hanehalklarının tüketimine yönelik mal ve hizmet fiyatlarının zaman içindeki değişimini ölçen bir yöntemdir. Bu çalışmada yapay zeka alanında TÜFE ile ilgili yapılan birbirinden bağımsız birçok çalışma araştırılmıştır. Yapılan literatür taraması neticesinde TÜFE hesaplanmasındaki yapay zeka kullanımı “verilerin toplanma ve işleme” ile “ürünlerin sınıflandırılması” olmak üzere iki süreçte olduğu bilgisine ulaşılmıştır. Verilerin toplanma ve işleme sürecinde resmi istatistik ofisleri tarafından anket aracılığıyla veri toplama işlemine alternatif yöntem çalışmaları gerçekleştirilmiştir. Bunun için çevrimiçi alışveriş sitelerindeki ürün bilgileri ve fiyatları web tarama tekniği gibi yöntemler kullanılarak derlenmiştir. Ürünlerin sınıflandırılması sürecinde ise çeşitli yöntemlerle firmalardan dijital ortamda alınan ürün bilgilerinin TÜFE kapsamındaki kullanılan sınıflandırma yöntemleri ile eşleşmesi için çeşitli makine öğrenme yöntemleri kullanılmıştır. Bunun için metin tabanlı verileri dijitalleştirme için doğal dil işleme yöntemleri uygulanmıştır. Sonuç olarak bu çalışma ile TÜFE kapsamında yapay zeka kullanan çalışmalar tek bir çatı altında toplanarak yeni alternatif hesaplama yöntemleri için öngörü oluşturulmuştur.

Anahtar Kelimeler

web tarama, ürün eşleştirme, yapay zeka, makine öğrenme, doğal dil işleme, tüfe, bilgi çıkarımı

Abstract

The Consumer Price Index (CPI) is a method that measures the change in the prices of goods and services for household consumption over time. In this study, many independent studies on CPI in the field of artificial intelligence were investigated. As a result of the literature review, it was learned that the use of artificial intelligence in the calculation of CPI is in two processes: "collection and processing of data" and "classification of products". During the data collection and processing process, alternative method studies were carried out by the official statistics offices to the data collection process by means of questionnaires. For this purpose, product information and prices on online shopping sites have been compiled using methods such as web crawling technique. In the process of classification of products, various machine learning methods were used to match the product information obtained from companies in digital environment with the classification methods used within the scope of CPI. For this, natural language processing methods have been applied to digitize text bases data. As a result, with this study, studies using artificial intelligence within the scope of CPI were gathered under a single roof and a prediction was created for new alternative calculation methods.

Key Words

web crawling, product matching, artificial intelligence, machine learning, natural language processing, cpi, information extraction

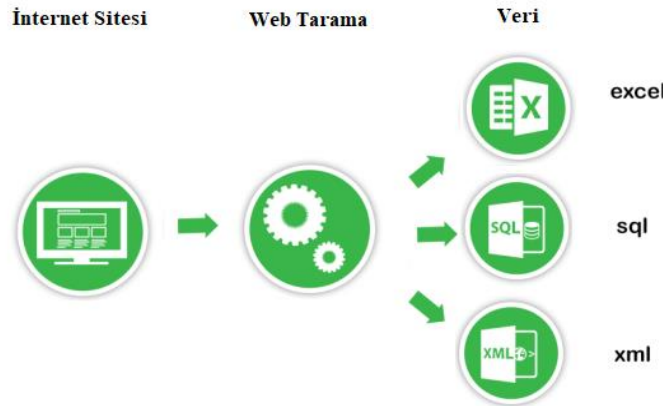
1. Giriş

Enflasyon fiyatlar genel düzeyinin etkili bir şekilde devamlı yükselmesi nedeniyle paranın sürekli olarak değer kaybetmesi, bunun sonucu olarak da tüketicilerin satın alma gücünü yitirmesidir. Enflasyon ölçümlerinden bir olan Tüketici Fiyat Endeksi (TÜFE), belirli bir referans döneminde bireylerin ortalama tüketimlerini yansıtan bir mal ve hizmet sepetinin zaman içinde fiyat değişimini gösteren bir endekstir. Bu endeks sepetinde yer alan mal ve hizmetlerin miktar ve kalite değişimleri göz önüne alınarak endeksin sadece fiyat hareketlerini yansıtmayı sağlamaktadır. Perakende fiyat değişimleri ve bu fiyat değişimlerinden bireylerin nasıl etkilendikleri ile ilgili enflasyon oranı, TÜFE değişim oranıdır.

TÜFE hesaplanması için öncelikle hangi ürünlerin fiyatlarının takip edileceği belirlenmektedir. Bunun için resmi istatistik ofisleri tarafından uygulanan hanehalkı bütçe anketlerinde, hanehalklarının referans ayında yaptıkları gıda, giyim, sağlık, ulaştırma, konut, ev eşyası ve benzeri konulardaki bütün tüketim harcamalarına ilişkin sorular sorulmaktadır. Hanehalkının yaptıkları tüketim harcamalarını günlük olarak kayıt ettiği “Hanehalkı Harcama Kayıt Defteri” ve hanede bulunan 15 yaşından büyük fertlerin bireysel harcamalarını kaydettikleri “Ferdî Harcama Kayıt Defteri” bilgileri ilgili ay içerisinde düzenli olarak anketörler tarafından toplanmaktadır. Toplanan bu bilgiler ile TÜFE kapsamına alınacak ürünler belirlenmektedir (TÜİK, 2003). 2019 yılı itibarıyla TÜFE madde sepeti 12 ana kategoriden oluşan amaca yönelik tüketim ürün gruplarına göre gerçekleştirilmektedir (United Nations et al. 2018). Bu ana kategoriler 418 alt gruba ayrılmaktadır (TÜFE, 2019).

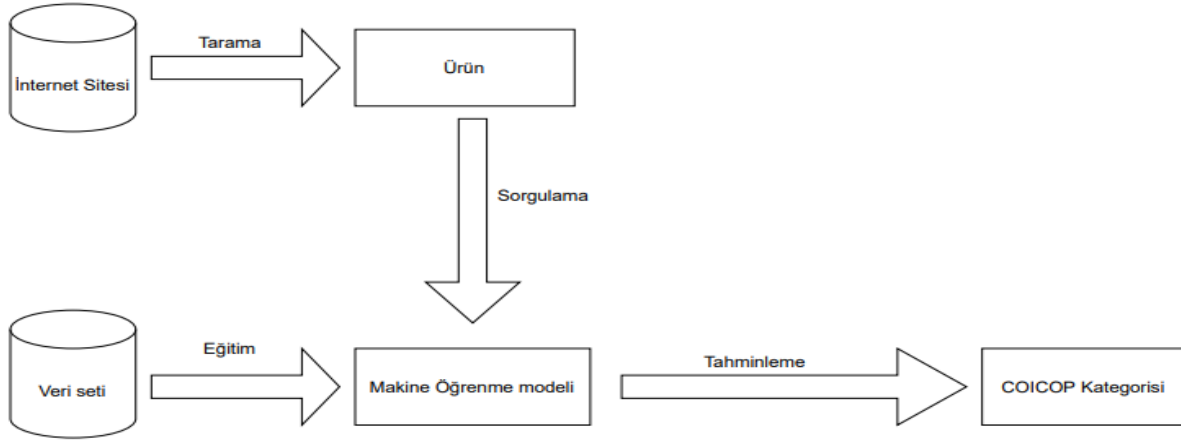
TÜFE kapsamına giren ürünler belirlendikten sonra ürünlerdeki fiyat dalgalanmasına göre fiyatlar aylık ya da haftalık şekilde anketörler aracılığıyla toplanmaktadır. İş yerlerine gidilerek tek tek fiyat alınarak yapılan bu yöntem yerine bazı ülkelerde istatistik ofisleri tarafından alternatif yollar denenmiştir. Örneğin günümüzde çevrimiçi sitelerdeki fiyatları kullanmaya yarayan web tarama yöntemi kullanılarak ürün fiyat bilgilerinin toplanabilmesi mümkün hale gelmiştir (Polzonetti&Re&Vaccari, 2013).

Web tarama, çevrimiçi sitelerde bulunan Hiper Metin İşaretleme Dili (Hyper Text Markup Language - HTML) kodunu ayrıştırma işlemidir. Bu kod web sayfasındaki bilgileri düzenli olarak gösterilmesini sağlar. Kodda yerleşim olarak kullanılan kısımlarda, her çevrimiçi site kendi standardına uygun olacak şekilde HTML içerisinde bir stil veya tanım kullanmaktadır. Ürün bilgileri bu kod kısımlarından elde edilmektedir. Geliştirilecek uygulamada tanımlanmış bir kural ile söz konusu kod kısımları kullanılarak web sayfasında bulunan tüm sayfaların taranmasıyla verilerin elde edilmesi sağlanmaktadır (Christopher&Najork, 2010; Cavallo, 2013). Elde edilen veriler yüksek boyutta, çeşitli ve karmaşıktır. Çünkü çevrimiçi sitelerde sunulan ürünler standart bir yapıda değildir. Bu verilerin TÜFE'ye uygun hale gelebilmesi için makine öğrenme, yapay zeka ve doğal dil işleme yöntemlerinin kullanılması gerekmektedir. Aksi durumda manuel işlemlerle çok yüksek insan işgücüne ve zamana ihtiyaç duyulur (Breton vd., 2016; Chuanyang vd., 2016).



Şekil 1. Web Tarama Yöntemi ile Verilerin Toplanması

TÜFE kapsamında derlenen fiyat verileri anketörler tarafından doğrudan sahadan toplanırken her ürüne ait bir kod bulunmakta ve ilgili ürünün kodu anketör tarafından sisteme girilmektedir. Ancak doğrudan veri toplama yerine çeşitli yapay zeka yöntemleri kullanarak web tarama ya da firmalardan çekilen verilerdeki ürün bilgileri ile TÜFE kapsamındaki ürün kodları örtüşmemektedir. Bunun için ürün bilgilerinin TÜFE kapsamındaki hangi koda ait olduğunu belirlemek için yapay zeka kullanılarak sınıflandırma işlemi gerçekleştirilmektedir.



Şekil 2. Ürün Bilgilerinin Makine Öğrenme Yöntemi ile Tahmin Edilmesi

Bu makalenin devamında Bölüm 2’de veri toplama ve işleme sürecinde kullanılan yapay zeka çalışmaları ele alınmıştır. Bölüm 3’de ise elde edilen ürün bilgilerinin yapay zeka kullanılarak sınıflandırılması ile ilgili yapılmış çalışmalar incelenmiştir.

2. Veri Toplama ve İşleme Sürecinde Yapay Zeka Kullanan Çalışmalar

Bu bölümde TÜFE hesaplamasının veri toplama ve işleme sürecinde yapay zeka kullanan çalışmalar ele alınmıştır. Firmalar tarafından tutulan fiyat verileri, çevrimiçi alışveriş siteleri gibi internet ortamında erişilebilen veriler son yıllarda arttığı için kaynakların büyük çoğunluğu yakın zamana aittir.

