

Küçük Ölçekli Verilerle Büyük Ölçekli Bir Üretim Tesisinin Makine Öğrenimine Dayalı Elektrik Enerjisi Tüketim Tahmini

Machine Learning Based Electric Energy Consumption Prediction of a Large-Scaled Production Plant with Small-Scaled Data

Volkan Özdemir¹, Anıl Çalışkan², Arif Yiğit³

¹Brisa Bridgestone Sabancı Lastik San. Tic. A.Ş., Kocaeli, Türkiye
v.ozdemir@brisa.com.tr

²Brisa Bridgestone Sabancı Lastik San. Tic. A.Ş., Kocaeli, Türkiye
anil.caliskan@brisa.com.tr

³Brisa Bridgestone Sabancı Lastik San. Tic. A.Ş., Kocaeli, Türkiye
a.yigit@brisa.com.tr

Özetçe—Bu çalışma bir lastik üretim fabrikası için tüketilen enerjiyi tahmin ve takip etmeye yönelik istatistiksel bir yaklaşımı kapsamaktadır. Çalışmanın arkasındaki nedenler, enerji tüketim bütçesini optimize etmek ve ISO 50001 Enerji Yönetim Sistemi Standardı için de kritik önem taşıyan üretim proses bilgisine dayalı performans indikatörlerini belirlemek ve takip etmektir. Bunun gerçekleşmesi için endüstriler arası veri madenciliği standart prosesinin (CRISP-DM) [1] adımları takip edilmiştir. Bu çalışmanın en önemli noktası, tarihsel verilerin küçük ölçekli olmasına rağmen, parametrelerin girdi örneklerine göre daha yüksek bir boyuta sahip olmasıdır. Böylelikle girdi olarak kullanılacak veriler bütçe döneminde kullanılacak basit değişkenlerle açıklanabilir. Çalışma, üretim alanına dayalı veri hazırlama adımlarını, çapraz geçerleme ile en iyi bağlantı algoritması ve parametrelerinin seçimini, makine öğrenimi modellerinin karşılaştırmasını ve sonuçlarını takip eden yedi aylık doğrulama adımını tanıtmaktadır.

Anahtar Kelimeler—Makine Öğrenimi; Bağlanım; Küçük Ölçekli Veri Analizi; Enerji Tüketim Tahmini.

Abstract—This report covers the statistical approach to predict consumed energy for a tire production plant. The reasons behind this study are also to optimize the energy consumption budget and to follow the production area wised

KPIs which is also vital for ISO 50001 Energy management system standard. In order to make it happen, writers clarify the main problem, then start to apply the steps of the cross industry standard process for data mining (CRISP-DM) [1] methodology. The most important point of this study was that although the historical data is small scaled, the parameters have a higher dimension according to input examples. Hence, the data to be used as input could be explained with simple variables to be used in the budget period. The study introduces data preparation steps based on the production area, grid search for best regression algorithm, comparison of models, and seven-month validation results.

Keywords—Machine Learning, Regression, Small Scaled Data Analytics, Energy Consumption Prediction.

I. GİRİŞ

İmalatçı firmalar için pazardaki rekabet her geçen gün artmakta ve bu yarışta öne geçmek için üretim maliyetlerinin düşürülmesi büyük önem taşımaktadır. Üretim maliyetlerindeki ana kalemler enerji, malzeme, işçilik ve lojistikten gelmektedir. Üretici firmalar, pazarda daha rekabetçi olabilmek adına, şirketin tanımladığı finansal hedeflere ulaşmayı amaçlamaktadırlar. Bunun için ihtiyaç duydukları enerji, malzeme, işçilik ve lojistik giderlerini doğru bir şekilde bütçelemeleri gerekmektedir. Yapılan bütçeleme çalışmaları, toplanan karmaşık verilerin kısa sürede yorumlanması ve hedeflenen rakamlar için

açıklanabilir bir senaryo oluşturulması bakımından büyük değer taşımaktadır.

Otomotiv ve lastik endüstrileri için bütçe döneminde tüketilecek elektrik enerjisinin doğru tahmin edilmesi ve ardından gerçekleşme döneminde bütçe hedefinden sapan parametrelerin belirlenerek önlem alınması finansal denge ve sürdürülebilirlik açısından önemlidir. Hedeflenen iyileşmeyi sağlamak için, bütçe döneminde verilebilecek frekansı aylık olan sınırlı sayıda parametre ile başarılı tahmin yapmak ve gerçekleşme dönemi içerisinde de takip etmek gerekmektedir. Bu noktada doğru çözüm veri biliminde gizlidir. Veri bilimi, karmaşık problemlerin üstesinden gelmek için bir dizi araç ile veri çıkarımı yapmayı sağlayan, istatistiksel modelleme, matematiksel optimizasyon ve algoritma geliştirme çalışmaları için çok disiplinli bir alandır. Veri odaklı yaklaşım, bir işletmenin kârını artırmasına ve finansallarını genişletmesine yardımcı olmak amacıyla enerji, malzeme, işçilik ve lojistik kalemleri odağında geçmiş verilerindeki gizli şablonları ortaya koymaktadır.

