

LSTM YÖNTEMİ İLE EKONOMİK GÖSTERGELER KULLANILARAK OTOMOBİL SATIŞ TAHMİNİ

AUTOMOBILE SALES FORECAST USING ECONOMIC INDICATORS WITH LSTM METHOD

Mustafa YURTSEVER

Dokuz Eylül Üniversitesi, Bilgi İşlem Dairesi Başkanlığı

mustafa.yurtsever@deu.edu.tr

ORCID: 0000-0003-2232-0542

ÖZ

ABSTRACT

Geliş Tarihi:

25.08.2021

Kabul Tarihi:

01.08.2022

Yayın Tarihi:

30.09.2022

Anahtar Kelimeler:

Ekonomik

Göstergeler

LSTM

Otomobil satış
tahmin

Zaman Serileri

Keywords:

Economic Indicators

LSTM

Automobile Sales
Forecast

Time Series

Otomotiv sanayi birçok ülke için en önemli sanayi kollarından birisidir. Bu nedenle araç satışlarına ilişkin tahminler otomotiv sanayisine ve tedarikçilerine değerli bilgiler sağlamaktadır. Otomobil satışları, piyasa ortamı, ekonomik kriz, petrol fiyatlarındaki artış, vergi avantajları, faiz oranları gibi dış faktörlerden etkilenmektedir. Otomobil endüstrisinin uzun vadeli tahminlerle ilgilendiği göz önüne alındığında, basit tek değişkenli modeller yeterli değildir. Çok değişkenli modeller araç satışlarını tahmin etmede daha iyi sonuçlar verebilmektedir. Derin öğrenmenin güçlü temsil yeteneği ve satış tahmini uygulamalarında kullanılması hem işletmeler hem de araştırmacılar tarafından büyük ilgi görmektedir. LSTM modelinin zaman serilerindeki başarısı göz önüne alınarak bu çalışmada çok değişkenli zaman serileri kullanılarak araç satış tahmini yapılmıştır. Çalışmada modelin girdileri olarak petrol fiyatı, işsizlik oranı, tüketici fiyat endeksi gibi ekonomik göstergeler kullanılmıştır. Sonuçlar LSTM'nin çok değişkenli zaman serilerinde tahmin doğruluğu açısından iyi performans sergilediğini göstermektedir.

The automotive industry is one of the most important industries in many countries. So forecasts about vehicle sales provide valuable information to the automotive industry and its suppliers. Automobile sales are affected by external factors such as market environment, economic crisis, increase in oil prices, tax advantages and interest rates. Given that the auto industry is interested in long-term predictions, simple univariate models are not enough. Multivariate models can give better results in predicting vehicle sales. The strong representational ability of deep learning and its use in sales forecasting applications are interests both businesses and researchers. In this study, considering the success of the LSTM model in time series, vehicle sales are estimated using multivariate time series. Economic indicators such as oil price, unemployment rate and consumer price index are used as inputs of the model in the study. The results show LSTM performs well with prediction accuracy in multivariate time series.

DOI: <https://doi.org/10.30783/nevsosbilen.987093>

Atıf/Cite as: Yurtsever, M. (2022). LSTM Yöntemi ile Ekonomik Göstergeler Kullanılarak Otomobil Satış Tahmini. *Nevşehir Hacı Bektaş Veli Üniversitesi SBE Dergisi*, 12(3), 1481-1492.

Giriş

Zaman serisi herhangi bir olaya ilişkin verilerin zamana bağlı olarak sıralı bir şekilde bir araya getirilmesidir. Zaman serisi analizi; zamana bağlı olarak önceden bilinen olayların, işlemlerin analiz edilmesi ve iç görüye dönüştürülmesi için kullanılan önemli bir tekniktir (Parmezan vd.,2019). Ardışık halde gözlemlenen değişen olaylardan elde edilen ilişkiler geleceği tahmin etme noktasında önemlidir.

Çok değişkenli zaman serisi, birden fazla zamana bağlı değişkeni olan bir zaman serisini ifade etmektedir (Nguyen vd., 2021). Bu durum tahminin sadece geçmiş verilere değil değişkenler arasındaki ilişkiye bağlı olmasıdır. Çok değişkenli zaman serisi verilerinde değişkenler arasında korelasyonlar vardır ve genellikle tek bir değişkeni modellemekten ziyade tüm ilgili değişkenleri birlikte modelleyerek daha iyi bir öngörü elde edilebilmektedir (Du vd., 2020). Girdi olarak uygun değişkenler kümesi seçmek önemlidir (Munkhdalai vd., 2019).

Zaman serisi analizi için literatürde farklı yöntemler bulunmaktadır. Doğrusal regresyon, üstel düzeltme ve otogresif entegre hareketli ortalama olarak adlandırılan ARIMA vb. modeller geleneksel istatistiksel modellerdir. Box ve Jenkins tarafından geliştirilen ARIMA modelleri tek değişkenli zaman serileri için en sık kullanılan yöntemdir. Doğrusal olmayan zaman serileri için bu modeller yeterli değildir (Zhang, 2003; Shen vd., 2020). Makine öğrenmesi ve yapay sinir ağları karmaşık girdi ve çıktılar arasındaki doğrusal olmayan ilişkileri anlamadaki üstün yetenekleri ile çok çeşitli uygulamalarda kullanılmaktadır (Topal, 2019; Nacar ve Erdebilli (b.d.rouyendegh), 2021). Derin öğrenmenin güçlü temsil yeteneği (Lecun vd., 2015) ile öne çıkması birçok alanda farklı uygulamaları beraberinde getirmiştir. Stokastik verilerden öğrenme ve eğilimleri tanıma kapasiteleri pazar tahminleri için geliştirilmiştir (Hatcher ve Yu, 2018).

Zamansal bağımlılıkları modellemek için tasarlanan tekrarlayan sinir ağı modeli (RNN) zaman serileri için bir yöntem olmakla birlikte gradyan sorunu bulunmaktadır (Karim vd., 2019). Hochreiter ve Schmidhuber (1997), bilgi akışını kontrol etmek için RNN katmanlarının içine geçit mekanizmasını ekleyen uzun kısa süreli bellek (LSTM) yöntemini sunmuşlardır. Bu, önceki birçok zaman adımının bilgisinin kalıcı olmasına yardımcı olan yöntem geleneksel bir RNN'ye kıyasla daha eski zaman adımlarının bilgilerinden ilişkilerin türetilmesini sağlamaktadır.

