



Long short-term memory network based deep transfer learning approach for sales forecasting

Begüm Erol^{1,2*}, Tülin İnkaya¹

¹Department of Industrial Engineering, Faculty of Engineering, Bursa Uludag University, Gorukle Campus, 16059, Nilufer, Bursa, Türkiye

²Department of Industrial Engineering, Faculty of Engineering, Beykent University, 34398, Ayazaga Campus, Sariyer, Istanbul, Türkiye

Highlights:

- A deep transfer learning approach is proposed for sales forecasting.
- A similarity-based transferability criterion is used to select the most appropriate source for the target data.
- The proposed approach provides increase in forecasting accuracy and decrease in the training time.

Keywords:

- Long short-term memory,
- sales forecasting,
- transfer learning,
- source selection,
- edit distance with real penalty

Graphical/Tabular Abstract

The general flow chart of the proposed approach is shown in Figure A.

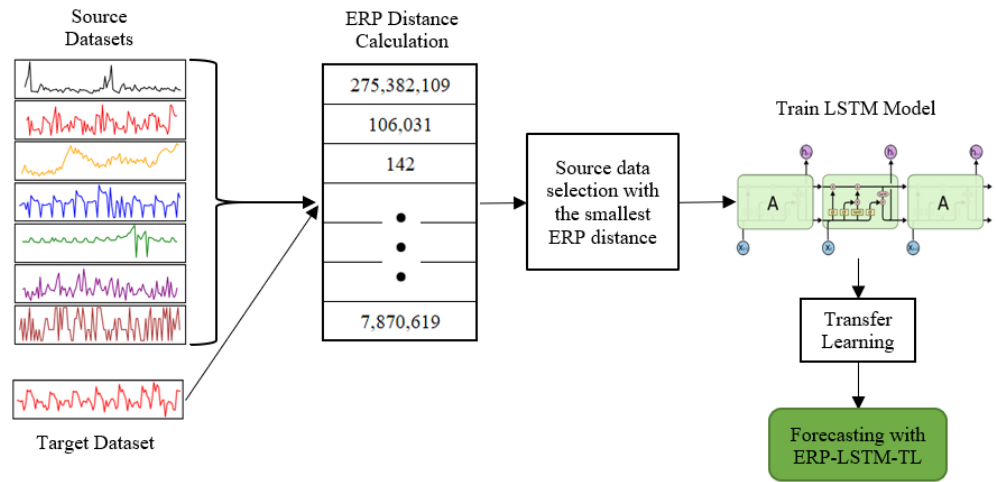


Figure A. General flow chart of the proposed forecasting model

Article Info:

Research Article

Received: 17.03.2022

Accepted: 15.01.2023

DOI:

10.17341/gazimmfd.1089173

Acknowledgement:

This research was supported by Bursa Uludag University Scientific Research Projects Unit (Project Code: FDK-2021-518).

Correspondence:

Author: Begüm Erol

e-mail:

begumerol@beykent.edu.tr

phone: +90 212 444 1997

Purpose:

The aim of this study is to increase the forecasting accuracy and reduce the computational cost of the deep learning models for sales forecasting. For this purpose, a long short-term memory (LSTM) based deep transfer learning approach is proposed.

Theory and Methods:

Deep transfer learning enables the transfer of the knowledge acquired in a source domain and task to a target domain and task. In the proposed approach, source selection is performed according to the similarities between the source and target sales datasets, and edit distance with real penalty (ERP) is adopted for this purpose. The most similar source dataset is used for training the LSTM network, which allows extracting the temporal dependencies within the dataset. After the parameter transfer, the LSTM network is re-trained with the target dataset. Eventually, the proposed ERP-LSTM-TL model is obtained for sales forecasting.

Results:

Experiments with various sales datasets showed that transfer learning improved the forecasting accuracy in 38 out of 46 source and target dataset combinations. On the other hand, negative transfer learning was observed in the remaining eight combinations. The proposed ERP-LSTM-TL method prevented the negative transfer in all target datasets. Also, it yielded superior performance compared to the traditional forecasting and machine learning methods, and reduced the training time of the deep learning models.

Conclusion:

Experimental results showed the effectiveness of ERP-LSTM-TL in sales forecasting for different products and different sectors. Manufacturers, retailers and distributor companies can obtain cost and time savings using the proposed approach.



Satış tahmini için uzun kısa-sürelili bellek ağı tabanlı derin transfer öğrenme yaklaşımı

Begüm Erol^{1,2*}, Tülin İnkaya¹

¹Bursa Uludağ Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, Görükle Kampüsü, 16059, Nilüfer, Bursa, Türkiye

²Beykent Üniversitesi, Mühendislik-Mimarlık Fakültesi, Endüstri Mühendisliği Bölümü, Ayazağa Kampüsü, 34398, Sarıyer, İstanbul, Türkiye

Ö N E Ç I K A N L A R

- Satış tahmini için derin transfer öğrenme yaklaşımı önerilmiştir
- Hedef veriye en uygun kaynak seçimi için benzerlik tabanlı bir aktarılabirlik ölçütü kullanılmıştır
- Önerilen yaklaşım, tahmin doğruluğunda artış sağlamıştır ve eğitim süresini kısaltmıştır

Makale Bilgileri

Araştırma Makalesi

Geliş: 17.03.2022

Kabul: 15.01.2023

DOI:

10.17341/gazimmfd.1089173

Anahtar Kelimeler:

Uzun kısa-sürelili bellek (UKSB), satış tahmini, transfer öğrenme, kaynak seçimi, gerçek cezalı düzenleme uzaklığı

ÖZ

Üretim ve hizmet sektörlerinde faaliyet gösteren firmalar, artan rekabet koşulları ile mücadele edebilmek için belirsizlik altında geleceğe yönelik çeşitli kararlar alırlar. Bu kritik kararlardan biri satış tahminidir. Dijital teknolojilerin yaygınlaşmasıyla derin öğrenme yaklaşımlarının satış tahmininde kullanımı artmaktadır. Derin öğrenme, başarılı sonuçlar vermesine rağmen büyük miktarda veriyle uzun eğitim sürelerine ihtiyaç duymaktadır. Bu duruma çözüm olarak problemler arası bilgi aktarımını sağlayan transfer öğrenme (TÖ) kullanılmaktadır. Transfer öğrenme, kaynak veriler ile modelin eğitimini ve hedef veriye aktarımını sağlamaktadır. Bu çalışmada, farklı ürünlerin satış tahmini modellerinden elde edilen bilginin gelecekteki tahmin modellerine aktarımını sağlamak üzere derin transfer öğrenme yaklaşımı önerilmiştir. Satış verisi tek değişkenli zaman serisi olarak ele alınmıştır. Kaynak veri seçiminde aktarılabirlik ölçütü olarak hedef ve kaynak veri arasındaki gerçek cezalı düzenleme uzaklığı (GCDU) kullanılmıştır. Seçilen kaynak veriyle zamansal bağımlılıkların modellenmesini sağlayan uzun kısa-sürelili bellek (UKSB) ağı eğitilmiştir. Ön eğitilen UKSB ağına parametre transferi yapılarak hedef veri için GCDU-UKSB-TÖ tahmin modeli oluşturulmuştur. Çeşitli sektörlerdeki satış verileri ile yapılan deneysel çalışmalarda GCDU-UKSB-TÖ, hedef veriyle eğitilen UKSB'ye kıyasla ortalama karesel hatanın karekökü ve ortalama mutlak hata ölçütlerine göre sırasıyla ortalama %5,77 ve %7,42 iyileşme sağlamıştır, ayrıca eğitim süresini ortalama %47,90 kısaltmıştır. GCDU-UKSB-TÖ, klasik tahmin ve makine öğrenmesi yöntemlerinden istatistiksel olarak daha iyi sonuç vermiştir.

Long short-term memory network based deep transfer learning approach for sales forecasting

H I G H L I G H T S

- A deep transfer learning approach is proposed for sales forecasting
- A similarity-based transferability criterion is used to select the most appropriate source for the target data
- The proposed approach provides increase in the forecasting accuracy and decrease in the training time

Article Info

Research Article

Received: 17.03.2022

Accepted: 15.01.2023

DOI:

10.17341/gazimmfd.1089173

Keywords:

Long short-term memory (LSTM), sales forecasting, transfer learning, source selection, edit distance with real penalty

ABSTRACT

Firms that operate in the production and service sectors take various decisions for the future under uncertainty in order to combat the increasing competitive conditions. One of these critical decisions is sales forecasting. With the spread of digital technologies, the use of deep learning approaches in sales forecasting is increasing. Although deep learning gives successful results, it needs long training time with large amounts of data. As a solution to this situation, transfer learning (TL), which provides information transfer between problems, is used. Transfer learning provides the training of the model with source data and its transfer to the target data. In this study, a deep transfer learning approach is proposed to transfer the information obtained from the sales forecasting models of different products to future forecasting models. Sales data are considered as univariate time series. The edit distance with real penalty (ERP) between the target and source data is used as a measure of transferability in the selection of source data. A long short-term memory (LSTM) network has been trained, which enables modeling of temporal dependencies with the selected source data. ERP-LSTM-TL forecasting model is created for the target data by transferring parameters from the pre-trained LSTM network. In the experimental studies on sales data from various industries, compared to LSTM trained with target data, ERP-LSTM-TL provided average 5.77% and 7.42% improvements in terms of root mean squared error and mean absolute error measures, respectively, also it shortened training time by average 47.90%. ERP-LSTM-TL gave statistically better results compared to classical forecasting and machine learning methods.

*Sorumlu Yazar/Yazarlar / Corresponding Author/Authors : *begumerol@beykent.edu.tr, tinkaya@uludag.edu.tr / Tel: +90 212 444 1997

1. Giriş (Introduction)

Satış tahmini, globalleşen tedarik zincirlerinde malzeme, ürün, para ve bilginin yönetiminde önemli unsurlardan biridir [1]. Doğru satış tahmini firmaların kapasite ile talebi dengelemesini sağlar. Doğru zamanda, doğru ürünün, doğru miktarda temini hem müşteri memnuniyetini hem de firmaların kârlılığını ve verimliliğini artırır. Diğer yandan artan rekabet, yüksek ürün çeşitliliği, afetler ve salgınlar nedeniyle yaşanan belirsizlikler satış tahminini zorlu bir problem haline getirmektedir.

Son yıllarda büyük veri, bulut bilişim ve hesaplama gücü yüksek donanımların gelişmesi ile derin öğrenme yöntemlerinin satış tahmininde kullanımı artmıştır [2]. Derin öğrenme yöntemleri çok sayıda katman ve nöron içeren, nöronlar arasında karmaşık bağlantıların bulunduğu derin sinirsel ağlardır. Bu yöntemler, satış tahminindeki karmaşık ve doğrusal olmayan örüntülerin modellenmesini sağlamaktadır. Özellikle günümüzde derin öğrenme uygulamaları için geliştirilen çok sayıda çerçeve ve kütüphane (TensorFlow, Knet, Torch, vb.) bulunmaktadır [3]. Ancak, derin öğrenme modelleri ile etkili sonuç elde etmek için büyük miktarda geçmiş veriye ve uzun eğitim sürelerine ihtiyaç duyulmaktadır. Ayrıca satış tahmini modellerinden elde edilen bilgi birikimi yeni geliştirilen modellere sistematik bir şekilde aktarılamamaktadır. Bu durumlara çözüm olarak farklı problemler arasında bilgi aktarımını sağlayan derin transfer öğrenme yaklaşımı öne çıkmaktadır [4].

