VBM 683

Makine Öğrenmesi

Gaz Türbinlerinde NOx ve CO Emisyon Tahminleme Modeli

Tunahan KANBAK

N22130182

3 Ocak 2022

Özet

Bu raporda 5 yıl boyunca toplanmış gaz türbini operasyon verileri kullanılarak *NOx* ve *CO* emisyon tahmin modelleri oluşturulmuştur. Modeller arasında yapılan seçimler sonucunda *ensemble* modellerinin bu çalışma için uygun olduğu görülmüştür. Daha sonra nitelik mühendisliği ve nitelik seçimleri uygulanarak elde edilen model performansları değerlendirilmiştir. Nitelik mühendisliğinin test verisi performansını hem *CO* hem de *NOx* emisyon tahminlemesi açısından iyileştirdiği görülmüştür. Nitelik seçimin ise iki modelde de performansı iyileştiremediği görülmüştür.

Çalışma sonucunda en iyi performans sağlayan modeller *NOx* emisyonu için *Histogram Gradient Boosting* ve *CO* emisyonu için *Random Forest* modeli olarak elde edilmiştir.

İçindekiler

Şekil Listesi

Tablo Listesi

1. Giriş

Bu rapor, enerji üretimi sektöründe yüksek kullanım oranlarına sahip gaz türbini ekipmanının operasyonu esnasında gerçekleşen NOx ve CO gaz emisyonunun tahminlenmesi üzerine yapılan makine öğrenmesi çalışmalarını içermektedir. Gaz emisyonu yasal mevzuatlar ile sıkı bir şekilde takip edilen bir çevre konusudur. Türkiye içerisinde yer alan enerji tesisleri Çevre ve Orman Bakanlığı tarafından yayınlanan Büyük Yakma Tesisleri Yönetmeliği kapsamında emisyon değerlerini belli bir limitin altında tutmalı ve devamlı olarak takip etmelidir.

Türkiye'de ve Dünya'da emisyon değerleri Sürekli Emisyon Ölçüm Sistemleri (ing. CEMS) adı verilen ve gaz türbini gaz çıkış hattına bağlı ekipmanlar aracılığı ile 7/24 takip edilmektedir. Yönetmelikler gereği bu sistemlerin düzenli olarak bakımı yapılmalı ve doğru ölçüm yapıldığından emin olunmalıdır. Türkiye'de geçerli olan mevzuat kapsamında (Sürekli Emisyon Ölçüm Sistemleri Tebliği) SEÖS sisteminin yılın %95'inde düzgün ölçüm aldığı kalite güvence sistemlerince teyit edilmelidir.

SEÖS'e alternatif veya paralel bir yöntem olarak Tahminsel Emisyon Ölçüm Sistemleri (ing. PEMS) literatürde çalışılan bir konudur. TEÖS kullanımı SEÖS'e göre hem daha kolay hem de daha ucuz bir yöntemdir ancak şu an için tek başına TEÖS kullanımı kabul gören bir yöntem değildir. Genel kullanımda, SEÖS sistemleri bakım veya arıza durumunda devre dışı iken TEÖS aracılığı ile ölçüm toplanmaya devam etmektedir. Böylece emisyon kontrolü kesintisiz bir şekilde sağlanabilir. Ayrıca, TEÖS kullanımı SEÖS sistemlerinin bakım zamanlarının daha sağlıklı tespit edilmesini sağlayabilecek yardımcı bir aracıdır[1].

2. Literatür

TEÖS çalışmaları literatürde de dikkat çeken konulardan biridir. Sanayinin emisyon takip ihtiyacını karşılamak amacı ile Kaya ve çalışma arkadaşları tarafından Türkiye'de yer alan bir gaz türbininde 5 yıl kadar veri toplanarak açık kaynaklarda paylaşılmıştır. İlgili çalışma "Extreme Learning Machine" adı verilen bir makine öğrenme metodunu kullanarak gaz türbinine ait sensör verileri üzerinden hem NOx hem de CO emisyon verilerini tahminlemeye çalışmıştır. Aynı çalışma ELM çıktılarını "Random Forest" ve "Averaging" gibi yöntemler ile iyileştirerek model performansını arttırmaya çalışmıştır. Çalışmanın bir çıktısı olarak da CO emisyonlarının NOx emisyonlarından daha yüksek doğruluk oranı ile tahmin edilebildiği belirtilmiştir. Son olarak model performansının iyileştirilebilmesi için sensör verileri ile termo-kimyasal ilişkilerin birleştirilmesi önerilmiştir[1].

Korpela ve çalışma arkadaşları da birbirine benzeyen iki adet doğal gaz ısıtması kullanılan sıcak su kazanındaki NOx emisyonlarını modellemeye çalışmıştır. Bu çalışmada da hem doğrusal hem de doğrusal olmayan modellerden faydalanılmıştır. Çalışma sonucunda iki sistemin farklı modellere ihtiyaç duyduğu ve sistemlerin birbirinden ayrı modellenmesi gerektiği savunulmuştur[2].

Lv ve çalışma arkadaşları ise kömür bazlı bir kazanın NOx emisyonlarını en küçük kareler destek vektör makineleri yöntemi ile modellemeye çalışmıştır. Yapılan çalışmada tahmin performansının ve öğrenme süresinin "ensemble" adı verilen model sınıfı ile ciddi oranda arttığı sonucuna varılmıştır. Aynı zamanda tahminleme performansının iyileştirilmesine yönelik kümelenme algoritmaları kullanarak verinin ön işleme adımlarına tabi tutulabileceğini önermişlerdir[3].

Bu çalışmada ise Kaya ve çalışma arkadaşlarının açık kaynak olarak paylaştığı veri seti UCL Makine Öğrenmesi Deposundan alınarak makine öğrenmesi modelleri geliştirilmiştir.

3. Metot

Bu çalışma kapsamında literatürde de kullanımı yer alan karar ağaçları, destek vektörleri ve “ensemble” yöntemleri kullanılarak kıyaslamalar yapılmıştır. Tahminlemenin iyileştirilmesi için Kaya ve arkadaşlarının da önerdiği üzere termodinamiksel ilişkilerden faydalanılmıştır. Aynı zamanda Spline yöntemi ile veride ön işleme operasyonları da gerçekleştirilmiştir. Alt başlıklarda bu operasyonları detayları aktarılmıştır.

3.1. Veri Seti İçeriği

Bu çalışmada kullanılan veri seti aşağıdaki nitelikleri içermektedir:

- AT: Hava Sıcaklığı, ºC
- AP: Hava Basıncı, mbar
- AH: Hava Nemi, %
- AFDP: Hava Filtresi Basınç Farkı, mbar
- GTEP: Gaz Türbini Çıkış Basıncı, mbar
- TIT: Türbin giriş sıcaklığı, ºC
- TAT: Türbin çıkış sıcaklığı, ºC
- CDP: Kompresör çıkış basıncı, bar
- TEY: Türbin Enerji Üretimi, MWH

Veri setinde yer alan hedef parametler ise aşağıda verildiği gibidir:

- CO: Karbon Monoksit Emisyonu, ppm
- NOx: Nitrojen Oksit Emisyonu, ppm

Kaya ve çalışma arkadaşları tarafından gaz türbini verileri 5 yıl boyunca saatlik ortalama olarak toplanmıştır. Toplamda 36733 adet gözlem ve 11 adet ölçüm kalemi bulunmaktadır [1].

3.2. Makine Öğrenmesi Modelleri

Bu çalışma kapsamında kullanılan makine öğrenmesi modelleri ve kısa açıklamaları aşağıda aktarılmıştır (anlaşılabilirlik adına model isimleri orijinal halleri ile kullanılmıştır).

- Epsilon-Support Vector Regressor
- Classification and Regression Trees (CART)
- Random Forest
- Adaboost
- Histogram Gradient Boosting

Epsilon-Support Vector Regressor algoritması *Support Vector Machines* olarak bilinen sınıflandırıcı algoritmasının bir türevi olan ve tahmin edilen değerlerin kategorik yerine gerçek sayı olduğu bir yöntemdir. *Support Vector Machine* algoritmasında olduğu gibi *Regressor* algoritması da sadece öğrenme verisinin bir kısmını dikkate alarak işlemlerini gerçekleştirilir. Hata miktarı *Epsilon* değerinin altında olan noktalar modelin eğitiminde dikkate alınmazken *Epsilon* değerinden yüksek hatalar ise modelin eğitimi için kullanılır. Özetle *Epsilon-Support Vector Regressorlar* en büyük hatanın belirlenen *Epsilon* değeri kadar olacağı ideal fonksiyonu bulmaya çalışır ve bunu yaparken de olabildiğince düz ayrımlar (düşük eğilim, ing. bias) yapar. *SVR* modelinin en büyük avantajlarından birinin de model boyutunun büyük veriler için bile oldukça küçük olmasıdır [4].

Classification and Regression Trees (CART) algoritması ise *ID3* karar ağacı algoritmasının gelişmiş versiyonu olan *C4.5* algoritmasının bir benzeridir. *CART*'ın *C4.5*'ten en büyük farkı tahminlenecek hedef parametrenin kategorik olabileceği gibi gerçek sayı da olabilmesidir. *ID3* ve *C4.5* algoritmaları *Gini Index* ve *Entropy* gibi saflık ölçüklerini kullanırken *CART* algoritması her düğümde ortalama ve gözlemlerin farkı için en küçük karelerin toplamını verecek şekilde nitelikleri ayırır. *CART* algoritması aynı zamanda bir ikili ağaçtır dolayısıyla her düğümde sadece bir eşik değeri belirlenir[5].

Random Forest algoritması bir "*ensemble*" metotudur. *Random Forest* algoritmasında hem veri seti (tekrarlı, ing. *bootstrapping*) hem de nitelikler (tekrarsız) rastgele bir şekilde seçilerek bir çok karar ağacı eğitilir. Daha sonra rastgele oluşturulan nitelik ve veri setinin alt kümeleri ile eğitilen karar ağaçlarının her birinin yaptığı tahminin ortalaması alınarak nihai tahmin belirlenir. Bu yöntem genelde karar ağaçlarından çok daha yüksek performansa sahiptir ancak düğümlerin ayrılması için belirlenen kurallar karar ağaçlarındaki kadar yorumlanabilir değildir[6].