Üretim planlama süreçlerinde, uzun vadeli tahmin, üretim planlaması için uygun bir insan gücü seviyesi belirlemek için ve üretimin kapasitesinin belirlenmesi açısından kritiktir (Sa-ngasoongsong vd., 2012). Talep tahminlerindeki hatalar çoğu zaman maliyetlere ve gelir kaybına yol açmaktadır. Satış tahminini etkileyen birçok değişken bulunmaktadır. Satışların güvenilir bir şekilde tahmin edilmesi, iş stratejisinin kalitesini artırabilmektedir (Yan ve Tu, 2012). Etkili bir satış tahmin modeli, tedarik zinciri yönetiminin iyi yönetilmesine ve karı artırmaya yardımcı olmaktadır. Aksi durumu ise ürün birikimine, envanter sıkıntısına ve memnuniyetsiz müşteri taleplerine neden olabilmektedir (Shahabuddin, 2009; Nunnari ve Nunnari, 2017; Topal, 2019). Bununla birlikte, rekabetçi bir iş ortamında doğru talep tahmini kolay değildir (Abbasimehr vd., 2020).

Otomotiv sanayi birçok ülke için imalat sanayinin en önemli dallarından biridir (Sánchez ve Pérez, 2005). Bu endüstri, demir-çelik ve petrokimya endüstrisi gibi ana sanayi dalları ile yakın ilişki içindedir ve diğer sektörlerdeki teknolojik gelişmelerin itici gücüdür. Otomotiv endüstrisi, gelir bakımından dünyanın en önemli ekonomik sektörlerinden biridir. Uzun geliştirme ve üretim süreçleri ile karakterizedir ve bu nedenle yeni otomobillerin satışına ilişkin uzun vadeli tahminler, birçok paydaşına değerli bilgiler sağlamaktadır. Otomobil geliştiricileri, otomobil üreticileri, otomobil satıcıları ve pazarlamacılar otomobillere yönelik gelecekteki talep için güvenilir tahminlerinden yararlanmak istemektedirler.

Doğru satış tahmini, otomobil pazarı rekabetinde çok önemli bir rol oynamaktadır. Otomobil satışları, piyasa ortamı, ekonomik kriz, petrol fiyatlarındaki artış, vergi avantajları, faiz oranları gibi dış faktörlerden de etkilenmektedir. Otomobil endüstrisinin uzun vadeli tahminlerle ilgilendiği göz önüne alındığında, basit tek değişkenli modeller yeterli değildir ve tüm değişkenler için çok aşamalı öngörülerini üretmek için çok değişkenli modellere ihtiyaç vardır (Fantazzini ve Toktamyssoya, 2015). Araştırmacılar modellerinde tahmin doğruluğunu artırmak adına ekonomik göstergeleri kullanmaktadırlar (Wang vd., 2011).

Ekonomik göstergelerin satış tahmininde çok değişkenli zaman serilerinde kullanımı literatürde farklı çalışmalarda görülmektedir. Bazı çalışmalarda araç özellikleri ve arama indeksleri de kullanılmıştır. Tablo 1'de literatürde otomotiv sektörü üzerine yapılan satış tahmin çalışmaları verilmiştir.

Tablo 1. Otomotiv sektöründe Satış Tahmin Çalışmaları

| Referans | Metodoloji | Veri/Sıklık | İndikatör | Sonuç |
|-----------------------------------|--|-----------------------|---|-----------------------------|
| (Shahabuddin, 2009) | Regresyon | 1959-2006 / Aylık | Dayanıklı endüstriyel talep, dayanıklı kişisel tüketim, iskonto oranı, dayanıksız sanayi malları talebi, kişisel tüketim, GSMH, GSYİH, nüfus, önde gelen ekonomik göstergeler, M1 (likit para), M2 (M1 artı kısa vadeli yatırılan varlıklar) ve M3 (M2 ve tüm kurumsal fonlar). | R ² : %75 |
| (Wang vd., 2011) | Uyarlanabilir ağ tabanlı bulanık çıkarım sistemi | 2002-2009/Aylık | Çalışanların ortalama kazançları, petrol fiyatları, inşaat izinleri, İnşaat proje sayısı | MSE:9359,14 |
| (Sa-ngasoongsong vd., 2012) | VECM | 1975-2010/Aylık | Tüketici Fiyat Endeksi (TÜFE), İşsizlik Oranı, Gaz Fiyatları ve Konut Başlangıçları. | MAPE: 0,0135 RMSE:0,1416 |
| (Karaath vd., 2012) | Yapay Ağları | Sinir 2007-2011/Aylık | Gayri safi yurtiçi hasıla, reel kesim güven endeksi, yatırım harcamaları, tüketim harcamaları, tüketici güven endeksi, dolar kuru | MAPE: %16,82 |
| (Fantazzini ve Toktamysova, 2015) | Doğrusal olmayan modeller | 2001-2014/Aylık | Bina inşaatı, tüketici güven göstergeleri, tüketici fiyat endeksi, euro bankalararası teklif oranı, gayri safi yurtiçi hasıla, üretim endeksi, işsizlik oranı, petrol fiyatı | <%10 |
| (Zhang vd., 2017) | VAR | 2011-2015/Aylık | Tüketici fiyat endeksi, tüketici güven endeksi, üretici fiyat endeksi, perakende yakıt fiyatı, araç fiyatı, Baidu verileri | MAPE: 0,291 RMSE:0,1631 |
| (Gao vd., 2018) | ARMA, VAR, VECM | 2007-2016/Aylık | Tüketici güven endeksi, çelik üretimi, tüketici fiyat endeksi ve benzin fiyatı | MAPE:0,16-0,26-0,12 |
| (Pai ve Lui, 2018) | En küçük kareler destek-vektör makineleri | 2008-2017/Aylık | Tweetlerin duyarlılık skorları, borsa değerleri ve hibrit veriler | <%10 |

