

CO₂ Emisyonunun Makine Öğrenmesi Metotları ile Tahmin Edilmesi

Forecasting CO₂ Emission with Machine Learning Methods

Evin GARİP

Mühendislik Yönetimi
Fen Bilimleri Enstitüsü
İstanbul Medeniyet Üniversitesi
İstanbul, Türkiye
evin.garip@medeniyet.edu.tr

Ayşe Betül OKTAY

Bilgisayar Mühendisliği Bölümü
Mühendislik ve Doğa Bilimleri Fakültesi
İstanbul Medeniyet Üniversitesi
İstanbul, Türkiye
abetul.oktay@medeniyet.edu.tr

Abstract— The amount of CO₂ emission has significantly increased because of the increase in industrial production, usage of fossil fuels such as petroleum and coal which is a danger for global warming. The countries measure CO₂ emissions and make plan for the future. In this study, Turkey's CO₂ emissions are estimated using random forest and support vector machine methods. Not only time, but also attributes such as fuel consumption and population are also employed for forecasting. It has been observed that the support vector machine method produces better forecasting results.

Index Terms— CO₂ emission, machine learning, random forest, support vector machines

Özet— Sanayii üretimindeki artış, petrol, kömür gibi fosil yakıtların kullanımının artması gibi nedenlerle atmosfere salınan CO₂ miktarı önemli derecede artmıştır ve küresel ısınma için bu önemli bir etkidir. Ülkeler, CO₂ emisyonlarını ölçmekte ve gelecek yıllar için de tahmin yaparak bir planlama yapmak durumundadır. Bu çalışmada, popüler makine öğrenmesi yöntemlerinden rastgele orman ve destek vektör makineleri yöntemleri ile yakıt tüketimi ve popülasyon gibi öznitelikler de kullanılarak Türkiye'nin CO₂ emisyonunu tahmin edilmektedir. Yapılan uygulamada destek vektör makinesi metodunun daha iyi sonuçlar ürettiği gözlemlenmiştir.

Anahtar kelimeler— CO₂ emisyonu, makine öğrenmesi, rastgele orman, destek vektör makineleri

I. GİRİŞ

Atmosfer yerküreyi saran gazlardan oluşan ve iklim olaylarının meydana geldiği katmandır. Atmosferin yapısındaki gazların yaklaşık % 78'i azot, % 21'i oksijen, % 0.9'u argon, % 0.03'ü karbon dioksittir (CO₂) ve bunlarının yanında az miktarda neon, kripton, ksenon ve helyum gibi soy gazları içerir.

Sera etkisi, atmosferdeki gazların yeryüzündeki ısının bir kısmını tutarak, ısı kaybına engel olmasıdır. Sera gazları, hem doğal, hem de insan kaynaklı olup atmosferdeki kızıl ötesi

radyasyonu emen ve tekrar yayan gaz oluşumlarıdır. Doğal sera gazları en büyük sera etkisine sahip olan su buharı, CO₂, metan (CH₄), diazot mono oksit (N₂O), ozon (O₃) ve endüstriyel üretim sonucu ortaya çıkan Hidroflorokarbon (HFC), Perflorokarbon (PFC), Sülfürheksaflorid (SF₆) dir.

CO₂ küresel ısınmada en önemli paya sahip gazdır. Sera etkisinin %80 oranında bu gaza bağlanmaktadır. Geçtiğimiz yüzyıllarda atmosferdeki oranı % 0.03 iken günümüzde bu oran % 0.04 'e kadar yükselmiştir [1] . Karbondioksit emisyonu, sanayii üretimindeki artış, petrol, kömür gibi fosil yakıtların kullanılması ile önemli derecede artmaktadır ve bu da iklim değişikliğine sebep olmaktadır.