Lastik üretimini ele aldığımızda, birbirini takip eden yedi proses karşımıza çıkmaktadır. Bu proseslerin doğru planlanması üretim maliyetlerinin optimize edilmesi ve enerji tüketiminin minimumda tutulması adına belirleyicidir [2]. Birbirini takip eden bu prosesler sırasıyla; (1) hammaddelerin karıştırılarak karışımların oluşturulduğu mikser sahasında, (2) karışımlar ile profilli bileşenlerin oluşturulduğu ekstruder sahasında, (3) takviye bileşenlerinin kaplandığı ve kesildiği stok sahasında, (4) malzemelerin birleştirildiği lastik imalat sahasında, (5) birleşen malzemelerin pişirildiği pişirme sahasında, (6) çıkan ürünün kontrol edildiği bitirme sahasında ve (7) tüm bu sahalar enerji sağlayan yardımcı işletmeler sahasında gerçekleşmektedir. Her bir sahanın kendine özgü makine parkuru ve proses adımları bulunmaktadır. Dolayısıyla ilgili sahanın tüketeceği enerjiyi belirleyen ana parametreler de farklılık göstermektedir. Bu doğrultuda ilgili parametrelerin belirlenerek, hedeflerin alan bazlı yapılması hem takip açısından hem de enerji yönetimi açısından önem taşımaktadır.

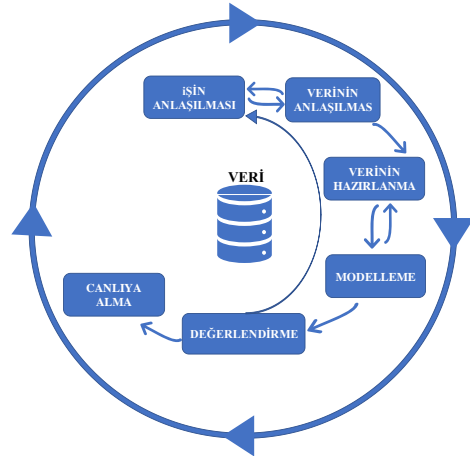
ISO 50001 Enerji Yönetim Standardı, kuruluşların enerji kullandıklarını daha iyi yönetmelerine yardımcı olarak üretkenliği arttırmayı ve maliyeti düşürmeyi amaçlayan bir sistemdir [3]. Standart, organizasyonların enerjilerini daha verimli kullanmaları için bir polişe oluşturmalarını ve bu polişeye uygun hedefleri tayin etmelerini istemektedir. Organizasyonlar bu hedeflere ulaşmak için enerji tüketimini doğru temsil eden ana parametreleri belirlemekte ve bu parametrelerin gerçekleşmesini ölçerek sonuçları değerlendirip enerji tüketiminin sürekli gelişimini sağlamaktadırlar. Standartlaştırılmış ve etkili enerji yönetim sistemi, kuruluşların ekolojik ve operasyonel faydalarla tasarruf etmesine yön verirken, kaynaklarını korumasını da sağlamaktadır [4]. ISO 50001, global ölçekte endüstriyel enerji tüketiminin %60'ına etki etme potansiyeline sahip olup, enerji üreticilerinin verimliliğine

de yılda %2.5'tan fazla bir iyileştirme şansı yaratmaktadır. [5].

Tüm bu faydalar ışığında Kocaeli, Türkiye'de bulunan lastik üretim fabrikasının alan bazlı olarak elektrik enerjisi tüketimine etkisi olan parametrelerinin belirlenmesi ve bütçe döneminde sonraki yılların enerji tüketimini bu parametreler aracılığıyla tahmin edebilecek modellerin geliştirilmesi ihtiyacı ortaya çıkmıştır.