Adaboost algoritması da bir *ensemle* metotu olup *boosting* sınıfının ilk örneklerinden birisidir. *Adaboost* algoritması zayıf öğrenicilerin (düşük derinlikli karar ağaçları gibi) peşpeşe eğitilmesi ile oluşturulan bir model türüdür. Her zayıf öğrenici yaptığı hatayı azaltması için bir sonraki öğreticinin girdileri farklı ağırlıklarda dikkate alması sağlar. Böylece model düşük performans göstermeye meyili olan noktaları daha çok dikkate alarak eğitilir.

Histogram Gradient Boosting ise *Adaboost* gibi boosting sınıfındaki bir algoritmadır. *Adaboost*'tan farklı olarak *Gradient Boosting*'te her öğrenici bir önceki öğrenicinin yaptığı hatayı en iyi şekilde tahmin etmeye çalışır. Böylece model en çok hatanın yapıldığı noktalara odaklanarak eğitilmiş olur. *Histogram Gradient Boosting*'te ise girdiler histogramlarda olduğu gibi gruplandırılarak belli değerler ile (örn. grup ortalaması veya gruptaki en büyük değer) temsil edilir. Böylece modelleme süresi ve boyutu ciddi oranda kısaltılmış olur.

3.3. Nitelik Mühendisliği

3.3.1. Sistem Verimliliği

Gaz türbinleri belirlenen çalışma modlarında gaz emisyonlarını istenilen seviyelerde kontrol edebilen optimum ekipmanlardır. Ancak proses koşullarında yaşanan değişimler gaz türbinin çalışma verimliliği etkileyerek içerideki yanma operasyonunu bozabilmektedir. Bu durum emisyon değerlerinin de yükselmesi ile sonuçlanabilmektedir. Özellikle CO emisyonu yakıttan gelen ve tam yanamayan karbonlardan ötürü ortaya çıkmaktadır bu da türbinin operasyon parametreleri ile ilişkili

bir durumdur. NOx emisyonu havadaki nitrojen atomunun yanması sonucunda ortaya çıkan bir emisyondur ve hem ortam koşullarından hem de operasyon parametrelerinden etkilenebilmektedir[7]. Bu iki durumu da açıklayabilecek bir parametre olarak gaz türbinlerinde termal verimlilik adı verilen bir katsayı hesaplanabilmektedir. Farklı kaynaklarda bu değer Brayton Döngüsü Verimi olarak da geçmektedir.

Brayton verimininin hesaplanabilmesi için veri setinde yer almayan kompressör çıkış sıcaklığına ihtiyaç duyulmaktadır. Kompressör çıkış sıcaklığı aşağıdaki denklem ile hesaplanabilmektedir[8]:

$$T_2 = T_1 * (P_2 / P_1)^{(n)}$$

Bu denklem T_1 ve P_1 kompressör giriş sıcaklık ve basınç, T_2 ve P_2 değerleri ise kompressör çıkış sıcaklık ve basınç değerlerini temsil etmektedir. Denklemde yer alan n değeri ise poliizotropik oran olarak bilinen bir katsayıdır. Bu katsayı Kaya ve çalışma arkadaşlarının çalışmasında yer alan ve anlık gaz türbini çalışma parametrelerinin gösterildiği görsel aracılığı ile 0.32 olarak hesaplanmıştır.

Kompressör çıkış sıcaklığının hesaplanmasının ardından aşağıdaki denklem aracılığı ile Brayton Döngüsü Verimi hesaplanabilir[8]:

$$\eta = 1 - [T_a (T_d / T_a - 1)] / [T_b (T_c / T_b - 1)]$$

Bu denklemde T_c türbin giriş sıcaklığı, T_b yanma öncesi sıcaklığı, T_a hava sıcaklığı, T_d ise türbin çıkış sıcaklığıdır. Bu ilişki ile hesaplanan verimlilik girdi matrisine eklenerek modellerin eğitilmesinde kullanılmıştır. Bu eklemenin yapılabilmesi ve kodlama işlemlerini kolaylaştırmak için özel bir dönüştürücü sınıfı oluşturulmuştur.

3.3.2. Spline Dönüşümü

Gaz türbinleri kompleks ekipmanlar olmak ile birlikte içerisinde gerçekleşen termodinamik operasyonlar çoğu zaman doğrusal olmayan ilişkiler ile açıklanmaktadır. Verim hesabında da görüldüğü üzere kompressör giriş çıkış verileri birbirleri ile doğrusal olmayan bir şekilde ilişkilendirilmektedir[7].

Doğrusal olmayan bu sistemin daha iyi açıklanabilmesi ve model performansının iyileştirilebilmesi için *Spline* dönüşümü kullanılarak veri setindeki girdiler kendilerini temsil eden parçalı polinom denklemlerine dönüştürülmüştür. Polinom dönüşümü yerine *Spline* dönüşümünün tercih edilmesinin temel sebebi ise *Spline* dönüşümünün polinomlara göre daha stabil olmasıdır. Polinom dönüşümünde belirlenen girdi uzayının dışına çıkıldığında stabil olmayan ani değişimler görülebilmektedir[6].

3.4. Nitelik Seçimi

Nitelik seçimi için permütasyon yöntemi tercih edilmiştir. Permütasyon yönteminin temel akışı aşağıda aktarılmıştır:

1) Önemi incelenecek nitelik belirlenir.
2) İlk veri satırında yer alan önemi incelenecek nitelik dışındaki nitelikler sabitlenir.

3) Önemi incelenecek veri, veri setinde yer alan bütün gerçek değerlerini alacak şekilde sırasıyla değiştirilir.
4) Her yeni permütasyonda önceden eğitilmiş model ile tahminleme yapılır ve tahmin kaydedilir.
5) Bu adımlar her nitelik için tekrarlanır.

Niteliklerin permütasyonları ile elde edilen tahminler incelendiğinde tahmin edilen verinin her veri satırı için bir trende sahip olması ilgili niteliğin model için önemli bir nitelik olduğunu gösterir. Nitelik önemi çıkarıldığında modele etki etmeyen nitelikler veri setinden elenerek modelin performansında iyileşme sağlanabilmektedir.

3.5. İş Akışı

Bu çalışma kapsamında hiç bir işlem yapılmadan önce veri seti test (0.30) ve öğrenme (0.7) olarak iki parçaya ayrılmıştır. Daha sonra 3.2 numaralı başlıkta yer alan makine öğrenmesi modelleri, öğrenme verisi kullanılarak, gerekli ön işlemeler ve hiper parametre iyileştirmeleri yapıldıktan sonra kıyaslanmıştır. Hiper parametre kıyaslamaları için rassal arama yöntemi kullanışmış ve her model için toplamda 75 farklı hiper parametre kombinasyonu rastgele olarak seçilmiştir. Rassal aramada en iyi modelin seçilebilmesi ve seçimin şansa dayalı olmaması için *k-fold* yöntemi tercih edilmiş ve *7-fold* uygulanarak hiper parametreler iyileştirilmiştir. Modellerin karşılaştırılabilmesi için *RMSE* ve R^2 değerleri her model için hem test hem de öğrenme verisinde hesaplanmıştır

Yapılan kıyaslamanın ardından bazı modeller iş akışından çıkartılmıştır. Daha sonra 3.3 numaralı başlıkta yer alan nitelik mühendislikleri uygulanarak tekrardan hiper parametre düzeltmeleri yapılarak modeller karşılaştırılmıştır.

Nitelik mühendisliğinin (türbin verimi eklendikten sonra *Spline* dönüşümü yapılmadan önce) ardından nitelik seçimi değerlendirmesi yapılarak modeller tekrar hiper parametre iyileştirmesi yapılarak kıyaslanmıştır.

En sonda seçilen model ile *Bayes* tipi arama yapılarak aynı hiper parametre uzay limitlerinde belli bir iterasyonda (75 adet) daha iyi bir modelin elde edilip edilemeyeceği değerlendirilmiştir.

Modeli seçimi tamamlandıktan sonra modelin test verisi üzerindeki performansı ve hataların dağılımı detaylıca incelenmiştir.

4. Sonuçlar

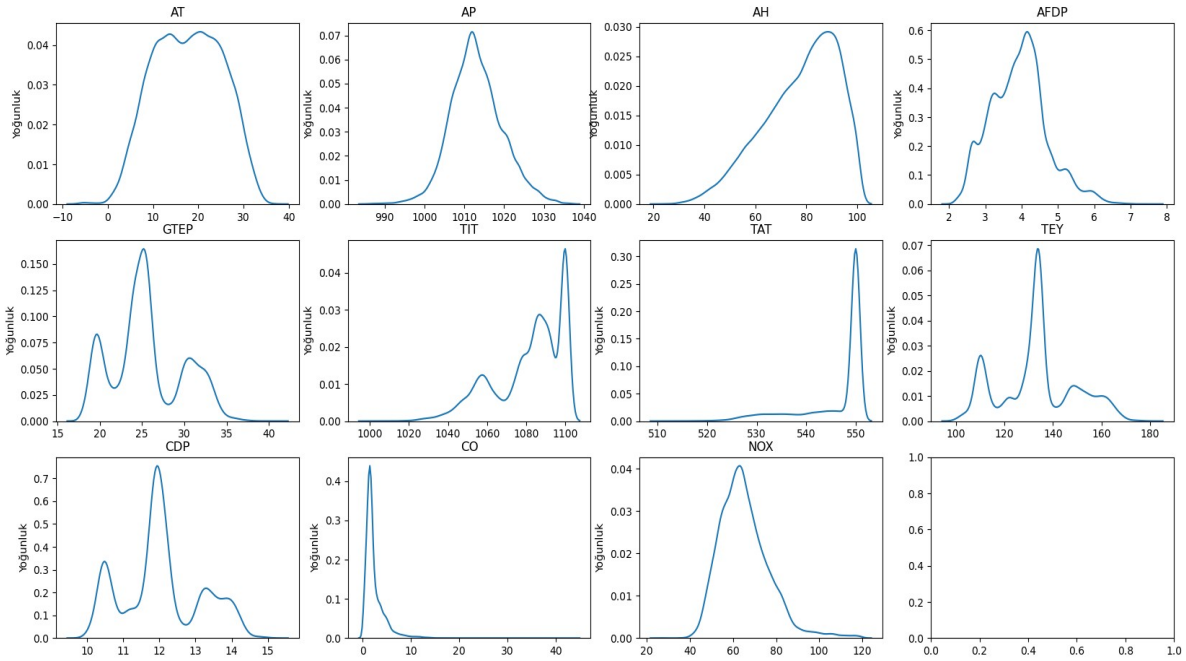
4.1. Veri Setinin İstatistiksel Özeti

Kullanılan veri setinde yer alan niteliklerin ortalama, standard sapma vb. istatistiksel özellikleri aşağıdaki tabloda verilmiştir.