Cavallo tarafından yapılan bir çalışmada web tarama verileri enflasyon ölçümü için ele alınmıştır. Veriler Ekim 2007’den Mart 2011’e kadar MIT (Massachusetts Institute of Technology) ’deki The Billion Prices Project (BPP) ile toplanmıştır (Cavallo&Rigobon, 2016). Toplanan veriler 6 süpermarket’in çevrimiçi sitelerinde sunduğu ürünlerden oluşmaktadır. Ürünler için, fiyat verisi, ürün kimlikleri ve bir kategori göstergesi bilgilerini içeren veri seti oluşturulmuştur. Bu veri seti kullanılarak çevrimiçi TÜFE hesaplanmıştır. Bu çalışmada ele alınan metodoloji 5 ülke için uygulanmıştır. Ülkelerin çevrimiçi süpermarket pazar payı hesaplanmış ve bu bilgiler ile ülkelerin aylık ve yıllık enflasyonları çevrimiçi TÜFE ile karşılaştırılmıştır. Regresyon analizi yöntemleri kullanılarak elde edilen sonuçlara göre Arjantin hariç diğer ülkelerin web tarama TÜFE’si ile resmi TÜFE’si arasında çok yakın benzerlikler olduğu ortaya çıkarılmıştır (Polzonetti vd., 2013).

Hang ve diğerleri yaptıkları bir çalışmada kullanıcıların web ortamında aramalar için kullandığı anahtar kelimelerin enflasyon endeksiyle ilişkilendirilmesi üzerine bir model geliştirmişlerdir. Anahtar kelime seçimi TÜFE ilgili terimlerden seçilmiştir. Tekrarlanan kelimeler kaldırıldıktan sonra 140 dan fazla anahtar kelime elde edilmiştir. Her anahtar kelimenin Pearson korelasyon katsayısı hesaplanmış ve Ulusal İstatistik Bürosundan TÜFE hesaplanarak, aralarındaki korelasyon test edilmiştir. Ortalama kişi korelasyon katsayısının 0.6’nın üzerinde olduğu 21 anahtar kelime seçilmiştir. Bu anahtar kelimeler ekonomik etkilerine göre arz ve talep kategorilerine ayrılmıştır. İlişkili olmayan anahtar kelimeler silinip granger testi yapılarak 4 farklı kategori oluşturulmuştur. Kategoriler; önde gelen, tutarlı, zaman gecikmeli ve ilişkisiz anahtar kelimeler olarak oluşturulmuştur. Anahtar kelimeler için Dickey-Fuller test yöntemi uygulanmıştır. Ayrıca bütünlük testi uygulanmıştır. Bunun için 8 anahtar kelime silinip diğer 13 anahtar kelime ile birlikte Johansen-Juselis bütünlük testi yapılmıştır. Bu çalışmada SPSS istatistiksel yazılımının kullanımı ile her bir anahtar kelimenin ağırlık oranına daha iyi bir çözüm sağlayan ve önde gelen bir enflasyon endeksini oluşturmak için temel bir bileşen analizi benimsenmiştir (Hang vd., 2013).

Yuan ve diğerleri tarafından yapılan bir çalışmada haber sitelerinde bulunan makale bilgileri kullanılarak TÜFE için tahmin modeli oluşturulmuştur. Ekonometrik veriler için Otoresif Entegre Hareketli Ortalama (Autoregressive Integrated Moving Average - ARIMA) modeli, taranmış makale bilgileri için Geri Yayımlı Sinir Ağları (Backpropagation Neural networks - BNN) yöntemi kullanılmıştır. Bu iki yöntem kullanılarak hibrit bir model oluşturulmuştur. Haber bilgileri verileri Çin’de en çok ziyaret edilen haber sitelerinden biri olan sina.com.cn adresinden alınmıştır. Elde edilen bulgular ile haber bilgilerinin TÜFE tahminine çok yakın değerde olduğu gösterilmiştir (Yuan vd., 2013).

Hollanda İstatistik Ofisi Metodoloji departmanında Griffioen ve diğerleri tarafından yapılan bir çalışmada web tarama ile giyim kategorisi için TÜFE hesaplamasını yapan bir proje geliştirilmiştir. Çalışmanın amacı, giyim fiyat endekslerini tahmin etmek için farklı yaklaşımları karşılaştırmak ya da tercih edilen bazı yöntemleri önermemektir. Web tarama ile elde edilen ürünlerden tekrarlı olanlar zamanla tespit edilip kaldırılmıştır. Ayrıca aynı internet adresinden erişim sağlayan farklı ürünler de tespit edilip endekse olan etkileri azaltılmıştır. Çevrimiçi verilerin TÜFE için araştırılmasına yönelik çalışmaların bir parçası olarak, giyim için standart bir

sınıflandırma sistemi geliştirilmiştir. Bu sınıflandırma sistemi kadınlar, erkekler, çocuklar ve bebek kıyafetleri gibi dört kategoriye ayrılmıştır. Sınıflandırma için ürünlerin kısa açıklamalarından faydalanılmıştır. Uzun açıklamalar sınıflandırmayı daha da kolaylaştırmıştır. Sınıflandırma işlemi web tarama ile elde edilen ürünlerin kategori bilgileriyle sağlanmıştır. Sınıflandırma başarısı yüzde 98 gibi bir doğruluk oranıyla gerçekleştirilmiştir. Sınıflandırma yapıldıktan sonra her ürün kategorisi için endeksler hesaplanmıştır. Web tarama ile satış miktarı elde edilemediği için ürün ağırlıklandırılması yapılmamıştır. Bu yüzden bir sınıflandırma için ortalama fiyat kullanılmıştır. Fiyatlar günlük olarak gözlemlendiğinden, bir ay içinde daha sık gözlenen öğelerin aylık ortalama fiyatları diğer öğelere göre daha fazla ağırlık kazanmıştır. Giyim kategorisi için en önemli sorunun çevrimiçi satılan ürünün iade edilme durumunun ortaya çıkması olduğu bilgisine ulaşılmıştır. Bu da endekse etki eden durumlardan bir tanesidir (Griffioen vd., 2014).

Büyük verinin ürün fiyatları istatistikleri üzerindeki etkisini araştıran bir çalışmada büyük veri yönteminin istatistik ofisleri için anlık olarak fiyat yükselişlerindeki değişimleri bildirebilecek bir potansiyeli olduğu araştırılmıştır. Veriler 3 farklı yöntemle elde edilmiştir. Bunlar marketlerden alınan veriler, web tarama ile elde edilen veriler ve mobil uygulama verileridir. Örnek bulgular için Brezilya ürün fiyatları incelenmiştir. Elde edilen sonuçlar ile büyük veri TÜFE'si ile Brezilya resmi TÜFE'si karşılaştırılmıştır (Dubey& Gennari, 2014).

D. P. Manik ve Albarda tarafından yapılan bir çalışmada büyük veri kapsamında web tarama verisi kullanılarak TÜFE hesabının günlük düzeye indirgenmesi hedeflemiştir. Bunun için web tarama yöntemi kullanılarak marketlerin çevrimiçi alışveriş sitelerinden günlük ürün fiyatları elde edilmiştir. Günlük TÜFE hesaplama akışı; çevrimiçi marketleri belirleme, web tarama gerçekleştirme, veri ayrıştırma, kayıt gerçekleştirme, TÜFE hesaplama ve veri görselleştirme aşamalarından oluşmuştur. Bu çalışmada 3 farklı veri toplama yöntemi kullanılmıştır; marketlerden veri toplanması, web tarama ve mobil uygulama verileri. Bu yöntemler ile elde edilen sonuçlar karşılaştırılmıştır ve web tarama yönteminin avantajları sunulmuştur. Bu çalışmada web tarama yönteminin veri toplama maliyetini azalttığı, verilerin TÜFE için günlük düzeyde hesaplanabildiği, veri çeşitliliğinin fazla olduğu ve detaylı veri özelliklerinin olduğu gösterilmiştir (Manik, 2015).

Swier yaptığı bir çalışmada web tarama ile çevrimiçi 3 marketten ürün bilgilerini toplamıştır. Bu çalışma ile elde ettiği sonuçların İngiltere İstatistik Ofisi için enflasyon ölçümünde etkili bir yöntem olabileceğini göstermiştir. Web tarama yönteminin maliyeti azalttığı, veri kapsamını genişlettiği, güncelleme sıklığını artırdığını ve yeni endeksler için araştırma sağladığı vb. avantajlarından bahsedilmiştir. TÜFE sepetinde bulunan 35 ürün kategorisi için ortalama 6500 fiyat teklifi toplanmıştır. Toplamda 424011 ürün verisi üzerinde işlemler gerçekleştirilmiştir. Elde edilen ürünler 'Munging' yöntemi kullanılarak sınıflandırılmıştır. Bu sınıflandırma işlemi TÜFE sepetinde bulunan ürün kategorilerine göre gerçekleştirilmiştir (Swier, 2015).

Norveç İstatistik Ofisi TÜFE'deki çevrimiçi fiyatların payının artırılması ve etkin çevrimiçi veri toplama çalışmalarının artırılması için bir çalışma gerçekleştirmiştir. Norveç'te çevrimiçi en çok satış işlemi gerçekleşen kategorilerin tatil ve eğlence amaçlı seyahatler olduğu bilgisine ulaşılmıştır. Bu da çevrimiçi satışların toplam yüzde 44'üne denktir. Yabancı sitelerden alınan ürün oranı ise yüzde 23'üne denk gelmektedir. Satın alınan ürünlerin yaklaşık yüzde 50'si düşük fiyatlı ürünlerdir (23 Euro'dan düşük). Bu çalışmada web tarama işlemi import.io web uygulamasıyla gerçekleştirilmiştir. Ayrıca özel yazılımlar da kullanılmıştır. Başlangıç için, kişisel bakım ürünleri ve ev elektroniği gibi çevrimiçi satın almaların yüksek bir kısmını temsil eden tüketici grupları seçilmiştir. Test hesaplamaları yapmak için İstatistiksel Analitik Sistem (Statistical Analysis System -SAS) programında tam ürün kapsamına sahip olan gıda ve alkolüz içeceklerin endeksinde kullanılan tarayıcı veri üretim sistemini büyük oranda kopyalayan bir üretim sistemi kurulmuştur (Nygaard, 2015).

İtalya İstatistik Ofisi'nde (ISTAT) yapılan bir çalışmada elektronik eşya ve uçak biletleri fiyatlarının web tarama ile elde edilmesini ve bu veriler ile TÜFE'nin hesaplanması hedeflenmiştir. Test işlemleri veri toplama sürecinde ISTAT tarafından (%21'den fazla ürün ile) yapılmış ve internet üzerinden gerçekleştirilen İtalyan tüketici fiyat anketi kapsamında yapılmıştır. 21 gün içerisinde hesaplanan resmi anket verileri web tarama ile 16 güne indirilebilmiştir. Ancak çoğu çevrimiçi site web tarama için engel koyduğundan sınırlı sayıda bilet fiyatları alınabilmiştir. İstat, engel koyulan siteler ile hukuki bir anlaşma yaparak ürün kapsamını genişletmiştir (Polidoro vd., 2015).