Transfer öğrenme, kaynak bir alanda ve görevde elde edinilen bilginin hedef bir alana ve göreve aktarılmasını sağlayan bir makine öğrenmesi yöntemidir [5]. Kullanılan tahmin modeli derin sinirsel ağ olduğunda bu bilgi aktarımı derin transfer öğrenme olarak adlandırılmaktadır [6]. Derin transfer öğrenmenin hedef alanda ihtiyaç duyulan eğitim verisini azaltma ve tahmin modelinin performansını iyileştirme potansiyeli bulunmaktadır [6]. Öte yandan, transfer öğrenme uygulamalarında önemli unsurlardan biri hedef ve kaynak alanlar arasındaki benzerliktir [4, 7]. Kaynak ve hedef alanlar arasında yeterli benzerlik olmadığı takdirde bilgi aktarımı tahmin performansını olumsuz etkileyebilmektedir. Bu durum, negatif öğrenme olarak ifade edilmektedir [4, 7]. Negatif öğrenmeyi önlemek amacıyla hedef ve kaynak alanlar arasında aktarılabilirlik ile kaynak seçimi problemi ele alınmaktadır [8]. Sınıflandırma problemleri için bu alanda çalışmalar [9, 10] bulunmakla birlikte bilindiği kadarıyla tahmin problemleri için çalışma sayısı sınırlıdır [11].

Bu çalışmada, farklı sektörler ve farklı ürünlerin satış tahmini modellerinden edinilen bilginin gelecekteki tahmin modellerine aktarımını sağlamak üzere derin transfer öğrenme yaklaşımı önerilmiştir. Problem tek değişkenli zaman serisi olarak ele alınmıştır ve satış verileri farklı uzunlukta olabilmektedir. Önerilen yaklaşım kapsamında öncelikle hedef veriye en uygun kaynak seçimi için benzerlik tabanlı bir aktarılabilirlik ölçütü kullanılmıştır. Sonrasında, seçilen kaynak veri ile zamansal bağımlılıkların modellenmesini sağlayan uzun kısa-süreli bellek (UKSB) (Long Short-Term Memory - LSTM) ağı eğitilmiştir. Geliştirilen UKSB ağına parametre transferi ile hedef alana bilgi aktarımı gerçekleştirilmiştir. Bu çalışma, farklı sektörler ve farklı uzunluklardaki satış verileri arasında UKSB tabanlı derin transfer öğrenme yaklaşımını sunması itibarıyla literatüre katkı sağlamaktadır. Ayrıca, bilindiği kadarıyla farklı uzunlukta zaman serilerinde aktarılabilirlik ölçütü olarak metrik bir uzunluk olan gerçek cezalı düzenleme uzaklığı (GCDU) (Edit Distance with Real Penalty - ERP) kullanan ilk çalışmadır.

Makalenin geri kalan bölümleri şu şekilde organize edilmiştir: İkinci bölümde konuyla ilgili yapılan literatür araştırmasına yer verilmiştir. Üçüncü bölümde önerilen yaklaşım açıklanmıştır. Dördüncü bölümde

kullanılan veri kümeleri, deneysel koşullar ve sonuçlar karşılaştırılmalı olarak sunulmuştur. Beşinci ve son bölümde ise sonuç ve gelecek çalışmalar verilmiştir.

2. Literatür Özeti (Literature Review)

Satış tahmininde üstel düzeltme, hareketli ortalamalar, doğrusal regresyon, otoregresif bütünlük hareketli ortalama gibi klasik tahmin yöntemlerinin yanı sıra son yıllarda yapay zeka yöntemleri etkin bir şekilde kullanılmaktadır [12]. Özellikle perakende [18, 22, 24], moda [19], otomotiv [14, 15], gıda [16, 20], sağlık [2, 17] ve e-ticaret [19, 21, 23] gibi sektörlerde satış tahmini için yapay zekanın bir alt dalı olan derin öğrenme yöntemleri öne çıkmaktadır. Bu kapsamda derin sinir ağı (DSA) (Deep Neural Network - DNN) [13-15, 24], konvolüsyonel sinir ağı (KSA) (Convolutional Neural Network - CNN) [23], tekrarlayan sinir ağı (TSA) (Recurrent Neural Network - RNN) [22, 23] ve UKSB [2, 16-23] yöntemleri kullanılmıştır. Satış tahmininde derin öğrenme yöntemlerini kullanan bu çalışmalar Tablo 1'de özetlenmiştir. Ayrıca Tablo 1'de Türkiye'deki e-ticaret ve perakende sektörleri için derin öğrenme modelleri ile başarılı uygulamalar yer almaktadır [23, 24].

Derin öğrenme modellerinde eğitim performansını arttırmak amacıyla kullanılan transfer öğrenme; görüntü işleme [25], ses tanıma [26] ve doğal dil işleme [27] alanlarında başarılı sonuçlar vermiştir. Bu motivasyonla, son yıllarda satış tahmini probleminde derin transfer öğrenme yaklaşımı uygulayan çalışmalarda artış görülmektedir. Bu alandaki çalışmaların bir kısmı aynı sektörde farklı ürünlerin ya da farklı bölgelerin satış tahminini ele almaktadır. Bu kapsamda, Zhao vd. [28] farklı bölgelerdeki satış tahmini için KSA geliştirmişlerdir. Her bölgedeki e-ticaret satış verisi ile KSA tekrar eğitilmiştir ve her bölge için tahmin modeli oluşturulmuştur. Pan vd. [29] e-ticaret satış hacmini tahmin etmek için KSA uygulayarak önemli öznitelikleri çıkarmışlardır ve farklı satış bölgeleri arasında bilgi aktarımı ile tahmin doğruluğunun iyileştirilebileceğini göstermişlerdir. Karb vd. [12] yeni ürün satış tahmininde çok katmanlı sinirsel ağlar ile transfer öğrenme yaklaşımını önermişlerdir. Parametre transferi yaparak farklı ağ konfigürasyonlarını incelemişlerdir. Ayrıca ürün ailesi, fiyat ve promosyon gibi ürün özelliklerinin transfer öğrenmeye katkısını ele almışlardır.

Yukarıdaki çalışmalardan farklı olarak Hirt vd. [30] satış tahmininde hedef ve kaynak veri arasında aktarılabilirlik konusunu ele almışlardır ve farklı restoranların satış tahmininde KSA yöntemini kullanmışlardır. Aktarılabilirlik için orijinal veri ile verinin projeksiyonlarına ait dağılımlar arasındaki maksimum ortalama tutarsızlık (maximum mean discrepancy) değerlerini kullanmışlardır. İki sinir ağı arasındaki aktarılabilirlik değerlendirmesini ise tekil değer kanonik korelasyon analizi ile yapmışlardır. Deneysel çalışmalarda önerilen ölçütlerin aktarılabilirlik ile negatif yönlü ve zayıf bir ilişki gösterdiğini tespit etmişlerdir. Ancak yapılan çalışmada bir hedef veri için hangi kaynak verinin seçilmesine ilişkin bir yöntem önerisi yapılmamıştır. Bu çalışmada, satış verileri zaman serisi şeklinde ele alındığı için zaman serileri ve transfer öğrenme kesişiminde yer alan çalışmalar da incelenmiştir. Bu alanda Weber vd. [8] kapsamlı bir derleme sunmuştur. Ayrıca, zaman serilerinde aktarılabilirlik ve kaynak seçimi konusunu ele alan çalışmalar bulunmaktadır [8]. Bu çalışmalar zaman serilerinin sınıflandırılması ve tahmini olmak üzere iki başlıkta incelenebilir. Zaman serilerinin sınıflandırılmasında Fawaz vd. [10] hedefe uygun kaynak verinin seçiminde benzerlik analizini kullanmışlardır. Zaman serileri arasında şekilsel benzerlikleri dinamik zaman bükmesi (DZB) (Dynamic Time Warping - DTW) ile hesaplamışlardır. Kaynak veri seçiminden sonra zaman serilerinin sınıflandırılmasında KSA yöntemi ile transfer öğrenme yaklaşımını uygulamışlardır. Bir diğer çalışmada, Meiseles

Tablo 1. Derin öğrenme yöntemleri ile satış tahmini yapan çalışmalar
(Studies in the literature that make sales forecasting with deep learning methods)

Çalışma	Amaç	Yöntem	Veri Tipi ve Kaynağı
[13]	Moda endüstrisindeki satışları ve gelecek sezon için yeni bireysel ürünlerin satışlarını tahmin etmek	DSA	Bir şirketin geçmiş satış verileri
[14]	Ekonomik değişkenler yardımıyla satış hacmi tahmin modeli geliştirmek	DSA	İki otomobil üreticisinin aylık satış hacimleri
[15]	Farklı katman sayılarına sahip derin sinir ağlarla otomobil satışlarını tahmin etmek	DSA	Bir ülkedeki otomotiv şirketlerinin üretim verileri
[16]	Sebze perakende satışlarında uygun tahmin modelini seçmek	UKSB	Üç farklı sebze için günlük satış verileri
[17]	Tek değişkenli ve çok değişkenli satış tahmini yapmak	UKSB	İlaç satan bir süpermarketin satış verileri
[18]	Bir perakende tedarik zincirinin tüm seviyelerinde tutarlı tahminler oluşturmak	UKSB	Çok kanallı bir perakendeciye ait satış verileri
[2]	Eczanelerin gün öncesi satış hacmini tahmin etmek	UKSB	1113 farklı eczaneye ait satış verileri
[19]	E-ticaret satışlarını UKSB ile tahmin etmek ve UKSB performansını değerlendirmek	UKSB	Profesyonel sporcular için dış mekan malzemeleri satan bir e-ticaret firmasının satış verileri
[20]	Bozulabilir ürünlerin pişirme planını oluşturabilmek için saatlik satış tahminlerini yapmak	UKSB	Dokuz mağazadan 14 farklı unlu mamulün satış verileri
[21]	Parametre optimizasyonuna dayalı bir UKSB modeli ile satış tahmini yapmak	UKSB	E-ticaret, ilaç ve araba satış verileri
[22]	Bir perakende zincirinde bozulabilir ve dayanıklı ürünler için talep tahmini yapmak	TSA, UKSB	Süpermarketlere ait satış veri kümeleri
[23]	E-perakende sektöründe belirli bir kategorideki ürünlere olan talebi yapay zeka algoritmaları ile tahmin etmek	KSA, TSA, UKSB, Derin Otokodlayıcı, Çift Yönlü UKSB	Türkiye'deki yerel bir süpermarkete ait internet ve mobil satış verileri
[24]	Derin öğrenmeye dayalı topluluk tabanlı talep tahmini sistemi geliştirmek	DSA	Türkiye'deki yerel bir süpermarkete ait satış verileri

ve Rokach [9] zaman serilerinin sınıflandırılmasında kaynak model seçimini incelemiştir. Önerdikleri yaklaşımda hedef veri, tam konvolüsyonel ağlardan (TKA) (Fully Convolutional Network - FCN) oluşan kaynak modeller ile kodlanmıştır. Bu kodlayıcı uzayında ortalama Silüet katsayısı ile hedef veri için kaynak model seçimi yapılmıştır. Zaman serilerinin tahmininde ise Ye ve Dai [11] eşit uzunluktaki zaman serilerinde kaynak ile hedef veri arasındaki aktarılabirlik ölçütü olarak DZB ve Jensen-Shannon yakınsamasını birlikte kullanmışlardır. Ayrıca zaman serisi tahmini için KSA mimarisine sahip bir transfer öğrenme yaklaşımı önermişlerdir. He vd. [31] zaman serisi tahmini için DSA ve UKSB tabanlı iki farklı transfer öğrenme yaklaşımı ile farklı uzunluktaki zaman serilerinde kaynak seçimi problemini ele almışlardır. Kaynak veri kümelerini seçmek için DZB benzerlik ölçütünü kullanmışlardır.