Tablo 4:1: Veri Setinin İstatistiksel Özeti

Nitelik İstatistik	AT	AP	AH	AFDP	GTEP	TIT	TAT	TEY	CDP	CO	NOX
Ortalama	17.71	1013.07	77.86	3.92	25.56	1081.42	546.15	133.50	12.06	2.37	65.29
Standard S.	7.44	6.46	14.46	0.77	4.19	17.53	6.84	15.61	1.08	2.26	11.67
Min	-6.23	985.85	24.08	2.08	17.69	1000.80	511.04	100.02	9.85	0.00	25.90
Medyan	17.80	1012.60	80.47	3.93	25.10	1085.90	549.88	133.73	11.96	1.71	63.84
Max	37.10	1036.60	100.20	7.61	40.71	1100.90	550.61	179.50	15.16	44.10	119.91

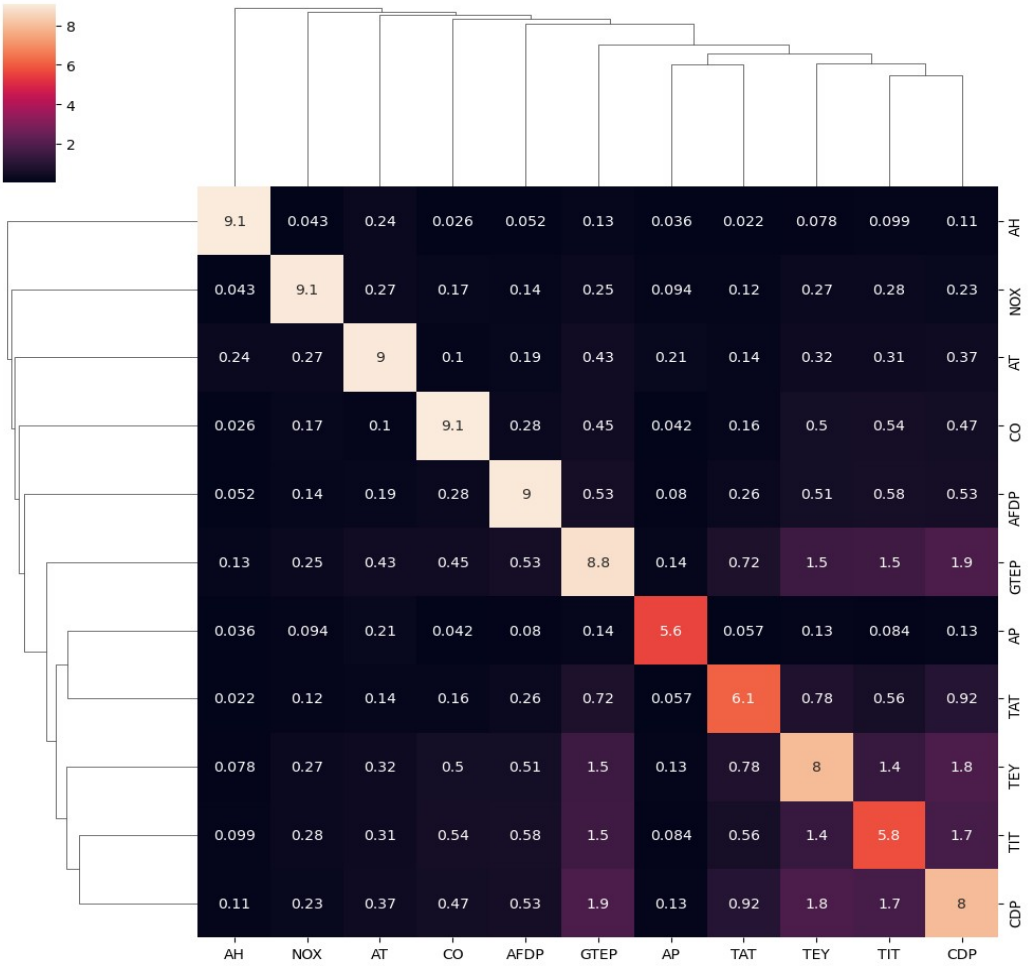
Tablo 4:1’de yer alan verilere bakıldığı yalnızca CO niteliğinde bir çarpıklık göstergesi bulunmaktadır. Ortalama değer medyan değerinden büyük olduğu ve minimum/maksimum genişliği ortalamanın 20 katı için CO niteliğinin pozitif çarpık olduğu söylenebilir. Bunun dışında verilerin minimum ve maksimum noktalarına bakıldığı zaman fiziksel olarak anlamsız bir değer olmadığı (örn negatif emisyon değeri) teyit edilmiştir. Son olarak verilerin genişliğine bakıldığında da anormal bir veri (örn 200ºC türbin giriş sıcaklığı) gözlemlenmemektedir.



Şekil 4:1: Veri Dağılımlarının KDE Grafikleri ile Gösterimi

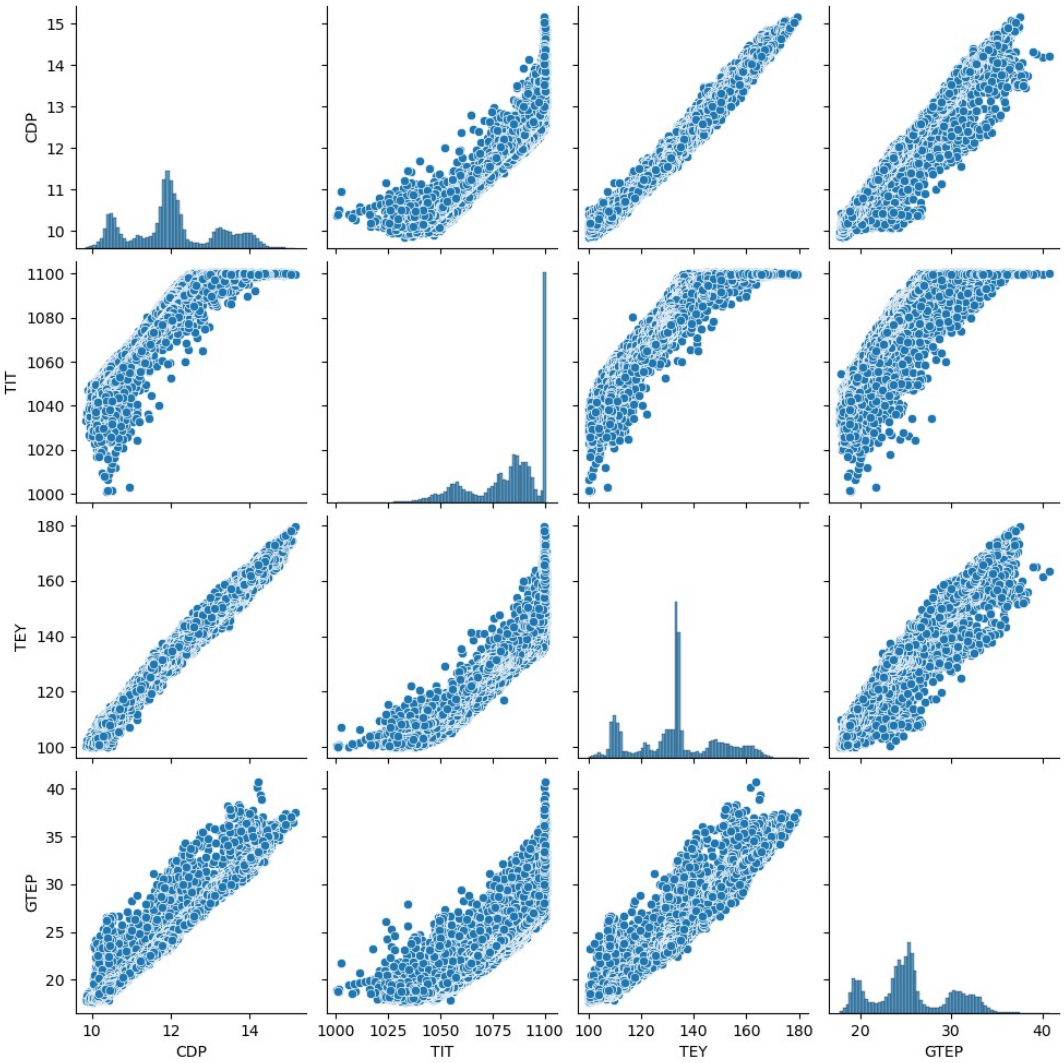
Şekil 4:1’de yer alan dağılım grafikleri incelendiği zaman Tablo 4:1’den farklı olarak TAT, TIT ve AH verilerinde de çarpıklık olduğu görülmektedir. Şekil 4:1’te dikkat çeken bir diğer özellik ise

GTEP, TIT, TEY, CDP verilerinin 3 modlu bir dağılıma sahip olmasıdır. Bu durum gaz türbinin düşük, orta ve yüksek enerji üretim modunda çalışıyor olması ile ilişkilendirilebilir.



Şekil 4:2: Veri Seti Ortak Bilgi Haritası

Şekil 4:2'de ortak bilgi (ing. mutual information) hesaplaması ile oluşturulmuş bir küme haritası gösterilmektedir. Normal korelasyon hesabına göre ortak bilgi hesabının tercih edilmesinin sebebi ise ortak bilgi hesaplamasının normal korelasyona kıyasla doğrusal olmayan ilişkileri de açıklayabilmesidir. Haritaya bakıldığı zaman GTEP, TEY, TIT ve CDP değerlerinin birbiri hakkında bilgi sağladığı görülebilmektedir. Bu dört parametre de türbinin operasyonu hakkında bilgiler sağlamakta olup ilişkili olmaları fiziken de mantıklı bir durumdur. Bu ilişkiler görselleştirildiği zaman ise genel olarak ilişkilerin doğrusala yakın olduğu görülmüştür.