Barcaroli ve diğerleri yaptıkları bir çalışmada istatistik ofislerine yönelik olarak web tarama yönteminin daha etkin kullanabilmeleri için bir strateji geliştirmişlerdir. Bu çalışmanın amacı işletmelerin kullandığı internet adres bilgilerinin doğruluklarının teyit edilmesi ve internet sayfasında bulunan verilerin sağlıklı bir şekilde elde edilmesidir (Barcaroli vd, 2015).

Bu çalışmada Java ile bir uygulama geliştirilmiştir. Uygulamada web tarama modülü geliştirilmiş ve 20000 e yakın işletmenin (çevrimiçi marketler de mevcut) internet sayfaları adres bilgileri kullanılmıştır. Makine öğrenme teknikleri kullanılarak işletmelerin internet adreslerinin doğrulukları test edilmiştir. İnternet adres bilgilerinin doğrulukları için yapay sinir ağları, Rastgele Orman (Random Forest – RF) ve Logistic Regresyon (LR) yöntemleri kullanılmıştır. RF yöntemi ile yüzde 82 doğruluk oranı elde edilmiştir (Barcaroli vd., 2016).

İngiltere İstatistik Ofisi tarafından yapılan bir çalışmada web tarama yöntemi ile TÜFE sepetinde bulunan 520 tane ürün kategorisi için ürün bilgileri elde edilmiştir. İngiltere'de gıda pazarının yüzde 50'sine denk gelen üç çevrimiçi market zincirinden web tarama

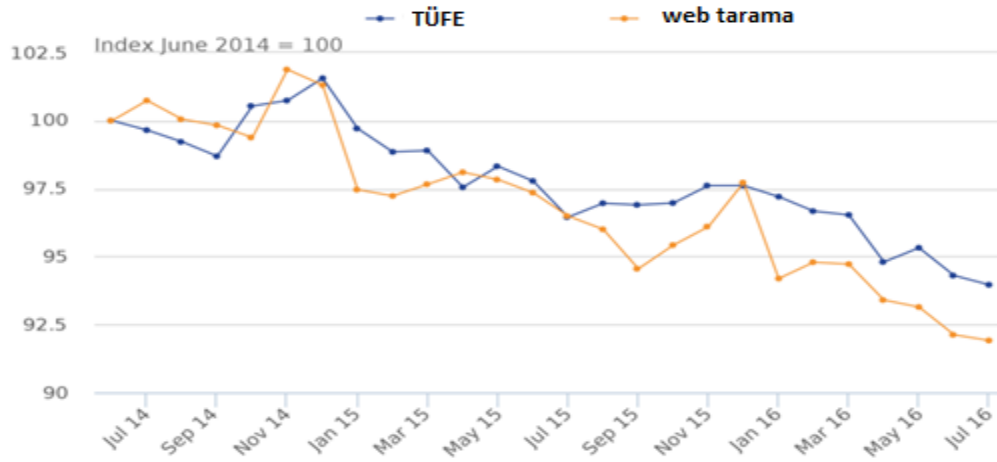
yöntemi ile ürün bilgileri toplanmıştır. Web tarama Python kütüphanesinde bulunan Scrapy ile gerçekleştirilmiştir. Bu çalışmada Web tarama, TÜFE ögesi açıklamasına uyan uygun ürünleri tanımlamak için web sitelerinin kendi sınıflandırma yapısını kullanmıştır. 33 TÜFE sınıflandırması için günde yaklaşık 150.000 fiyat toplanmıştır. Bu da her bir sınıf için yaklaşık 5000 farklı ürün fiyatı tekabül etmiştir. Elde edilen ürünlerin sınıflandırılması aşamasında TÜFE sınıflandırma yapısına uygunluğu için fiyatların toplanabileceği temsili ögeler seçilmiştir. Bu aşamada denetlenen makine öğrenme yöntemlerinden Destek Vektör Makinesi (Support Vector Machine - SVM) tekniği kullanılarak ürünlerin sınıflandırılması hedeflenmiştir. Bu çalışmada kullanılan sınıflandırma sistemine en yakın market ürün kategorileri seçilmiştir. Bu kategoriler SVM’de veri seti olarak kullanılmıştır. Verilerin yüzde 80’i eğitim yüzde 20’si test verileri için kullanılmıştır. Bu eğitim verilerini oluşturmak için toplanmış ürünlerin TÜFE madde endekslerinin her biri için uzman fiyat toplayıcılarına binlerce örnek verilmiştir. SVM ile sınıflandırma eğitimi tamamlandıktan sonra hatalı sınıflandırılmış veriler için denetlenmeyen makine öğrenme yöntemi kullanılmıştır. Bu yöntem ile sınıflandırması yapılan ürünlerin fiyat uyumsuzlukları tespit edilmiştir. Bu işlemlerden sonra SVM yöntemi ile yüzde 85 ile sonuç alınmıştır (Chuanyang vd., 2016).

Cavallo ve Rigobon yaptıkları bir çalışmada TÜFE’den başlayarak geleneksel ekonomik göstergelerin hesaplamasını iyileştirmek için yeni bilgi kaynaklarının kullanılmasını hedeflemişler. Web tarama için kullanılacak çevrimiçi marketler özenle seçilmiştir. Bu marketler hem çevrimiçi hem de çevrimdışı satışlar yapabilmektedir. Ayrıca TÜFE sepetinde bulunabilecek ürünlerin kategorileri baz alınarak çalışmalara devam edilmiştir. Daha sonra, çevrimiçi fiyat endekslerini hesaplamak ve çoğu ülkedeki tüketici fiyatları endeksleriyle nasıl birlikte hareket ettiğini göstermek için kullanılan metodoloji tanımlanmıştır. Çevrimiçi fiyatların, resmi fiyat endeksleri için ve tarayıcı fiyatları için toplanan fiyatlara kıyasla çok farklı fiyat değişiklikleri sergilediği gösterilmiştir. Veriler toplandıktan sonra veri temizleme aşamasına geçilmiştir. Sonra ürünler ortak bir veritabanında TÜFE kategorilerine uyacak şekilde sınıflandırılmıştır. Zaman içinde ürün özellikleri ve performanslarını gözlemlemek için göstergeler hesaplanmıştır. Bir çevrimiçi market endekse dahil edilmeden önce 1 yıl boyunca verileri izlenmiştir. Yaklaşık 25 ülke için veri seti, TÜFE sepetinin en az yüzde 70’ini kapsayan kategorilerden oluşmaktadır. Bu çalışmada hedonik(kalite ayarlama) teknikler de ele alınmıştır. İki mal bir süredir birlikte mevcutsa, üst üste gelen fiyatlar, kalite değişikliği tahminini elde etmek için kullanılabilir. Uygulamada bu yaklaşımın, yeni ürün çeşidinin piyasaya sürülmesindeki fiyat aralığının daha çok kalite farkını yansıttığını varsaymaya eğilimli olduğu sonucuna varılmıştır (Cavallo&Rigobon, 2016).

Gabrielli ve diğerleri yaptıkları bir çalışmada veri toplama maliyetlerinin ve endeksin gerçek zamanlı olarak kullanılabilirliğinin azaltılması amacıyla TÜFE hesaplanması için perakende veri piyasası verileriyle büyük veri yöntemini kullanmışlardır. Endeks için kullanılan büyük veri seti, süpermarket zinciri Coop’un kullanıcılarına ait tüm işlemlerden oluşan verilerdir. Bu veri kümesindeki her kayıt, satın alma tarihi, işlem numarası, ürün numarası, satın alınan öğelerin sayısı, öge fiyatı hakkında bilgiler içermektedir. Test işlemleri için 2011 ve 2012’nin ilk çeyreğine karşılık gelen veriler kullanılmıştır. Veri kümesi 56 milyon işlemden oluşmaktadır (ortalama 80 milyon ürün için). Coop ürünleri resmi TÜFE’ye uygun olarak sınıflandırılmıştır. Ürün fiyatları birim üzerinden hesaplanmış ve TÜFE sepeti ağırlıklandırılması için her kategorinin satın alma sıklığı hesaplanmıştır (Gabrielli vd., 2016).

Avusturya İstatistik Ofisi web tarama ile TÜFE hesaplamasını hedeflemektedir. Web tarama için en önemli hedef kategoriler, nakliye(örneğin uçak bileti, tren biletleri, tatil paketi turları), teknik ekipman, giysi ve oteller olmuştur. Giyim haricinde, bu kategoriler esas olarak merkezi fiyat toplama ile karşılanmaktadır. Anlık olarak, Avusturya’daki TÜFE sepetinin yaklaşık %10’u web tarama ile sağlanabilmektedir. Bu çalışma, TÜFE derlemesi için tarayıcı ve web tarama verileri kullanılırken ortaya çıkan zorlukları göstermiştir. Web tarama için import.io sitesi kullanılmıştır (Auer& Boettcher, 2016).

Metcalf ve diğerleri İngiltere İstatistik Ofisinde yaptıkları bir çalışmada web tarama ile elde edilen verilerin ürün fiyatları endeksleri için veri kümelenmesini sağlayan bir model geliştirmişlerdir. Bu model benzer ürün gruplarının oluşturulmasını ve bu grupların zaman içindeki fiyat değişimlerinin izlenmesini sağlamaktadır. Bu modeli bir tüketicinin belirli bir üründen ziyade belirli ürün türlerini satın alma eğilimi teorisine dayanmaktadır. Benzer ürün gruplarını bir araya getirmek için denetlenmeyen bir makine öğrenme yaklaşımlarından Mean-Shift algoritması kullanılmıştır. Çalışmanın sonucunda web tarama ile hesaplanan TÜFE ile resmi hesaplanan TÜFE arasındaki ilişkiye ait nicel bir buğu açıklanmamıştır. Ancak iki endeks arasındaki ilişkiye ait şekil 3’de gösterilen grafik sunulmuştur. Grafikten de anlaşılacağı üzere iki endeks arasında anlamlı bir ilişki saptanmıştır (Metcalf vd., 2016).



Şekil 3. Web Tarama ile Hesaplanan TÜFE ile Resmi Yayınlanan TÜFE Arasındaki İlişki

Thakur ve diğerleri yaptıkları bir çalışmada Geri Yayımlı Çok Katmanlı İleri Beslemeli Sinir Ağı (Multilayer Feedforward Neural Network - MLFFN) yöntemi kullanarak Hindistan için enflasyonu tahmin eden bir model geliştirmişlerdir. Hindistan için 2000 ve 2012 yılları arasındaki geçmişteki veriler incelenmiştir. Sinir ağı için kullanılan girdi değerleri gayri safi yurt içi hasıla, faiz oranları, ithalat ve ihracat bilgileri, altın ve petrol fiyatları, cari açık bilgileri ve uluslararası rezervler bilgileri olmuştur. Matlab ortamında verilerin yüzde 70'i eğitim, yüzde 15'i test ve yüzde 15'i de doğrulama için kullanılmıştır. Bu çalışmada oluşturulan enflasyon tahmin modeli ile gerçek enflasyonu arasında 2014 yılı için yüzde 6,37 fark oluşmuştur (Thakur vd., 2016).