II. VERİ MADENCİLİĞİ SÜRECİ

Veri madenciliği için sektörler arası standart süreç (CRISP-DM) metodolojisi, verileri değerlendirmek ve veriler için en iyi modeli geliştirmek amacıyla veri madenciliğini genelden özele giden hiyerarşik bir süreç modeli olarak tanımlamaktadır [1]. Süreç, işin anlaşılması ile başlamakta, ardından veriyi anlama, veri hazırlama, modelleme, değerlendirme ve canlıya alma adımları ile tekrarlı bir şekilde devam etmektedir (Şekil 1). Bu süreç, karşılık gelen verileri müşterinin tanımladığı ölçülebilir kabul kriterlerinde ifade eden en iyi modeli uygulamaya koymayı amaçlamaktadır. CRISP-DM metodolojisi banka dolandırıcılığı tespiti [6], tıbbi veri madenciliği [7], mühendislik uygulamaları [8] gibi çok çeşitli alanlarda kullanılmıştır. Enerji tüketimi tahmini için de bu metodoloji adımları takip edilerek süreç ele alınmıştır.



Şekil 1. CRISP-DM Şeması ve Akışı

A. İşin Anlaşılması

Lastik üretim tesisi, elektrik enerjisi tüketimi ölçülebilir alt üretim alanlarına bölünmüştür. Bu alanlar proses adımları, çıkan son ürünün benzerliği ve makine parkuruna göre kümelenmiştir. Böylelikle mikser, ekstruder, stok hazırlama, lastik imalat, pişirme, bitirme ve yardımcı işletmeler olmak üzere 7 adet alan için elektrik enerjisi tüketimini tahmin edecek bağımlı (regression) modelleri kurulması ve takip edilmesi amaçlanmaktadır. Bu temel alanların her biri için elektrik enerji tüketimini etkileyen temel parametreleri belirlemek ve bu parametrelerin geçmiş

verileri ile başarıyı yüksek makine öğrenmesi modeli geliştirmek ana hedeflerdir. İlave olarak sonraki yıllar için tahmin yapabilecek bir sistem geliştirilerek elektrik enerjisi tüketimi de takip edilecektir.

B. Verinin Anlaşılması

Bu adımda elektrik enerjisi tüketimini etkileyen ve bütçeleme dönemindeki tahmine zemin oluşturacak parametrelerin belirlenmesi adına ilgili alana giren ilk ürün, çıkan ürün, makine parkuru ve üretim temel parametreleri incelenmiştir. Her bir sahanın 2013 yılından itibaren tükettiği aylık elektrik enerjisi (kWh) veri tabanından toplanmıştır. Bu veri hedef verisi olarak kaydedilmiştir. Diğer veriler ortak parametreler, sahaya özgü parametreler ve sahaya özgü makine değişimleri olarak 2013 yılından itibaren aylık olarak SAP sisteminden ve SQL veri tabanlarından elde edilmiştir. Son 7 yıldan elde edilen 84 aylık veri içerisinde 70'den farklı parametreye ulaşılmıştır. Hedef çıktı, elektrik enerjisi tüketimi (kWh) olarak belirlenmiştir. Ortak veriler üretim tonajı (ton/ay), çalışma günü, ortalama sıcaklık (C), ürün bazında üretim oranları (%) gibi tüm sahalara etkileyecek gerçek verilerdir. Sahalara özgü veriler ise o sahada bulunan makinelerin çalışma süreleri, malzeme özellikleri gibi sahaya özgü gerçek verilerdir. Son olarak makine değişimi verileri de 2013 yılından itibaren elektrik enerjisi tüketimini doğrudan etkileyebilecek ilave makine kurulumu veya çıkarılması temsil eden kategorik verilerdir. Tüm bu veriler ile oluşturulan veri havuzu Şekil 2'de verilmiş. Veri havuzu içerisinde ilgili sahanın elektrik enerjisi tüketimini etkileyen parametrelerin seçilmesi için verilerin analiz edilmesi ve hazırlanması gerekmektedir. Eksik verilerin doldurulması, çoklu eş doğrusallık sorununun (multicollinearity) ele alınması ve çok boyutluluk sorununun (multidimensionality) giderilmesi adımlarının ardından veriler hazırlanmış olacaktır.