Sonuç olarak zaman serilerinde kaynak ve hedef veri arasında benzerlik olması transfer öğrenme için fayda sağlamaktadır [10]. Bu nedenle, bu çalışmada farklı sektörler ve farklı ürünler arasında satış tahmini için bilgi aktarımını sağlamak amacıyla benzerlik analizi kullanılmıştır. Bu çalışma, farklı sektörlerdeki ve farklı uzunluktaki satış verileri arasında kaynak seçimi yapması itibarıyla literatürdeki çalışmalardan farklılaşmaktadır. Ayrıca bilindiği kadarıyla derin transfer öğrenmede farklı uzunluktaki zaman serilerinde kaynak seçimi için GCDU'yu kullanan ilk çalışmadır. Bu aktarılabirlik ölçütü ile seçilen kaynak veri, satış ve talep tahmininde zamansal bağımlılıkları yakalaması nedeniyle etkili sonuçlar veren UKSB [2, 16-23] ile eğitilerek transfer öğrenme uygulanmıştır.

3. Yöntem (Method)

Bu bölümde, öncelikle önerilen çözüm yaklaşımında kullanılan UKSB, derin transfer öğrenme ve kaynak seçimi açıklanmıştır. Sonrasında önerilen çözüm yaklaşımının adımları verilmiştir.

3.1. Uzun Kısa-Süreli Bellek (Long Short-Term Memory)

UKSB, tekrarlayan sinir ağlarının (TSA) özel bir versiyonudur [32]. TSA'da kaybolan/patlayan gradyan nedeniyle uzun dönem bağımlılıkların öğrenilmesinde problem görülebilmektedir. Buna karşılık, UKSB'deki hafıza hücreleri sıralı dizideki değerlerin rastgele aralıklarla hatırlanmasını sağlamaktadır [33]. Bu özelliği ile UKSB zaman serilerinin tahmininde başarılı sonuçlar vermektedir.

UKSB mimarisi bloklardan oluşmaktadır. Şekil 1'de bir UKSB bloğu yer almaktadır. UKSB'nin anahtarı hafıza hücresidir. Hafıza hücre durumu, bilgiyi taşıyan ve tutan bir ağ hafızasıdır. UKSB kapıları, hücre durumuna bilgi ekleme veya çıkarma yapılmasını sağlar. Kapılar, sigmoid sinir ağı katmanı ile noktasal çarpma işleminden oluşan ve isteğe bağlı olarak bilgilerin geçmesine izin veren yapılardır. UKSB, hücre durumunu korumak ve kontrol etmek için üç kapıya sahiptir [34]. UKSB ağı matematiksel olarak Eş. 1-Eş. 7 ile ifade edilebilir.

$$f_t = \sigma(W_f \otimes [h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i \otimes [h_{t-1}, x_t] + b_i) \quad (2)$$

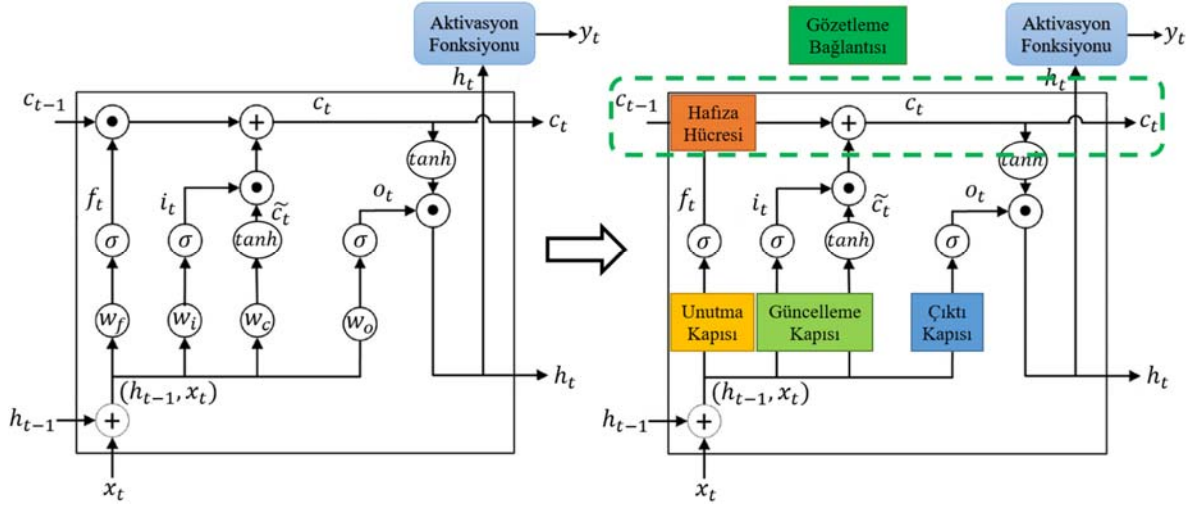
$$\tilde{c}_t = \tanh(W_c \otimes [h_{t-1}, x_t] + b_c) \quad (3)$$

$$c_t = f_t \times c_{t-1} + i_t \times \tilde{c}_t \quad (4)$$

$$o_t = \sigma(W_o \otimes [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \times \tanh(c_t) \quad (6)$$

$$y_t = \sigma(h_t) \quad (7)$$



Şekil 1. UKSB bloğu (LSTM block)

Eş. 1-Eş. 7 denklem setinde t bir zaman adımı; i, f, c ve o sırasıyla girdi kapısını, unutma kapısını, hafıza hücrecini ve çıktı kapısını temsil etmektedir. σ ve \tanh aktivasyon fonksiyonlarını, \otimes eleman bazında çarpımı, \times ise noktasal çarpımı ifade etmektedir. $x_t, y_t, c_t, \tilde{c}_t, f_t, o_t$ ve h_t sırasıyla t zaman adımındaki girdi vektörünü, çıktı vektörünü, hafıza hücreci durumunu, girdi kapısı aday değerini, unutma kapısı aday değerini, çıktı kapısı aday değerini ve gizli katman aday değerini göstermektedir. b_f, b_i, b_c, b_o sapma vektörlerini ve W_f, W_i, W_c, W_o ağırlık matrislerini ifade etmektedir [34].

UKSB'de birinci adımda hücre durumunda hangi bilgilerin unutulacağına karar verilir. Eş. 1'de yer alan ve unutma kapısı olarak adlandırılan sigmoid katman ile bu karar verilir. h_{t-1} ve x_t alınır, ardından c_{t-1} hücre durumunda yer alan her sayı için 0 ile 1 arasında bir sayı verilir. Burada, 1 değeri "bilginin tamamını sakla" ifadesini, 0 değeri ise "bilginin tamamını unut" ifadesini temsil eder [34].

İkinci adımda, hücre durumunda hangi bilgilerin depolanacağına karar verilir. Öncelikle Eş. 2'de girdi kapısı olarak adlandırılan bir sigmoid katman ile hangi değerlerin güncelleneceği belirlenir. Ardından, Eş. 3'te verilen bir \tanh katmanı ile duruma eklenebilecek yeni aday değerleri içeren \tilde{c}_t vektörü oluşturulur. Üçüncü adımda, Eş. 4'te geçmiş hücre durumu c_{t-1} ve yeni hücre durumu \tilde{c}_t sırasıyla unutma ve girdi kapısı değerleriyle çarpılarak birleştirilir, böylece mevcut durum güncellenir. Son adımda Eş. 5-Eş. 7 ile çıktı hücre durumuna bağlı olarak belirlenir [34]. UKSB ağında ileri yayılımda gerçekleştirilen matematiksel işlemler Eş. 1-Eş. 7 denklem setinde gösterilmektedir. UKSB ağına eğitimi için geri yayılım algoritması kullanılır [35]. Geri yayılım algoritması, ağıdaki hatanın (kayıp fonksiyonu) en küçüklenmesi için çıktı katmanından girdi katmanına doğru güncellemeler yapar. Tahmin problemlerinde genellikle kayıp fonksiyonu olarak ortalama karesel hata (Mean Squared Error - MSE) kullanılır. Her zaman adımında kayıp fonksiyonun kısmi türevleri (gradyanları) hesaplanarak her katmandaki ağırlıklar ve sapma değerleri güncellenir. Bu ileri ve geri yönlü güncelleme döngüsü belirli bir durdurma kriteri sağlanana kadar devam eder [35].

3.2. Derin Transfer Öğrenme ve Kaynak Seçimi (Deep Transfer Learning and Source Selection)

Derin transfer öğrenme, verilen bir hedef alan ve hedef görevde bir derin öğrenme modelinin performansını arttırmak için kaynak alan ve kaynak görevde yer alan bilginin keşfedilerek hedefe aktarımını ifade eder [7]. TÖ sürecinde hangi bilginin ne zaman aktarılacağı önem

taşımaktadır. Bu çalışmada, ağ bazlı aktarım ele alınmıştır [7]. Buna göre, ilk olarak kaynak veri ve kaynak görev ile kaynak bir ağ oluşturulur. Sonrasında, kaynak ağda öğrenilen öznelilik, ağırlık, parametre gibi bilgiler hedef veri ve hedef görevde kullanılmak üzere hedef ağa aktarılır.

Kaynak ve hedef veri arasında benzerlik olması transfer öğrenme için fayda sağlamaktadır [9]. Bu nedenle, bu çalışmada aktarılabirlik ve kaynak seçimi için benzerlik analizi kullanılmıştır. Elastik uzaklık ölçütleri farklı uzunluktaki tek değişkenli zaman serileri arasında doğrusal olmayan eşleştirmeler yaparak şekilsel benzerliklerini hesaplar [36]. Bu nedenle, bu çalışmada Chen ve Ng [37] tarafından önerilen GCDU satış verileri arasındaki benzerlik analizinde kullanılmıştır. GCDU uzaklığı, L1 normu ile gerçek sıralı düzenleme uzaklığı [38] birleştirir. Farklı uzunluktaki zaman serilerinde nokta eşleştirme yöntemini kullanır ve zamansal hizalama yaparak seriler arasındaki uzaklığı hesaplar. GCDU, DZB'den farklı olarak metrik bir uzaklıktır ve üçgen eşitsizliğini sağlar.