Griffioen ve diğerleri Hollanda İstatistik ofisinde yaptıkları bir araştırmada web tarama verilerinin TÜFE için bir kaynak olabileceğini göstermişlerdir. 16 web tarama robotu ile 19 çevrimiçi siteden giyim ürünleri toplanmıştır. Çalışmaya göre giyim kategorisi daha homojen bir yapıya sahip olduğu için ürün sınıflandırılması manuel olarak gerçekleştirilmiştir. Ancak çalışmanın sonraki aşamalarında makine öğrenme teknikleri kullanılarak ürün sınıflandırılması hedeflenmiştir (Griffioen vd., 2016).

Avrupa İstatistik Ofisi tarafından 2016 yılında ESSNet Big Data projesi oluşturulmuştur. Bu projenin amacı büyük veri kaynaklarının resmi istatistiklerin üretilmesine katkı sağlamasıdır. Proje kapsamında TÜFE için kullanılacak veri kaynaklarına web tarama verilerinin eklenmesi hedeflenmiştir (Struijs vd., 2017).

Nugroho ve diğerleri yaptıkları bir çalışmada büyük veri ve bulut teknolojisi yöntemlerini kullanarak günlük TÜFE tahmin modelinin geliştirilmesine yönelik bir araştırma yapmışlardır. Bu çalışmada bulut teknolojisi kullanılarak Destek Vektör Regresyonu (Support Vektör Regression - SVR) ve Rastgele Orman Regresyonu (Random Forest Regression - RFR) yöntemleri ile günlük TÜFE tahmini modeli geliştirilmesi hedeflenmiştir. Sonuçların doğruluğunu artırmak için GridSearch ve Random yöntemlerinden faydalanılmıştır. Ürün bilgileri 2012-2016 yılları arasında web tarama yöntemi ile çevrimiçi sitelerden toplanmıştır. Verilerin yüzde 70'i eğitim yüzde 30'u test için kullanılarak TÜFE tahmini için bir sistem oluşturulmuştur. Geliştirilen sistem kişisel bilgisayar ortamında 837 saniyede gerçekleşirken bulut ortamında 522 saniye iken gerçekleşmiştir. Bu çalışmada TÜFE hesabının kişisel bilgisayar ve bulut teknolojisine aktarıldığındaki performans karşılaştırması üzerinde de durulmuştur. Bu çalışma ile SVR'nin aylık tahminde 0.3454, günlük tahminlerde 0.0095 daha az hata oranına sahip olduğu gösterilmiştir. RFR yöntemi için sonuçlar günlük tahmin değeri için 0.0171 hata oranında olmuştur (Nugroho vd., 2017).

Powell ve diğerleri tarafından yapılan çalışmada kategori düzeyinde ayrıştırılmış TÜFE'yi resmi anketlerle karşılaştırarak gizli enflasyon durumunu ortaya çıkarılmıştır. Bu çalışmanın amacı bir ürünün piyasaya sürülmeden önce ait olduğu TÜFE kategorisinin durumunu tahmin etmeye yardımcı olmaktadır. Zamanla web tarama ile toplanan ürünler ile en iyi anket verileriyle birleştirildiğinde, daha sık ara dönem toplam TÜFE tahmini elde edilmesi hedeflenmiştir. Bu çalışmanın ikincil amacı yayınlanan TÜFE için öngörüler üretmektir; böylece ilgili taraflar için resmi enflasyon istatistiklerinin açıklanmasından önce, tahminleri değerlendirebilir hale gelmesi sağlanmıştır. BPP çalışmalarını temel alan bu çalışma, enflasyon için tahmin aralıkları üretilmesine ve modele uygun parametrelerinin belirlenmesine yardımcı olmaktadır. Parametreler için Ampirik Bayes tahminlerini hesaplamak için log-posterior yoğunluğun sayısal optimizasyonu, R programlama dilinde uygulanan Nelder-Meadsimpleks yöntemi kullanılarak gerçekleştirilmiştir. Tahmin öncesinde model uyumunun bir yan ürünü olarak, bu çalışmadaki metodoloji aynı zamanda farklı ürünlerin fiyat değişkenliğini kodlayan hiperparametreler için tahminler üretmiştir. Bu çalışmada web tarama ürünlerinin fiyat endeksi yüksek olduğu için anketlerin ötesine geçebileceği ancak web tarama ile elde edilmiş ürünlerin kullanılması durumunda riskler olabileceği belirtilmiştir. Çünkü bu çalışmada parekendecilerin TÜFE'yi etkilemek üzere çevrimiçi fiyatlar üzerinde manipülasyon yaptıklarından bahsedilmiştir (Powell vd., 2018).

Sutiawan ve Nugraha yaptıkları bir çalışmada tüketim malları (TÜFE emtia) için çevrimiçi fiyat tahmini sistemi geliştirmişlerdir. Hypermart ve Alfacart çevrimiçi marketlerden web tarama ile veri toplanmıştır. Web tarama ile elde edilen veriler Weka ile geliştirilmiş bir tahmin modeline sunulmuştur. Tahmin için önceden veri setleri hazır hale getirilmiştir. Veritabanına kayıt edilen ürünlerin ortalama fiyatı veri kapsamını oluşturmaktadır. Tahmin modeli için yapay sinir ağlarından Çok Katmanlı Algılayıcı (Multi Layer Perceptron - MLP) yöntemi kullanılmıştır. Uygulama için kullanılan alt yapı teknolojileri PHP programlama dili ve Yii Framework 2.0 ve Mysql veritabanıdır. Bu araştırmadaki kullanıcı yanıtları, Endonezya İstatistik Ofisi çalışanlarından elde edilmiştir. Toplam 33 katılımcı anketleri doldurmuştur. Kullanıcılardan veri toplandıktan sonra, istatistiksel test olarak tanımlayıcı ve regresyon analizi kullanılmıştır. Ankete katılanların %83,34'ünün bu sistemin geçerli bir karar destek sistemi olabileceği gösterilmiştir (Sutiawan&Nugraha, 2017).

Aparicio ve Bertolotto yaptıkları bir çalışmada çevrimiçi marketlerden toplanan fiyatların önceki ürün fiyatlarına göre TÜFE'nin gelecekteki seviyesini tahmin etmede daha etkili araçlar olduğunu araştırmıştır. Bir ay öncesinde TÜFE enflasyon oranını tahmin edip ardından iki ve üç aylık önde tahminler hesaplanmıştır. Bu tahminler Bloomberg tahmincilerine göre daha başarılı sonuçlar ortaya çıkarmıştır. MIT'deki BPP'den ayrılan özel bir şirket olan PriceStats tarafından sağlanan çevrimiçi endeks verileri kullanılmıştır. Resmi TÜFE'ye öncesinde yol göstericilik yapan çevrimiçi endeksler oluşturulmuş ve bu endeksleri oluşturan bir model geliştirilmiştir (Aparicio&Bertolotto, 2020).

Huang ve diğerleri yaptıkları bir makale araştırmasında büyük veri kapsamında web tarama verilerinin TÜFE için alternatif bir kaynak olabileceğini ifade etmişlerdir. Web tarama ile toplanan verilerin TÜFE için belirlenen bazı ana kategorileri sağlayabileceğini savunmuşlardır. Bu çalışmada herhangi bir nicel sonuca ulaşılmamış olup genel perspektiften önerilerden bahsedilmiştir (Huang vd., 2017).

Yeni Zelanda İstatistik Ofisi tarafından yapılan bir araştırmada büyük veri modelleri kullanılarak TÜFE hesaplanması için hangi aşamaların ihtiyaç olabileceği üzerine çalışma yapılmıştır. İstatistik Ofisinde Sabit Etkiler endeksi geliştirilmiş ve bu endekse bağlı olarak büyük veri yöntemleri TÜFE için kullanılabilir hale getirilmiştir. BPP tarafından toplanan web tarama çevrimiçi verileri kullanılmıştır. Bu veriler ile birlikte Yeni Zelanda'da çevrimiçi satış yapan perakendecilerin potansiyeli hakkında fikir sağlamıştır. Bu çevrimiçi veriler anket ve marketlerden alınan tarayıcı verileri ile kullanıldığında istatistik ofisinde endeks hesaplaması için büyük imkânlar sağlamıştır. Ayrıca bu çalışmada büyük verilerin elektronik eşya ve ikinci el arabalarda kullanılan kalite ayarlamalarındaki etkisi araştırılmıştır (Bentley&Krsinich, 2017).

Cavallo tarafından yapılan bir çalışmada web tarama ile elde edilen çevrimiçi fiyatlar ile normal perakende fiyatları arasındaki değişimler gözlemlenmiştir. Bu çalışmada zaman serisinde çevrimiçi fiyatların gerçek fiyatlarla yüzde 72 oranında aynı olduğu tespit edilmiştir. Yapılan incelemeler ile çoklu perakendeci bulunduran çevrimiçi sitelerin fiyat değişimlerinde genelde daha tutarlı olduğu gösterilmiştir (Cavallo, 2017).

Hull ve diğerleri yaptıkları bir çalışmada İsveç'te bulunan çevrimiçi marketlerden meyve ve sebze ürünlerini web tarama yöntemiyle toplayarak kısa vadeli enflasyon tahmini için bir sistem geliştirmişlerdir. Geliştirilen sistem Linux sunucusunda kurulmuştur. Bu sunucuda 4 çevrimiçi market için veri toplayan uygulama geliştirilmiştir. Ürünlerin kategorilerinin oluşturulması için sözlük tabanlı bir veri tabanı kurulmuştur. Bu sözlük, toplanan web tarama ürün bilgilerini TÜFE kategorileriyle eşleştirilmesini sağlamıştır (Hull vd., 2017).

Zhou yaptığı bir çalışmada market verilerini kullanarak İsveç enflasyonu için bir modelleme araştırması gerçekleştirmiştir. Bu modelleme sürecinde tüketici bakış açısından TÜFE için web tarama verilerinin kullanılabilirliği vurgulanmıştır (Zhou, 2017).

Whitaker tarafından yapılan bir çalışmada istatistik ofisleri için anket verileri ile büyük veri bilgileri karşılaştırılmıştır. Büyük veri, anket verilerinin ne kadarını karşılayabileceği tartışılmış ve bulgular ile başarılı sonuçlar alınmıştır. Büyük veri için Equifax tüketici kredisi kayıtları ve New York Tüketici Kredisi Paneli / Equifax Federal Reserve Bank bilgileri kullanılmıştır (Analytic Dataset, 2018). Bu çalışmada her iki veri kümesinde yer alan değişkenlerin yanı sıra kredi kayıtlarına dahil edilen nüfus sayımı toplam demografi verilerine sahip modellerde kullanarak hanehalkı borç modellerinin tahmini oluşturulmuştur. Politik öneriler için genelde anket verilerinden faydalanılmıştır. Bu çalışmada yapılan analizlerle borç dengelerinin yaş, gelir ve aile yapısı arasındaki ilişkiler nasıl etkisinin olduğu araştırılmıştır (Whitaker, 2018).