| | | | | | |
|------------------------------------|--|-------|-----------------|--|------------------------------|
| (Arslankaya ve Vildan, 2018) | Yapay Ağları, Regresyon, Basit üstel düzeltme, ortalama averaj | Sinir | 2011-2016/Aylık | Tüketici fiyat endeksi, gayrisafi yurt içi hasıla, dolar kuru, aylık çalışma saatleri, tüketici güven endeksi, reel sektör güven endeksi | MAPE: 7,44- 12,66-24,2-23,08 |
| (Kaya 2019) | vd., Yapay Ağları | Sinir | 2012-2017/Aylık | Otomobil fiyatı, Euro ve Dolar kur, istihdam oranı, tüketici güven endeksi, petrol fiyatları ve sanayi üretim güven endeksi, otomobil satın alma olasılığı, kadın istihdam oranı, genel ekonomik durum, genel ekonomik durum beklentisi, hanehalkının mali durumu ve hanehalkının mali durumu beklentisi | MAPE:%1,18 RMSE:782 |
| (Wachter 2019) | vd., Çok değişkenli regresyon analizi | | 2004-2019/Aylık | Tüketici güven endeksi, yeni araçlar için tüketici fiyat endeksi, benzin fiyatı, standard & Poor's 500 Endeksi | RMSE:11815 |
| (Xia vd., 2020) | ForeXGBoost | | 2013-2017/Aylık | Araç markası, modeli, gücü, yakıt tipi, silindir hacmi, vb. | MAPE:18,54067 |
| (Kuvvetli 2015) | vd., Yapay Ağları | Sinir | 2008-2012/Aylık | Geçmiş satış miktarı, taşıt kredi faiz oranı, gayrisafi milli hasıla, CO ₂ emisyonu, yakıt tüketim miktarları ve fiyatı | MSE:58.081,2 |
| (Kayapınar Kaya ve Yıldırım, 2020) | Derin Sinir Ağları | | 2011-2018/Aylık | Döviz kuru, gayrisafi yurt içi hasıla, tüketici güven endeksi, tüketici fiyat endeksi | MSE:79,83 |
| (Yang ve Zhang, 2020) | Makine öğrenmesi modelleri | | 2011-2019/Aylık | Baidu arama dizini, Weibo metin miktarı indexi | MAPE: %9,42 |

Bu makale, ekonomik göstergelere göre otomobil satışlarının tahmin edilmesi hakkında bazı yararlı bilgiler sağlamaktadır. Bu amaçla, otomobil satışlarını tahmin etmek için LSTM yöntemi kullanılmıştır. 2003-2021 yılları arasında Amerika'daki bazı ekonomik göstergeleri ve otomobil satış sayıları kullanılarak oluşturulan veri seti kullanılmıştır. Literatürde farklı yöntemler ile otomobil satışlarını tahmin edilmekle beraber kullanılan yeni değişkenler çalışmanın farklılığını ortaya koymaktadır. Bu çalışmada ekonomik göstergeler ile oluşturulan çok değişkenli zaman serisinin tahmin sürecine etkisi incelenmiştir. LSTM modeli eğitilirken hiper parametrelerin tahmin doğruluğu üzerindeki etkileri de araştırılmıştır.

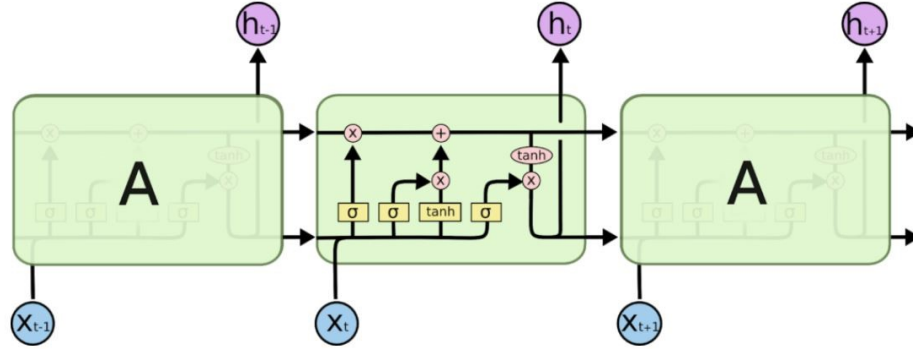
Çalışmanın giriş kısmında zaman serileri hakkında genel bilgi verildikten sonra literatürde çok değişkenli otomobil satış verileri ile yapılan çalışmalar incelenmiştir. İkinci bölümde çalışmada kullanılan metodoloji açıklanmıştır. Üçüncü bölümde çalışmada kullanılan veri seti tanıtılmıştır. Dördüncü bölümde çalışmadan elde edilen bulgular sunulmuştur. Son bölüm olan tartışma ve sonuç kısmında çalışmadan elde edilen sonuçlar literatür kapsamında tartışılarak teorik ve pratiğe yönelik katkıları açıklanmıştır.

Yöntem

LSTM

LSTM, gelişmiş bir RNN mimarisidir. LSTM, geleneksel RNN'lerden daha doğru olmasını sağlayan uzun menzilli bir bağımlılığa sahiptir. LSTM'nin temel yapısı, birim çıktılarını farklı zaman adımlarında hatırlamak ve açıkça yaymak için bir bellek hücresi kullanmasıdır. LSTM'nin bellek hücresi, zamansal bağlamların bilgilerini hatırlamak için hücre durumlarını kullanır. Aynı zamanda, farklı zaman adımları arasındaki bilgi akışını kontrol etmek için bir unutma geçidi, bir giriş geçidi ve bir çıkış geçidine sahiptir.

RNN'in yapısındaki uzun vadeli bağımlılıkları öğrenmeyle ilgili matematiksel zorluk, kaybolan gradyan problemi olarak adlandırılır. Giriş dizisinin uzunluğu arttıkça, ilk aşamaların etkisini yakalamak zorlaşır ve ilk birkaç giriş noktasına giden gradyanlar kaybolur ve sıfıra eşit olur. Bir unutma geçidi, bir giriş geçidi ve bir çıkış geçidinden oluşan LSTM modeli, RNN'nin dezavantajlarını ortadan kaldırmaktadır. LSTM'nin üç kapısı, uzun süreli belleğin düzenlenmesini kolaylaştırır (Yang vd., 2019). LSTM hücre bloğu, bu bileşenlere bağlı olarak bilgileri unutma veya hafızaya alma, bir sonraki hücreye ne kadar bilginin aktarılması gerektiğini belirleme yeteneğine sahip üç kapı ve bir bellek hücre birimi içerir. Şekil 1'de LSTM yapısı görülmektedir.