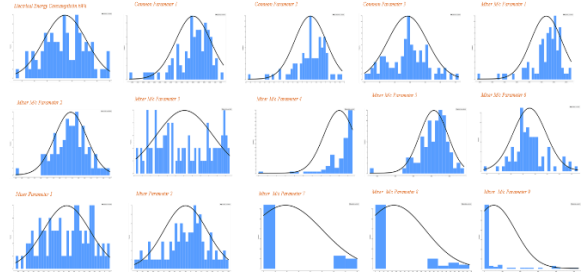
İLGİLİ SAHA	MİKSER	STOK HAZIRLAMA	EKSTRUDER	LASTİK İMALAT	PIŞIRMA	BİTİRME	YARDIMCI İŞLETMELER
Y = HEDEF	MİKSER ELEKTRİK TÜKETİMİ (kWh)	STOK HAZIRLAMA ELEKTRİK TÜKETİMİ (kWh)	EKSTRUDER ELEKTRİK TÜKETİMİ (kWh)	LASTİK İMALAT ELEKTRİK TÜKETİMİ (kWh)	PIŞIRMA ELEKTRİK TÜKETİMİ (kWh)	BİTİRME ELEKTRİK TÜKETİMİ (kWh)	YARDIMCI İŞLETMELER ELEKTRİK TÜKETİMİ (kWh)
X = GİRİŞ	ORTAK PARAMETRELER						
	MİKSER PARAMETRELERİ	STOK HAZIRLAMA PARAMETRELERİ	EKSTRUDER PARAMETRELERİ	LASTİK İMALAT PARAMETRELERİ	PIŞIRMA PARAMETRELERİ	BİTİRME PARAMETRELERİ	YARDIMCI İŞLETMELER PARAMETRELERİ
	MİKSER M DEĞİŞİMİ	STOK HAZIRLAMA M DEĞİŞİMİ	EKSTRUDER M DEĞİŞİMİ	LASTİK İMALAT M DEĞİŞİMİ	PIŞIRMA M DEĞİŞİMİ	BİTİRME M DEĞİŞİMİ	YARDIMCI İŞLETMELER M DEĞİŞİMİ

Şekil 2. Veri Havuzu

C. Verinin Hazırlanması

Hazırlanan veri havuzu içerisinde her bir saha ile ilgili parametreler teker teker ele alınmıştır. Bu adımda ilk odaklanılan, her bir saha için hazırlanan parametrelerin histogram dağılımlarına bakmak olmuştur. Parametrelerin dağılım şekilleri (normal dağılım, sola çarpık dağılım, sağa çarpık dağılım) incelenerek hedef çıktıyı açıklayabilir olup olmadıkları incelenmiştir. Bu adım, tamamlanması gereken

eksik verilerin saptanması ve çok boyutluluk sorunu oluşmaması için yüksek korelasyon ilişkisinde olabilecek parametrelerin belirlenmesi için de öncül bir kontrol olmuştur. Örnek bir saha için seçilen parametrelerin dağılımı Şekil 3'de verilmiştir.



Şekil 3. Örnek bir Saha için parametrelerin histogram dağılımları

Literatürde eksik verilerin doldurulması için birçok farklı yöntem (bayesian estimation, hot-deck imputation, regression imputation, mean substitution, expectation maximization imputation, multiple imputation) bulunmaktadır [9-12]. Veri havuzu içerisinde eksik veri bulunduran parametreler de bu yaklaşımla dağılımlarına göre doldurulmuştur. Normal dağılım gösteren parametreler için çoklu doldurma tekniği olarak Rubin kuralı [13] kullanılmıştır. Sola çarpık dağılım ve sağa çarpık dağılım için ise bağlanım doldurma (regression imputation) yöntemi tercih edilmiştir. Veri havuzunda eksik veri kalmayacak şekilde bu yaklaşım tüm üretim alanları içerisindeki tüm parametreler için uygulanmıştır.

Veri hazırlığının diğer bir önemli etabı olan çoklu eş doğrusallık, iki veya daha fazla değişken arasındaki doğrusal ilişkiyi ifade etmektedir. Bu durum tahminlerin güvenilirliğinde ciddi zorluklara neden olabilecek parametre ilişkisine dikkati çekmektedir. Çoklu eş doğrusallığın varlığının incelenmesi, her zaman herhangi bir çoklu regresyon analizinde ilk adım olarak gerçekleştirilmelidir [14-16]. Bu doğrultuda, kurulacak modelin başarıyı artırmak ve çoklu eş doğrusallık sorunundan kurtulmak için her alan parametre seti içerisinde yer alan her iki bağlanım değişkeni arasındaki aşırı ikili korelasyon incelenmiştir [17]. Aşırı ikili korelasyon içerisinde olan parametreler sadeleştirilerek veri havuzu yalınlaştırılmıştır.