Türkiye, CO₂ emisyonunu azaltma açısından en temel uluslararası strateji belgelerinden olan Birleşmiş Milletler'in "İklim Değişikliği Çerçeve Sözleşmesini imzalamıştır ve bu sözleşme doğrultusunda CO₂ emisyonunun azaltılması için yapılması gerekenler kararlaştırılmıştır. Şekil 1'de Türkiye'de 1965-2014 yılları arasındaki toplam CO₂ emisyonu gösterilmiştir.

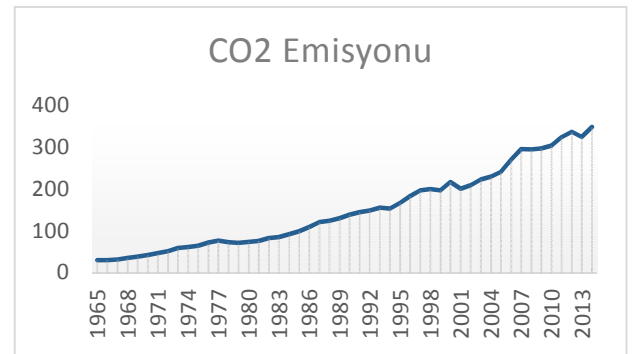


Fig. 1. 1965-2014 yılları arasında Türkiye'nin toplam CO₂ emisyonu

CO₂ emisyonu, yıllık ölçümler halinde raporlanmaktadır. Bu yıllık veri içeren bir zaman serisi olup, gelecek yılların çeşitli istatistiki yöntemlerle tahmin edilmesi mümkündür. Literatürde son yıllarda makine öğrenmesi teknikleri ile farklı alanlar için başarılı tahminler yapılmıştır.

Bu çalışmada, Türkiye'nin CO₂ emisyonu rastgele orman metodu ve Destek Vektör Makinesi (DVM) yöntemleri ile tahmin edilmektedir. Klasik zaman seri tahmin yöntemlerinde verinin zamana bağlı olarak değişimi modellenmektedir. Rastgele orman ve DVM yöntemlerinde ise sadece zaman değil, yıllık kömür, petrol, doğalgaz tüketimi, hidroelektrik ve yenilenebilir enerji verileri de modele dahil edilerek daha başarılı bir tahmin yapılması sağlanmaktadır.

II. LİTERATÜR

Literatürde CO₂ emisyonunu tahmin etmek ve değerlendirmek amacıyla yapılmış birçok çalışma bulunmaktadır. Akçasoğlu ve diğerleri [3] çalışmasında Türkiye'nin 2010 CO₂ salınımı tahmininde bulunmuştur. Bu çalışmada trend üzerinden yapılan tahminler yapılmıştır.

Say ve Yücel tarafından yapılan çalışmada [4] toplam enerji tüketimi ile CO₂ emisyonu arasındaki ilişki toplam regresyon analizi ile incelenmiş ve 2006-2015 yılları için CO₂ emisyonu tahmininde bulunulmuştur. [5] çalışmasında ise 1990,2000 ve 2004 yılları için Türkiye, tüm dünya ve OECD ülkelerindeki CO₂ emisyonları açısından kıyaslamalar yapılmıştır.

Sözen ve diğerleri [6] yapay sinir ağları kullanarak CO₂ tahmininde bulunmuştur. Sektörel olarak oluşan CO₂ emisyonları için yapay sinir ağı eğitilmiş ve bu ağ ile CO₂ emisyonu tahminleri yapılmıştır.

Shin ve ve Tsokos [7]1981- 2003 arasındaki aylık CO₂ emisyon dataları kullanarak zaman serisi yöntemi olan ARİMA ile tahmin yapmışlardır. Bu çalışmanın yanılma oranı 0,17'dir.

Röne ve Büke [8] istatistiki olarak CO₂ emisyon miktarlarını incelemiş ve trend analizi metodu ile CO₂ tahmininde bulunmuşlardır. Tezde incelenen ülkeler en çok sera gazı üreten 25 ülkedir.