Harchaoui ve Janssen yaptıkları bir çalışmada farklı frekanslarda örneklenen verileri barındıran bir model kullanarak BPP TÜFE'sinin öngörücü içeriğini değerlendirmiştir. Bu model zaman periyotlarının çeşitliliği için iyi tahminler üretmektedir. Resmi TÜFE rakamlarının tahmini doğruluk oranına önemli bir katkıda bulunmuştur. Bu çalışmada uygun bir modelleme stratejisiyle günlük BPP TÜFE'sinin, resmi TÜFE'nin zamanlamasını güvenilir bir şekilde arttırmak için verimli bir şekilde kullanılabilirliğini savunulmuştur. Bu model ile ABD enflasyonunun bir uygulama ile değerlendirilmesi gerçekleştirilmiştir. Alınan sonuçlar ile 12 aylık tahminin 0,4 oranında iyileşmesi ve 0,6 gibi bir doğruluk gibi sonuçlar alınmıştır. Elde edilen bulgular ile bu modelin günlük BPP TÜFE'sinin zaman aralıklarında genel olarak sağlam sonuçlar elde ettiği gösterilmiştir (Harchaoui&Janssen, 2018).

Cavallo yaptığı bir çalışmada web tarama yöntemi ile elde edilen ürün bilgilerini, marketlerden toplanan tarayıcı ürün bilgileri ile karşılaştırmıştır. Bu karşılaştırma ile fiyat değişim süreleri ve fiyat değişim büyüklüğü analizleri yapılmıştır. Web tarama verileri PriceStat tarafından elde edilmiştir. Söz konusu veriler 31 ülkeye ait 181 perakendecinin ürün bilgileridir. Büyük veri seti ile istatistiksel analizler yapılmıştır. Yapılan analiz sonuçlarına göre ülkelerin TÜFE oranı, fiyat değişim oranı ve ortalama fiyat büyüklüğü oranı vb. bilgiler elde edilmiştir (Cavallo, 2018).

Abe ve Shinozaki yaptıkları bir çalışmada dayanıklı ürünlerin kalite değişimi sonucu oluşan fiyat değişiklikleri için makine öğrenme ve doğal dil işleme tekniklerini kullanmışlardır. Eski ürünlerin yerine geçen yeni ürünler eşleştirilerek kalite değişimleri tespit edilmiştir. Ürünler web tarama yöntemi ile Japonya'da bulunan kakaku.com çevrimiçi alışveriş sitesinden elde edilmiştir. Metin eşleştirme için Jaro-Winkler Distance yöntemi, makine öğrenme için SVM yöntemi kullanılmıştır. Web tarama ürünleri ve resmi olarak anketler ile elde edilen ürünler karşılaştırılmıştır. Bu karşılaştırma sonucunda 20 elektronik ev eşyası için kaliteden kaynaklı fiyat değişimleri gözlemlenebilmiştir (Abe&Shinozaki, 2018).

Cavallo ve diğerleri yaptıkları bir çalışmada BPP verilerini, Uluslararası Karşılaştırmalar Programı (International Comparison Program - ICP) ve Ekonomik İşbirliği ve Kalkınma Örgütü (Organisation for Economic Co-operation and Development - OECD) tarafından sunulan tüketim ürün verileriyle karşılaştırmışlardır. Bu çalışmada kullanılan veriler, 2011'den 2017'ye kadar üç ana tüketim kategorisinde, gıda ve içecek, yakıt ve elektronik olmak üzere toplam 11 ülkeyi kapsamaktadır. BPP tarama verilerinin ICP ve OECD tarafından sunulan istatistik verileri arasında ülkeler için ortalama yüzde 15 farklılık göstermiştir (Cavallo vd., 2018).

Bosch ve diğerleri yaptıkları bir çalışmada web tarama ve anket verilerini birleştirerek istatistik ofisleri için genel bir metodoloji önermişlerdir. Web tarama verileri hakkında detaylı bilgileri elde etmek için birden fazla çevrimiçi siteden veriler toplanmıştır. Bu süreçte edinilen tecrübeye bağlı olarak web tarama yöntemi için izlenecek adımlar belirlenmiştir. Ayrıca web tarama verilerinin anket verilerine uyumluluğunu ölçmek için belli aşamalar tanımlanmıştır. Süreç içerisinde web tarama yönteminin istatistik ofisleri için anket yanıt yükünü hafifletmesi ile istatistik üretim süreçlerini iyileştirip hızlandırdığı ele alınmıştır. Dezavantaj olarak web tarama ile işlenmemiş verilerin elde edildiği ve bunun bir maliyeti olabileceği vurgulanmıştır (Bosch vd., 2018).

TÜFE için yapılan büyük veri çalışmalarının literatür karşılaştırmaları Tablo 1'de gösterilmiştir.

Tablo 1. Büyük veri çalışmalarının karşılaştırılması

Yazarlar	Konu	Özellikler	Yöntem	Veri Seti	Hassasiyet
Cavallo	Çevrimiçi ve resmi fiyat endeksleri: Arjantin'in enflasyonunu ölçmek	Çevrimiçi TÜFE hesaplanması	Web tarama, Büyük veri	Web verileri	tarama
Hang	Web arama verilerine göre enflasyon endeksinin oluşturulmasına ilişkin bir çalışma	Anahtar kelimeler ile TÜFE tahmini	Büyük veri, Granger Test, Dickey-Fuller Test, SPSS	Arama logları	motoru
Yuan ve diğerleri	İnternet haber makaleleriyle hibrit Sarima ve Sinir Ağı Modelini Kullanarak TÜFE Tahmini	TÜFE tahmini	Web tarama, Büyük veri, ARIMA	Haber sitelerindeki haber ve makale bilgileri	
Griffioen ve diğerleri	İnternette giyim verisi toplamak	Web tarama ile giyim kategorisi için TÜFE hesaplanması	Web tarama, Büyük veri	Çevrimiçi alışveriş site verileri (Giyim)	98
Dubey ve Gennari	Büyük miktarda veri içeren tüketici fiyat endeksleri: Kamu-özel tamamlayıcılıklar	Büyük veri TÜFE'si ile resmi TÜFE karşılaştırılması	Web tarama, Büyük veri	Web verileri, Verileri	tarama Market Mobil
Manik ve Albarda	İstatistik Endonezya'da büyük veri yöntemiyle günlük TÜFE oluşturma stratejisi	Günlük TÜFE	Web tarama, Büyük veri	Web verileri, Verileri	tarama Market Mobil

Tablo 1 (devam). Büyük veri çalışmalarının karşılaştırılması

Yazarlar	Konu	Özellikler	Yöntem	Veri Seti	Hassasiyet
Breton diğerleri	ve Tüketici fiyat endekslerini oluşturmak için web tarama verileri kullanma	Aylık ve günlük Çevrimiçi hesaplanması	Web tarama, Büyük veri, Munging	Tesco, Sainsbury ve Waitrose çevrimiçi market verileri	
Nygaard	Norveç Tüketici Fiyatları Endeksinde çevrimiçi fiyatların kullanımı	Çevrimiçi tahmini	TÜFE Web tarama, Büyük veri, import.io	Çevrimiçi market verileri	
Federico diğerleri	ve İtalyadaki Harmonik TÜFE derlemesi için tüketici elektroniği ve uçak bileti verileri toplamak için web tarama teknikleri	TÜFE hesaplanması	Web tarama, Büyük veri	Çevrimiçi alışveriş site verileri (Elektronik Eşya ve Uçak Biletleri)	
Barcaroli diğerleri	ve İnternetin Resmi İstatistikler İçin Veri Kaynağı Olarak Kullanılması: Web tarama teknolojisinin Karşılaştırmalı Analizi	Web tarama yönteminin daha etkin kullanılması	Büyük veri, Web tarama	Çevrimiçi alışveriş site verileri	
Barcaroli diğerleri	ve İstatistiğinin Veri Kaynağı Olarak İnternet Kullanımı: İnternetteki İşletmeleri Belirleme Stratejisi	Çevrimiçi marketlerin internet adreslerinin doğruluklarının ölçülmesi	Büyük veri, Web tarama Random Forest, Logistic Model, Web Tarama	Çevrimiçi alışveriş site verileri	82
Cavallo Rigobon	ve Milyar Fiyatlar Projesi: Ölçüm ve Araştırma İçin Çevrimiçi Fiyatları Kullanma	TÜFE tahmini iyileştirmek	Web tarama, Büyük veri	Çevrimiçi market verileri	
Gabrielli diğerleri	ve Perakende Fiyat piyasasını kullanarak TÜFE için Büyük Veri yöntemi kullanılması	TÜFE hesaplanması	Web tarama, Büyük veri	Coop çevrimiçi market verileri	
Auer diğerleri	ve Yeni büyük veri kaynakları fiyat istatistikçilerinin endeks derleme prosedürlerini tekrar ele alınması	TÜFE hesaplanması	Web tarama, Büyük veri, import.io	Çevrimiçi alışveriş site verileri (Tren ve Uçak Biletleri, tatil siteleri)	
Metcalf diğerleri	ve Web'deki fiyat verilerini kullanarak araştırma endeksleri: büyük veri kümelerini fiyat endekslerine kümelemek	Tüketicinin ihtiyaç ürünlere olan eğiliminin ölçülmesi	Web tarama, Büyük veri, Mean-Shift	Çevrimiçi alışveriş site verileri	
Thakur diğerleri	ve Hindistan'da Enflasyon Tahmini için Yapay Sinir Ağı Tabanlı Model	Enflasyon Tahmini	Büyük veri, Geri Yayılımlı Çok Katmanlı İleri Beslemeli Sinir Ağı	2000-2012 yılları arası resmi veriler(GSYH, Faiz oranları, ithalat ve ihracat, altıp ve petrol fiyatları)	93.63

Tablo 1 (devam). Büyük veri çalışmalarının karşılaştırılması

Yazarlar	Konu	Özellikler	Yöntem	Veri Seti	Hassasiyet
Griffioen ve diğerleri	Hollanda TÜFE'si için İnternet verilerinin kullanımı	Web tarama verilerinin TÜFE için kaynak olabilmesi	Web tarama, Büyük veri	Çevrimiçi alışveriş site verileri (Giyim)	
Struijs	ESSnet Big Veri projesi	Büyük veri yönteminin İstatistik ofisleri için istatistikleri üretmesi	Büyük veri, Web tarama		
Nugroho ve diğerleri	Destek vektör regresyon yöntemi tabanlı bulut bilişim kullanarak günlük tüketici fiyat endeksini tahmin etmek	Günlük TÜFE tahmini	Web tarama, Destek Vektör Regresyon (SVR), Random Forest, Bulut teknolojisi, Grid Search, Random Search, Büyük veri	Web tarama verileri	
Powell ve diğerleri	Web tarama fiyat mikro verilerini kullanarak fiyatları izleme ve modelleme: otomatik günlük tüketici fiyat endeksi tahminine doğru	Günlük TÜFE tahmin modeli	Web tarama, Log-posterior yoğunluğu sayısal optimizasyonu, Büyük veri	BPP Verileri	
Sutiawan ve Nugraha	Tüketim mallarının çevrimiçi fiyat tahmin sistemi	Çevrimiçi fiyat tahmini	Web tarama, Big Data, Weka, MLP	Hypermart ve Alfacart çevrimiçi market verileri	83.34
Aparicio ve Bertolotto	Çevrimiçi Fiyatlarla Enflasyon Tahmini	TÜFE tahmini	Web tarama, Büyük veri	BPP Verileri	
Huang ve diğerleri	Sepet Güncellemelerinin Frekansı ve Uygulama Gecikmesinin Kanada TÜFE'sine Etkileri	Web tarama yönteminin TÜFE için alternatif olması	Web tarama, Büyük veri		
Bentley ve Krsinich	Yeni Zelanda için büyük veri kapsamında TÜFE	TÜFE hesaplanması, ürünleri kalite ayarlamaları	Web tarama, Büyük veri	BPP Verileri, Çevrimiçi alışveriş site verileri, Market Verileri	
Cavallo	Çevrimiçi ve Çevrimdışı Fiyatlar Benzer misiniz? Büyük Çok Kanallı Perakendecilerin Kanıtları	TÜFE için fiyat istikrarının hesaplanması	Web tarama, Büyük veri	BPP Verileri	
Hull ve diğerleri	Çevrimiçi olarak toplanan fiyat bilgileri ve kısa vadeli enflasyon tahminleri	Kısa vadeli enflasyon tahmini	Web tarama, Büyük veri	Çevrimiçi market verileri (meyve ve sebze)	
Zhou	Piyasa Verilerini Kullanarak İsveç Enflasyonunu Modellemek	Market verilerinin TÜFE için kaynak olabilmesi	Büyük veri	Market verileri	