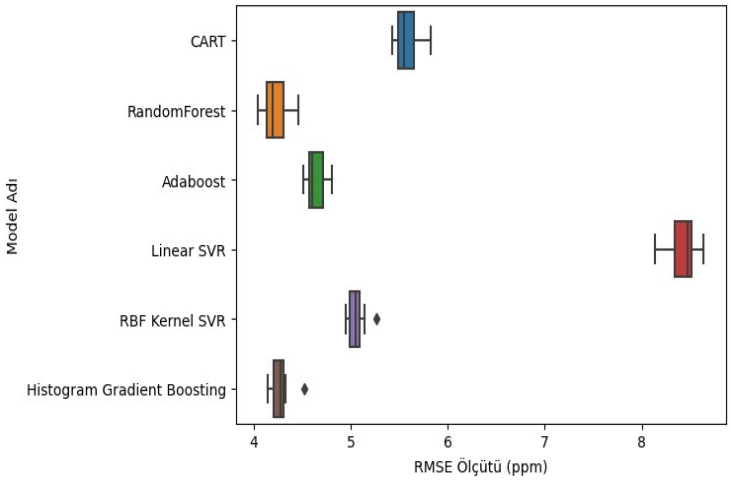
Şekil 4:3: CDP, TIT, TEY ve GTEP Verilerinin Çapraz İlişkileri

Şekil 4:3'te de görülmekte olan nitelikler içerisindeki çapraz doğrusal ilişkiler eş doğrusallık olarak da (ing. Colinearity) bilinen ve bazı makine öğrenmesi modellerinin performansını kötü yönde (örn. düzenleyicisiz en küçük kareler modeli) etkileyebilecek nümerik bir olgudur. Bu durumun önüne geçmek için PCA gibi gözetimsiz makine öğrenmesi yöntemleri kullanılabilmektedir ancak bu çalışma kapsamında eş doğrusallıktan etkilenmeyen ağaç modelleri ve destek vektörleri (Ridge düzenleyicisine benzer bir düzenleyiciye sahiptir.) kullanıldığı için eş doğrusallık üzerinde durulmamıştır.

4.2. NOx Emisyon Modelleri

NOx emisyon modellemesinin ilk adımı olarak *CART, Random Forest, Adaboost, Histogram Gradient Boosting, Linear Kernel SVR* ve *RBF Kernel SVR* algoritmaları için hiper parametre iyileştirmesi yapılmıştır. Her bir model için belirlenen parametre uzayında 75 farklı kombinasyon rastgele olarak seçilmiş *7-fold* çapraz doğrulama yöntemi kullanılarak 75 kombinasyon arasından en

iyi parametrelere sahip modeller seçilerek kıyaslanmıştır. Elde edilen sonuçlar ile oluşturulan kutu grafiği Şekil 4:4‘de gösterilmiştir.



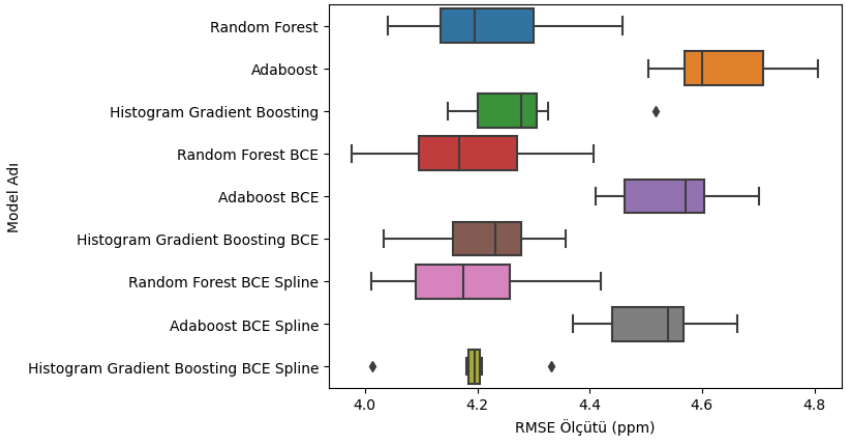
Şekil 4:4: NOx Emisyonunda Tahminlemesi için Model Karşılaştırması

Şekil 4:4’de gösterilen kutu grafiklerine bakıldığı zaman *Linear SVR* modelinin diğer modellere kıyasla oldukça kötü bir performans sergilediği görülmüştür. *CART* modeli *Linear SVR* modeline göre iyi bir performans sağlasa da kendisinden daha iyi sonuç gösteren modeller olduğu için çalışmanın devamında dahil edilmemiştir. Öte yandan *ensemble* metotlarının birbirlerine yakın ve diğer modellerden üstün bir performansa sahip olduğu görülebilmektedir. *SVR* algoritmasında *RBF Kernel* kullanımı model performansını ciddi oranda iyileştirmiş hatta ensemble modelleri ile neredeyse kıyaslanabilir bir noktaya taşımıştır. Ancak hiper parametre iyileştirmede hesaplama süresi çok uzun olduğu için *RBF Kernel SVR* da çalışmanın devamından çıkartılmıştır.

İlk hiper parametre iyileştirmesinden sonra *Random Forest, Adaboost* ve *Histogram Gradient Boosting* ile çalışmaya devam edilmiştir. Model performansının arttırılabilmesi için veri setine türbin verimi niteliği türetilerek eklenmiştir. Türetilmiş veri setinde ortak bilgi haritası oluşturulduğunda türbin verimi (*BCE*) ile *NOx* arasındaki ortak bilgi değeri 0.32 olarak hesaplanmıştır. Bu değer veri seti içerisinde *NOx* niteliği ile başka bir nitelik arasında elde edilebilen en yüksek ortak bilgi değeridir. Daha sonra karmaşık ilişkilerin daha iyi açıklanabilmesi için türetilmiş veri setinde *Spline* dönüşümü kullanılarak modeller bir daha karşılaştırılmıştır. Modeller hem sadece türbin verimi eklenen veri seti hem de türbin verimi eklendikten sonra *Spline* uygulanmış veri setinde eğitilerek karşılaştırılmıştır. Böylece sadece yeni bir nitelik eklemenin etkisi de gözlemlenebilmiştir. Her iki durumda da modeller sabit tutulan parametre uzayında hiper parametre iyileştirmesi yapılarak eğitilmiştir.

Şekil 4:5’te yer alan karşılaştırmalara bakıldığı zaman genel olarak her iki işleminde model performansını iyileştirmeye yönelik bir etkisi var gibi görünmektedir. Ancak kutu grafiklerinin limitleri keskin bir şekilde ayrıştırılamadığı için orijinal veri seti ile eğitilmiş modeller ile (R*andom Forest, Adaboost, Histogram Gradient Boosting*) nitelik mühendisliği sonucunda ortaya çıkan veri setinde eğitilmiş modeller (*Random Forest BCE Spline, Adaboost BCE Spline, Histogram Gradient*

Boosting BCE Spline) istatistiksel olarak kıyaslanmıştır. İstatistiksel olarak t-testinin parametresiz bir eşleneği olan permütasyon testi kullanılmıştır. Permütasyon tercih edilmesinin sebebi hem hesaplanan model performansı sonuç sayısının az olması hem de model performans sonuçlarının normal bir dağılımından geldiğinin bilinmemesidir.



Şekil 4:5: NOx Emisyonunda Nitelik Mühendisliği Uygulanmış Veri Setlerinde Model Karşılaştırmaları

Permütasyon testi için kullanılan model performansları aşağıdaki tabloda yer almaktadır. Tablo 4:2'de yer alan sonuçlara baktığımız zaman her üç model içinde nitelik mühedisliği ortalama ve standard sapma değerlerini azaltarak daha iyi bir model eğitilebilmesine olanak sağlamıştır. Ek olarak performans iyileşmesinin test verisine de etki ettiği görülebilmektedir.

Tablo 4:2: NOx Emisyonunda 7-fold Öğrenme ve Test Setinde Model Performans (RMSE) Değerleri

	Random Forest	*Adaboost*	*Histogram Gradient Boosting*	*Random Forest BCE Spline*	*Adaboost BCE Spline*	*Histogram Gradient Boosting BCE Spline*
ÇD*-1	4.25	4.56	4.20	4.13	4.40	4.18
ÇD-2	4.04	4.57	4.32	4.05	4.47	4.19
ÇD-3	4.18	4.60	4.28	4.21	4.53	4.19
ÇD-4	4.34	4.77	4.27	4.29	4.58	4.20
ÇD-5	4.45	4.80	4.51	4.41	4.66	4.33
ÇD-6	4.08	4.50	4.14	4.01	4.37	4.01
ÇD-7	4.19	4.64	4.19	4.17	4.55	4.18
Ortalama	4.22	4.63	4.27	4.18	4.51	4.18
SS**	0.13	0.10	0.11	0.13	0.09	0.08
Test Verisi *RMSE*	4.14	4.55	4.17	4.11	4.47	4.10
Test Verisi R^2	0.88	0.85	0.88	0.88	0.86	0.88

*ÇD: Çapraz Doğrulama
**SS: Standard Sapma

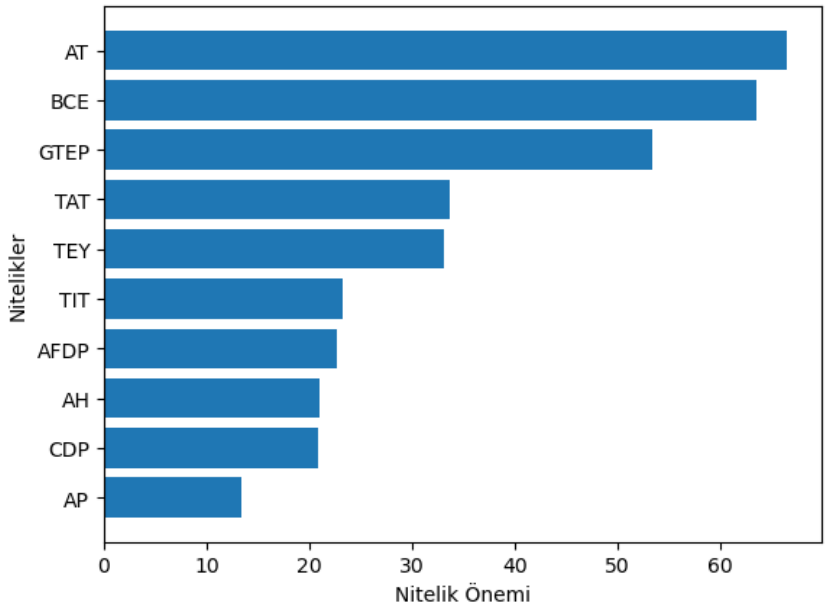
Modeller için orijinal veri ve nitelik mühendisliği ile türetilen veri arasında tek kuyruklu permüstasyon testi yapıldığında da aşağıdaki p-değeri sonuçları elde edilmiştir. P-değerinin 0.05'in altında kaldığı durumlarda modeldeki iyileşmelerin istatistiksel olarak da manalı olduğu değerlendirilebilmektedir.