Pau ve diğerleri [9] Grey metodunu kullanarak Brezilya'nın CO₂ emisyonunu, enerji tüketimi ve ekonomik büyümesini 2008-2013 yılları için tahmin etmiştir.

Başka bir çalışmada, bulanık mantık yaklaşımını kullanarak CO₂ tahmininde yapılmıştır [10]. Grey model kullanarak CO₂ tahmini yapan bir metot [12]'de sunulmuştur. Grey modelde örnek sayısının çoğaltılmasıyla yanlış tahminde bulunma oranları artmaktadır.

Baareh ve diğerleri [13] Yapay sinir ağları kullanarak tüm dünya için CO₂ emisyonu tahmininde bulunmuştur.

Loftalipour ve diğerleri [15] İran'ın CO₂ emisyonunu ARİMA ve Grey metoduyla tahmin edip, yapılan tahminleri kıyaslanmıştır. Sonuç olarak Grey metodu daha doğru sonuçlar vermiş ve 2020 yılı için İran'ın CO₂ emisyon miktarı 925.68 milyon ton olarak hesaplanmıştır. Saleh ve diğerleri [16] DVM ile CO₂ tahmininde bulunmuştur. Bu çalışmada elektrik enerjisi ve kömür bilgisi kullanılarak CO₂ değerleri tahmin edilmeye çalışılmıştır.

III. MAKİNE ÖĞRENMESİ İLE TAHMİN

Zaman serileri için istatistiksel yöntemlerin dışında makine öğrenmesi metotları da kullanılmaktadır. Verinin zamansal olarak matematiksel modelini çıkarmak için eğitim verisindeki farklı veriler de kullanılmakta ve böylece zaman serisinin gelecekteki değerleri tahmin edilmektedir.

Bu çalışmada, popüler makine öğrenmesi tekniklerinden rastgele orman ve DVM yöntemleri ile CO₂ emisyonu tahmin edilmektedir.

A. RASTGELE ORMAN METODU

Rastgele orman yöntemi karar ağaçlarının daha gelişmiş bir türü olup, eğitim kümesi için birçok karar ağacı oluşturup, bunların sonuçlarını birleştirme prensibiyle çalışmaktadır.

Rastgele orman, her bir karar ağacının girdi vektöründen bağımsız olarak örneklenmesi ile ürettiği rasgele ağaçlar topluluğudur [17]. Bu bir topluluk öğrenme metodu olup, her bir karar ağacı farklı veri örneğinden oluşmaktadır.

Rastgele orman metodu şu şekilde çalışmaktadır:

- İlk olarak girdi verisinden n adet farklı örnek küme seçilir
- Her bir örnek küme için bir karar ağacı oluşturulur
- Yeni test verisi için n adet ağacın her biri tarafından tahmin gerçekleştirilir.
- N adet tahmin ortalama almak gibi bir teknikle birleştirilir ve böylece en son tahmin bulunur.

Başka bir deyişle, bir x öznitelik vektörü rastgele ormana girdi olarak verildiğinde K adet regresyon ağacı oluşturulur ve bu ağaçların ürettiği sonuçların ortalaması bulunur [17]. K adet ağaç $T(x)^K$ oluşturulduktan sonra rastgele orman tahmini aşağıdaki formülle hesaplanır:

$$f_{rf}^K(x) = \frac{1}{K} \sum_{k=1}^K T(x). \quad (1)$$

Rastgele orman metodunun yüksek başarımlı sonuçlar üretmesi, eğitim kümesine göre iyi genelleşmesi, hızlı olması ve az sayıda parametre içermesi gibi avantajları bulunmaktadır. Dezavantajı ise eksik veri ile çalışmamasıdır.