$a = [a_1, a_2, \dots, a_{T_1}]$ ve $b = [b_1, b_2, \dots, b_{T_2}]$ sırasıyla T_1 ve T_2 uzunluğunda iki zaman serisi olsun. GCDU'nun hesaplanmasında zaman serisi a 'nın i . elemanı ile b 'nin j . elemanı boşluk değil ise gerçek ceza kullanılır, bir eleman boşluk ise sabit bir sayı g kullanılır ve g değerinin 0 seçilmesi önerilmiştir [38]. Buna göre uzaklık formülü Eş. 8'de verilmiştir.

$$dist_{GCDU}(a_i, b_j) = \begin{cases} |a_i - b_j|, & \text{eğer } a_i \text{ ve } b_j \text{ boşluk değil ise;} \\ |a_i - g|, & \text{eğer } b_j \text{ boşluk ise;} \\ |b_j - g|, & \text{eğer } a_i \text{ boşluk ise.} \end{cases} \quad (8)$$

Zaman serileri a ve b arasındaki GCDU değeri $GCDU(a, b)$ ile gösterilir ve Eş. 9'a göre hesaplanır.

$$GCDU(a, b) = \begin{cases} \begin{cases} \sum_1^{T_1} |a_i - g| & \text{eğer } T_2 = 0, \\ \sum_1^{T_2} |b_j - g| & \text{eğer } T_1 = 0, \end{cases} \\ \min \left\{ \begin{array}{l} GCDU(Kalan(a), Kalan(b)) \\ + dist_{GCDU}(a_1, b_1), \\ GCDU(Kalan(a), b) \\ + dist_{GCDU}(a_1, [.]), \\ GCDU(a, Kalan(b)) \\ + dist_{GCDU}(b_1, [.]) \end{array} \right\} & \text{d.d.} \end{cases} \quad (9)$$

Burada, [...] zaman serisinin ilgili elemanın boşluk olduğunu gösterir ve T uzunluğundaki x zaman serisi için $K(x) = Kalan(x) = [x_2, \dots, x_T]$ olarak tanımlanır.

3.3. UKSB Tabanlı Derin Transfer Öğrenme (LSTM Based Deep Transfer Learning)

UKSB tabanlı derin transfer öğrenme yaklaşımı (GCDU-UKSB-TÖ), bir hedef satış verisi için en benzer kaynak satış verisine UKSB ile derin transfer öğrenme uygulayarak tahmin doğruluğunu arttırmayı amaçlamaktadır. Önerilen yaklaşımın genel akış şeması Şekil 2'de verilmiştir. Kaynak veri kümesi $X_K = \{x_1, x_2, \dots, x_n\}$, n adet potansiyel kaynak satış verisinden oluşmaktadır ve l . potansiyel kaynak satış verisinde $x_l = \{x_{lt} | t = 1, 2, \dots, T_l\}$, T_l adet dönem bulunmaktadır. Hedef satış verisi $y_h = \{y_{ht} | t = 1, 2, \dots, T_h\}$, T_h döneme ait satış verisini içermektedir. Buna göre, ilk olarak, potansiyel kaynak ve hedef satış verileri $[0, 1]$ aralığına normalize edilir. Sonrasında, hedef ve potansiyel kaynak verileri arasında GCDU uzaklığı hesaplanır ve hedef veriye en küçük uzaklığa sahip (en benzer) satış verisi kaynak veri olarak seçilir. Ardından, seçilen kaynak veri ile kaynak UKSB ağı eğitilir. Son olarak, parametre transferi uygulanır ve ön eğitilmiş kaynak UKSB ağı hedef veri ile tekrar eğitilir. Böylece, hedef veri için UKSB tahmin modeli elde edilir. GCDU-UKSB-TÖ algoritması Tablo 2'de verilmiştir.

4. Deneysel Çalışmalar (Experimental Studies)

4.1. Veri Kümeleri ve Performans Kriterleri (Datasets and Performance Criteria)

Bu çalışmada, önerilen derin transfer öğrenme yaklaşımının performansını değerlendirmek için Tablo 3'te yer alan farklı

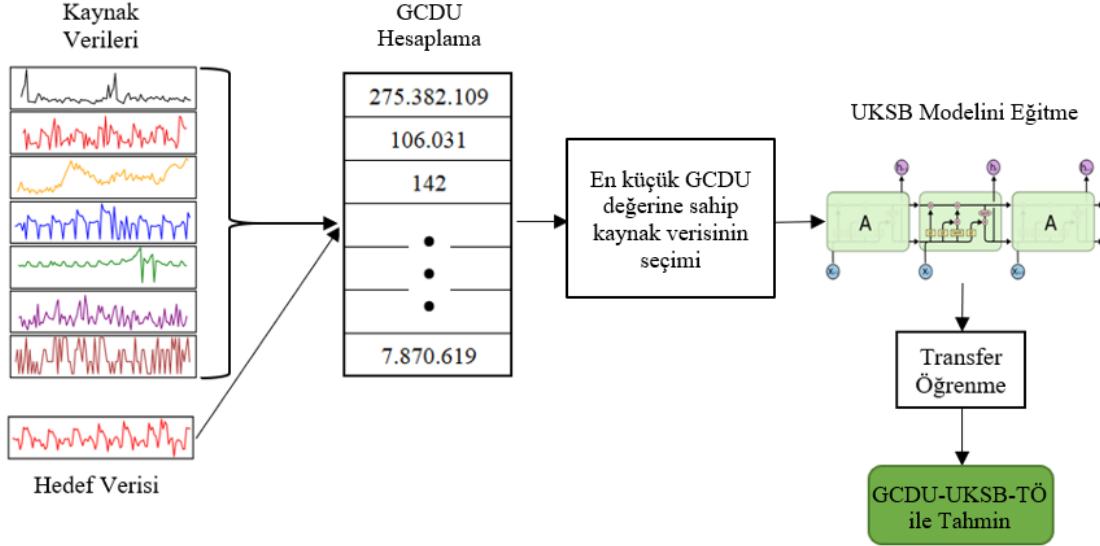
sektörlere ait satış verileri kullanılmıştır [39, 40]. Tüm veri kümeleri haftalık satış verilerini içerecek şekilde düzenlenmiştir. Tablo 3'te yer alan Walmart, Forklift, LearnX ve Pharma birden fazla bölge, mağaza ya da ürünlere ait satış verileri içermektedir. Bu nedenle, bu çalışma kapsamında bu veri kümelerinden bir kaynak veri ve bir hedef veri olmak üzere iki adet satış verisi (seri) seçilmiştir. Rossmann, Tv_Sales ve Favorita veri kümeleri ise tek seri içermektedir. Walmart, Forklift, LearnX ve Pharma veri kümelerindeki her bir hedef veriye karşılık Tablo 3'te yedi adet kaynak veri bulunmaktadır. Rossmann, Tv_Sales ve Favorita veri kümelerindeki serilerden biri hedef veri olarak alındığında ise altı adet kaynak veri vardır. Örneğin, Rossmann veri kümesindeki tek seri hedef veri olarak alındığında Walmart, Forklift, LearnX, Pharma, Tv_Sales ve Favorita veri kümelerindeki seriler kaynak veri olarak alınmıştır, aynı veri kümesinden (Rossmann) kaynak veri seçimi ise tanımlı değildir. Bu nedenle, deneysel çalışmalarda toplam 46 adet kaynak ve hedef veri kombinasyonu incelenmiştir.

Tahmin performansını değerlendirmek için Eş. 10-Eş. 11'de verilen ortalama karesel hatanın karekökü (OKHK) (Root Mean Square Error - RMSE) ve ortalama mutlak hata (OMH) (Mean Absolute Error - MAE) kullanılmıştır.

$$OKHK = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{y}_t - y_t)^2} \quad (10)$$

$$OMH = \frac{1}{n} \sum_{t=1}^n |\hat{y}_t - y_t| \quad (11)$$

Burada, \hat{y}_t ve y_t sırasıyla t . dönemdeki tahmin değeri ile gerçek değeri, n ise dönem sayısını ifade etmektedir.



Şekil 2. Önerilen yaklaşımın genel akış şeması (General flow chart of the proposed approach)

Tablo 2. GCDU-UKSB-TÖ algoritması (ERP-LSTM-TL algorithm)

Girdi: $X_K = \{x_1, x_2, \dots, x_n\}$ kaynak veri kümesi ve $y_h = \{y_{ht} | t = 1, 2, \dots, T_h\}$ hedef veri

Çıktı: Hedef veri için satış tahmini modeli

Adım 1. Kaynak veri kümesi X_K ve hedef satış verisi y_h , $[0, 1]$ aralığına normalize edilir.

Adım 2. Hedef veri ve potansiyel kaynak verileri arasındaki GCDU uzaklığı $GCDU(x_l, y_h), \forall x_l \in X_K$ hesaplanır.

Adım 3. Hedef veriye en küçük GCDU uzaklığına sahip kaynak veri $x_{l^*}, l^* = \text{m} \min \{GCDU(x_l, y_h)\}$, seçilir.

Adım 4. Kaynak UKSB modeli, seçilen kaynak veri x_{l^*} ile hiperparametre optimizasyonu yapılarak eğitilir.

Adım 5. Kaynak UKSB modelindeki hiperparametre ve ağırlıklar hedef UKSB modeline aktarılır.

Adım 6. Hedef UKSB modeli, hedef veri y_h ile eğitilir.

4.2. Karşılaştırmalı Değerlendirme Yöntemleri ve Hiperparametre Optimizasyonu (Benchmark Methods and Hyperparameters Optimization)

Öncelikle kaynak seçimi ve transfer öğrenmenin etkisini görmek amacıyla üç model karşılaştırılmıştır: İlk model, temel UKSB modelidir. Bu modelde, her bir hedef veri UKSB ile eğitilir ve transfer öğrenme yoktur. İkinci model, transfer öğrenmenin uygulandığı UKSB-TÖ modelidir. Bu modelde, UKSB önce bir kaynak veri ile eğitilir, sonrasında ise ön eğitilmiş UKSB modeli hedef veri ile tekrar eğitilir. Bu modelde, kaynak seçimi bulunmamaktadır. Üçüncü model ise hem transfer öğrenme hem de kaynak seçiminin yer aldığı GCDU-UKSB-TÖ modelidir ve Bölüm 3.3'te açıklanmıştır.

Bunlara ek olarak, GCDU-UKSB-TÖ'nün tahmin performansı klasik tahmin yöntemleri ve makine öğrenmesi yöntemleri ile karşılaştırılmıştır. Klasik tahmin yöntemlerinde basit hareketli ortalama (BHO) (Simple Moving Average - SMA), basit üstel düzeltme (BÜD) (Single Exponential Smoothing - SES), lineer (doğrusal) regresyon (LR) (Linear Regression - LR) ve otoregresif bütünleşik hareketli ortalama (Autoregressive Integrated Moving Average - ARIMA) ele alınmıştır. Makine öğrenmesi yöntemlerinde ise k-en yakın komşuluk regresyon (KYK-R) (k Nearest Neighbor

Regression - KNN-R), destek vektör regresyon (DVR) (Support Vector Regression - SVR), regresyon ağacı (RA) (Regression Tree - RT) ve yapay sinir ağı (YSA) (Artificial Neural Network - ANN) kullanılmıştır.