Tablo 1 (devam). Büyük veri çalışmalarının karşılaştırılması

Yazarlar	Konu	Özellikler	Yöntem	Veri Seti	Hassasiyet
Whitaker	Ankete Karşı Büyük Veri	Büyük	Büyük veri	Equifax tüketici kredisi kayıtları, New York Tüketici Kredisi Paneli / Equifax Federal Reserve Bank bilgileri	
Harchaoui ve Janssen	Büyük veriler resmi istatistiklerin zamanlamasını nasıl artırabilir?: ABD tüketici fiyat endeksi örneği	TÜFE tahmini için zamanlama modeli	Web tarama, Büyük veri, MIDAS	BPP verileri	
Cavallo	Web tarama verileri ve yapışkan fiyatlar	Market verileri ve Web tarama verilerinin karşılaştırılması	Web tarama, Büyük veri	BPP Verileri, Çevrimiçi alışveriş site verileri	
Abe ve Shinozaki	Büyük Veri ve Makine Öğrenimi Kullanarak Deneysel Fiyat Endekslerinin Derlenmesi	Tüketici ürünlerinin kalite değişimlerine bağlı olarak fiyat değişimleri	Büyük veri, Web tarama, SVM, Jaro-Winkler	Çevrimiçi alışveriş site verileri (kakaku.com: elektronik eşya)	
Cavallo ve diğerleri	Ülkeler Arasında Gerçek Tüketimi Ölçmek İçin Çevrimiçi Fiyatların Kullanılması	BPP tüketici ürünlerinin ICP ve OECD tüketici ürünleri ile karşılaştırılması	Web tarama, Büyük veri	BPP Verileri	85
Bosch ve diğerleri	Web tarama anket tasarımı karşılar: güçlerin birleştirilmesi	Web tarama ve anket verilerinin birleştirilmesi TÜFE hesaplanması	Büyük veri, Web tarama	Çevrimiçi alışveriş site verileri, Anket verileri	

3. Ürünlerin Sınıflandırılmasında Yapay Zeka Kullanan Çalışmalar

Bu bölümde TÜFE hesaplaması için ürün sınıflamasında kullanılan yapay zeka çalışmalarına yer verilmiştir.

Ghani ve diğerleri yaptıkları bir çalışmada çevrimiçi sitelerden toplanan ürün metin bilgileriyle nitelik değer çıkarma yöntemi geliştirmişlerdir. Çalışmanın amacı özellik-değer çiftleri ile bir veritabanı oluşturarak ürünlerin sınıflandırmasını gerçekleştirmektir. Spor malzemeleri ve giyim kategorisi ürünleri kullanarak farklı yöntemler ile sınıflandırma işlemi gerçekleştirilmiştir. Metin sınıflandırma işlemi için Naive Bayes teorimi, etiketsiz ürün bilgileri için ise Expectation-Maximization (EM) yöntemi kullanılmıştır. Ürün markaları ile yapılan deneyde Naive Bayes yöntemi kullanılarak yüzde 87 ile en iyi sonuç elde edilmiştir (Ghani vd., 2006).

Kannan ve diğerleri yaptıkları bir çalışmada Bing arama motoru için ürün eşleştirme yöntemi geliştirmişlerdir. Çalışmanın amacı arama motoru üzerinden alınan tekliflerin hedef ürünlere uyacak şekilde uygulanabilir bir sistem geliştirmektir. Eşleştirme için ürün açıklamaları kullanılmıştır. Veri seti için 54 kategoriden oluşan 20 bin ürün kullanılmıştır. Geliştirilen Li & Wong subtractMM (LVMM) modeli ile ortalama yüzde 85 doğruluk elde edilmiştir (Kannan vd., 2011).

Kannan ve diğerleri yaptıkları bir çalışmada ürün sınıflandırmasını iyileştirmek için ürün özellikleri ve ürün resimlerini kullanarak kategorilerin oluşturulmasını hedeflemişlerdir. Ürün açıklaması, tanımlanmış alt kategoriyi bulmada kullanılmıştır. Ürün resimleri eşleştirme başarı oranını artırmak için kullanılmıştır. Bu çalışmada görüntü sinyallerinin nasıl etkin bir şekilde kullanılabilceğini araştıran nispeten keşfedilmemiş ortam üzerinde bir araştırma gerçekleştirilmiştir. Eğitim için 17989, test için 10026 ürün kullanılmıştır (Kannan vd., 2011).

Köpcke ve diğerleri yaptıkları bir çalışmada e-ticaret sitelerinde bulunan elektronik ürünlerin eşleştirmelerini iyileştirmek için ürün tanımları kullanılarak özellik çıkarma tekniği uygulamışlardır. Eşleştirme için makine öğrenme ve doğal dil işleme teknikleri kullanılmıştır. Makine öğrenmesi için SVM ve doğal dil işleme için Term Frequency - Inverse Document Frequency (TF-IDF), Trigram ve Jaccard yöntemleri kullanılmıştır. Bir e-ticaret sitesi tarafından sağlanan yaklaşık 100 bin ürün üzerinde yapılan eşleştirmede ortalama yüzde 85 oranında en iyi sonuç elde edilmiştir (Köpcke vd., 2012).

Zahavy ve diğerleri yaptıkları bir çalışmada web tarama ile çevrimiçi alışveriş sitesinden(Walmart.com) ürün verilerini toplamışlardır. Bu çalışmada çok modlu ürün sınıflandırması için karar düzeyinde füzyon yaklaşımı modeli önerilmiştir. Bu model için ürünlerin metin ve görüntü bilgileri girdi olarak kullanılmıştır. Model içeriğinde girdi olarak kullanılan bilgiler ile derin sinir ağları eğitilmiştir ve girdi türü için aralarında seçim yapmayı öğrenen yeni bir yöntem ele alınmıştır. Bu iki yöntemle Walmart çevrimiçi sitesinde bulunan ürünlerin sınıflandırması için doğruluk oranı yüzde 1 olarak artırılmıştır (Zahavy vd., 2016).

Cevahir ve Murakami yaptıkları bir çalışmada e-ticaret siteleri için belirli bir ürün başlığı veya tanımlaması ile ürünün kategorisini tahmin eden bir model geliştirmişlerdir. Bunun için iki sinir ağı(Deep Belief Nets and Deep Autoencoders) modeli birlikte kullanılmıştır. Sinir ağları modelleri grafik işlemci birimleri tarafından eğitilmiştir. E-ticaret sitelerinden elde edilen 150 milyon ürünün 28338 kategoriye eşleştirilmesi yapılmıştır. Yapılan testler ile e-ticaret sitelerinin ürün kategorisi eşleştirme doğruluk oranı yüzde 81 olarak gerçekleşmiştir (Cevahir&Murakami, 2016).

Ristoski ve diğerleri yaptıkları bir çalışmada derin öğrenme tekniklerini, ürün eşleme ve sınıflandırma için standart sınıflandırma yaklaşımlarıyla birlikte kullanan bir yaklaşım ele almışlardır. Metinsel ürün açıklamalardan özellik-değer çifti çıkarma yeteneğine sahip eğitim özelliği çıkarma modelleri için denetimli olarak yapılandırılmış ürün verileri kullanılmıştır. Büyük miktarda kamuya açık ürün verilerinden kelime yerleştirmeleri üretmek için sinirsel dil modelleri kullanılmıştır. Ayrıca, ürün görüntülerinden görüntü yerleştirmeleri üretmek için derin Konvansiyonel Sinir Ağı kullanılmış; bu da her iki aşamada daha iyi sonuçları ortaya çıkarmıştır. Aynı ürün birçok farklı e-mağazada bulunabilir, ancak ürün teklifleri hakkındaki bilgiler farklı e-mağazalar arasında büyük farklılık gösterir. Ürün sınıflandırılmasında kullanılan resimler için derin sinir ağları görüntü bindirmeleri (deep neural nets image embeddings) yöntemi kullanılmıştır (Ristoski vd., 2018).

Xia ve diğerleri yaptıkları bir çalışmada Japonya'da bulunan Rakuten Ichiba çevrimiçi alışveriş sitesi ürünleri için 35 ana kategoriye sınıflandıran bir model geliştirmişlerdir. Japon alfabesine uyacak şekilde Dikkat Evrimsel Sinir Ağı (Attention Convolutional Neural Network - ACNN) sinir ağı kullanılmıştır. Sinir ağı modelini eğitmek için geçen süre 3 haftadan 3 güne indirilmiştir. Rakuten çevrimiçi sitesinde 44000 satıcı için yaklaşık olarak 239 milyon ürün bulunmaktadır. Bu ürün verileri için yapılan eşleştirmede doğruluk oranı yüzde 96 olarak gerçekleşmiştir (Xia vd., 2017).

Roels ve Loon'un Belçika İstatistik Ofisi'nde (Statbel) yaptıkları bir çalışmada web tarama ile tüketici elektroniği, ayakkabı, otel, ikinci el araba ve öğrenci pansiyonları kategorileri için TÜFE hesaplaması gerçekleştirmişler. Elde edilen bu ürünler için veri temizleme ve veri sınıflandırma işlemleri gerçekleştirilmiştir. Veri sınıflandırması için En Yakın Komşu (K-Nearest Neighbors - KNN), SVM, Naive Bayes ve Random Forests makine öğrenme yöntemleri kullanılmıştır. Verilerin yüzde 85'i eğitim ve yüzde 15'i test için kullanılmıştır. Yüzde 86 doğruluk oranıyla Random Forests yöntemi en iyi sonucu vermiştir (Loon&Roels, 2018).