B. DESTEK VEKTÖR MAKİNELERİ

Destek vektör makineleri, iki sınıfı birbirinden farklı sınıfları birbirinden ayıran en uygun karar fonksiyonun tahmin edilmesi ve bunları ayıran optimum hiperdüzlemin bulunması esasına dayanmaktadır [18]. Farklı sınıflar arasındaki uzaklığı maksimum yapan hiperdüzlem DVM ile bulunmakta ve bu düzleme uzaklığa göre de tahmin yapılmaktadır. DVM'nin en büyük avantajlarından birisi farklı çekirdek fonksiyonları kullanılarak doğrusal olmayan hiper-düzlemlerin elde edilebilmesidir. DVM sınıflandırma için kullanılmakla birlikte, Ardışık Minimal Optimizasyon adı verilen teknik ile hızlı bir şekilde eğitilmektedir.

Eğitim kümesi ile en uygun hiper-düzlem bulunduktan sonra test kümesindeki örneğin özniteliklerinin hiper-düzlemin hangi tarafında kaldığı bulunarak hangi sınıfa ait olduğu bulunur. Bu

çalışmadaki tahmin için ise hiper-düzleme olan uzaklığa göre göre değer bulunur.

IV. VERİ KÜMESİ

CO₂ emisyonunun enerji tüketim ve daha çok kullanılan fosil yakıtlar, nüfus, gayri safi milli hasıla gibi değerlerle ilişkili olduğu bilinmektedir. TÜİK 2014 raporuna göre toplam CO₂ emisyonlarının 2014 yılında %85.2'si enerjiden, %14.6'sı endüstriyel işlemler ve ürün kullanımından, %0.2'si ise tarımsal faaliyetler ve atıktan kaynaklanmıştır [2].

Yaptığımız çalışmada, CO₂ emisyonu tahmini için BP enerji şirketinin hazırladığı enerji raporu [19] ve Tüik verileri kullanılmıştır. Bu verilerin birimleri, minimum, maksimum ve ortalama değerleri Tablo I'de verilmiştir.

TABLO I. CO₂ EMİSYON VERİ KÜMESİ

	Birimi	Minimum	Maksimum	Ortalama
Zaman	Yıl	1965	2014	1989
Petrol	Mton	4.5	33.8	21.6
Doğal Gaz	Mton	0	43.7	10.0
Kömür	Mton	3.6	35.9	15.65
HidroElektrik	Mton	0.5	13.4	5.3
Yenilenebilir	Mton	0	0.9	0.085
Popülasyon	Milyon	34	78	55
CO ₂ Salınımı	Mton	29.9	348.5	151.52

Türkiye'nin CO₂ üreten kömür, petrol tüketimi ve doğalgaz tüketimi her yıl artmaktadır. Buna karşın CO₂ emisyonu daha az olan yenilenebilir enerji kullanımı ise 2000li yıllarda başlamış olup, bu yolla üretilen enerji 2014 yılında 0.88 Mtona ulaşmıştır. Şekil 2'de Türkiye'nin yıllık petrol, kömür, doğalgaz tüketimi gösterilmiştir.

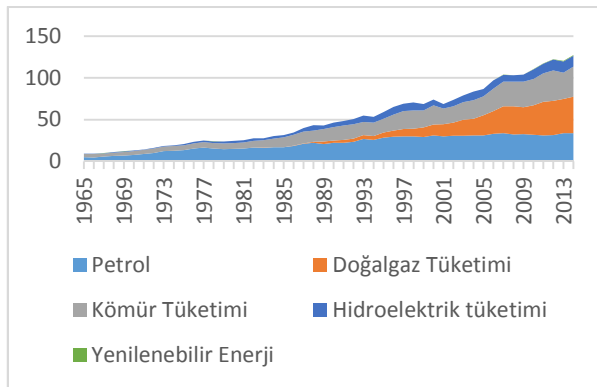


Fig. 2. 1965'ten günümüze Türkiye'de çeşitli yakıt tiplerinden enerji tüketimi

V. DENEYSEL SONUÇLAR

Veri kümesi oluşturulduktan sonra Rastgele Ağaç ve DVM ile tahminler yapılmıştır. Bunun için 1965-2014 arasındaki verilerin %70'i eğitim, %30'u ise test amaçlı kullanılmıştır. Sonuçların başarımını ölçmek için Ortalama Mutlak Hata (OMH) ve Ortalama Mutlak Yüzde Hata (OMYH) metrikleri kullanılmıştır. Denklem 2 ve 3'te, P_t t örneği için tahmin edilen değer, Z_t gerçek değer ve T de toplam örnek sayısı olmak üzere bu metriklerin hesaplanması gösterilmiştir:

$$OMH = \sum_{t=1}^T \frac{|P_t - Z_t|}{T} \quad (2)$$

$$OMYH = \frac{1}{T} \sum_{t=1}^T \frac{P_t - Z_t}{Z_t} \quad (3)$$

Şekil 3'te 2004-2014 yılları için gerçek CO₂ değerleri, DVM tarafından tahmin edilen değerler ve rastgele orman tarafından tahmin edilen CO₂ değerleri gösterilmiştir. Grafiğe göre ilk yıllar için daha başarılı tahminler yapılırken, 2013 yılındaki düşüş ve 2014 yılındaki artış iyi tahmin edilememiştir.

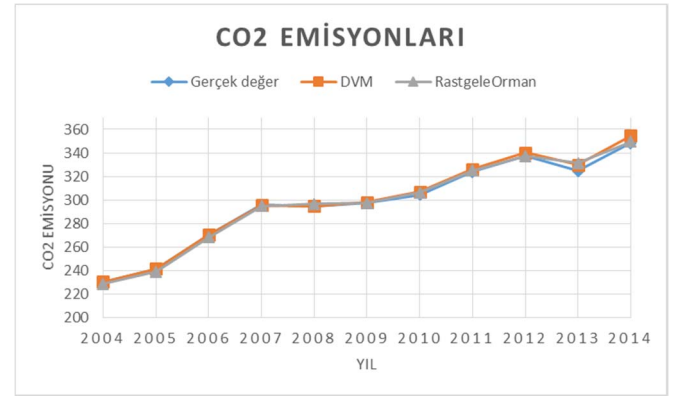


Fig. 3. 1965'ten günümüze Türkiye'de çeşitli yakıt tiplerinden enerji tüketimi

Tablo II'de yapılan 10 tahminin OMH ve OMYH oranları verilmiştir. Buna göre DVM, rastgele orman metodundan çok daha başarılı bir şekilde CO₂ emisyonunun tahmin etmiştir. DVM'nin OMH oranı 2.04 iken, OMYH oranı ise 0.63'tür. Bu oranlar oldukça küçük değerler olup, DVM'nin CO₂ emisyon tahmini için başarılı sonuçlar ürettiği gözlemlenmiştir.

TABLO I. CO₂ EMİSYONU TAHMIN VERİ KÜMESİ

	OMH	OMYH
DVM	2.04	0.63
Rastgele Orman	5.97	1.45

VI. SONUÇLAR

CO₂ emisyonu özellikle artan fosil yakıt tüketimi ile ilişkili olarak her geçen yıl artmakta ve iklim değişikliğinde büyük rol oynamaktadır. Bu nedenle, ülkelerin gelecekteki gaz

emisyollarını iyi planlamaları gerekmektedir. Bu çalışmada, popüler makine öğrenmesi metodlarından rastgele orman ve DVM ile CO₂ emisyon tahmini yapılmıştır. 1965 yılından 2003 yılına kadar olan veri eğitim için kullanılmış ve 2004-2014 yılları için tahminler yapılmıştır. DVM, rastgele orman metoduna göre daha başarılı sonuçlar üretmiştir. DVM ile tahmin yönteminin CO₂ emisyon tahmini için kullanılabileceğini göstermiştir.