Tablo 4:3: Nitelik Mühendisliğinin Model Performansı Üzerindeki Etkisinin P-Değeri Olarak İncelenmesi

Model Adı	P-Değeri
Random Forest	0.322
Adaboost	0.065
Histogram Gradient Boosting	0.047

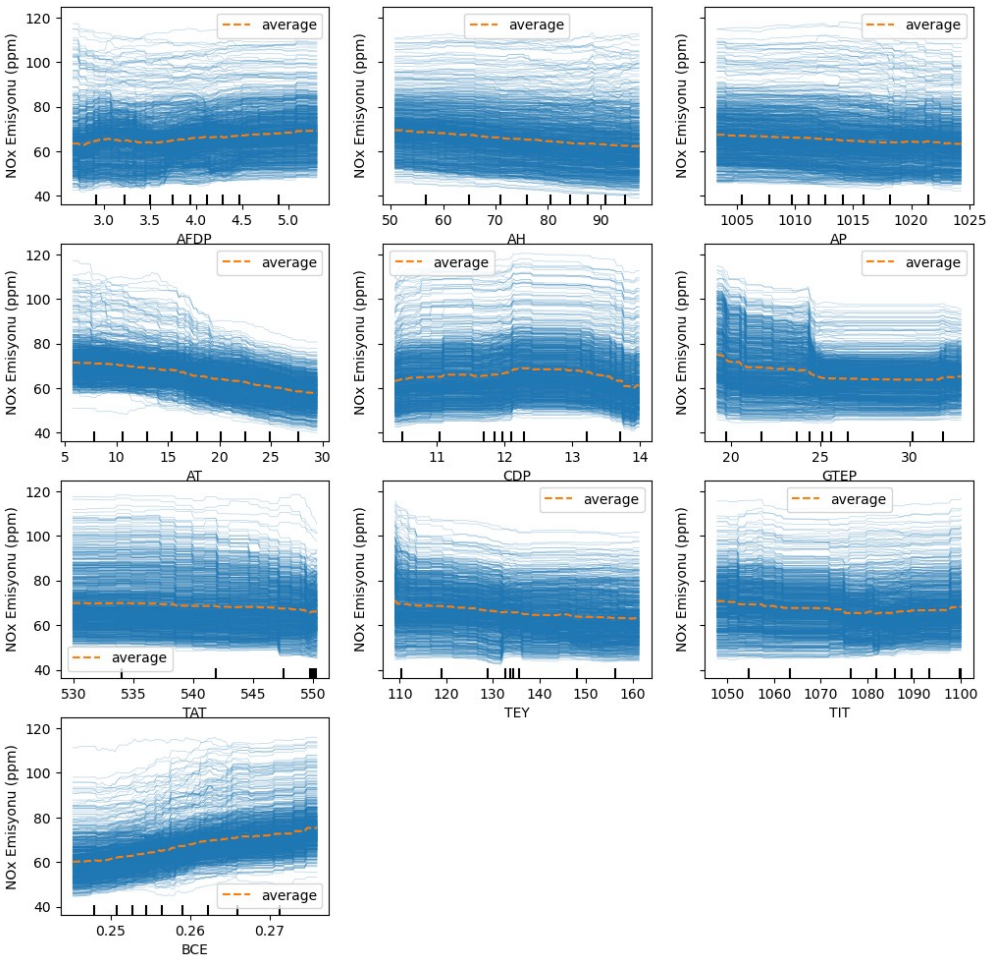
Tablo 4:3'te yer alan sonuçlara bakıldığı zaman *Adaboost* için 0.05 sınır değerine çok yakın bir p-değeri elde edildiği ve *Histogram Gradient Boosting* için ise 0.05 sınır değerinin altında bir sonuç alındığı görülmüştür. Bu durumda *Adaboost* ve *Histogram Gradient Boosting* modellerinin performansı nitelik mühendisliği sonucunda istatistiksel manada da kabul edilir bir iyileşme göstermiştir. Dolayısıyla modelleme çalışmaların devamında nitelik mühendisliği yapılmış veri seti kullanılmıştır.

Model performansını iyileştirme için nitelik seçimi yapılması değerlendirilmiştir. Bunun için öncelikle niteliklerin seçilen model için ne kadar etkili olduğunun incelenmesi gerekmektedir. Bu kapsamda da permütasyon yönteminden faydalanarak niteliklerin model üzerindeki etkileri göreceli olarak kıyaslanabilmiştir. Permütasyon yöntemi için türbin verimi eklenen veri setinde eğitilmiş *Histogram Gradient Boosting* modeli kullanılmıştır.



Şekil 4:6: NOx Emisyonunda Permütasyon Yöntemine Göre Bağıl Nitelik Önemi

Şekil 4:6’da yer alan nitelik önemi sıralamasına bakıldığı zaman en önemli niteliklerin atmosfer hava sıcaklığı (*AT*) ve türbin verimi (*BCE*) olduğu görülmüştür. Bu da *NOx* emisyonunun, beklendiği üzere, hem doğal ortam koşullarından hem de türbin çalışma koşullarından etkilendiği görülmüştür. En önemli nitelik olan hava sıcaklığı ile en etkisiz nitelik olan atmosfer hava basıncı (*AP*) arasında 6 kat önem farkı bulunmaktadır. Önem farkı belirgin olduğu için atmosfer hava basıncının etkisiz bir parametre olabileceği kısmı bağımlılık grafikleri ile çapraz olarak tekrar incelenmiştir.



Şekil 4:7: NOx Emisyonunda Nitelikler İçin Kısmı Bağımlılık Grafikleri

Şekil 4:7’de yer alan kısmı bağımlılık görselleri incelendiğinde atmosfer basıncının (*AP*) *NOX* emisyonu üzerinde negatif eğimli belli belirsiz bir etkisi olduğu görülmüştür. Diğer niteliklerin ise *NOx* emisyonunu daha belirgin bir şekilde etkilediği görülebilmektedir. Bu sebep ile atmosfer hava basıncı veri setinden çıkartılarak veri setine *Spline* dönüşümü uygulanmış ve modeller bir kez daha eğitilmiştir.

Yine benzer hiper parametre uzayında yapılan eğitim süreci sonunda modellerin performansında ciddi bir düşüş gözlemlenmiştir. Elde edilen performans sonuçları Tablo 4:4‘te verilmiştir.

Tablo 4:4: NOx Emisyonunda Nitelik Seçimi Sonrasında Eğitilen Modellerin Performans Çıktıları

Model Adı	Öğrenme Seti Ortalama *RMSE*	Öğrenme Seti Standard Sapma	Test Seti *RMSE*	Test Seti R^2
Random Forest	4.37	0.14	4.30	0.87
Adaboost	4.80	0.13	4.79	0.84
Histogram Gradient Boosting	4.45	0.14	4.46	0.86

Tablo 4:4'te yer alan sonuçlar Tablo 4:2'de yer alan sonuçlar ile kıyaslandığında nitelik seçimi en kötü sonuçları vermiştir.

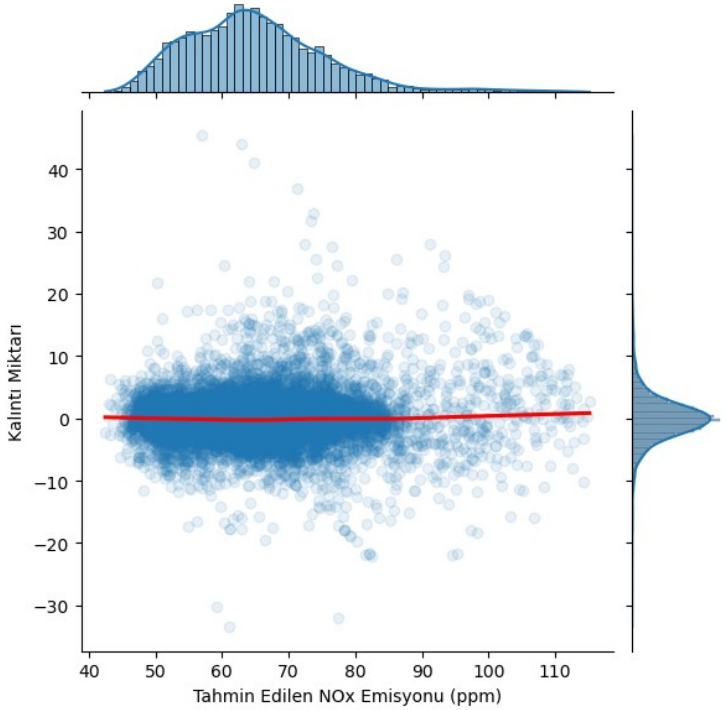
Bu noktaya kadar eğitilen modeller karşılaştırıldığında test setindeki en düşük *RMSE* skorunun nitelik mühendisliği yapılmış veri üzerinde eğitilen *Histogram Gradient Boosting* olduğu görülmüştür. Son olarak *Histogram Gradient Boosting* algoritmasının hiper parametre uzayını daha efektif tarayabilmek adına aynı uzay limiti içerisinde Bayes tipi arama yapılmıştır. Ancak bu arama sonucunda öğrenme setindeki ortalama *RMSE* değeri 4.26 (SS: 0.08) gibi yüksek bir değere sahip olup *Histogram Gradient Boosting* algoritması için en düşük performansa sahip hiper parametre kombinasyonu elde edilmiştir.

NOx emisyonu için modeller kıyaslandığında en iyi performansa sahip modelin aşağıdaki parametreler (geri kalan parametreler varsayılan değerlerine sahiptir.) ile nitelik mühendisliği verisi üzerinde eğitilen *Histogram Gradient Boosting* modeli olduğu bulunmuştur.