UKSB ve YSA modellerinde Tablo 4'te yer alan hiperparametreler ile ızgara arama yöntemi uygulanmıştır. Modellerde kayıp fonksiyonu olarak ortalama karesel hata ve optimizasyon algoritması olarak ADAM [41] algoritması kullanılmıştır. Transfer öğrenme olan UKSB modellerinde dönem sayısı UKSB'deki dönem sayısının %50'si olarak alınmıştır. Buna göre, Tablo 5'te Walmart-Seri_1 veri kümesi için seçilen UKSB mimarisi gösterilmektedir. Ayrıca klasik tahmin ve makine öğrenmesi yöntemlerinin parametre değerleri, Tablo 6'daki parametreler kullanılarak ızgara arama ile belirlenmiştir.

Deneyisel çalışmalarda her satış verisi öncelikle %80'i eğitim ve %20'si test kümesi olarak ayrılmıştır. Parametre optimizasyonu için eğitim verisinin %20'si doğrulama amacıyla kullanılmıştır.

Deneyisel çalışmalar Google Colab ortamında Python kodlama dilinde yürütülmüştür. UKSB modelleri keras kütüphanesinde geliştirilmiştir. Klasik tahmin yöntemleri için statsmodels ve makine öğrenmesi yöntemleri için scikit-learn (Sklearn) kütüphanesi kullanılmıştır.

Tablo 3. Veri kümeleri ve özellikleri (Datasets and properties)

Veri Kümesi	Hafta Sayısı	Kaynak Veri	Hedef Veri	Verinin Özellikleri	Açıklama
Walmart	143	Seri_1	Seri_2	Mevsimsellik vardır.	Farklı bölgelerdeki Walmart mağazalarının 2010-2012 yılları arasındaki satış verileri [36]
Forklift	276	Seri_1	Seri_2	Çok sayıda sıfır içermektedir ve azalma eğilimi vardır.	Bir forklift üreticisinin 2016-2020 yılları arasındaki satış verileri [37]
LearnX	882	Seri_1	Seri_2	Mevsimsellik vardır.	LearnX çevrimiçi öğrenme platformunun iki yıllık satış verileri [36]
Rossmann	942	Seri		Mevsimsellik vardır.	Rossmann mağazalarının 2010-2012 yılları arasındaki satış verileri [36]
Tv_Sales	974	Seri		Artma eğilimi vardır.	Bangladeş'in 2014-2016 yılları arasındaki TV satış verileri [36]
Favorita	1688	Seri		Mevsimsellik vardır.	Favorita süpermarketlerinin 2013-2017 yılları arasındaki satış verileri [36]
Pharma	2106	Seri_1	Seri_2	Mevsimsellik ile önce azalma sonra artma eğilimi vardır.	2014-2019 yılları arasında eczane ilaç satış verileri [36]

Tablo 4. UKSB ve YSA modellerinin hiperparametreleri (Hyperparameters of LSTM and ANN models)

Hiperparametre	Değer
Girdi boyutu	[1, 10, 25, 50, 75, 100]
Gizli katman sayısı (YSA)	[1]
Gizli katman sayısı (UKSB)	[2-5]
Gizli katman nöron sayıları	[1, 10, 25, 50, 75, 100, 150, 200]
Öğrenme oranı	[0,1; 0,01; 0,001]
Parti (Batch) büyüklüğü	1
Dönem (Epoch) sayısı	[25, 50, 100, 150, 200]
Aktivasyon Fonksiyonu	tanh, ReLU, leaky ReLU

Tablo 5. Walmart-Seri_1 için UKSB ağ mimarisi (LSTM network architecture for Walmart-Series_1)

Model: "sıralı"		
Katman (türü)	Çıktı Boyutu	Parametre Sayısı
UKSB (Girdi Katmanı)	(1, 100, 100)	40.800
UKSB_1 (1. Gizli Katman)	(1, 100, 50)	30.200
UKSB_2 (2. Gizli Katman)	(1, 25)	7.600
Tam Bağlantı (Çıktı Katmanı)	(1, 1)	26
Toplam parametreler: 78.626		
Eğitilebilir parametreler: 78.626		
Eğitilemez parametreler: 0		

Deneyler Intel(R) Core(TM) i7-6700HQ CPU 2.60GHz 16,0 GB RAM özelliklerine sahip bir bilgisayarda yapılmıştır.

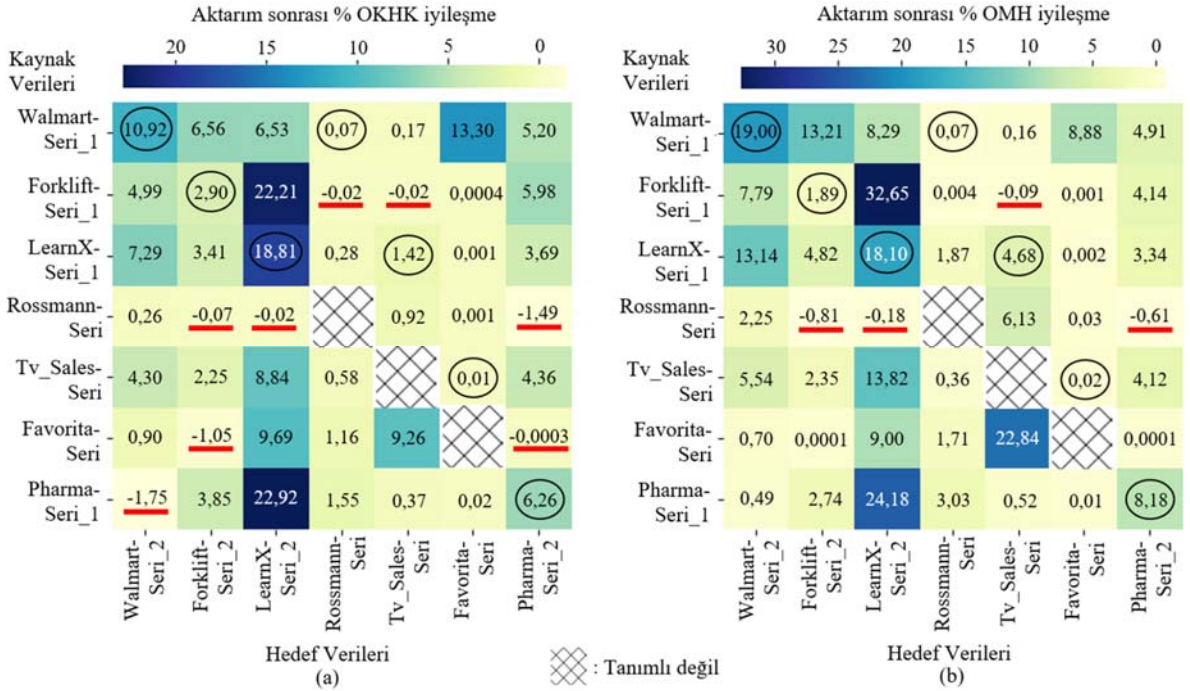
4.3. Transfer Öğrenmenin Etkisi (The Effect of Transfer Learning)

Öncelikle kaynak seçimi ve transfer öğrenmenin etkisini görmek için UKSB, UKSB-TÖ ve GCDU-UKSB-TÖ yöntemleri karşılaştırılmıştır. Her bir hedef veri için UKSB ve UKSB-TÖ'ye ait OKHK ve OMH değerleri arasındaki farkların UKSB'ye göre yüzde değişimi sırasıyla Şekil 3a ve Şekil 3b'de ısı haritalarında yer almaktadır. Isı haritalarındaki koyu renkler UKSB-TÖ'nün UKSB'ye göre OKHK ve OMH performans kriterlerinde iyileşme sağladığı durumları, açık renkler bu performans kriterlerinde yüzde değişimin sıfıra yakın ya da negatif olduğu durumları göstermektedir. Şekil 3'teki taralı alanlar ise tek seri içeren veri kümeleri için kaynak-hedef veri eşleşmesinin tanımlı olmadığı kombinasyonları belirtmektedir.

Buna göre, toplam 46 adet kaynak ve hedef veri kombinasyonu için deney yapılmıştır. Bu kombinasyonların 38 adedinde UKSB-TÖ modeli UKSB'ye göre her iki performans kriterinde iyileşme sağlamıştır. Geri kalan sekiz adet kombinasyon için ise negatif transfer öğrenme gözlenmiştir. Bu sekiz adet kombinasyon Şekil 3a ve Şekil 3b'deki ısı haritalarında kırmızı renk ile altı çizili olarak belirtilmiştir. OKHK'ya göre sırayla Pharma-Walmart, Rossmann-Forklift, Favorita-Forklift, Rossmann-LearnX, Forklift-Rossmann, Forklift-Tv_Sales, Rossmann-Pharma ve Favorita-Pharma olmak üzere sekiz adet kaynak-hedef veri eşleşmesinde; OMH'ye göre ise Rossmann-Forklift, Rossmann-LearnX, Forklift-Tv_Sales ve Rossmann-Pharma olmak üzere dört adet kaynak-hedef veri eşleşmesinde negatif transfer öğrenme meydana gelmiştir. Her iki performans kriterine göre ise toplamda sekiz adet kaynak-hedef veri eşleşmesinde negatif transfer öğrenme meydana gelmiştir.

Tablo 6. Klasik tahmin ve makine öğrenmesi yöntemlerinin parametreleri (Parameters of classical forecasting and machine learning methods)

Yöntem	Parametreler ve Değer Aralıkları
BHO	Geçmiş [1-10] dönem ortalaması
BÜD	$0 \leq \alpha \leq 1$ (α : üstel düzeltme katsayısı)
ARIMA	$p \in \{0, 1, 2, 3, 4\}$, $q \in \{0, 1, 2, 3, 4\}$, $d \in \{0, 1, 2, 3\}$ (p : otoregresif model derecesi, q : hareketli ortalama model derecesi, d : fark alma derecesi)
KYK-R	$1 \leq k \leq 20$ (k : komşu sayısı) Maksimum derinlik $\in \{3, 5, 10, 15\}$,
RA	Maksimum yaprak düğümü $\in \{3, 5, 7, 10, 15\}$, Yapraktaki minimum örnek sayısı = 1, Bölünme için minimum örnek sayısı = 2 $C \in \{0,01, 0,05, 0,1, \dots, 1, \dots, 200, 500, 1000\}$, $\varepsilon \in \{0,0001, 0,0005, \dots, 0,01, \dots, 8, 10\}$,
DVR	$\gamma \in \{0,0001, 0,001, 0,005, \dots, 0,1, \dots, 7, 10\}$, Kernel çekirdek fonksiyonu $\in \{\text{radyal tabanlı fonksiyon, polinom, sigmoid, doğrusal}\}$ (C : düzenleme katsayısı, γ : dağılım genişliği kontrol parametresi, ε : duyarsız Laplace kayıp fonksiyonu parametresi)



Şekil 3. (a)-(b) UKSB ve UKSB-TÖ'nün OKHK ve OMH değerleri arasındaki farkın UKSB'ye göre yüzde değişimi (Yuvarlak ile işaretlenen GCDU-UKSB-TÖ ile seçilen kombinasyondur.)

((a)-(b) Percentage change of the difference between RMSE and MAE values of LSTM and LSTM-TL compared to LSTM (The combination selected with ERP-LSTM-TL is marked with a round.))