REFERANSLAR

- [1] Kasım Koçak, "İklim Değişiminde İnsan Faktörü", İstanbul Teknik Üniversitesi, <http://www3.itu.edu.tr/~kkocak/iklim.html> - 81k (12.02.2006 tarihinde erişilmiştir.)
- [2] Aynur Tokel, "Sera Gazları Salınımı, 2014", TÜİK Haber Bülteni, <http://www.tuik.gov.tr/PreHaberBultenleri.do?id=21582> (28.04.2017 tarihinde erişilmiştir.)
- [3] Akçasoy, K., Önder, F., Güven, S., 2000. Statistical Evaluation of Greenhouse Gas Emissions of Turkey between the years of 1970 2010, <http://www.unescap.org> stat/envstat/stwes - 08.pdf, erişim: 12 - 09 - 2010.
- [4] Say, N.P. and Yücel, M. (2006) "Energy Consumption and CO₂ Emissions in Turkey: Empirical Analysis and Future Projection Based on An Economic Growth". Energy Policy 34, ss. 3870-3876.
- [5] Yamaoğlu, G. Ç. (2006). Türkiye'de Küresel Isınmaya Yol Açan Sera Gazı Emisyonlarındaki Artış İle Mücadelede İktisadi Araçların Rolü Yüksek Lisans Tezi, AÜ.
- [6] Sozen, A., Gulseven, Z., Arcaklioglu, E., 2007. Forecasting Based on Sectoral Energy Consumption of GHGs in Turkey and Mitigation Policies. Energy Policy 35, 6491 - 6505.
- [7] Shou Hsing Shih and Chris P. Tsokos (2008). Prediction Models for Carbon Dioxide Emissions and the Atmosphere, The International Journal Neural, Paralel scientific Computations, Volume 16
- [8] Aylin Cigdem Kone ve Tayfun Buke (2010). "Forecasting Of CO₂ Emissions From Fuel Combustion Using Trend Analysis," Renewable and Sustainable Energy Reviews Science Direct.
- [9] Pao, H.T., Tsai, C.M. (2011). "Modeling and forecasting the CO₂ emissions, energy consumption, and economic growth in Brazil". Energy ,36(5), 2450-2458.
- [10] Chen and Y.-C. Wang, "A fuzzy-neural approach for global CO₂ concentration forecasting," Intell. Data Anal., vol. 15, no. 5, pp. 763-777, Sep. 2011..
- [11] Cui Wan-Zhao, Zhu Chang-Chun, Bao Wen-Xing and Liu Jun-Hua. "Chaotic Time Series Prediction Using Mean-Field Theory For Support Vector Machine" Chinese Physics, Vol 14, Num.5
- [12] Lifeng Wu, Sifeng Liu, Dinglin Liu, Zhigeng Fang, Haiyan Xu . "Modelling And Forecasting CO₂ Emissions In The BRICS (Brazil, Russia, India, China, And South Africa) Countries Using A Novel Multi-Variable Grey", Model. Enerjy ,vol 79, pp 489-495, 2015.
- [13] Baareh, Abdel Karim, "Solving the Carbon Dioxide Emission Estimation Problem: An Artificial Neural Network Model."
- [14] Journal of Software Engineering and Applications,, 2013: 338-342T.
- [15] M. Lotfalipour, M. Falahi and M. Bastam (2013). Prediction of CO₂ Emissions in Iran using Grey and ARIMA Models, 2146-4553
- [16] Chairul Saleh, Nur Rachman Dzakiyullah, Jonathan Bayu Nugroho (2015). "Carbon Dioxide Emission Prediction Using Support Vector Machine" iMEC-APCOMS 2015
- [17] Leo Breiman (2001) "Random Forests" Machine Learning, 45, 5-32.
- [18] Cortes, C., Vapnik, V. (1995). Support vector networks, Machine Learning, 20, 1-25.
- [19] BP Statistical Review of World Energy Report, <http://www.bp.com/content/dam/bp/pdf/energy-economics/statistical-review-2016/bp-statistical-review-of-world-energy-2016-full-report.pdf>