Şekil 1. LSTM Modelinin Yapısı (Olah, 2015)

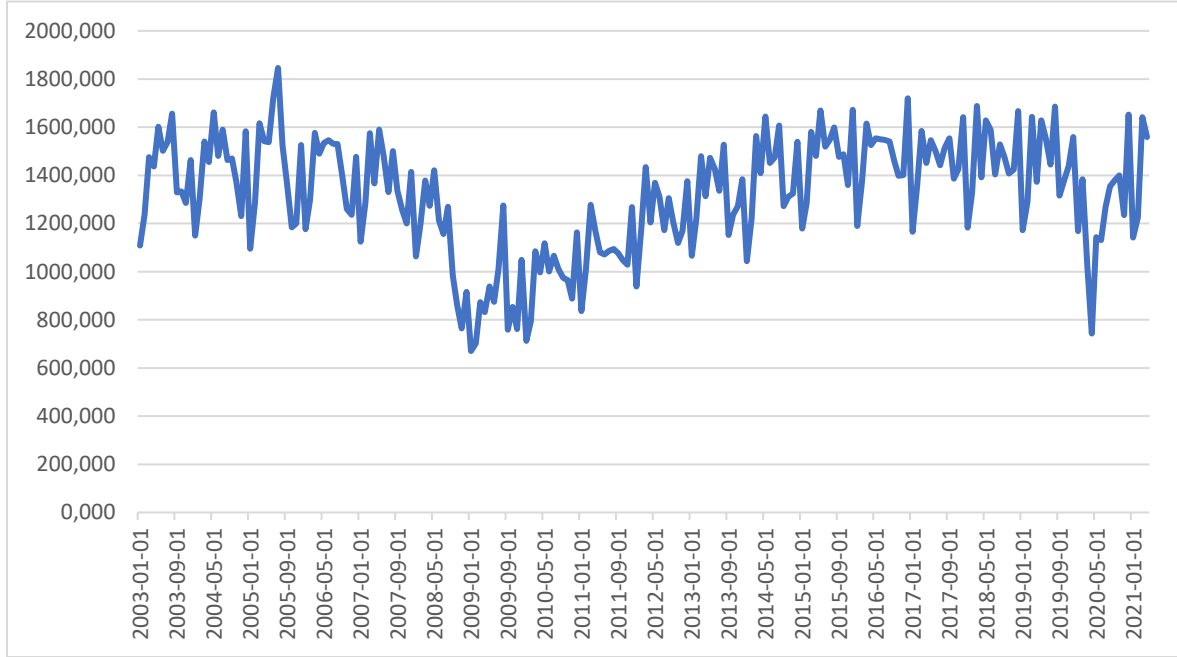
Girdi dizisi $x = (x_1, x_2, \dots, x_t)$ verildiğinde, ağ $y = (y_1, y_2, \dots, y_t)$ çıktısına bir eşleme hesaplar. Aşağıdaki denklemler birim etkinleştirmelerini belirler:

$$\begin{aligned}
 f_t &= \sigma(W_f * [h_{t-1}, x_t] + b_f) \\
 i_t &= \sigma(W_i * [h_{t-1}, x_t] + b_i) \\
 \hat{C}_t &= \tanh(W_C * [h_{t-1}, x_t] + b_C) \\
 C_t &= f_t * C_{t-1} + i_t * \hat{C}_t \\
 o_t &= \sigma(W_o * [h_{t-1}, x_t] + b_o) \\
 h_t &= o_t * \tanh(C_t)
 \end{aligned}$$

σ lojistik sigmoid fonksiyonudur. \dot{I} , f , o ve c kapıları sırasıyla giriş kapısı, unutmama kapısı, çıkış kapısı ve hücre aktivasyon vektörleridir. Tüm bu vektörler, h gizli vektörü ile aynı boyuta sahiptir. W terimleri, hücreden kapı vektörlerine ağırlık matrislerini belirtmektedir. \tanh , çıkış aktivasyon fonksiyonudur. LSTM ağında yaygın olarak kullanılan aktivasyon fonksiyonları sigmoid ve hiperbolik tanjanttır.

Veri Seti

Bu makalede ele alınan ana zaman serisi, Ocak 2003 - Nisan 2021 döneminde ABD'deki aylık otomobil satışları verileridir (U.S. Bureau of Economic Analysis, 2021). Çalışmada kullanılan ekonomik göstergeler; ham petrol fiyatı (U.S. Energy Information Administration, 2021), tüketici fiyat endeksi (Organization for Economic Co-operation and Development, 2021), gıda ve enerji hariç kişisel tüketim harcamaları (U.S. Bureau of Economic Analysis, 2021), 10 yıllık başa baş enflasyon oranı (Federal Reserve Bank of St. Louis, 2021), işsizlik oranı (U.S. Bureau of Labor Statistics, 2021), emtia bazında üretici fiyat endeksidir (U.S. Bureau of Labor Statistics, 2021). Şekil 2'de 18 yıllık dönemdeki araç satış grafiği görülmektedir.



Şekil 2. Otomobil Satış Grafiği

Araç satışlarını etkileyen çeşitli faktörlerin tespiti ve satış adetlerinin tahmin edilmesi üzerine çok sayıda ampirik çalışma bulunmaktadır. Çeşitli istatistiksel ve ekonometri analizleri içeren çalışmalar incelendiğinde birbirinden farklı açıklayıcı değişkenler olduğu görülmektedir.

Wang vd. (2011) çalışmalarında araç satışlarının petrol fiyatlarından etkilendiğini ifade etmektedir. Sa-Ngasoongsong vd. (2012), tüketici fiyat endeksi ve işsizlik oranını araç satışlarını tahmin modellemesinde kullanmıştır. Karaatlı vd. (2012), tüketim harcamaları, tüketici fiyat endeksini araç satış tahmin modelinde girdi olarak kullanmıştır. Kaya ve Yıldırım'ın (2020) çalışmalarında tüketici fiyat endeksi, gayrisafı yurtiçi hasıla, döviz kuru ve tüketici güven endeksi kullandıkları göstergelerdir. Ampirik modelleri kullanan sonuçlar, bu göstergelerin otomobil satışları üzerinde önemli bir etkiye sahip olduğunu göstermektedir. Çalışmada ek olarak üretici fiyat endeksi ve 10 yıllık başa baş enflasyon oranının da araç satışlarını etkilediği varsayılarak girdi değişkenleri olarak kullanılmıştır. Üretici fiyat endeksi üretim sürecinde girdi olarak kullanılan ürünlerin fiyat değişimleri ile ilgili olduğu için maliyetler ile birlikte satış fiyatlarını etkilemektedir. 10 yıllık başabaş enflasyon oranı, piyasaların 10

yıllık ufuktaki enflasyon beklentilerinin bir göstergesi olarak hizmet etmektedir. Çalışmada kullanılan değişkenler ile ilgili ayrıntılar Tablo 2'de gösterilmektedir.