Shah ve diğerleri yaptıkları bir çalışmada eBay çevrimiçi sitesinin ürün bilgileri için eşleştirme yöntemi önermişlerdir. Elektronik eşya, giyim ve aksesuar kategorilerinde bulunan ürünlerin eşleştirilmesi için Sığ Sinir Ağı-FastText (Shallow Neural Network) ve Siyam Ağları (Siamese Networks) yöntemleri kullanılmıştır. Kullanılan ürün veri setinin yüzde 80'ini uluslararası standart etiketleri olan ürünlerdir. Geriye kalan ürünler satıcılar tarafından veya üçüncü kaynaklardan sağlanmıştır. Sığ sinir ağı yöntemi kullanılarak yapılan eşleştirmelerde elektronik eşya için yüzde 88.8 giyim için yüzde 78.8 aksesuar için yüzde 96.4 doğruluk değerleri elde edilmiştir. Siyam sinir ağı yöntemi kullanılarak yapılan eşleştirmelerde elektronik eşya için yüzde 89.9 giyim için yüzde 82.4 aksesuar için yüzde 97.1 değerleri elde edilmiştir (Shah vd., 2018).

Bonnett yaptığı bir çalışmada Lynks çevrimiçi alışveriş sitesinde satılan ürünler için ürün açıklamaları ve ürün resimleri bilgilerini kullanarak bir sınıflandırma modeli geliştirmiştir. Giyim ve moda ürünleri kategorileri için VGG sinir ağı yöntemi kullanılmıştır. Sadece ürün açıklamaları kullanılarak yüzde 85 oranıyla eşleştirme gerçekleştirilmiştir. Ürün resimlerini eşleştirmek için ImageNet eğitilmiş veri seti kullanılmıştır. Bu veri seti için öğrenme transferi yöntemi kullanılarak yapılan eşleştirmede yüzde 85 doğruluk oranı elde edilmiştir. Son aşamada hem ürün açıklamaları hem de ürün resimleri veri seti olarak kullanılmış ve yüzde 93 doğruluk oranı elde edilmiştir (Bonnett, 2016).

More yaptığı bir çalışmada Walmart çevrimiçi alışveriş sitesinde bulunan ürün bilgileri ile benzer ürünler için eşleştirme işlemini gerçekleştirmiştir. Ürün eşleşmesi için ürün isimleri, ürün özellikleri, ürün fiyatları ve ürün resimleri kullanılmıştır. Ürün isimleri ve özellikleri kullanılarak yapılan eşleştirme için tekrarlayan sinir ağları (Long Short Term Memory - LSTM) ve evrimsel sinir ağları (Convolutional Neural Network - CNN) yöntemleri kullanılmıştır. Bu eşleştirme ile Walmart sitesindeki ürünler yüzde 99 doğruluk oranıyla eşleşebilmiştir. Ürün resimleri için Residual Networks (ResNet), Visual Geometry Group (VGG) ve Inception sinir ağları kullanılmıştır. Aynı sitede farklı satıcıların aynı ürün için sunduğu farklı açılardan, farklı yakınlaştırmalardan ve farklı renk tonlarından resimleri bulunmaktadır. Bu şekilde farklı sunulan resimler için elde edilen doğruluk oranları farklı olmuştur. Farklı

yakınlaştırmaları olan resimler için ResNet sinir ağı ile yüzde 99, farklı renk tonları ve farklı açıları olan ürünler için Inception sinir ağı ile yüzde 99, farklı renk tonları olan resimler için VGG sinir ağı ile yüzde 88 gibi en iyi doğruluk oranları elde edilmiştir. Ürün fiyatları üzerinden yapılan eşleştirme için Grubs' Test yöntemi kullanılmıştır. Bu yöntem ile fiyat aralığını aşan ürünler, eşleştirme dışında tutulmuştur (More, 2017).

Ürün eşleştirme için yapılan büyük veri çalışmalarının literatür karşılaştırmaları Tablo 2’de gösterilmiştir.

Tablo 2. Ürün Eşleştirme için Yapılan Çalışmaların Karşılaştırılması

Yazarlar	Konu	Özellikler	Yöntem	Veri Seti	Hassasiyet
Ghani ve diğerleri	Ürün özellik çıkarımı için metin madenciliği	Çevrimiçi sitelerden toplanan ürün metin bilgileriyle ürün nitelik değerini çıkararak sınıflandırma işlemi	Naive Bayes, Expectation-Maximization	Çevrimiçi site verileri	87
Kannan ve diğerleri	Yapılandırılmamış ürün tekliflerini yapısal ürün özelliklerine uygun hale getirmek	Arama motoru üzerinden alınan tekliflerin hedef ürünlere uyacak şekilde uygulanması	LVMM	Bing arama motoru, 54 kategoriden oluşan 20 bin ürün	85
Kannan ve diğerleri	Görüntüleri Kullanarak Ürün Sınıflandırmasını İyileştirme	Ürün özellikleri ve ürün resimlerini kullanarak kategorilerin oluşturulması	NLP, ANN	28015 ürün veri seti	
Köpcke ve diğerleri	Eşleşen ürün teklifleri için varlık çözümü tasarlama	Çevrimiçi sitelerde bulunan elektronik ürünlerin eşleştirmelerini iyileştirmek	SVM, TF-IDF, Trigram, Jaccard	Bir e-ticaret sitesi tarafından sağlanan yaklaşık 100 bin ürün	85
Zahavy ve diğerleri	E-ticarette Ürün Sınıflandırmasına Yönelik Çok Amaçlı Bir Füzyon Mimarisi	ürün özellikleri ve ürün resimlerini kullanarak eşleştirmenin gerçekleştirilmesi	Decision Level Fusion	Wallmart alışveriş sitesi verileri	
Cevahir ve Murakami	Bir E-Ticaret Devri İçin Büyük Ölçekli Çok Sınıflı ve Hiyerarşik Ürün Sınıflandırması	Web tarama verileriyle ürünün kategorisini tahmin eden bir model	Deep Belief Nets, Deep Auto encoders	Çevrimiçi alışveriş sitesi verileri (150 milyon ürün)	81
Ristoskia ve diğerleri	Ürün Eşleştirme ve Sınıflandırma için Makine Öğrenmesi Yaklaşımı	Farklı çevrimiçi alışveriş sitelerinin ürünlerinin eşleştirilmesi	CNN, Doğal Dil İşleme (Natural Language Processing - NLP), Deep Neural Nets Image Embeddings	Çevrimiçi alışveriş site verileri	
Xia ve diğerleri	Sinir Dikkat Modelleri Kullanarak Japon Ürün Başlıklarının Büyük Ölçekli Sınıflandırılması	Çevrimiçi alışveriş sitesi ürünleri için 35 ana kategoriye sınıflandırma	ACNN	Rakuten çevrimiçi sitesinde 44000 satıcı için yaklaşık olarak 239 milyon	96

Tablo 2 (devam). Ürün Eşleştirme için Yapılan Çalışmaların Karşılaştırılması

Yazarlar	Konu	Özellikler	Yöntem	Veri Seti	Hassasiyet
Roels ve Loon	Tüketici fiyat endeksi için web tarama, çevrimiçi veri toplama ve işleme	Ürün eşleştirme sonucunda TÜFE hesaplanması	KNN, SVM, Naive Bayes, Random Forests	Çevrimiçi alışveriş sitesi verileri (elektronik eşya, ayakkabı, otel, 2.el araba, öğrenci pansiyonları)	86
Shah ve diğerleri	Ürün Eşleştirme için Sinir Ağı Tabanlı Aşırı Sınıflandırma ve Benzerlik Modelleri	Çevrimiçi alışveriş sitede bulunan ürün bilgileriyle eşleştirme	Sığ Sinir Ağı-FastText(Shallow Neural Network), Siyam Ağları(Siamese Networks)	eBay çevrimiçi alışveriş sitesi verileri	97.1
Bonnett	E-ticaret ürünlerinin görüntülere ve metne göre sınıflandırılması	Çevrimiçi alışveriş sitede bulunan ürünler için ürün açıklamaları ve ürün resimleriyle sınıflandırma	VGG	Lynks çevrimiçi alışveriş sitesi verileri, ImageNet verileri	93
More	Derin öğrenmeyi kullanarak e-Ticarette Ürün Eşleştirme	Çevrimiçi alışveriş sitede bulunan ürün bilgileriyle eşleştirme	Tekrarlayan sinir ağları (LSTM), CNN, Residual Networks(ResNet), VGG, Grubs' Test	Wallmart çevrimiçi alışveriş sitesi verileri	99

4. Sonuçlar

Bu makalede TÜFE için web tarama ve yapay zeka ile ilgili çalışmalar ve farklı yöntemlerle elde edilen ürünlerin sınıflandırılmaları ile ilgili çalışmalar ele alınmıştır. Web tarama verileri istatistik ofisleri için büyük bir veri kaynağı olmuştur. Genel olarak veri toplama maliyetini azaltan bir yöntem olarak ele alınmıştır. Ancak web tarama verileri, veri işleme hacmini artırmaktadır. Çünkü veri toplama işlemi günlük olarak çevrimiçi marketlerden toplanarak veri tabanlarına kaydedilmektedir. Ve bu veri toplama yöntemi her çevrimiçi markete özel olarak yapılmaktadır. Bu veriler çevrimiçi marketlerde farklı olduğu için çeşitliliği ve karmaşıklığı fazladır. Buna bağlı olarak istatistik ofislerinin bu verileri işlemesi için yüksek işlemcili ve yüksek kapasiteli bilgisayar ortamlarına ihtiyacı vardır. Bu şartlar oluştuğunda web tarama verileri daha faydalı hale gelmektedir. İstatistik ofisleri web tarama verilerinin anketlerin bir kısmını karşılayabilecek şekilde işlemeye çalışmaktadır. Bu süreçte web tarama ile elde edilen tüketici ürün bilgileri karmaşık ve çeşitli olduğundan dolayı ürünlerin düzenli hale getirilmesine ihtiyaç vardır. Ürünlerin düzenli bir şekilde sınıflandırılması sonucunda resmi istatistik ofisleri TÜFE hesaplayabilmektedir. Ürünlerin sınıflandırılması yapıldıktan sonra ayrıca bu sınıflandırmaların TÜFE ürün kategorilerine uygun hale getirilmesi gerekmektedir. Bu durum ürün eşleştirmenin önemini ortaya koymaktadır. Büyük veri kapsamında TÜFE işlemleri ile ürün sınıflandırma işlemleri birleştirildiğinde resmi istatistik ofisleri için faydalı bir sonuca dönüşmektedir. Yapılan literatür çalışmaları ile genel bir tasarım oluşturulduğunda web tarama verilerinin elde edilmesinden ürün eşleştirmeye doğru işlem adımlarının uygulanmasının ihtiyaç olduğu ortaya çıkmaktadır.