- max_iter: 10000,
- max_bins: 64,
- learning_rate: 0.1,
- l2_regularization: 10.0

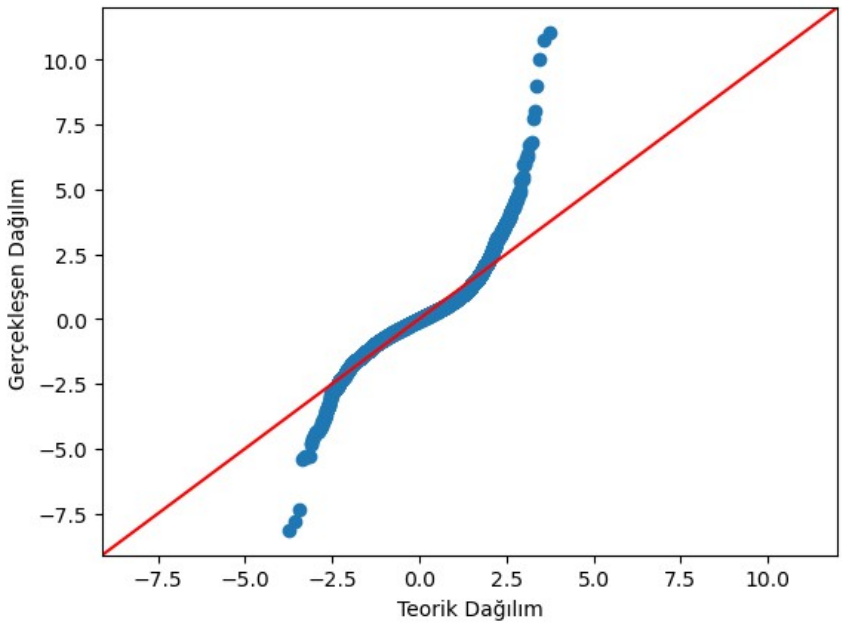
Yukarıda belirtilen parametrelere sahip model ile tahminleme yapılarak gerçek veri ile kıyaslanmış ve elde edilen kalıntıların dağılımı incelenmiştir. Şekil 4:8'e bakıldığında kalıntıların beklendiği şekilde 0 değeri etrafında dağıldığı görülmüştür. Aksi durumlarda model ön yargılı (ing. *Bias*) bir şekilde tahminleme yapacağı için sağlıklı olmayan bir performs sergileyebilmektedir. Ek olarak bu modelin ortalamada ±4.10 hata payına (*RMSE*) sahip olması modelin iyi kabul edilebilecek bir performansa sahip olduğu göstermektedir. Tablo 4:1'de yer alan istatistiksel özete bakıldığında *NOx* verisinin 12 ppm standard sapma değerine sahip olduğu görülmektedir. Bu durumda standard sapmanın üçte biri kadar ortalama hataya sahip bir model elde edildiği görülmüştür.

İyi tahminleme yapabilen bir modelde istatistiksel açıdan kalıntıların normal dağılıma benzer bir dağılım göstermesi beklenmektedir çünkü bu durum hatanın rastgelelik ile açıklanabilmesini sağlamaktadır. Şekil 4:9'te ise kalıntı dağılımının normal dağılım ile olan karşılaştırmasına yer verilmiştir. Bu grafiğe bakıldığı zaman hata dağılımının normal dağılıma göre daha dar bir şekle sahip olduğu görülmektedir.



Şekil 4:8: NOx Emisyonunda Kalıntı ve Gerçek Veri Karşılaştırma Grafiği

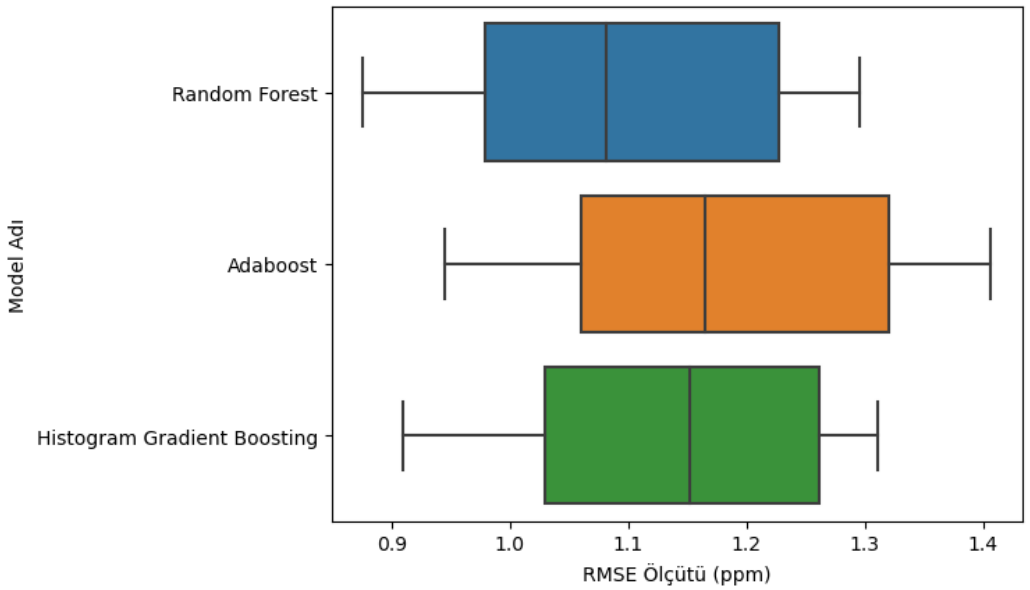
Ancak hatalar ±2.5 standard sapmaya kadar (>%95) normal dağılıma yakın bir dağılım gösterirken bu limitler dışarısında normal dışı dağılım göstermektedir. Bu da hata dağılımının ciddi oranda normale yakın bir dağılıma sahip olduğunu ve az miktarda tahmin hatasının rastgelelelik ile açıklanamayacağını göstermektedir.



Şekil 4:9: NOx Emisyonunda Kalıntı Dağılım Grafiği

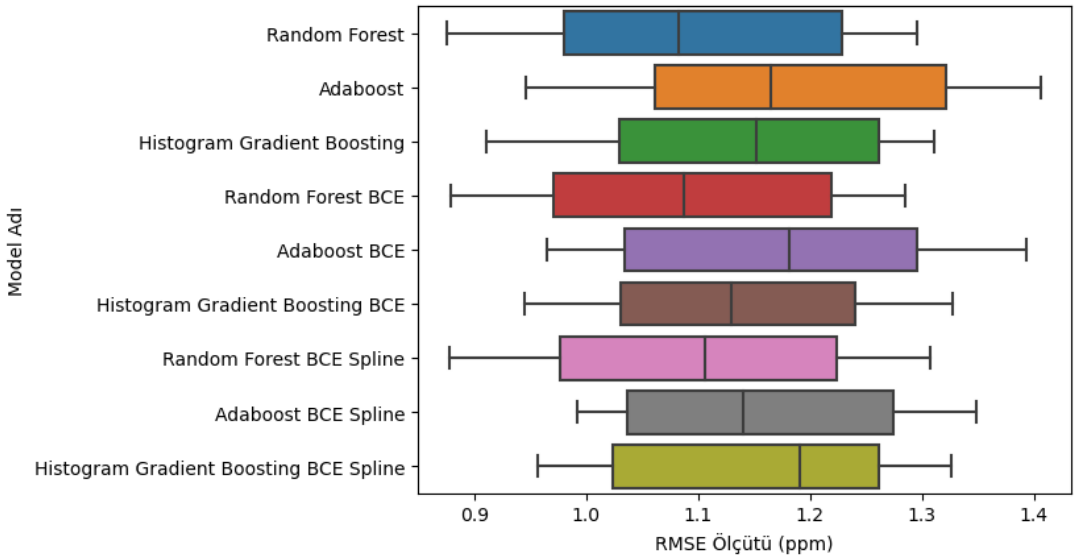
4.3. CO Emisyon Modelleri

NOx emisyonunda gerçekleştirilen modelleme çalışmalarının ardından aynı iş akışı *CO* emisyonu modelleme çalışmaları için de gerçekleştirilmiştir. *NOx* emisyonu çalışmaları sırasında *SVR* ve *CART* algoritmalarının düşük performans göstermesinden kaynaklı olarak *CO* emisyonu çalışmalarında da sadece *ensemble* metotları kullanılmıştır.



Şekil 4:10: CO Emisyonunda Tahminleme Modellerinin Kıyaslanması

Şekil 4:10'te yer alan hiper parametre iyileştirme sonrasında yapılan kıyaslamada üç modelin de benzer performanslara sahip olduğu görülmektedir. *Random Forest* algoritmasının az bir fark ile daha iyi performans sergilediği değerlendirilebilir ancak bu farklılık, yapılan permütasyon testi sonucunda, istatistiksel açıdan önemsiz bulunmuştur.



Şekil 4:11: CO Emisyonunda Nitelik Mühendisliği Uygulanmış Veri Setlerinde Model Karşılaştırmaları

Daha sonra *NOx* emisyonunda olduğu gibi nitelik mühendisliği uygulanan veri setinde *CO* emisyonu değeri için ortak bilgi haritası oluşturulmuştur. Ortak bilgi değerlerine bakıldığında

türetilerek eklenen türbin veriminin *CO* emisyonu hakkında az miktarda (0.12 değeri ile sondan üçüncü) bir bilgi sağladığı görülmüştür.

Şekil 4:11'de yer alan model performans dağılımları incelendiğinde *NOx* emisyonunda olduğu gibi model performanslarında nitelik mühendisliği sonrası belirgin bir iyileşme gerçekleşmemiştir.