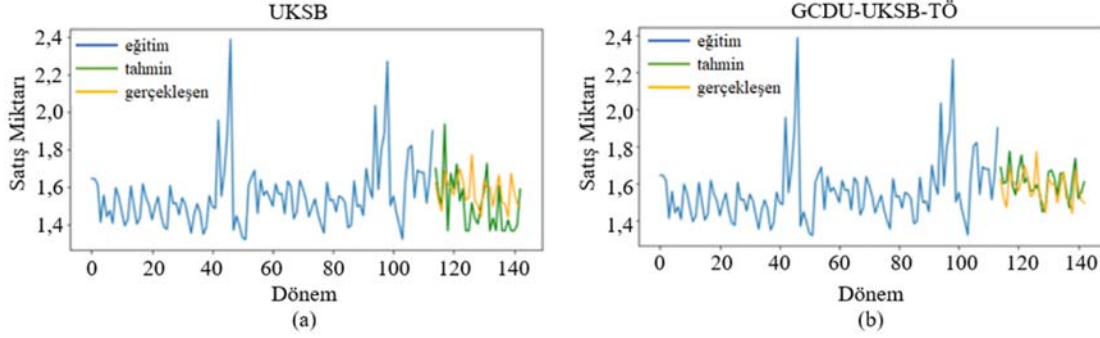
GCDU-UKSB-TÖ yöntemi uygulandığında elde edilen değişim yüzdeleri ise Şekil 3a ve Şekil 3b'de yuvarlak ile işaretlenmiştir. GCDU-UKSB-TÖ yöntemi ile seçilen kaynak-hedef eşleşmeleriyle tüm hedef verilerde UKSB'ye göre iyileşme elde edilmiştir ve negatif transfer öğrenme gözlenmemiştir. Örneğin, GCDU-UKSB-TÖ ile LearnX-Seri_2 hedef verisi için LearnX-Seri_1 kaynak verisi seçilmiştir, UKSB'ye göre OKHK ve OMH değerlerinde sırasıyla %18,81 ile %18,10 oranında iyileşme elde edilmiştir. LearnX-Seri_2 hedef verisine GCDU-UKSB-TÖ uygulandığında OKHK değeri 22,88'dir, UKSB uygulandığında ise OKHK değeri 28,18'dir. Buna göre, GCDU-UKSB-TÖ yönteminin UKSB'ye kıyasla tahmin hatasındaki iyileşme $((28,18 - 22,88) / 28,18) * 100 = \% 18,81$ olarak hesaplanmıştır. Şekil 3'te yer alan diğer iyileşme yüzdelere hesaplamaları da benzer şekildedir. LearnX-Seri_2 hedef verisi için UKSB ve GCDU-UKSB-TÖ ile elde edilen tahmin sonuçları Şekil 4a ve Şekil 4b'de gösterilmiştir. Aynı hedef veride, Rossmann kaynak verisi ile transfer öğrenme uygulandığında ise Şekil 3'te görüldüğü gibi OKHK ve OMH değerleri artmakta ve negatif öğrenme gözlenmektedir.

Özetle, GCDU-UKSB-TÖ'deki benzerlik tabanlı kaynak seçimi negatif öğrenmenin engellenmesinde etkilidir. GCDU-UKSB-TÖ

yöntemi hedef veriyle eğitilen UKSB'ye kıyasla OKHK ölçütüne göre ortalama %5,77 ve OMH ölçütüne göre ortalama %7,42 iyileşme sağlamıştır. Bu amaçla, GCDU-UKSB-TÖ yöntemi ile seçilen kaynak veriler için Şekil 3'te yer alan iyileşme yüzdelere ortalaması alınmıştır. Örneğin, ele alınan yedi hedef verisinde OKHK ölçütüne göre ortalama iyileşme $(10,92 + 2,90 + 18,81 + 0,07 + 1,42 + 0,01 + 6,26) / 7 = \% 5,77$ olarak hesaplanmıştır. Ayrıca her bir hedef verisi için eğitim süresindeki iyileşme, UKSB ağı ile UKSB-TÖ'nün eğitim süreleri arasındaki farkın UKSB'nin eğitim süresine oranı olarak hesaplanmıştır. Böylece, UKSB-TÖ ve UKSB'nin eğitim süreleri karşılaştırıldığında UKSB-TÖ ile eğitim sürelerinde %40,28 ile %53,19 arasında azalma elde edilmiştir. Eğitim sürelerindeki iyileşme değerleri benzer şekilde hesaplandığında GCDU-UKSB-TÖ yönteminin eğitim sürelerinde ortalama %47,90 azalma sağladığı görülmüştür.

4.4. Tahmin Yöntemlerinin Karşılaştırılması (Comparison of Forecasting Methods)

Her bir hedef veri için tahmin yöntemlerinin sonuçları Tablo 7'de verilmiştir. Buna göre, GCDU-UKSB-TÖ yedi adet veri kümesinin beş tanesinde her iki performans kriterine göre en iyi sonuçları



Şekil 4. (a)-(b) LearnX-Seri_2 hedef verisi için UKSB ve GCDU-UKSB-TÖ sonuçları
(LSTM and ERP-LSTM-TL results for LearnX-Series_2 target data)

Tablo 7. Tahmin yöntemlerinin karşılaştırılması (Comparison of forecasting methods)

Yöntem	Kriter	Hedef Veri Kümesi						
		Walmart-Seri 2	Forklift-Seri 2	LearnX-Seri 2	Rossmann-Seri	Tv_Sales-Seri	Favorita-Seri	Pharma-Seri 2
BHO	OKHK	115.770,64	0,75	45,68	3.329.590,09	2.908,25	14.670,57	19,67
	OMH	98.716,04	0,71	38,44	2.711.614,43	1.920,06	9.191,51	14,25
BÜD	OKHK	116.166,98	0,68	32,46	3.263.554,06	2.897,19	14.514,52	19,10
	OMH	98.550,29	0,65	25,38	2.502.639,81	1.915,83	9.095,75	13,80
ARIMA	OKHK	122.601,63	0,66	28,04	3.264.280,31	2.897,20	14.421,66	19,10
	OMH	104.878,25	0,64	21,43	2.506.686,66	1.915,83	9.047,29	13,80
LR	OKHK	112.991,92	0,46*	32,48	3.310.276,02	2.271,73	13.790,60	15,79
	OMH	93.347,53	0,39**	25,89	2.656.003,10	1.829,05	9.130,77	12,32
KYK-R	OKHK	98.477,90	2,74	29,65	2.902.171,84	1.495,86	13.721,35	14,57
	OMH	78.923,85	2,67	22,36	2.262.102,26	895,17	8.938,76	11,12
RA	OKHK	206.813,33	1,29	31,57	3.704.580,63	1.807,19	19.971,05	16,62
	OMH	141.551,82	1,24	24,27	2.496.191,80	1.214,33	10.967,13	12,37
DVR	OKHK	85.313,99	1,35	29,30	3.256.237,12	2.628,90	14.058,08	13,94
	OMH	68.845,31	1,30	21,35	2.491.571,59	1.478,29	8.950,41	10,65
YSA	OKHK	97.200,29	0,54	24,41	1.760.873,74	1.485,04	12.141,89*	13,10
	OMH	79.663,94	0,46	18,94	1.282.503,70	789,19	7.147,44**	9,85
UKSB	OKHK	90.000,71	0,69	28,18	1.758.434,87	1.379,42	13.771,48	11,47
	OMH	74.722,67	0,65	23,07	1.282.280,57	715,58	9.124,39	9,61
GCDU-UKSB-TÖ	OKHK	80.169,47*	0,67	22,88*	1.757.254,14*	1.359,90*	13.770,33	10,75*
	OMH	60.526,93**	0,64	18,89**	1.281.410,33**	682,13**	9.122,89	8,82**

* : Hedef veri kümesindeki en küçük OKHK değeridir.

** : Hedef veri kümesindeki en küçük OMH değeridir.

vermiştir. Forklift verisinde en iyi sonucu LR, Favorita verisinde ise YSA vermiştir. Tahmin yöntemlerinin performanslarının karşılaştırılması için Demsar [42]'de önerilen parametrik olmayan istatistiksel testler kullanılmıştır. Bu testlerin normallik veya varyans homojenliği gibi varsayımları yoktur. Bu kapsamda, tahmin yöntemleri için ikili ve çoklu karşılaştırmalar yapılmıştır [42]. Çoklu karşılaştırmalar için Friedman ile Nemenyi post-hoc testleri, ikili karşılaştırmalar için Wilcoxon eşleştirilmiş işaretli sıra testi uygulanmıştır [42].

Friedman testinde hipotezler şöyledir: (i) H_0 : Tahmin yöntemlerinin performanslarında farklılık yoktur. (ii) H_1 : Tahmin yöntemlerinden en az birinin performansında farklılık vardır. Friedman testinde Demsar [42]'de önerildiği üzere Iman ve Davenport [43] istatistiği kullanılmıştır.

OKHK ve OMH performans kriterleri için %5 anlamlılık düzeyinde Friedman testlerinin sonuçları Tablo 8'de verilmektedir. Buna göre, tahmin yöntemlerinden en az birine ait OKHK ve OMH değerlerinde istatistiksel olarak anlamlı bir fark vardır. İstatistiksel olarak farklı performansa sahip tahmin yöntemlerinin belirlenmesi için Nemenyi post-hoc testi yapılmıştır. Nemenyi post-hoc testinde yöntemlerin ortalama sıralama değerleri (OSD) arasındaki farklar kritik uzaklık ile karşılaştırılır. Eğer iki yöntemin ortalama sıralama değerleri arasındaki fark kritik uzaklıktan küçük ise bu iki yöntem arasında istatistiksel olarak anlamlı fark olmadığına kanaat getirilir. Bu nedenle, Nemenyi test sonuçlarında bu iki yöntem aynı doğru üzerinde ve birbirleriyle bağlantılı olarak gösterilmektedir. Tahmin yöntemlerinin ortalama sıralama değerleri Şekil 5a ve Şekil 5b'de yer almaktadır ve kritik uzaklık (Kritik Fark - KF) %5 anlamlılık düzeyinde 5,12 olarak hesaplanmıştır. Bir yöntemin ortalama sıralama değerinin küçük olması yöntemin performansının diğer yöntemlerden üstün olduğunu göstermektedir. Buna göre, Şekil 5a ve Şekil 5b'de GCDU-UKSB-TÖ yöntemi ile OKHK ve OMH performans

kriterlerine göre en iyi ortalama sıralama değerleri (sırasıyla 1,75 ve 2,09) elde edilmiştir. GCDU-UKSB-TÖ yöntemi OKHK'ya göre BHO (OSD = 8,84), BÜD (OSD = 7,42) ve RA (OSD = 8,21)'dan; OMH'ye göre ise BHO (OSD = 9,27) ve RA (OSD = 7,58)'dan istatistiksel olarak daha iyi sonuç vermiştir. Örneğin, OKHK'ya göre GCDU-UKSB-TÖ (OSD = 1,75) ile BHO (OSD = 8,84) yöntemlerinin ortalama sıralama değerleri arasındaki fark $8,84 - 1,75 = 7,09$ 'dur ve bu değer kritik farktan ($7,09 > 5,12$) büyüktür. Bu nedenle, OKHK'ya göre GCDU-UKSB-TÖ yönteminin performansı BHO yönteminin performansından istatistiksel olarak üstündür. Diğer yöntemler ile GCDU-UKSB-TÖ arasında istatistiksel olarak anlamlı fark bulunmadığı gözlenmiştir.

