

LSTM Derin Öğrenme Yaklaşımı ile Hava Kalitesi Verilerinin Tahmini: Sakarya Örneği

^{1*}Ayşegül ATALI, ^{1,2}Beytullah EREN, ³Caner ERDEN, ⁴Gökhan ATALI

¹Mühendislik Fakültesi, Çevre Mühendisliği Bölümü, Sakarya Üniversitesi, Sakarya, Türkiye

²Halfeti Meslek Yüksekokulu, Harran Üniversitesi, Halfeti, Şanlıurfa, Türkiye

³Uygulamalı Bilimler Fakültesi, Uluslararası Ticaret ve Finansman Bölümü, Sakarya Uygulamalı Bilimler Üniversitesi, Sakarya, Türkiye

⁴Teknoloji Fakültesi, Mekatronik Mühendisliği Bölümü, Sakarya Uygulamalı Bilimler Üniversitesi, Sakarya, Türkiye

Özet

Son yıllarda, dünya çapındaki birçok kentsel alan, ciddi hava kirliliğine ve bununla birlikte gelen sağlık tehlikelerine maruz kaldığından dolayı hava kalitesi tahmini ve önleyici ve düzeltici önlemler alınması önemli bir durum haline gelmiştir. Bu çalışmada Sakarya ilinde hava kalitesi istasyonlarında izlenen hava kalitesi parametreleri göz önüne alınarak, makine öğrenimi ile hava kalitesi tahmini yapılmıştır. Bu amaçla hava kalitesi parametrelerinden PM₁₀, CO, NO_x ve NO₂ konsantrasyonlarının tahmini için Uzun Kısa Süreli Bellek (LSTM) derin öğrenme yaklaşımı uygulanmış ve performans sonuçları elde edilmiştir. Sonuç olarak uygulanan yöntemin iyi bir tahmin performansına sahip olduğu görülmüştür.

Anahtar kelimeler: Hava kirliliği, hava kalitesi tahmin, derin öğrenme, uzun kısa süreli bellek (LSTM)

Abstract

In recent years, air quality forecasting and taking preventive and corrective measures have become important as many urban areas worldwide are exposed to severe air pollution and the health hazards that come with it. In this study, air quality estimation was made with machine learning, considering the air quality parameters monitored at air quality stations in Sakarya province. For this purpose, the Long Short-Term Memory (LSTM) deep learning approach was applied to estimate PM10, CO, NOX, and NO2 concentrations from air quality parameters, and performance results were obtained. As a result, it has been seen that the applied method has good estimation performance.

Keywords: Air pollution, air quality prediction, deep learning, long short-term memory (LSTM)

1. Giriş

Hava kirliliği, insan sağlığını, yaşamsal faaliyetler için hayati önem taşıyan doğal döngüleri ve ekosistemi olumsuz etkilemektedir. Türkiye’de hava kirliliğinin insan sağlığına olumsuz etkileri ile ilgili birçok çalışma yapılmıştır. Sakarya sanayi alanında her geçen gün gelişmekte olan bir şehirdir, bu durum Sakarya’nın jeopolitik konumundan ve gelişime elverişli geniş alanlara sahip olmasından kaynaklanmaktadır. Sanayinin gelişimi şehrin göç almasına sebep olmakta ve bu da hava kirliliğine neden olan bir diğer etken haline gelmektedir.

Hava kalitesi parametrelerinin ulusal hava kalitesi standartları içerisinde tutulabilmesi için baca gazı emisyonlarında çeşitli önlemlerin alınmış olsa da nüfus artışına bağlı olarak kirleticilerin salınımında artış gözlemlenmektedir. Bu durum havada kükürt dioksit (SO_2), azotoksitler (NO_x), hidrokarbon, partikül madde (PM_{10}) gibi zararlı emisyonların artmasına sebep olmaktadır. Bu tip kirletici emisyonların kaynakları değişkenlik göstermektedir her biri için farklı sınıflandırma yapılabilmek mümkündür.

1.1. Kirletici kaynaklar

Bu çalışmada yer verilen, hava kalitesi değerlerinin değişimine sebep olan kirletici parametreler, kirleticilerin kaynakları ve sebep oldukları etkiler Tablo 1’de gösterilmiştir [1].