Tablo 2. Veriler ile ilgili özet istatistikler

| Değişken | Ortalama | Standart Sapma | Maksimum | Minimum |
|--|----------|----------------|----------|---------|
| Ham petrol fiyatı | 66,09 | 23,78 | 133,88 | 16,55 |
| Tüketici fiyat endeksi | 0,18 | 0,38 | 1,22 | -1,92 |
| Gıda ve enerji hariç kişisel tüketim harcamaları | 99,46 | 8,74 | 116,02 | 84,29 |
| 10 yıllık başa baş enflasyon oranı | 2,04 | 0,4 | 2,71 | 0,25 |
| İşsizlik oranı | 6,18 | 2,04 | 14,40 | 3,30 |
| Emtia bazında üretici fiyat endeksi | 184,08 | 20,45 | 217,50 | 135,30 |

Veri setinin ilk %70lik kısmı eğitim için kullanılmıştır ve %30lık kısmı test amaçlı kullanılmıştır. Veri seti eğitim ve test için kullanılmadan önce normalize edilmiştir. Bu çalışmadaki tüm deneyler, Microsoft Windows 10 işletim sistemli bir bilgisayarda Python programlama dili 3.6 versiyonu kullanılarak yapılmıştır. LSTM modeli, derin öğrenmenin "Keras" kütüphanesi kullanılarak uygulanmıştır. Bu makalenin doğrulama deneyleri sırasında, seçilen hiper parametreler Tablo 3'de belirtilmiştir. Spesifik olarak bu parametreler Tablo 3'de belirtilmiştir.

Tablo 3. LSTM modellerinde kullanılan hiper parametreler

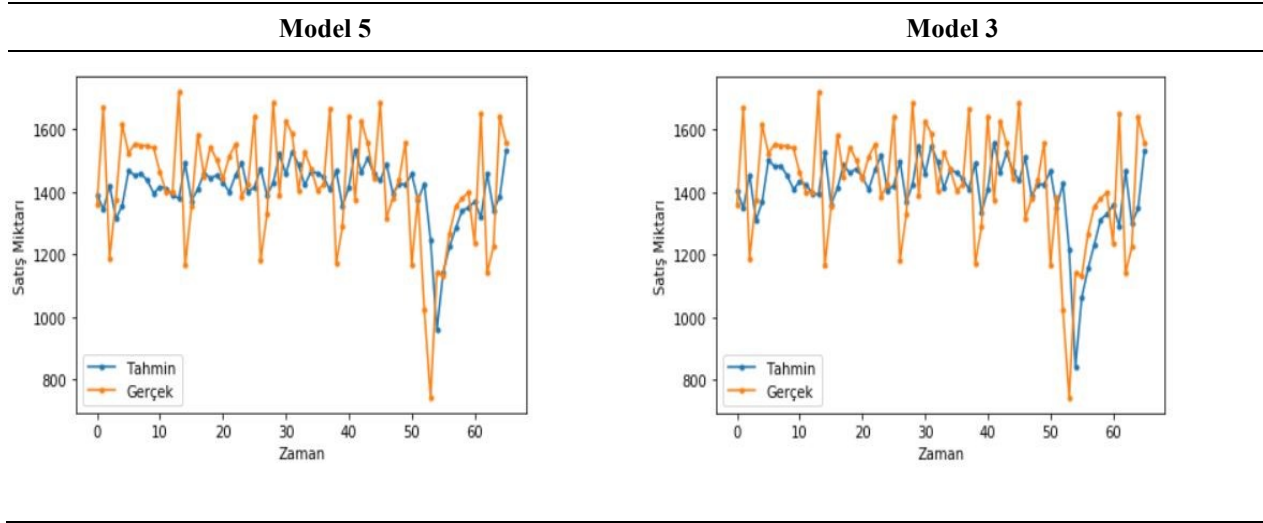
| Model No | Gizli katman sayısı | Gizli katman nöron sayısı | Parti boyutu | Döngü Sayısı |
|----------|---------------------|---------------------------|--------------|--------------|
| 1 | 2 | 16,8 | 32 | 100 |
| 2 | 3 | 32,16,8 | 32 | 100 |
| 3 | 2 | 16,8 | 64 | 100 |
| 4 | 3 | 32,16,8 | 64 | 100 |
| 5 | 2 | 16,8 | 128 | 100 |
| 6 | 3 | 32,16,8 | 128 | 100 |
| 7 | 2 | 16,8 | 128 | 250 |
| 8 | 2 | 16,8 | 128 | 500 |
| 9 | 2 | 16,8 | 128 | 1000 |

Bulgular

Deneyisel çalışmalar Ocak 2003 - Nisan 2021 yılları arasında elde edilen ekonomik göstergelere ve araç satışlarına ilişkin 220×7 boyutlarında bir veri seti kullanılmıştır. Modelin girdisi olarak kullanılan ekonomik göstergeler doğrultusunda otomobil satış tahmini yapılmıştır. Çalışmada farklı hiper parametrelerin sonuca etkisi de test edilmiştir. Parti boyutu, gizli katman sayısı ve döngü sayısının tahmin doğruluğu üzerinde direkt etkisi olduğu görülmektedir.

Modeli eğittikten sonra, test veri seti ile modelin başarısı ölçülmüştür. Tablo 4 gerçek satışlar ile LSTM kullanılarak öngörülen satışlar arasındaki karşılaştırmayı göstermektedir. Tabloda en iyi iki sonucu veren Model 5 ve Model 3'ün grafikleri gösterilmiştir. Grafiklere bakıldığında tahmin edilen değerlerin gerçek satışların değişen trendini yakaladığı görülmektedir.