Tablo 4:5: CO Emisyonunda 7-fold Öğrenme ve Test Setinde Model Performans (RMSE) Değerleri

	Random Forest	*Adaboost*	*Histogram Gradient Boosting*	*Random Forest BCE Spline*	*Adaboost BCE Spline*	*Histogram Gradient Boosting BCE Spline*
ÇD*-1	1.29	1.39	1.27	1.27	1.34	1.30
ÇD-2	0.87	0.94	0.91	0.87	0.99	0.95
ÇD-3	0.92	1.02	1.01	0.91	0.99	0.97
ÇD-4	1.03	1.09	1.04	1.03	1.07	1.07
ÇD-5	1.16	1.24	1.24	1.17	1.21	1.21
ÇD-6	1.29	1.40	1.31	1.30	1.33	1.32
ÇD-7	1.08	1.16	1.15	1.10	1.14	1.19
Ortalama	1.09	1.18	1.13	1.09	1.15	1.14
SS**	0.15	0.16	0.13	0.15	0.13	0.13
Test Verisi *RMSE*	1.02	1.15	1.10	1.00	1.09	1.09
Test Verisi R^2	0.78	0.72	0.74	0.79	0.75	0.75

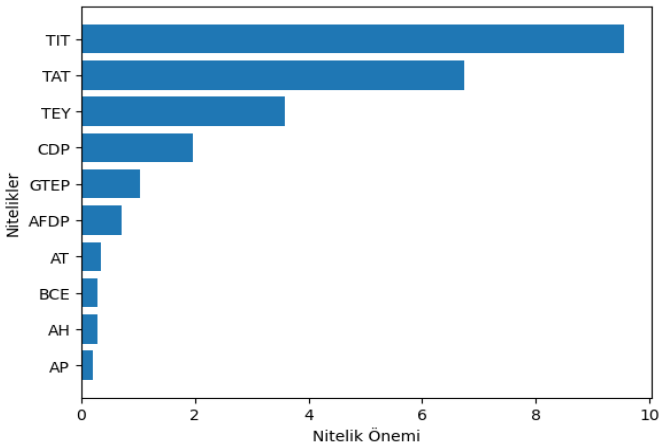
*ÇD: Çapraz Doğrulama
**SS: Standard Sapma

Tablo 4:5'te yer alan öğrenme veri setindeki performans sonuçları permütasyon testi ile incelendiğinde bütün modeller için p-değeri 0.35 değerinden yüksek gelmiştir. Elde edilen p-değeri 0.05 sınır değerinden çok yüksek olduğu için performansın iyileşmediği istatistiksel olarak da teyit edilmiştir. Ancak test verisinde NOx modeline benzer bir iyileşme (yaklaşık %2) gerçekleştiği için nitelik mühendisliği ile dönüştürülen veri setinin kullanımına devam edilmiştir. Ek olarak *CO* emisyonu tahmin modellerinin *NOx* tahmin modellerinden daha düşük bir R^2 değerine sahip olduğu bilgisi de elde edilmiştir.

Nitelik öneminin incelenmesi için bu noktaya kadar en iyi performansı sağlayan türetilmiş türbin verisinin de yer aldığı veri setinde eğitilen *Random Forest* modeli kullanılmıştır.

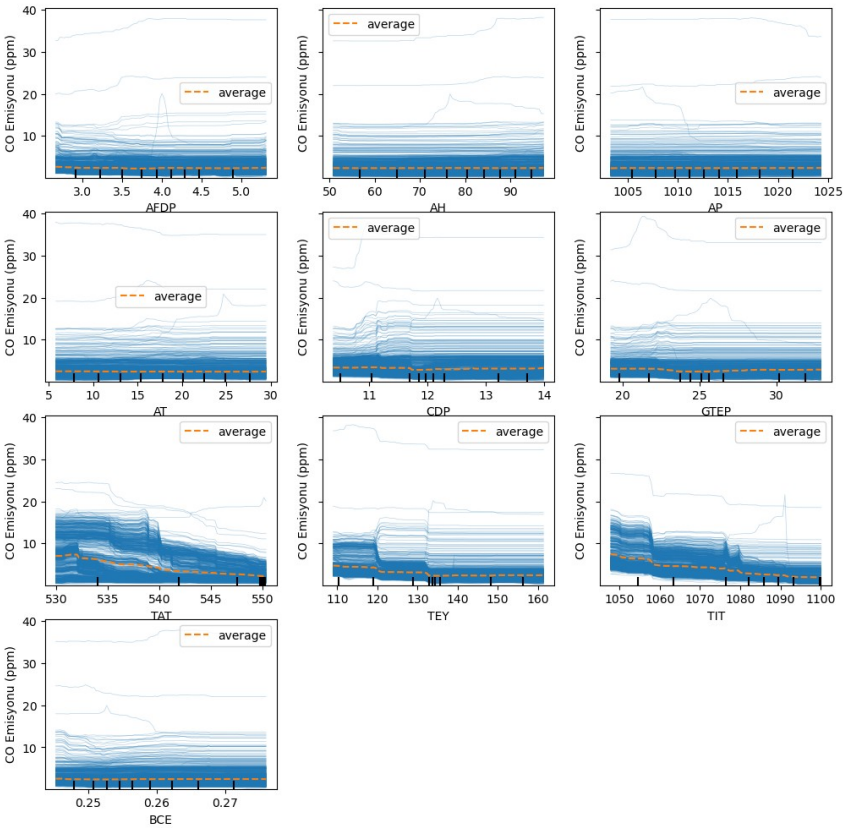
Şekil 4:12'de yer alan nitelik önemi sıralamasına bakıldığı zaman *CO* emisyon değerlerinin sadece operasyon koşullarından etkilendiği (*TIT, TAT, TEY*) ve atmosfer sıcaklığı, basıncı, nemi gibi ortam koşullarının önemsiz olduğu görülmüştür. *NOx* emisyonundan farklı olarak önemi düşük olarak nitelendirilebilecek nitelik sayısı oldukça fazladır. Nitelik mühendisliği sonucunda türetilen türbin

veriminin de *NOx* emisyonunun aksine *CO* emisyonu için göreceli olarak önemsiz bir nitelik olduğu görülmektedir.



Şekil 4:12: CO Emisyonunda Permütasyon Yöntemine Göre Bağıl Nitelik Önemi

CO emisyon değerinin sadece operasyon koşullarına bağlı olması literatür ile de tutarlı bir durum ortaya koymuştur. *CO* oranı yakıtın düzgün yanmamasından kaynaklandığı için temelde yakıt odası özellikleri (*TIT, TAY, TEY*) gibi niteliklerden etkilenmesi fiziken de manalıdır. Niteliklerin kısmı bağımlılıkları da Şekil 4:13'te gösterilmiştir.



Şekil 4:13: CO Emisyonunda Nitelikler İçin Kısmı Bağımlılık Grafikleri

Şekil 4:13'te gösterilen kısmı bağımlılık grafiklerine bakıldığı zaman türbin sıcaklık verilerinin (*TAT*, *TIT*) çok belirgin bir şekilde *CO* emisyonlarını etkilediği görülebilmektedir. *NOx* emisyonundan farklı olarak çoğu nitelik için kısmı bağımlılığın da değişkenlik göstermediği görülebilmektedir. *CO* emisyonunun tahmininde önemsiz olduğu değerlendirilen nitelikler arasından hava basıncı (*AP*) niteliği veri setinden çıkartılarak modeller tekrar eğitilmiştir.

Tablo 4:6: NOx Emisyonunda Nitelik Seçimi Sonrasında Eğitilen Modellerin Performans Çıktıları

Model Adı	Öğrenme Seti Ortalama *RMSE*	Öğrenme Seti Standard Sapma	Test Seti *RMSE*	Test Seti R^2
Random Forest	1.11	0.16	1.02	0.78
Adaboost	1.20	0.14	1.13	0.73
Histogram Gradient Boosting	1.16	0.15	1.08	0.75

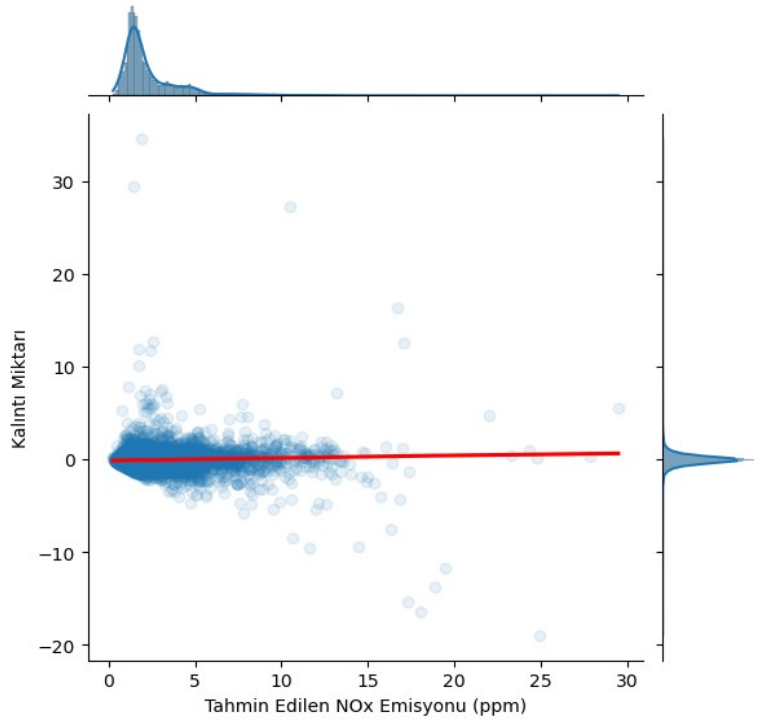
Tablo 4:6'da yer alan sonuçlara bakıldığı zaman nitelik seçiminin öğrenme setinde her modelin daha düşük performans ile çalışmasına sebep olduğu görülmüştür. Test veri setindeki performanslara bakıldığında ise sadece *Histogram Gradient Boosting* modelinde az da olsa bir iyileşme sağlanabilmiştir.

Bu noktaya kadar eğitilen modeller karşılaştırıldığında test setindeki en düşük *RMSE* skorunun nitelik mühendisliği yapılmış veri üzerinde eğitilen *Random Forest* modeli olduğu görülmüştür. Son olarak *Random Forest* algoritmasının hiper parametre uzayını daha efektif tarayabilmek adına aynı uzay limiti içerisinde *Bayes* tipi arama yapılmıştır. Ancak bu arama sonucunda öğrenme setindeki ortalama *RMSE* değeri 1.09 (SS: 0.15) olarak elde edilmiş ve *Random Forest* algoritması için bir performans iyileştirmesi sağlanamamıştır.