Yapılan literatür çalışmalarına göre tasarlanacak işlem adımları özet olarak şu şekilde ele alınabilir;

- Çevrimiçi marketlerden ürün bilgilerinin web tarama ile elde edilmesi
- Elde edilen ürün bilgilerinin sınıflandırılması
- Sınıflandırılan ürünlerin TÜFE için belirlenmiş Uluslararası amaca yönelik Kişisel Tüketim sınıflamasına eşleştirilmesi
- Resmi olarak belirlenmiş tüketim ürünlerinin harcama ağırlıklandırmaları kullanılarak yapılan çalışmalarla elde edilmiş ürün sınıflaması ile günlük veya aylık periyotta olacak şekilde TÜFE'nin hesaplanması

Bu adımlardan en zorlu olanı 2.adımda gösterilen ürünlerin sınıflandırılmasıdır. Çünkü yapılan araştırmalarda ürün sınıflandırma için makine öğrenme, yapay zeka ve doğal dil işleme gibi yöntemler kullanılmıştır. Bu yöntemlere ihtiyaç duyulmasının temel nedeni aynı ürünlerin farklı çevrimiçi marketlerde farklı şekilde sunulabilmesidir. Ya da aynı ürünlerin farklı varyantları olabilmektedir. Ürün

eşleştirme başarısı TÜFE göstergesi için önemli bir adımdır. Çünkü istatistik ofisleri TÜFE'yi hesaplarken tüketici ürünlerinin ana kategorileri(COICOP) ile ve bu kategorilere bağlı olan ürün sınıfları üzerinde ölçümü gerçekleştirmektedir. Bu durumda web tarama ürünlerinin sınıflandırılması ve bu sınıflandırılmış ürünlerin TÜFE için resmi olarak belirlenen ürün sınıflandırılmalarına eşleştirilmesi önemli bir adım olarak görülebilir. Eşleştirme performansına bağlı olarak TÜFE ölçümü daha kısa sürede elde edilebilecektir.

Referanslar

Fiyat Endekleri ve Enflasyon, Sorularla Resmi İstatistik Dizisi-3, TÜİK, 2008

United Nations. (2018). Classification of Individual Consumption According to Purpose (COICOP)

<https://data.tuik.gov.tr/Bulten/Index?p=Tuketici-Fiyat-Endeksi-Ocak-2019-30849>, "Tüketici Fiyat Endeksi, Ocak 2019", (2019)

Polzonetti, A., Re, B., & Vaccari, C. Big Data in Official Statistics 2013-2014. Available on: http://www.academia.edu/7571682/PhD_Thesis_on_Big_Data_in_Official_Statistics_, 26-43.

Olston, C., & Najork, M. (2010). Web crawling. Now Publishers Inc.

Cavallo, A. (2013). Online and official price indexes: Measuring Argentina's inflation. *Journal of Monetary Economics*, 60(2), 152-165.

Breton, R., Flower, T., Mayhew, M., Metcalfe, E., Milliken, N., Payne, C., ... & Woods, A. (2016). Research indices using web scraped data: May 2016 update. Newport: Office for National Statistics. Available from www.ons.gov.uk/file.

Chuanyang, F., & Hao, J. L. W. (2016). Experiences with the Use of Online Prices in Consumer Price Index.

Cavallo, A., & Rigobon, R. (2016). The Billion Prices Project. *Journal of Economic Perspectives*. Forthcoming.

Hang, C., Yi, S., Xin, Y., & Benfu, L. (2013, July). A study on construction of inflation index based on web search data. In 2013 International Conference on e-Business (ICE-B) (pp. 1-8). IEEE.

Yuan, H., Zhang, D., Xu, W., Wang, M., & Dong, W. (2013, November). Forecasting the CPI Using a Hybrid Sarima and Neural Network Model with Web News Articles. In 2013 Sixth International Conference on Business Intelligence and Financial Engineering (pp. 84-88). IEEE.

Griffioen, R., de Haan, J., & Willenborg, L. (2014, May). Collecting clothing data from the Internet. In Proceedings of Meeting of the Group of Experts on Consumer Price Indexes, May (Vol. 2628).

Dubey, S., & Gennari, P. (2014). Now-casting food consumer price indexes with big data: Public-private complementarities. FAO Working Paper.

Manik, D. P. (2015, November). A strategy to create daily Consumer Price Index by using big data in Statistics Indonesia. In 2015 International Conference on Information Technology Systems and Innovation (ICITSI) (pp. 1-5). IEEE.

Swier, N. (2015). Using Web Scraped Data to Construct Consumer Price Indices, referat wygłoszony na konferencji NTTS 2015, 10-12 marca 2015, Bruksela.

Nygaard, R. (2015). The use of online prices in the Norwegian Consumer Price Index. Statistics Norway.

Polidoro, F., Giannini, R., Conte, R. L., Mosca, S., & Rossetti, F. (2015). Web scraping techniques to collect data on consumer electronics and airfares for Italian HICP compilation. *Statistical Journal of the IAOS*, 31(2), 165-176.

Barcaroli, G., Scannapieco, M., Scarno, M., & Summa, D. (2015). Using internet as a data source for official statistics: a comparative analysis of web scraping technologies. In Proceedings of Proceedings of the New Techniques and Technologies for Statistics Conference (NTTS).

Barcaroli, G., Scannapieco, M., & Summa, D. (2016). On the use of internet as a data source for official statistics: a strategy for identifying enterprises on the web. *Rivista italiana di economia, demografia e statistica*, 70(4), 20-41.

Cavallo, A., & Rigobon, R. (2016). The billion prices project: Using online prices for measurement and research. *Journal of Economic Perspectives*, 30(2), 151-78.

Gabrielli, L., Riccardi, G., & Pappalardo, L. Using retail market Big Data to nowcast Customer Price Index.

Auer, J., & Boettcher, I. (2016). From price collection to price data analytics. How new large data sources require price statisticians to re-think their index compilation procedures. Experiences from web-scraped and scanner data. Statistics Austria.

Metcalfe, E., Flower, T., Lewis, T., Mayhew, M., & Rowland, M. (2017, May). Research indices using web scraped price data: clustering large datasets into price indices (CLIP). In th meeting of the Ottawa Group (pp. 10-12).

Thakur, G. S. M., Bhattacharyya, R., & Mondal, S. S. (2016). Artificial neural network based model for forecasting of inflation in India. Fuzzy Information and Engineering, 8(1), 87-100.

Griffioen, A. R., & Ten Bosch, O. (2016, May). On the use of internet data for the Dutch CPI. In Statistics Netherlands. Paper presented at the UNECE/ILO Meeting of the Group of Experts on Consumer Price Indices, Geneva, May (pp. 2-4).

Struijs, P., Consten, A., Daas, P., Debusschere, M., Ilves, M., Nikic, B., ... & Swier, N. (2017, May). The ESSnet Big Data: Experimental Results. In STATISTICS AND DATA SCIENCE: NEW CHALLENGES, NEW GENERATIONS.

Nugroho, S. M. S., Budiastuti, I. A., & Hariadi, M. (2017, August). Predicting daily consumer price index using support vector regression method based cloud computing. In 2017 International Seminar on Intelligent Technology and Its Applications (ISITIA) (pp. 313-318). IEEE.

Powell, B. J., Nason, G., Elliott, D., Mayhew, M., Davies, J., & Winton, J. (2018). Tracking and modelling prices using web-scraped price microdata: towards automated daily consumer price index forecasting. Journal of the Royal Statistical Society: Series A (Statistics in Society), 737-756.

Sutiawan, N. S., & Nugraha, I. B. (2017, October). Online price prediction system of consumption commodities. In 2017 International Conference on Information Technology Systems and Innovation (ICITSI) (pp. 145-150). IEEE.

Aparicio, D., & Bertolotto, M. I. (2020). Forecasting inflation with online prices. International Journal of Forecasting, 36(2), 232-247.

Huang, N., Wimalaratne, W., & Pollard, B. (2017). The Effects of the Frequency and Implementation Lag of Basket Updates on the Canadian CPI. Journal of Official Statistics, 33(4), 979-1004.

Bentley, A., & Krsinich, F. (2017). Towards a big data CPI for New Zealand. In 15th Meeting of the Ottawa Group on Price Indices, Eltville am Rhein.

Cavallo, A. (2017). Are online and offline prices similar? Evidence from large multi-channel retailers. American Economic Review, 107(1), 283-303.

Hull, I., Löf, M., Tibblin, M., & Riksbank, S. (2017). Price information collected online and short-term inflation forecasts. In Proc. IFC-Bank Indonesia Satellite Seminar on " Big Data" at the ISI Regional Statistics Conference.

Zhou, Y. (2017). Modelling Swedish Inflation Using Market Data.

<https://www.equifax.com/business/analytic-dataset/>, "Analytic Dataset", (2018)

Whitaker, S. D. (2018). Big data versus a survey. The Quarterly Review of Economics and Finance, 67, 285-296.

Harchaoui, T. M., & Janssen, R. V. (2018). How can big data enhance the timeliness of official statistics?: The case of the US consumer price index. International Journal of Forecasting, 34(2), 225-234.

Cavallo, A. (2018). Scraped data and sticky prices. Review of Economics and Statistics, 100(1), 105-119.

Abe, N., & Shinozaki, K. (2018). Compilation of Experimental Price Indices Using Big Data and Machine Learning: A Comparative Analysis and Validity Verification of Quality Adjustments (No. 18-E-13). Bank of Japan.

Cavallo, A., Diewert, W. E., Feenstra, R. C., Inklaar, R., & Timmer, M. P. (2018, May). Using online prices for measuring real consumption across countries. In AEA Papers and Proceedings (Vol. 108, pp. 483-87).

- ten Bosch, O., Windmeijer, D., van Delden, A., & van den Heuvel, G. (2018, October). Web scraping meets survey design: combining forces. In Big Data Meets Survey Science Conference, Barcelona, Spain.
- Ghani, R., Probst, K., Liu, Y., Krema, M., & Fano, A. (2006). Text mining for product attribute extraction. ACM SIGKDD Explorations Newsletter, 8(1), 41-48.
- Kannan, A., Givoni, I. E., Agrawal, R., & Fuxman, A. (2011, August). Matching unstructured product offers to structured product specifications. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 404-412).
- Kannan, A., Talukdar, P. P., Rasiwasia, N., & Ke, Q. (2011, December). Improving product classification using images. In 2011 IEEE 11th International Conference on Data Mining (pp. 310-319). IEEE.
- Köpcke, H., Thor, A., Thomas, S., & Rahm, E. (2012, March). Tailoring entity resolution for matching product offers. In Proceedings of the 15th International Conference on Extending Database Technology (pp. 545-550).
- Zahavy, T., Magnani, A., Krishnan, A., & Mannor, S. (2016). Is a picture worth a thousand words? a deep multi-modal fusion architecture for product classification in e-commerce. arXiv preprint arXiv:1611.09534.
- Cevahir, A., & Murakami, K. (2016, December). Large-scale Multi-class and Hierarchical Product Categorization for an E-commerce Giant. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers (pp. 525-535).
- Ristoski, P., Petrovski, P., Mika, P., & Paulheim, H. (2018). A machine learning approach for product matching and categorization. Semantic web, 9(5), 707-728.
- Xia, Y., Levine, A., Das, P., Di Fabrizio, G., Shinzato, K., & Datta, A. (2017, April). Large-scale categorization of japanese product titles using neural attention models. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers (pp. 663-668).
- Loon, K.V., & Roels D., (2018). Le webscraping, la collecte et le traitement de données en ligne pour l'indice des prix à la consommation.
- Shah, K., Kopru, S., & Ruvini, J. D. (2018, June). Neural network based extreme classification and similarity models for product matching. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 3 (Industry Papers) (pp. 8-15).
- Bonnett, C. (2016). Classifying e-commerce products based on images and text. Adventures in Machine Learning.
- More, A. (2017). Product matching in ecommerce using deep learning.