Tablo 1. Temel Kirletici Parametreler

Parametre Türü	Kaynakları	Etkileri
Karbonmonoksit (CO)	Taşıt emisyonları, orman yangınları, fosil yakıt kullanımı	Oksijen taşıma kapasitesinin azalması, solunum ve akciğer hastalıkları
Azotoksitler (NO_x)	Yanma reaksiyonları, Patlayıcı madde imalatı ve imhası	Akciğer işlevlerinin bozulması, solunum yolları hastalıkları
Partikül madde (PM_{10})	Endüstriyel salınım, yakıt kullanımı ve kimyasal reaksiyonlar	Kanser, dolaşım ve solunum yolu rahatsızlıkları, bebek ölümleri

1981 yılında Recihle ve arkadaşlarının yaptığı bir çalışmada uydu üzerinden yapılan hava kirliliği ölçümü sonrası, CO kontrasyonunun ne kadar yüksek seviyelere ulaştığı ve bu durumun küresel bir tehlike teşkil ettiği ortaya çıkmıştır [2]. Bölgesel ve küresel hava kalitesi bileşenlerinin olumsuz olarak etkilenmesinin sebebinin, yalnızca endüstriyel kaynaklı olmadığını aynı zamanda biyokütle yanmasıyla da (orman yangınları, bitkisel yakıtlar ve tarımsal atıkların yakılması) ortaya çıkabildiğini kanıtlanmıştır. Ayrıca yapılan bazı çalışmalarda yanmadan kaynaklanan dumanın hava kalitesinin olumsuz değişimlerine sebep olduğu, ulusal hava kalitesi standartlarının aşılmasına ve kırsal nüfus üzerinde önemli sağlık etkilerine yol açtığı gözlemlenmiştir [3].

Kirletici parametrelerin bir diğeri olan Azot oksitler (NO_x), doğal kaynaklardan, motorlu taşıtlardan ve diğer yakıt yakma işlemlerinden üretilen gazlar olan azot monoksit (NO) ve azot dioksit (NO_2) karışımını tanımlamaktadır. NO renksizdir ve atmosferde oksitlenerek azot dioksitleri oluşturur. NO_2 ’nin bir kokusu vardır ve sağlığımızı ve çevremizi etkileyebilecek asidik ve oldukça aşındırıcı bir gazdır. Yüksek azot dioksit seviyeleri, insan solunum yollarına zarar verebilir ve bir kişinin solunum yolu enfeksiyonları ve astıma karşı hassasiyetini ve şiddetini

artırabilir. Yüksek konsantrasyonlarda azot dioksite uzun süre maruz kalmak kronik akciğer hastalığına neden olabilir. Aynı zamanda, örneğin bir kişinin koku alma yeteneğini azaltarak duyuları da etkileyebilir. Ayrıca yüksek konsantrasyonlarda azot dioksit bitki örtüsüne zararlıdır; yapraklara zarar verir, büyümeyi azaltır veya mahsul verimini düşürür. Benzer şekilde azot dioksit, mobilyaların ve kumaşların solmasına neden olur ve rengini bozabilir, görünürlüğü azaltabilir ve yüzeylerle reaksiyona girebilir [4]. Atmosferdeki diğer bir kirletici parametre olan PM₁₀ endüstriyel salınımlardan, motorlu taşıtlardan ve fosil yakıt kullanımından kaynaklanmakta ve atmosfere yayılmaktadır [5]. PM₁₀ üzerine yapılan çalışmalar, hava kalitesi için bu parametrenin önemli olduğunu ve iklim değişiklikleri ile biyojeokimyasal döngülere etkisinin olduğunu vurgulamaktadır [6].

Hava kirliliğine sebep olan kirletici emisyon ve bunların etkileri hakkında birçok çalışma yapılmıştır ve bu çalışmaların asıl amacı hava kirliliğini azaltmaya yardımcı olabilmektir. Hava kirliliği, kirletici emisyonlarının yönetimi ve mevcut kirliliğin azaltılması gibi birçok etkenden dolayı karmaşık bir meseledir. Bu sebeple hava kirliliği sorunlarına etkili bir yaklaşımda bulunabilmek için öncelikle buna neden olan kaynakların iyi anlaşılması ve hava kalitesindeki parametrelerin artış ve azalış değişiminin analizinin doğru bir şekilde yapılabilmesi, bu değişimlerin insanlar ve ekosistemler üzerindeki etkisinin iyi anlaşılması gerekmektedir.