Tablo 4. Modellere ait tahmin grafikleri



Bu modellerin performansını değerlendirmek için, ortalama mutlak yüzde hata (MAPE) ve kök ortalama kare hata (RMSE) olmak üzere iki gösterge kullanılmıştır. MAPE, tahmin edilen ve gerçek değerlerin ne kadar yakın olduğunu göstermek için ortalama mutlak hatanın hesaplanmasıdır. Ft, tahmin değeri olduğunda, Yt gerçek değerdir ve n, zaman serilerinin uzunluğudur. MAPE her hatayı talebe göre ayrı ayrı böler, bu nedenle çarpıktır. Düşük talep dönemlerindeki yüksek hatalar MAPE'yi önemli ölçüde etkilemektedir.

$$MAPE = \left(\sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \right) \frac{100}{n}$$

MSE, gerçek gözlemlenen değer ile tahmini değer arasındaki farktır, RMSE ise MSE değerinin kareköküdür. . Daha düşük RMSE değerleri daha iyi uyumu gösterir. RMSE, modelin ne kadar doğru tahmin ettiğinin iyi bir ölçüsüdür ve modelin temel amacı tahmin ise, uyum için en önemli kriterdir. Tablo 5'de çalışmada uygulanan deneysel modellerin sonuçları görülmektedir.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Tablo 5. Modellere ait MAPE ve RMSE sonuçları

| MODEL NO | MAPE | RMSE |
|----------|--------------|----------------|
| 1 | 11,55 | 197,729 |
| 2 | 13,02 | 218,256 |
| 3 | 10,55 | 184,837 |
| 4 | 12,79 | 215,164 |
| 5 | 10,15 | 178,459 |
| 6 | 12,45 | 209,764 |
| 7 | 13,17 | 219,880 |
| 8 | 14,96 | 242,759 |
| 9 | 19,12 | 303,600 |

Çalıştırılan dokuz model içerisinde beşinci model en düşük RMSE ve MAPE değerini elde ederek en iyi sonucu vermiştir. Dokuzuncu model en başarısız model olarak gözükmektedir. İlk altı model oluşturulurken döngü sayısı sabit tutularak gizli katman sayısı ve parti boyutları değiştirilmiştir. Bu altı model içinde en iyi sonucu veren beşinci modelin gizli katman sayısı (2) ve parti boyutu (128) sabit tutularak döngü sayılarının sonuca etkisini gözlemek için üç model (7.,8.,9. model) daha oluşturulmuştur.

Çalışmada döngü sayısı arttıkça modelin başarısının düştüğü tespit edilmiştir. Yüz döngü sayısının iki yüz elli, beş yüz ve bin döngü sayısına göre daha iyi sonuç verdiği tespit edilmiştir. En iyi sonucu veren beşinci model de parti boyutu da yüz yirmi sekiz olarak belirlenmiştir.

Tartışma, Sonuç ve Öneriler

Bu çalışmada, araç satışlarının gelecek aylara yönelik satış sayısını tahmin etmek için derin öğrenme yöntemlerinden LSTM modeli kullanılmıştır. Çalışmada Ocak 2003 ile Nisan 2021 arasında aylık veriler kullanılmıştır. Çalışmada kullanılan ekonomik göstergeler; ham petrol fiyatı, tüketici fiyat endeksi, gıda ve enerji hariç kişisel tüketim harcamaları, 10 yıllık başa baş enflasyon oranı, işsizlik oranı, emtia bazında üretici fiyat endeksidir. Tahmin edilen değerler gerçek değerlerle karşılaştırıldığında LSTM yönteminde tahmin edilen ve gerçek değerlerin birbirine yakın olduğu görülmüştür. Çalışmada ekonomik göstergelerin araç satış sayılarını etkilediği görülmüştür. Bu bulgu önceki araştırmaların sonuçlarıyla tutarlı gözükmektedir (Fantazzini ve Toktamsoya, 2015; Zhang vd., 2017; Gao vd., 2018).

Türkiye’deki otomobil satış tahmin çalışmaları incelendiğinde dolar kuru ve gayri safi yurtiçi hasılanın (Kayapınar Kaya ve Yıldırım, 2020), marka ve CO2 emisyonunun (Kuvvetli vd., 2015) gayri safi yurtiçi hasıla, reel kesim güven endeksi, tüketici güven endeksi, yatırım harcamaları, tüketim harcamaları, tüketici fiyat endeksi, araç alım satımı, doların (Karaatlı vd., 2012); gayri safi yurtiçi hasıla, tüketici fiyat endeksi, dolar kuru, reel sektör güven endeksi, tüketici güven endeksinin (Arslankaya ve Vildan, 2018); dolar kurundaki değişimler, hanehalkının finansal durum beklentisi, mevsimsellikten arındırılmış sanayi üretim endeksi ve bir ay öncesi otomobil satışlarının logaritmik şeklinin (Kaya vd., 2019) en önemli değişkenler olduğu görülmektedir. Bu çalışmada kullanılan tüketici fiyat endeksi, üretici fiyat endeksi, kişisel tüketim harcamaları Türkiye’deki çalışmalarda kullanılan değişkenler ile ortaktır. Ham petrol fiyatı, 10 yıllık başa baş enflasyon oranı, işsizlik oranı değişkenleri ise bu çalışma özelinde farklı değişkenler olarak belirlenmiştir.

LSTM modelleri eğitilirken seçilen hiper parametrelerin tahmin sonuçları üzerindeki etkisi de ölçülmüştür. Bu çalışmada parti boyutu, gizli katman sayısı ve döngü sayısının da model başarısını direkt etkilediği görülmektedir. Bu parametreler genelde başlangıçta belli değildir ve seçimi modeli tasarlayan kişi belirlemektedir. Parametre ayarı LSTM modelinin performansı için çok önemlidir, bu nedenle yetersiz eğitim ve aşırı uyumdan kaçınmak için devir sayısına dikkat edilmelidir (Song vd., 2020).

Satışları tahmin etmek işletmeler için kritiktir çünkü etkili satış tahmini, bir karar destek sisteminde hedeflere zamanında ulaşmaya yardımcı olabilmektedir (Lee vd., 2012). Tersine zayıf tahmin, bir işletmenin rekabet gücüne ve karlılığına zarar vermektedir (Ramos vd., 2015). Bu açıdan çalışma karar vericilerin belirsizliğini azaltma noktasında önemlidir.