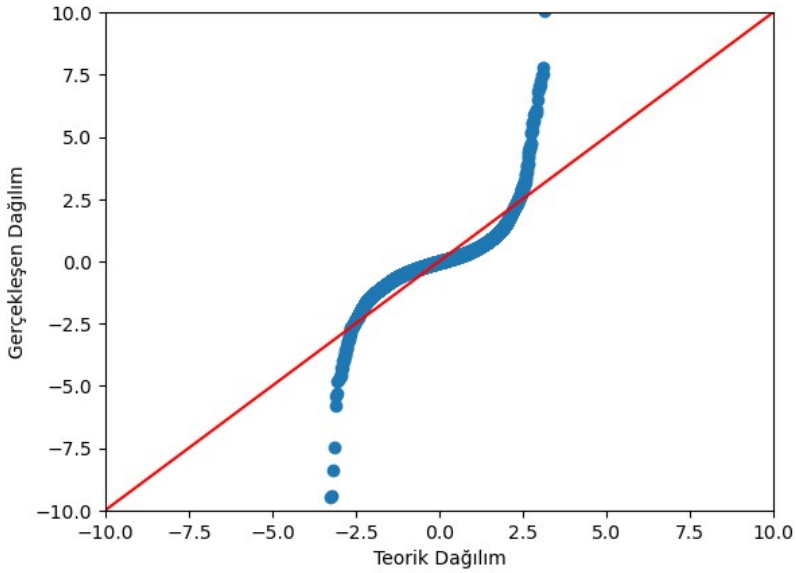
CO emisyonu için modeller kıyaslandığında en iyi performansa sahip modelin aşağıdaki parametreler (geri kalan parametreler varsayılan değerlerine sahiptir.) ile nitelik mühendisliği uygulanmış veri üzerinde eğitilen *Random Forest* modeli olduğu bulunmuştur.

- n_estimators: 235,
- min_samples_leaf: 1,
- max_features: 0.1,
- max_depth: 50

Yukarıda belirtilen parametrelere sahip model ile tahminleme yapılarak tahminler gerçek veri ile kıyaslanmış ve elde edilen kalıntıların dağılımı incelenmiştir. Şekil 4:14'e bakıldığında kalıntıların *NOx* emisyonuna benzer şekilde 0 değeri etrafında dağıldığı görülmüştür. Ancak *CO* modelinde *NOx* modelinden farklı olarak hataların yüksek emisyon değerleri için ortalamada 0 değerinin üzerine çıkmaya yönelik bir davranışı olduğu görülmektedir. Bu da modelin yüksek emisyon değerleri için pozitif bir ön yargıya sahip olduğunu göstermektedir.



Şekil 4:14: CO Emisyonunda Kalıntı ve Gerçek Veri Karşılaştırma Grafiği



Şekil 4:15: CO Emisyonunda Kalıntı Dağılım Grafiği

Öte yandan bu modelin test verisinde ortalama ±1.00 hata payına (*RMSE*) sahip olması modelin iyi kabul edilebilecek bir performansa sahip olduğu göstermektedir.

Tablo 4:1'de yer alan istatistiksel özete bakıldığında *CO* verisinin 2.2 ppm standard sapma değerine sahip olduğu görülmektedir. Bu durumda standard sapmanın yarısı kadar ortalama hataya sahip bir model elde edildiği görülmüştür.

Şekil 4:15’te yer alan kalıntı dağılım grafiğine bakıldığı zaman *NOx* emisyonuna benzer şekilde dar bir dağılım grafiği elde edildiği görülmüştür. Ancak *CO* emisyon tahmin kalıntılarında *NOx*’tan farklı olarak orta bölgede de normal dağılımdan sapma söz konusudur. Bu durum *CO* hata kalıntılarının rastgelelik ile açıklanmasını daha zor bir hale getirmektedir. Bu durum sahip olunan veri setinin *CO* emisyonunu fiziksel olarak açıklamak konusunda yetersiz kaldığı şeklinde değerlendirilebilir.

5. Değerlendirme

Bu raporda yer alan çalışmalar incelendiğinde gaz türbini sistemine ait sensör verilerinden oluşan niteliklerin *NOx* ve *CO* emisyon değerlerini genel manada tahminlemeye yeterli olabileceği görülmüştür. *Ensemble* modellerinin *SVR* ve *CART* modellerinden daha iyi ve verimli bir performansa sahip olduğu değerlendirilmiştir.

NOx emisyonunun tahminlemesinde önerilen nitelik mühendisliği yöntemlerinin model performansı açısından yararlı olduğu sonucu çıkarılmıştır. Nitelik seçiminin ise model performansını kötü yönde etkilediği görülmüştür. Öte yandan Bayes tipi arama yönteminin model performansına katkı sağlamadığı değerlendirilmiştir. *NOx* emisyonu için en iyi tahminleme modelinin *Histogram Gradient Boosting* modeli olduğu değerlendirilmiştir. *NOx* emisyon tahmin kalıntıları incelendiğinde modelin hem yüksek bir R^2 değerine hem de düşük bir *RMSE* değerine sahip olduğu değerlendirilmiştir. Kalıntı dağılımlarına bakıldığında ise dağılımın büyük oranda normal dağılıma benzerlik gösterdiği görülmüş ve bu durumun da modelin açıklayamadığı hataların rastgelelik ile açıklanabilmesine olanak sunduğu değerlendirilmiştir.

CO emisyonunun tahminlemesinde ise önerilen nitelik mühendisliğinin öğrenme verisindeki model performansını *NOx* emisyon modelleri kadar iyileştiremediği görülmüştür. Ancak test verisindeki iyileşmenin *NOx* emisyon modeline benzer olduğu görülmüştür. NOx modeline benzer şekilde nitelik seçiminin modeller üzerinde pozitif bir etkisi olmamıştır. *CO* emisyonu için en iyi tahminleme modelinin *Random Forest* modeli olduğu değerlendirilmiştir. *CO* emisyon tahmin kalıntıları incelendiğinde *NOx* emisyonuna göre daha düşük bir R^2 değeri elde edilmiştir. *RMSE* değerinin ise kabul edilebilecek seviyelerde (*SS*/2) olduğu değerlendirilmiştir. Ancak kalıntıların dağılımı hakkında incelemeler yapıldığında hem yüksek emisyonlarda pozitif sapmaların olduğu hem de dağılımın normalden uzak olduğu görülmüştür. Bu da eldeki niteliklerin *CO* emisyonunu tahminlemek için *NOx* emisyonu kadar yeterli olmadığı şeklinde değerlendirilmiştir. Kısmı bağımlılık grafiğinden elde edilen çoğu niteliğin etkisiz olduğu bilgisinin de bu sonucu desteklediği değerlendirilmiştir.

NOx ve *CO* emisyon tahmin modelleri arasındaki en büyük farklardan birisi de *CO* emisyon modelinin yaklaşık 800 MB yer kaplıyorken *NOx* emisyon modelinin yaklaşık 3.5 MB yer kaplıyor olmasıdır. Çalışmalar yeteri kadar alanın bulunduğu bir bilgisayar ortamında yapıldığı için bu boyutlar bir problem teşkil etmemektedir ancak modellerin gömülü bir sistemde çalışması değerlendirildiği takdir de *CO* emisyon modelinde performanstan ödün verilerek *Histogram Gradient Boosting* modeline geçilerek oldukça efektif bir model kullanımı değerlendirilebilir. *Random Forest* ve *Histogram Gradient Boosting* arasındaki boyut farklılığı *RAM* kullanım miktarlarında da gözlemlenmiştir.

Veri setinin alındığı çalışmanın aksine bu çalışmada *NOx* emisyonu *CO* emisyonlarından daha iyi bir şekilde tahmin edilebilmiştir[1]. Ek olarak literatürde de gösterildiği üzere *ensemble* metotları bu tip bir olgunun modellenmesi için uygun görülmektedir[3].

Gaz türbinleri karmaşık ekipmanlar olduğu için ve yanma reaksiyonları doğrusal olmayan karışık ilişkiler ile açıklanabildiğinden ötürü her gaz türbini ekipmanının kendi verisi ile eğitilerek özel bir modele sahip olması gerekliliği değerlendirilmiştir[2].

Bu çalışmadan elde edilen sonuçlar doğrultusunda daha iyi bir tahminleme modeli için aşağıdaki çalışmaların yapılması önerilmektedir:

- Türbin operasyon modları hakkında (düşük güç, orta güç, yüksek güç) nitelik mühendisliği yapılarak sistemi tanımlamaya yardımcı yeni niteliklerin oluşturulması
- Hem *CO* hem de *NOx* tahminlerinin iyileştirilebilmesi için türbin operasyonu hakkında daha çok niteliğin veri setine eklenmesi
- Zamana bağlı yıpranmayı temsil edebilecek son bakımdan beri geçen süre, toplam çalışma süresi gibi bazı niteliklerin veri setine eklenmesi
- Farklı *ensemble* modellerinden (örn. *XGBoost*) faydalanılarak model performansının iyileştirilmesi
- Hiper parametre aramasının yapıldığı uzayların optimize edilmesi.

6. Kaynakça

[1] Kaya, H., Tüfekci, P., & Uzun, E. (2019). Predicting CO and NOx emissions from gas turbines: novel data and a benchmark PEMS. Turkish Journal of Electrical Engineering and Computer Sciences, 27(6), 4783-4796.

[2] Korpela, T., Kumpulainen, P., Majanne, Y., & Häyrinen, A. (2015). Model based NOx emission monitoring in natural gas fired hot water boilers. IFAC-PapersOnLine, 48(30), 385–390. doi:10.1016/j.ifacol.2015.12.409

[3] Lv, Y., Liu, J., Yang, T., & Zeng, D. (2013). A novel least squares support vector machine ensemble model for NOx emission prediction of a coal-fired boiler. Energy, 55, 319–329. doi:10.1016/j.energy.2013.02.062

[4] Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and computing, 14(3), 199-222.

[5] Loh, W.-Y. (2011). Classification and regression trees. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(1), 14–23. doi:10.1002/widm.8

[6] Bruce P. C. & Bruce A. (2017). Practical statistics for data scientists : 50 essential concepts

(First). O'Reilly.

[7] Pavri, R., & Moore, G. D. (2001). Gas turbine emissions and control. Atlanta: GE Energy Services, 1, 1-20.

[8] Tosun, I. (2020). Thermodynamics: Principles and Applications. World Scientific.

7. Ekler

Bu çalışma kapsamında yazılan *Python* kodlarına https://github.com/TunahanKanbak/VBM683 adresinden erişilebilir.