1.2. Derin öğrenme algoritmaları ve hava kalitesi tahmini

Dünyamız gelişmekte ve her geçen gün yeni mühendislik uygulamalarıyla birlikte birçok yeniliği de beraberinde getirmektedir. Hava kalitesi verilerinin ölçümü de bu tür teknolojik gelişmeler ile kolaylaşmıştır. Ancak nesnel cihaz ölçümleri yanıltıcı olabilmekte oluşabilecek hatalar değerlendirmeleri güçleştirmekte ve geleceğe yönelik tahminler noktasında yetkilileri çözüm arayışına itmektedir [7].

Son zamanlarda, hava kalitesi verilerinin önceden tahmini için yapay zekâ giderek daha yaygın bir şekilde kullanılmaya başlanmıştır. Bunun nedeni, her geçen gün artan veri sayısının anlamlandırılması ve yorumlanmasında yapay zekanın sağladığı kolaylıklardır. Özellikle yapay zekâ alanlarından birisi olan derin öğrenme ile kentsel hava kalitesi konsantrasyonu tahmini çalışmaları, disiplinler arası araştırmalarda popüler hale gelmiştir [8].

Zaman serisi tahmini, geçmiş verileri kullanarak belirli bir dizinin değerlerini tahmin etmektedir. Son yıllarda, geleneksel zaman tahmin yöntemleri zaman alıcı ve karmaşık sonuçlar ortaya çıkarmaktadır. Bu sebeple daha akılcı ve tutarlı veri tahmini için araştırmacılar makine öğrenimini ele almakta ve bu konuda yapılan çalışmalara dikkat çekmektedir [9]. Derin öğrenme yapay sinir ağlarına dayalı bir makine öğrenmesi konseptidir. Birçok uygulama için derin öğrenme modelleri, sıg makine öğrenimi modellerinden ve geleneksel veri analizi yaklaşımlarından daha iyi performans göstermektedir [10]. Derin öğrenme yapay sinir ağları yaklaşımlarının arasında ileri düzey bir yöntemdir. Tek katmanlı yapay sinir ağları gelişerek, çok katmanlı sinir ağlarına ve daha da gelişerek derin öğrenme algoritmaları için gerekli olan yapıya dönüşmüştür [11]. Derin öğrenme sayesinde, verilerin aynı türden ve birimden tek bir özelliğini öğrenmekten ziyade, alt düzeyde verilerin birçok özelliğinin bileşiminden oluşan hiyerarşi ile üst düzey verilerin hiyerarşisini

öğrenebilmek mümkün olmaktadır. Üst düzey veri setinin özellikleri, alt düzey veri setinin özelliklerinden türetilmektedir. Bu şekilde türetilen üst düzey veriler birden çok veri özelliğini temsil etmektedir. Bu bağlamda, veri miktarı arttıkça ve makine öğrenimi yöntemleri geliştikçe, derin öğrenme yöntemlerinin güçlü öğrenme yeteneklerinin, gittikçe daha da önemli hale geleceği düşünülmektedir [12].

Bu çalışmada, yukarıda temel olarak türleri ve zararları üzerinde kısaca durulan ve sonuç olarak her birinin aşırı maruziyetinin çok önemli sağlık ve çevresel etkilere sahip olduğu bilinen kirleticilerin derin öğrenme yaklaşımı ile tahmini üzerine durulmaktadır. Bu amaçla geleceğe yönelik yapılan tahminlerin yenilikçi yöntemler ile gerçekleştirilmesi amacıyla çalışmada LSTM derin öğrenme yaklaşımı kullanılmıştır.

2. Materyal ve Metot

2.1. Veri seti ve Veri ön işleme

Çalışma kapsamında kullanılacak Sakarya iline ait hava kalitesi verileri Çevre ve Şehircilik Bakanlığı'nın hava kalitesi izleme istasyonlarından temin edilmiştir. Bu amaçla Sakarya'da bulunan 4 adet hava kalitesi ölçüm istasyonu incelenmiş ve çalışmada veri temini için bu istasyonlardan en merkezi konumda ve veri detayı açısından en zengin istasyon olan Sakarya Merkez hava kalitesi izleme istasyonu seçilmiştir. Bu istasyondan Ocak 2020-Eylül 2022 dönemine ait saatlik PM₁₀, CO, NO_x ve NO₂ hava kalitesi verileri temin edilmiştir.