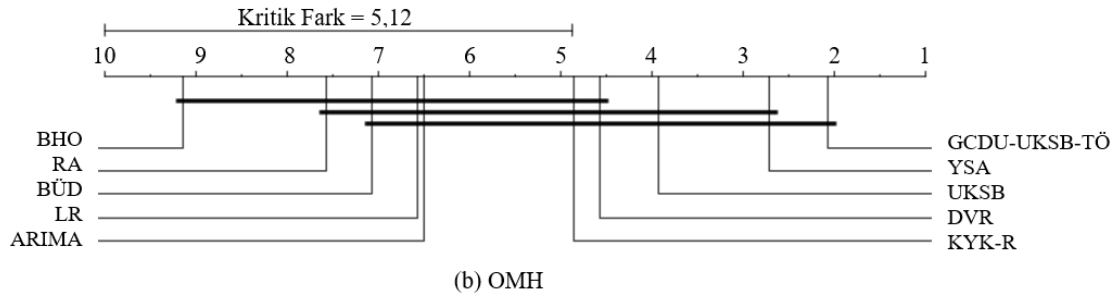
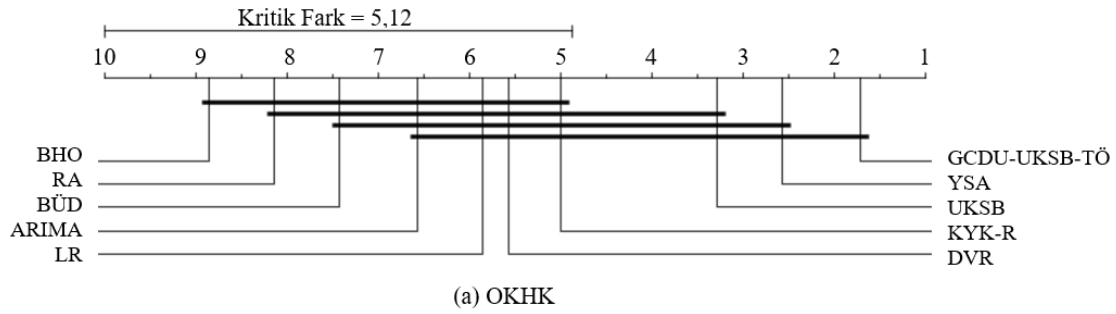
Tahmin yöntemlerinin ikili karşılaştırmaları için Wilcoxon eşleştirilmiş işaretli sıra testi uygulanmıştır. Tek taraflı kurulan hipotez testlerinde, H_0 ve H_1 hipotezleri şöyledir: (i) H_0 : İki tahmin yönteminin hata miktarları eşit veya birinci yöntemin hata miktarı, ikinci yöntemin hata miktarından daha fazladır. (ii) H_1 : Birinci tahmin yönteminin hata miktarı, ikinci tahmin yönteminin hata miktarından azdır. Tablo 9'da tahmin yöntemlerinin ikili karşılaştırmalarına ilişkin Wilcoxon eşleştirilmiş işaretli sıra testi sonuçları verilmektedir. Bu tabloda satırlar birinci tahmin yöntemini, sütunlar ise ikinci tahmin yöntemini göstermektedir ve %5 anlamlılık düzeyinde ($\alpha = 0,05$) H_0 hipotezi reddedilen ikili karşılaştırmalar * işareti ile belirtilmiştir.

Tablo 9'da verilen Wilcoxon eşleştirilmiş işaretli sıra testi sonuçlarına göre, aşağıdaki bulgular gözlenmiştir:

- Klasik tahmin yöntemleri analiz edildiğinde BÜD yöntemi, OMH performans kriterine göre BHO yönteminden; LR ise her iki performans kriterine göre BHO yönteminden istatistiksel olarak daha iyi performans göstermektedir (p-değeri $\leq 0,05$). BHO ve ARIMA'nın performansı diğer tahmin yöntemlerinin performansından düşüktür (p-değeri $> 0,05$).

Tablo 8. Friedman test sonuçları (Friedman test results)

Performans Kriteri	Hesaplanan F değeri (F_F)	F tablo değeri ($F_{\alpha=0,05, k-1=9, (k-1)*(N-1)=54}$)	Sonuç
OKHK	9,81	2,27	$F_F > F_{0,05; 9; 54}$, H_0 hipotezi reddedilir.
OMH	7,39	2,27	



Şekil 5. (a)-(b) Nemenyi testi sonuçları ($\alpha = 0,05$) ((a)-(b) Nemenyi test results ($\alpha = 0,05$))

Tablo 9. Wilcoxon eşleştirilmiş işaretli sıra testi testlerinin p-değerleri (p-values of Wilcoxon paired signed-rank tests)

Yöntem	Kriter	BHO	BÜD	ARIMA	LR	KYK-R	RA	DVR	YSA	UKSB	GCDU-UKSB-TÖ
BHO	OKHK		0,91	0,91	0,99	0,99	0,20	0,99	0,99	0,99	0,99
	OMH		0,99	0,91	0,99	0,99	0,63	0,99	0,99	0,99	0,99
BÜD	OKHK	0,09		0,43	0,84	0,99	0,20	0,99	0,99	0,99	0,99
	OMH	0,01*		0,43	0,57	0,99	0,57	0,99	0,99	0,95	0,95
ARIMA	OKHK	0,09	0,57		0,75	0,97	0,09	0,97	0,99	0,97	0,99
	OMH	0,09	0,57		0,50	0,97	0,37	0,98	0,99	0,88	0,95
LR	OKHK	0,01*	0,16	0,25		0,98	0,12	0,75	0,99	0,99	0,99
	OMH	0,01*	0,43	0,50		0,98	0,50	0,99	0,99	0,99	0,99
KYK-R	OKHK	0,01*	0,01*	0,03*	0,02*		0,01*	0,37	0,99	0,95	0,95
	OMH	0,01*	0,01*	0,03*	0,02*		0,02*	0,37	0,94	0,91	0,95
RA	OKHK	0,80	0,80	0,91	0,88	0,99		0,94	0,99	0,99	0,99
	OMH	0,37	0,43	0,63	0,50	0,98		0,94	0,99	0,99	0,99
DVR	OKHK	0,01*	0,01*	0,03*	0,25	0,63	0,06		0,91	0,91	0,99
	OMH	0,01*	0,01*	0,02*	0,01*	0,63	0,06		0,91	0,57	0,95
YSA	OKHK	0,01*	0,01*	0,01*	0,01*	0,01*	0,01*	0,09		0,80	0,91
	OMH	0,01*	0,01*	0,01*	0,01*	0,06	0,01*	0,09		0,75	0,84
UKSB	OKHK	0,01*	0,01*	0,03*	0,01*	0,05*	0,01*	0,09	0,20		0,99
	OMH	0,01*	0,05*	0,12	0,01*	0,09	0,01*	0,43	0,25		0,99
GCDU-UKSB-TÖ	OKHK	0,01*	0,01*	0,01*	0,01*	0,05*	0,01*	0,01*	0,09	0,01*	
	OMH	0,01*	0,05*	0,05*	0,01*	0,05*	0,01*	0,05*	0,16	0,01*	

* : %5 anlamlılık düzeyinde H_0 hipotezi reddedilmiştir ($\alpha = 0,05$).

- Makine öğrenmesi yöntemleri incelendiğinde KYK-R yöntemi; klasik tahmin yöntemleri ile RA'dan istatistiksel olarak daha iyi performans göstermektedir (p-değeri $\leq 0,05$). DVR yönteminin performansı BHO, BÜD ve ARIMA yöntemlerinin performansından istatistiksel olarak üstündür (p-değeri $\leq 0,05$). YSA ile BHO, BÜD, ARIMA, LR ve RA yöntemlerinin aralarındaki farklar istatistiksel olarak anlamlı olup YSA'nın performansı bu yöntemlerden daha iyidir (p-değeri $\leq 0,05$). Ayrıca, KYK-R ile YSA'nın OKHK değerleri arasındaki fark istatistiksel olarak anlamlıdır ve YSA yöntemi KYK-R'ye göre daha iyi performans göstermektedir (p-değeri $\leq 0,05$). RA yöntemi ise diğer tahmin yöntemlerinden istatistiksel olarak daha düşük performans göstermektedir (p-değeri $> 0,05$). Özetle, YSA yöntemi makine öğrenmesi yöntemleri içinde en iyi performansa sahip yöntemdir.

Derin öğrenme yöntemleri içerisinde yer alan UKSB ise BHO, BÜD, LR ve RA yöntemlerinden her iki performans kriterine göre; ARIMA ve KYK-R yöntemlerinden ise OKHK performans kriterine göre istatistiksel olarak performans üstünlüğüne sahiptir (p-değeri $\leq 0,05$). Önerilen GCDU-UKSB-TÖ yöntemi, YSA dışındaki tüm tahmin yöntemlerine göre istatistiksel olarak daha düşük hata değerleri vermektedir. Ayrıca anlamlılık düzeyi %10 ($\alpha = 0,10$) olduğunda, GCDU-UKSB-TÖ yöntemi OKHK'ya göre diğer tüm tahmin yöntemlerine istatistiksel olarak performans üstünlüğü sağlamaktadır. Önerilen yaklaşımın performansı Tablo 3'te yer alan veri kümelerinin özelliklerine göre incelendiğinde mevsimsellik içeren beş veri kümesinden dördünde (Walmart, LearnX, Rossmann ve Pharma) önerilen yaklaşım ile en doğru tahmin sonuçlarının elde edildiği görülmüştür. Diğer dört veri kümesinden daha uzun mevsimsel periyoda sahip Favorita'da ise 10 adet yöntem içerisinde, önerilen yaklaşım üçüncü en iyi performansa sahip yöntemdir. Ayrıca artma eğilimi olan Tv_Sales ile hem artma hem de azalma eğilimi olan Pharma veri kümelerinde en iyi performans önerilen yaklaşıma aittir. Ancak çok sayıda sıfır içeren, düzensiz ve aralıklı (intermittent) bir veri kümesi olan Forklift için önerilen yaklaşım OKHK'ya göre dördüncü, OMH'ye göre üçüncü en iyi performansa sahiptir. Sonuç olarak, yapılan deneysel çalışmalarda farklı sektörler ve farklı ürünlerdeki satış verilerine benzerlik tabanlı kaynak seçimi ve derin transfer öğrenme uygulandığında tahmin doğruluğu ve eğitim süresinde iyileşme sağlanmıştır. Özellikle hedef veriye benzer bir kaynak verinin seçilmesinin derin transfer öğrenmede aktarılabilirlik açısından önem taşıdığı gözlenmiştir.