Uygulanan yöntem ve sonuçlar tatmin edici olsa da, kullanılan ekonomik değişkenlerin sayısı nedeniyle hala sınırlıdır. Bu nedenle gelecekteki çalışma, otomobil satışlarını etkileyen daha fazla ekonomik gösterge kullanılarak oluşturulan hibrit modeller ile yapılabilir.

Kaynakça

- Abbasimehr, H., Shabani, M., & Yousefi, M. (2020). An optimized model using LSTM network for demand forecasting. *Computers & industrial engineering*, 143.
- Arslankaya, S., & Öz, V. (2018). Time series analysis of sales quantity in an automotive company and estimation by Artificial Neural Networks. *Sakarya Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 22(5), 1482-1492.
- Du, S., Li, T., Yang, Y., & Horng, S. J. (2020). Multivariate time series forecasting via attention-based encoder-decoder framework. *Neurocomputing*, 388, 269-279.
- Fantazzini, D., & Toktamysova, Z. (2015). Forecasting German car sales using Google data and multivariate models. *International Journal of Production Economics*, 170, 97-135.
- Federal Reserve Bank of St. Louis. (2021). *10-Year Breakeven Inflation Rate [T10YIEM]*. Federal Reserve Bank of St. Louis. <https://fred.stlouisfed.org/series/T10YIEM>
- Gao, J., Xie, Y., Cui, X., Yu, H., & Gu, F. (2018). Chinese automobile sales forecasting using economic indicators and typical domestic brand automobile sales data: A method based on econometric model. *Advances in Mechanical Engineering*, 10(2), 1-11.
- Hatcher, W. G., & Yu, W. (2018). A survey of deep learning: Platforms, applications and emerging research trends. *IEEE Access*, 6, 24411-24432.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.
- Karaatlı, M., Helvacıoğlu, Ö. C., Ömürbek, N., & Tokgöz, G. (2012). Yapay Sinir Ağları Yöntemi ile Otomobil Satış Tahmini. *Uluslararası Yönetim İktisat ve İşletme Dergisi*, 8(17), 87-100.
- Karim, F., Majumdar, S., Darabi, H., & Harford, S. (2019). Multivariate LSTM-FCNs for time series classification. *Neural Networks*, 116, 237-245.
- Kaya, A., Kaya, G., & Çebi, F. (2019). Forecasting automobile sales in Turkey with artificial neural networks. *International Journal of Business Analytics (IJBAN)*, 6(4), 50-60.
- Kayapınar Kaya, S. K., & Yıldırım, Ö. (2020). A Prediction Model For Automobile Sales In Turkey Using Deep Neural Networks. *Endüstri Mühendisliği*, 31(1), 57-74.
- Kuvvetli, Y., Dağsuyu, C., & Oturakci, M. (2015). Türkiye'deki Araç Satışları İçin Ekonomik ve Çevresel Faktörleri Göz Önüne Alan Yapay Sinir Ağı Tabanlı Bir Tahmin Yaklaşımı. *Endüstri Mühendisliği*, 26(3), 23-31.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- Lee, W. I., Shih, B. Y., & Chen, C. Y. (2012). Retracted: A hybrid artificial intelligence sales-forecasting system in the convenience store industry. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 22(3), 188-196.
- Munkhdalai, L., Munkhdalai, T., Park, K. H., Amarbayasgalan, T., Batbaatar, E., Park, H. W., & Ryu, K. H. (2019). An end-to-end adaptive input selection with dynamic weights for forecasting multivariate time series. *IEEE Access*, 7, 99099-99114.
- Nacar, E. N. & Erdebili (b.d.rouyendegh), B. (2021). Makine Öğrenmesi Algoritmaları ile Satış Tahmini. *Endüstri Mühendisliği*, 32 (2) , 307-320.
- Nguyen, H. D., Tran, K. P., Thomassey, S., & Hamad, M. (2021). Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management. *International Journal of Information Management*, 57, 1-13.
- Nunnari, G., & Nunnari, V. (2017). Forecasting monthly sales retail time series: a case study. *In 2017 IEEE 19th conference on business informatics (CBI)*, 1, 1-6.
- Olah, C. (2015). *Understanding lstm networks*. <https://colah.github.io/posts/2015-08-Understanding-LSTMs>.