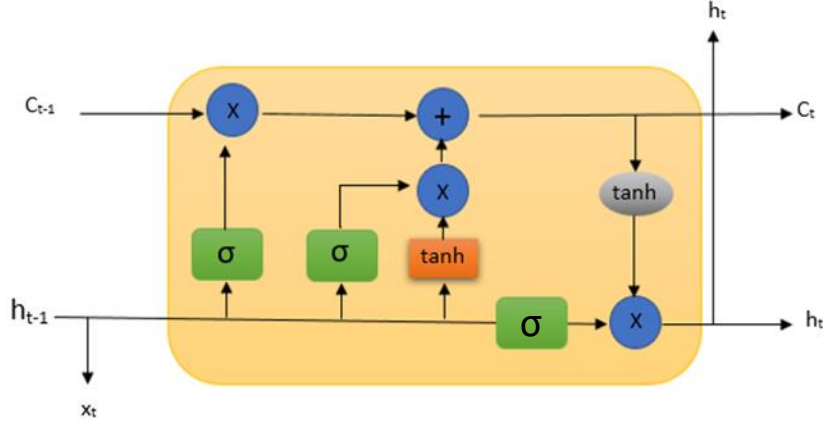
Veri işleme, ham veriler elde edildikten sonra modelleme çalışması için verilerin uygun bir formata dönüştürme işlemlerini kapsamaktadır. Veri ön işleme, makine öğrenimi algoritmasının genelleme yapabilmesi amacıyla gerekli olan işlemleri içeren bir adımdır [13]. Veri ön işleme, genellikle eksik veri atamasını, olağandışı gözlemleri kaldırmayı veya değiştirmeyi, veri dönüştürme ve öznitelik çıkarma işlemlerini kapsamaktadır [14]. Bu çalışmada veri ön işleme adımında hava kalitesi gözlem istasyonunda temine edilen verilerin temizleme çalışması Python programlama dili kullanılarak yapılmıştır. Bu amaçla öncelikle veri setinde boş değer olup olmadığı kontrol edilmiştir. Belirlenen tarih aralığında veri setinde her bir parametre için 24.282 olmak üzere toplam 97.128 veri olması gerekmektedir. Fakat sensör hataları ve benzeri durumlara bağlı olarak kısmi veri eksiklikleri gözlemlenmiştir. Tespit edilen eksik değerler ortalama ve en yakın komşu (kNN) algoritması doldurma yöntemleri kullanılarak doldurulmuştur. Tablo 2'de parametre bazlı gözlemlenen hataların giderilmesi için veri ön işleme ile doldurulan eksik veri adetleri sunulmuştur.

Tablo 2. PM₁₀, CO, NO_x, NO₂ hava kalitesi veri seti istatistikleri

Parametre	İstasyon	Birim	Ortalama	Eksik Veri Adeti	Eksik Veri Yüzdesi (%)
PM ₁₀			35.8	1402	5.77
CO	Sakarya – Merkez – MTHM	µg/m ³	1177.78	1615	6.65
NO ₂			31.76	1647	6.78
NO _x			65.80	1888	7.77

2.2. Uzun Kısa Süreli Bellek (LSTM)

Bu çalışmada derin öğrenme alanında kullanılan yapay bir yinelemeli sinir ağı mimarisi olan Uzun Kısa Süreli Bellek (LSTM) kullanılmıştır. LSTM ilk olarak 1997 yılında Hochreiter & Schmidhuber tarafından geliştirilmiştir [15]. Bu mimari uzun vadeli bağımlılık sorununun önüne geçmek için tasarlanmıştır. Standart ileri beslemeli sinir ağlarının aksine, LSTM geri bildirim bağlantısı içermektedir. Çalışmada, zaman serisi verilerine dayanarak sınıflandırma ve tahminlerde bulunma işlemleri yer aldığı için bu alanda başarılı olan LSTM tercih edilmiştir. Şekil 1'de LSTM'nin bilgi akış diyagramı sunulmuştur.