5. Sonuçlar (Conclusions)

Satış tahmini, firmalardaki kaynak ve envanter kararlarının etkili bir şekilde yönetilmesi için önemli bir girdidir. Günümüzde artan dijitalleşme ile firmaların bu alanda oluşan bilgi birikimini gelecek tahmin modellerine aktarımını sağlayan yaklaşımlar önem arz etmektedir. Özellikle eğitim aşamasında büyük miktarda veri ve süreye gereksinimi olan derin öğrenme yöntemlerinde bilgi aktarımı ve transfer öğrenme yaklaşımlarına ihtiyaç duyulmaktadır. Bu çalışmada, satış tahmininde farklı sektör ve farklı ürünler arasında bilgi aktarımını sağlamak amacıyla derin transfer öğrenme modeli geliştirilmiştir. Bu kapsamda, farklı uzunluktaki ve farklı sektörlerdeki satış verileri arasında kaynak seçimi için benzerlik tabanlı bir aktarılabilirlik ölçütü önerilmiştir. Bu ölçüte göre seçilen kaynak satış veri kümesi ile UKSB ağı eğitilmiştir ve parametre transferi ile hedef veri için UKSB modeli geliştirilmiştir. Çeşitli veri kümeleri ile yapılan deneysel çalışmalar önerilen GCDU-UKSB-TÖ yönteminin hem UKSB'ye göre hem de klasik tahmin ve makine öğrenmesi yöntemlerine göre istatistiksel olarak tahmin doğruluğunu iyileştirdiğini göstermiştir. Benzerlik tabanlı kaynak seçimi, negatif öğrenmeyi önlemiştir. Ayrıca, önerilen yaklaşım ile derin öğrenme modellerindeki eğitim süresinin kısalması gözlenmiştir. Önerilen bu yaklaşımın üretim ve hizmet sektörlerinde kullanılması ile satış tahmini kaynaklı maliyetlerde ve zaman kullanımında tasarruf sağlama potansiyeli bulunmaktadır.

Çalışmanın bazı kısıtları bulunmaktadır. İlk olarak, satış tahmin problemi tek değişkenli zaman serisi tahmini olarak ele alındığı için satış değerini etkileyecek diğer faktörlerin (hava durumu, ürünün özellikleri, ekonomik göstergeler, vb.) etkisi dikkate alınmamıştır. Ayrıca düzensiz ve aralıklı veride önerilen yaklaşımın performansı düşüktür. Gelecek çalışmalarda farklı özellikteki veride performansı arttırmak için öznitelik mühendisliği ele alınabilir ve problem çok değişkenli zaman serisi olarak incelenebilir. Bu kapsamda, öznitelik çıkarımı konusunda güçlü olan KSA yöntemi kullanılabilir. Bir diğer gelecek çalışma çoklu kaynak seçimi problemi olabilir. Ayrıca son yıllarda derin öğrenme modelleri açıklanabilirlik ve yorumlanabilirlik açısından sorgulanmaktadır. Bu nedenle, derin transfer öğrenme modellerinde bu özelliklerin geliştirilmesi ele alınabilir.

Teşekkür (Acknowledgement)

Bu araştırma Bursa Uludağ Üniversitesi (BUÜ) Bilimsel Araştırma Projeleri (BAP) Birimi tarafından desteklenmiştir (Proje Kodu: FDK-2021-518).

Kaynaklar (References)

1. Chopra S., Meindl P., Kalra D.V., Supply chain management: Strategy, planning, and operation, MA: Pearson, 269, Boston, 232, 2013.
2. Kraus M., Feuerriegel S., Oztekin A., Deep learning in business analytics and operations research: Models, applications and managerial implications, *European Journal of Operational Research*, 281 (3), 628-641, 2020.
3. Şeker A., Diri B., Balık H.H., Derin öğrenme yöntemleri ve uygulamaları hakkında bir inceleme, *Gazi Mühendislik Bilimleri Dergisi*, 3 (3), 47-64, 2017.
4. Pan S.J., Yang Q., A survey on transfer learning, *IEEE Transactions on Knowledge and Data Engineering*, 22 (10), 1345-1359, 2009.
5. Zhuang F., Qi Z., Duan K., Xi D., Zhu Y., Zhu H., Xiong H., He, Q., A comprehensive survey on transfer learning, *Proceedings of the IEEE*, 109 (1), 43-76, 2020.
6. Tan C., Sun F., Kong T., Zhang W., Yang C., Liu C., A survey on deep transfer learning, In *International Conference on Artificial Neural Networks, Rodos-Greece*, 270-279, 4-7 October, 2018.
7. Niu S., Liu Y., Wang J., Song H., A decade survey of transfer learning (2010-2020), *IEEE Transactions on Artificial Intelligence*, 1 (2), 151-166, 2020.
8. Weber M., Auch M., Doblander C., Mandl P., Jacobsen H.A., Transfer Learning with time series data: A systematic mapping study, *IEEE Access*, 165409-165432, 2021.
9. Meiseles A., Rokach L., Source model selection for deep learning in the time series domain, *IEEE Access*, 8, 6190-6200, 2020.
10. Fawaz H.I., Forestier G., Weber J., Idoumghar L., Muller P.A., Transfer learning for time series classification, In *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA, 1367-1376, 10-13 December, 2018.
11. Ye R., Dai Q., Implementing transfer learning across different datasets for time series forecasting, *Pattern Recognition*, 109, 107617, 2021.
12. Karb T., Kühl N., Hirt R., Glivici-Cotruta V., A network-based transfer learning approach to improve sales forecasting of new products, In *Proceedings of the 28th European Conference on Information Systems (ECIS)*, An Online AIS Conference, Marrakech-Morocco, 15-17 June, 2020.
13. Loureiro A.L., Miguéis V.L., da Silva L.F., Exploring the use of deep neural networks for sales forecasting in fashion retail, *Decision Support Systems*, 114, 81-93, 2018.
14. Yuan F.C., Lee C.H., Intelligent sales volume forecasting using Google search engine data, *Soft Computing*, 24 (3), 2033-2047, 2020.
15. Kaya S.K., Yıldırım Ö., A prediction model for automobile sales in Turkey using deep neural networks, *Endüstri Mühendisliği*, 31 (1), 57-74, 2020.
16. Priyadarshi R., Panigrahi A., Routroy S., Garg G.K., Demand forecasting at retail stage for selected vegetables: A performance analysis, *Journal of Modelling in Management*, 2019.
17. Helmini S., Jihan N., Jayasinghe M., Perera S., Sales forecasting using multivariate long short term memory network models, *PeerJ PrePrints*, 7, e27712v1, 2019.
18. Punia S., Singh S.P., Madaan J.K., A cross-temporal hierarchical framework and deep learning for supply chain forecasting, *Computers & Industrial Engineering*, 149, 106796, 2020.
19. Kolková A., Navrátil M., Demand forecasting in Python: Deep learning model based on LSTM architecture versus statistical models, *Acta Polytechnica Hungarica*, 18 (8), 2021.
20. Huber J., Stuckenschmidt H., Intraday shelf replenishment decision support for perishable goods, *International Journal of Production Economics*, 231, 107828, 2021.
21. He Q.Q., Wu C., Si Y.W., LSTM with Particle swarm optimization for sales forecasting, *Electronic Commerce Research and Applications*, 101118, 2022.
22. Wang J., Liu G.Q., Liu L., A selection of advanced technologies for demand forecasting in the retail industry, In *2019 IEEE 4th International Conference on Big Data Analytics (ICBDA)*, Suzhou-China, 317-320, 15-18 March, 2019.
23. Aci M., Doğansoy G.A., Demand forecasting for e-retail sector using machine learning and deep learning methods, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 37 (3), 1325-1339, 2022.
24. Kilimci Z.H., Akyuz A.O., Uysal M., Akyokus S., Uysal M.O., Atak B.B., Ekmis M.A., An improved demand forecasting model using deep learning approach and proposed decision integration strategy for supply chain, *Complexity*, 2019.
25. Thenmozhi K., Reddy U.S., Crop pest classification based on deep convolutional neural network and transfer learning, *Computers and Electronics in Agriculture*, 164, 104906, 2019.
26. Sargano A.B., Wang X., Angelov P., Habib Z., Human action recognition using transfer learning with deep representations, In *2017 International Joint Conference on Neural Networks (IJCNN)*, Anchorage, Alaska, USA, 463-469, 14-19 May, 2017.
27. Ruder S., Peters M.E., Swayamdipta S., Wolf T., Transfer learning in natural language processing, In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorials*, Minneapolis, MN, USA, 15-18, 2 June, 2019.
28. Zhao K., Wang C., Sales forecast in e-commerce using convolutional neural network, *arXiv preprint arXiv:1708.07946*, 2017.
29. Pan H., Zhou H., Study on convolutional neural network and its application in data mining and sales forecasting for E-commerce, *Electronic Commerce Research*, 20 (2), 297-320, 2020.
30. Hirt R., Srivastava A., Berg C., Kühl N., Sequential transfer machine learning in networks: Measuring the impact of data and neural net similarity on transferability, In *Hawaii International Conference on Systems Sciences (HICSS-54)*, Grand Wailea, Maui, Hawaii, 7078-7087, 5-8 January, 2021.
31. He Q.Q., Pang P.C.I., Si Y.W., Transfer learning for financial time series forecasting, *The 16th Pacific Rim International Conference on Artificial Intelligence (PRICAI-2019)*, Cuvu-Fiji, 24-36, 26-30 August, 2019.
32. Hochreiter S., Schmidhuber J., LSTM can solve hard long time lag problems, *Advances in Neural Information Processing Systems*, MIT Press, 473-479, 1997.
33. Bengio Y., Simard P., Frasconi P., Learning long-term dependencies with gradient descent is difficult, *IEEE Transactions on Neural Networks*, 5 (2), 157-166, 1994.
34. Abbasimehr H., Shabani M., Yousefi M., An optimized model using LSTM network for demand forecasting, *Computers & Industrial Engineering*, 143, 106435, 2020.
35. Sherstinsky A., Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network, *Physica D: Nonlinear Phenomena*, 404, 132306, 2020.
36. Abanda A., Mori U., Lozano J.A., A review on distance based time series classification, *Data Mining and Knowledge Discovery*, 33 (2), 378-412, 2019.
37. Chen L., Ng R., On the marriage of lp-norms and edit distance, In *Proceedings of the Thirtieth International Conference on Very Large Data Bases*, Toronto-Canada, 30, 792-803, 31 August-3 September, 2004.
38. Levenshtein V.I., Binary codes capable of correcting deletions, insertions, and reversals, In *Soviet Physics Doklady*, 10 (8), 707-710, 1966.
39. Kaggle: Makine Öğrenimi ve Veri Bilimi Topluluğunuz, <https://www.kaggle.com/datasets>. Erişim tarihi Ekim 5, 2020.
40. Puspita P.E., İnkaya T., Akansel M., Clustering-based sales forecasting in a forklift distributor, *International Journal of Engineering Research and Development*, 11 (1), 25-40, 2019.
41. Kingma D.P., Ba J., Adam: A method for stochastic optimization, *arXiv preprint, arXiv:1412.6980*, 2014.
42. Demšar J., Statistical comparisons of classifiers over multiple data sets, *The Journal of Machine Learning Research*, 7, 1-30, 2006.
43. Iman R.L., Davenport J.M., Approximations of the critical region of the Friedman statistic, *Communications in Statistics-Theory and Methods*, 9 (6), 571-595, 1980.