- Organization for Economic Co-operation and Development. (2021). *Consumer Price Index: Total All Items for the United States* [CPALTT01USM657N]. Federal Reserve Bank of St. Louis. <https://fred.stlouisfed.org/series/CPALTT01USM657N>
- Pai, P. F., & Liu, C. H. (2018). Predicting vehicle sales by sentiment analysis of Twitter data and stock market values. *IEEE Access*, 6, 57655-57662.
- Parmezan, A. R. S., Souza, V. M., & Batista, G. E. (2019). Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. *Information sciences*, 484, 302-337.
- Ramos, P., Santos, N., & Rebelo, R. (2015). Performance of state space and ARIMA models for consumer retail sales forecasting. *Robotics and computer-integrated manufacturing*, 34, 151-163.
- Sa-ngasoongsong, A., Bukkapatnam, S. T., Kim, J., Iyer, P. S., & Suresh, R. P. (2012). Multi-step sales forecasting in automotive industry based on structural relationship identification. *International Journal of Production Economics*, 140(2), 875-887.
- Sánchez, A. M., & Pérez, M. P. (2005). Supply chain flexibility and firm performance: a conceptual model and empirical study in the automotive industry. *International Journal of Operations & Production Management*, 25(7), 681-700.
- Shahabuddin, S. (2009), "Forecasting automobile sales", *Management Research News*, 32(7), 670-682.
- Shen, Z., Zhang, Y., Lu, J., Xu, J., & Xiao, G. (2020). A novel time series forecasting model with deep learning. *Neurocomputing*, 396, 302-313.
- Song, X., Liu, Y., Xue, L., Wang, J., Zhang, J., Wang, J., ... & Cheng, Z. (2020). Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model. *Journal of Petroleum Science and Engineering*, 186.
- Topal, İ. Çevrimiçi Tüketici Bütünleşmesi Ve Arama Motoru Verileri Kullanılarak Yapay Sinir Ağları İle Otomobil Satış Tahmini. *Neşehir Hacı Bektaş Veli Üniversitesi SBE Dergisi*, 9(2), 534-551.
- U.S. Bureau of Economic Analysis. (2021). *Total Vehicle Sales* [TOTALNSA]. Federal Reserve Bank of St. Louis. <https://fred.stlouisfed.org/series/TOTALNSA>.
- U.S. Bureau of Economic Analysis. (2021). *Personal Consumption Expenditures Excluding Food and Energy (Chain-Type Price Index)* [PCEPILFE]. Federal Reserve Bank of St. Louis. <https://fred.stlouisfed.org/series/PCEPILFE>.
- U.S. Energy Information Administration. (2021). *Crude Oil Prices: West Texas Intermediate (WTI) - Cushing, Oklahoma* [DCOILWTICO]. Federal Reserve Bank of St. Louis. <https://fred.stlouisfed.org/series/DCOILWTICO>.
- U.S. Bureau of Labor Statistics. (2021). *Unemployment Rate* [UNRATENSA]. Federal Reserve Bank of St. Louis. <https://fred.stlouisfed.org/series/UNRATENSA>.
- U.S. Bureau of Labor Statistics. (2021). *Producer Price Index by Commodity: All Commodities* [PPIACO]. Federal Reserve Bank of St. Louis. <https://fred.stlouisfed.org/series/PPIACO>.
- Xia, Z., Xue, S., Wu, L., Sun, J., Chen, Y., & Zhang, R. (2020). ForeXGBoost: passenger car sales prediction based on XGBoost. *Distributed and Parallel Databases*, 38, 713-738.
- Wachter, P., Widmer, T., & Klein, A. (2019). Predicting automotive sales using pre-purchase online search data. *In 2019 Federated Conference on Computer Science and Information Systems (FedCSIS)*, 18, 569-577.
- Wang, F. K., Chang, K. K., & Tzeng, C. W. (2011). Using adaptive network-based fuzzy inference system to forecast automobile sales. *Expert Systems with Applications*, 38(8), 10587-10593.
- Yan, H. S., & Tu, X. (2012). Short-term sales forecasting with change-point evaluation and pattern matching algorithms. *Expert systems with applications*, 39(5), 5426-5439.
- Yang, B., Sun, S., Li, J., Lin, X., & Tian, Y. (2019). Traffic flow prediction using LSTM with feature enhancement. *Neurocomputing*, 332, 320-327.
- Yang, Z., & Zhang, C. (2020, July). Automobile Sales Forecast Based on Web Search and Social Network Data. *The 11th International Conference on E-business, Management and Economics*, 37-41.
- Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159-175.
- Zhang, Y., Zhong, M., Geng, N., & Jiang, Y. (2017). Forecasting electric vehicles sales with univariate and multivariate time series models: The case of China. *PLoS one*, 12(5), 1-15.

EXTENDED SUMMARY

Automotive industry is one of the most important industrial branches for many countries. For this reason, forecasts for vehicle sales provide valuable information to the automotive industry and its suppliers. Automobile sales are affected by external factors such as the market environment, economic crisis, increase in oil prices, tax advantages and interest rates. Given that the auto industry is concerned with long-term forecasting, simple univariate models are not enough. Multivariate models can give better results in predicting vehicle sales. The use of deep learning in strong representation and sales forecasting applications is of great interest to both businesses and researchers. Considering the success of the LSTM model in time series, vehicle sales forecasting was made using multivariate time series in this study.

LSTM is an advanced RNN architecture. LSTM has a long range dependency that makes it more accurate than traditional RNNs. The basic nature of LSTM is that it uses a memory cell to remember and explicitly span unit outputs in different time steps. LSTM's memory cell uses cell states to remember information of temporal contexts. It also has a forget gate, an input gate and an output gate to control the flow of information between different time steps.

The mathematical difficulty associated with learning the long-term dependencies in the structure of the RNN is called the vanishing gradient problem. As the length of the input sequence increases, it becomes harder to capture the effect of the first stages and the gradients to the first few entry points disappear and become equal to zero. The LSTM model, which consists of a forget gate, an input gate, and an output gate, eliminates the disadvantages of RNN.

The first 70% of the data set was used for training and 30% was used for testing. The dataset was normalized before being used for training and testing. All experiments in this study were performed using Python programming language version 3.6 on a computer with Microsoft Windows 10 operating system. The LSTM model has been implemented using the "Keras" library of deep learning.

Experimental studies A 7×220 sized data set on economic indicators and vehicle sales obtained between January 2003 and April 2021 was used. In line with the economic indicators used as the input of the model, automobile sales were estimated. In the study, the effect of different hyperparameters on the result was also tested. It is seen that batch size, number of hidden layers and number of cycles have a direct effect on the prediction accuracy. After training the model, the success of the model was measured with the test dataset. Looking at the graphs, it is seen that the estimated values catch the changing trend of real sales. Two indicators, mean absolute percent error (MAPE) and root mean square error (RMSE), were used to evaluate the performance of these models. Among the six models that were run, the fifth model achieved the lowest RMSE and MAPE values and gave the best results. The ninth model seems to be the most unsuccessful model. While creating the first six models, the number of hidden layers and batch sizes were changed by keeping the number of loops constant. Three more models (7th, 8th, 9th models) were created to observe the effect of cycle numbers on the result by keeping the number of hidden layers (2) and batch size (128) constant for the fifth model, which gave the best results among these six models.

In the study, it was determined that the success of the model decreased as the number of cycles increased. It has been determined that the number of hundred cycles gives better results than the number of two hundred and fifty, five hundred and one thousand cycles. In the fifth model that gave the best results, the lot size was determined as one hundred and twenty-eight.

In this study, the LSTM model, one of the deep learning methods, was used to predict the number of vehicle sales for the next months. Monthly data from January 2003 to April 2021 were used in the study. Economic indicators used in the study; crude oil price, consumer price index, personal consumption expenditures excluding food and energy, 10-year break-even inflation rate, unemployment rate, producer price index on commodity basis. When the estimated values were compared with the actual values, it was seen that the estimated and actual values in the LSTM method were close to each other. While the LSTM models were trained, the effect of the selected hyperparameters on the prediction results was also measured. In this study, it is seen that the batch size, the number of hidden layers and the number of loops also directly affect the model success. These parameters are usually not clear at the beginning and the person who designed the model determines the selection. Parameter setting is very important for the performance of the LSTM model. Although the applied methods and results are satisfactory, they are still limited due to the number of economic variables used. Therefore, future work can be done with hybrid models created using more economic indicators that affect automobile sales.