Şekil 1. LSTM bilgi akışı diyagramı

Burada unutmama kapısı olan bir LSTM hücresinin ileri geçişi için kullanılan denklem yapısı Denklem 1'de sunulmuştur;

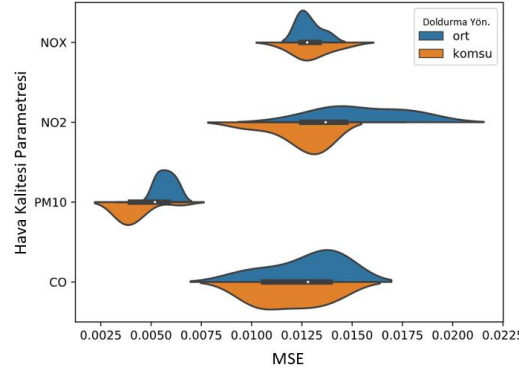
$$\begin{aligned}
 f_t &= \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \\
 i_t &= \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \\
 o_t &= \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \\
 \tilde{c}_t &= \sigma_c(W_c x_t + U_c h_{t-1} + b_c) \\
 c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \\
 h_t &= o_t \odot \sigma_h(c_t)
 \end{aligned}$$

Burada $c_0 = 0$ ve $h_0 = 0$ başlangıç değerlerini belirlerken t ise zaman adımlarını indekslemede kullanılmaktadır.

3. Deneysel Çalışma ve Sonuçları

Bu çalışmada öncelikle hangi doldurma yönteminin daha iyi sonuç verdiği araştırılmıştır. Ortalama ve kNN ile doldurulan hava kirleticilerine ait veri setleri LSTM algoritması ile eğitilmiş ve ortalama karesel hata (MSE) performans sonuçları kaydedilmiştir. Geliştirilen modeller rastgele üretilen sonuçlar içermemesi adına farklı rassal sayı üreteçleri ile 20 kez çalıştırılmış ve dağılım grafikleri violin grafiği olarak Şekil 2'de gösterildiği gibi verilmiştir. Elde edilen sonuçlara göre komşu arama algoritmasının dağılımı ve ortalaması NO_x dışında daha iyi olduğu gözlemlenmiştir.

NO_x parametresinin ise ortalama ile doldurulduğunda daha iyi sonuç vermektedir. Böylece ikinci aşamada kullanılacak olan veri setleri için doldurma yöntemi olarak NO_x için ortalama ve NO₂, PM₁₀ ve CO için ise kNN algoritması doldurma yöntemi olarak belirlenmiştir.



Şekil 2. Parametre dağılımlarını gösteren violin grafiği

Çalışmanın ikinci aşamasında ise hiperparametre optimizasyonu ile geliştirilen LSTM modelinin daha iyi bir performansa sahip olması hedeflenmiştir. Bu amaçla hiperparametre arama algoritmalarından “Hyperband Search” algoritması kullanılmıştır.

Hyperband Search arama algoritması kullanılarak parametre optimizasyonu yapılan LSTM modeline ait deneysel çalışmalar sonucunda elde edilen performans sonuçları Tablo 2’de verilmiştir. Buna göre tüm hava kirletici parametreleri için optimize edilen LSTM modelinin çok iyi performans sonuçlarına sahip olduğu, en iyi performansı ise CO hava kalitesi için verdiği görülmüştür.

Tablo 2. LSTM modeli Hyperband Search algoritması tahmin performansları (test seti)

Hava Kirletici	Arama Algoritması	RMSE	MSE	MAE	R ²
NO _x	Hyperband Search	0,000146612	0,01211	0,00755	0,943
NO ₂	Hyperband Search	0,000231648	0,01522	0,00955	0,961
PM ₁₀	Hyperband Search	2,79738E-05	0,00529	0,00291	0,983
CO	Hyperband Search	4,36383E-05	0,00661	0,00369	0,992

Çalışmada kullanılan LSTM tabanlı derin sinir ağı yaklaşımının, hava kalitesi parametrelerinin tahmini için uygun bir çıkarım gerçekleştirdiği görülmektedir. Sonuç olarak seçilen veri setinin yeterli olduğu ve LSTM modelinin ileriye dönük tahmin için kullanılabileceği görülmüştür.

4. Sonuç ve Öneriler

Sakarya’daki hava kalitesi verilerini içeren bu çalışma, hava kirliliğine sebep olan PM₁₀, CO, NO_x ve NO₂ dört kirleticinin konsantrasyon değişim eğilimi ve her kirletici LSTM derin öğrenme yöntemi ile ileriye dönük tahmini analiz edilmiştir. Bu amaçla eksik verilerin tamamlanması için

uygun doldurma yöntemi seçilmiş ve arama algoritması kullanılarak model parametrelerinin optimizasyonu yapılmıştır. Sonuçlar modelin yüksek tahmin doğruluğuna sahip olduğunu ve geliştirilen LSTM modelinin ileriye yönelik hava kalitesi tahmininde kullanılabileceğini göstermiştir.

Kaynakça

- [1] Eyüp Yahşi, “Ulusal Hava Kalitesi İzleme Ağı” *Hava Kirliliği ve Kontrolü Ulusal Sempozyumu-2008, 22-25 Ekim 2008, HATAY*
- [2] Reichle Jr, Henry G., Vickie S. Connors, J. Alvin Holland, Warren D. Hypes, H. Andrew Wallio, Joseph C. Casas, Barbara B. Gormsen, Mary S. Saylor, ve Wilfred D. Hesketh. “Middle and upper tropospheric carbon monoxide mixing ratios as measured by a satellite-borne remote sensor during November 1981”. *Journal of Geophysical Research: Atmospheres* 91, sy D10 (1986): 10865-87.
- [3] Reisen, Fabienne, CP Mick Meyer, ve Melita D. Keywood. “Impact of biomass burning sources on seasonal aerosol air quality”. *Atmospheric Environment* 67 (2013): 437-47.
- [4] Rahimi A. Short-term prediction of NO₂ and NO_x concentrations using multilayer perceptron neural network: a case study of Tabriz, Iran. *Ecol Process*. Aralık 2017;6(1):4.
- [5] Yun, Sug-gyeong, ve Changhyun Yoo. “The effects of spring and winter blocking on PM10 concentration in Korea”. *Atmosphere* 10, sy 7 (2019): 410.
- [6] Ramgolam, Kiran, Olivier Favez, Hélène Cachier, Annie Gaudichet, Francelyne Marano, Laurent Martinon, ve Armelle Baeza-Squiban. “Size-partitioning of an urban aerosol to identify particle determinants involved in the proinflammatory response induced in airway epithelial cells”. *Particle and Fibre Toxicology* 6, sy 1 (2009): 1-12.
- [7] Seng, Dwen, Qiyang Zhang, Xuefeng Zhang, Guangsen Chen, ve Xiyuan Chen. “Spatiotemporal prediction of air quality based on LSTM neural network”. *Alexandria Engineering Journal* 60, sy 2 (2021).
- [8] Kuremoto, Takashi, Shinsuke Kimura, Kunikazu Kobayashi, ve Masanao Obayashi. “Time series forecasting using a deep belief network with restricted Boltzmann machines”. *Neurocomputing* 137 (2014): 47-56.
- [9] Sagheer, Alaa, ve Mostafa Kotb. “Time series forecasting of petroleum production using deep LSTM recurrent networks”. *Neurocomputing* 323 (2019): 203-13.

- [10] Janiesch, Christian, Patrick Zschech, ve Kai Heinrich. “Machine learning and deep learning”. *Electronic Markets* 31, sy 3 (2021): 685-95.
- [11] Liu, Weibo, Zidong Wang, Xiaohui Liu, Nianyin Zeng, Yurong Liu, ve Fuad E. Alsaadi. “A survey of deep neural network architectures and their applications”. *Neurocomputing* 234 (2017): 11-26.
- [12] Bengio, Yoshua. “Learning deep architectures for AI”. *Foundations and trends® in Machine Learning* 2, sy 1 (2009): 1-127.
- [13] Kotsiantis, Sotiris B., Dimitris Kanellopoulos, ve Panagiotis E. Pintelas. “Data preprocessing for supervised leaning”. *International journal of computer science* 1, sy 2 (2006): 111-17.
- [14] Castelli, Mauro, Fabiana Martins Clemente, Aleš Popovič, Sara Silva, ve Leonardo Vanneschi. “A machine learning approach to predict air quality in California”. *Complexity* 2020 (2020).
- [15] Hochreiter, Sepp, ve Jürgen Schmidhuber. “Long short-term memory”. *Neural computation* 9, sy 8 (1997): 1735-80.