Çok boyutluluk sorunu ise, parametre sayısına göre girdi sayısının çok olması anlamına gelmektedir. Bu durum bağlanım problemlerinde başarıyı azaltmakta, veri setine istenmeyen gürültü eklemektedir. Bu noktada, çok değişkenli veriyi temsil eden daha az boyutlu temel bileşenlerin belirlenmesine dayanan bir boyut azaltma tekniği olan temel bileşenler analizi, (principal component analysis) kullanılmaktadır [18]. Azaltılacak boyut sayısı olan k değerine denklem (1)'de verilen varyans oranına (proportion of variance) göre karar verilmektedir. Bu çalışmada azaltılacak boyut sayısı varyans oranı değerinin

0.96 olduğu k değeri olarak seçilmiştir [19]. Daha sonrasında denklem (1) ve denklem (2)'deki adımlar izlenilerek boyut azaltma ihtiyacı olan sahalara için temel bileşenler analizi gerçekleştirilmiştir.

$$\frac{\alpha_1 + \alpha_2 + \dots + \alpha_k}{\alpha_1 + \alpha_2 + \dots + \alpha_k + \dots + \alpha_d} \quad (1)$$

$$z = \mathbf{W}^T(x - m) \quad (2)$$

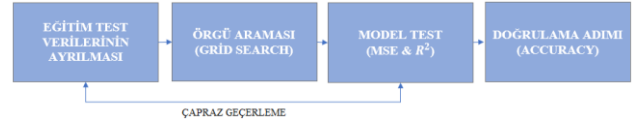
$$\max_{W_i} (W_i^T \sum W_i - \alpha (W_i^T W_i - 1)) \quad (3)$$

İlgili denklemlerde; m örneklem ortalamasını, Σ kovaryansı, \mathbf{W} kovaryansın öz vektörlerini, W_i ilgili α öz değerine ait öz vektörleri, d toplam boyut sayısını ve k azaltılan boyut sayısını ifade etmektedir. Elde edilen veri havuzu için, bağımsız değişkenlerin aralığı veri ölçeklendirme ile normalleştirilmiştir. Ortalama değerin 0, standart sapmanın ise 1 değerini aldığı standardizasyon yöntemi kullanılarak veriler ölçeklendirilmiştir. Denklem (4)'de m örneklem ortalamasını, σ ise standart sapmayı ifade etmektedir.

$$z = \frac{x - m}{\sigma} \quad (4)$$

D. Modelleme

Modelleme süreci içerisinde her bir üretim sahası için hazırlanan veri setleri ile makine öğrenmesi algoritmaları eğitecek olup, sahanın elektrik enerjisi tüketimini en doğru şekilde yansıtan makine öğrenmesi algoritması ve hiperparametre (hyperparameter) seti seçilecektir. Bu süreç boyunca en çok dikkat edilmesi gereken husus eğitim verileri ile ezberleme (overfitting) meydana gelmesinin engellenmesi olacaktır. Bundan dolayı tüm süreç boyunca çapraz geçirme (k-fold cross-validation) kullanılmıştır. Elde bulunan veri, eğitim ve test verisi olarak 5-kez rastgele şekilde ayrılmış ve sonuç test performansı, 5-kez çapraz geçirilen testin ortalama değeri olacak şekilde değerlendirilmiştir. Çözüme ulaşmak istenilen konu eğitici öğrenme (supervised learning) ve bağlantım problemi olduğundan dolayı çoklu doğrusal bağlantım (multiple linear regression), karar ağaçları (decision trees), rastlantısal orman algoritması (random forest), bağlantım için destek vektör makineleri (support vector machines) algoritmaları kullanılmıştır. Destek vektör makineleri için çekirdek fonksiyonlarından (kernel trick) faydalanılmıştır. Doğrusal, polinomal, sigmoid, yarı çapsal temelli (radial basis function) fonksiyonlar kullanılan çekirdek fonksiyonlarıdır. En iyi algoritmayı ve o algoritmaya ait hiper-parametreleri seçmek adına örgü araması (grid search) yapılmıştır [20]. Modelin gerçek performansını takip etmek için ise eğitim ve test verisinde kullanılmayan 2020 yılının ilk 7 ayının gerçekleştirme verisi kullanılmıştır.



Şekil 4. Modelleme ve Değerlendirme Süreci Adımları

E. Değerlendirme

Değerlendirme aşaması test ve doğrulama (validation) adımları olarak ikiye ayrılmıştır. Test adımı çapraz geçirme içerisindeki her eğitim-test verisi ayrışması için eğitilen modelin test verileri üzerindeki tahmin performansını baz almaktadır. Tüm çapraz geçirme sonrasındaki ortalama R^2 ve ortalama kare hata (mean squared error) metrikleri ile algoritma ve hiperparametre setinin başarımını göstermektedir. Denklem (5), denklem (6) ve denklem (7)'de y gerçek değeri, \hat{y} tahmin edilen değeri ve \bar{y} gözlem değerlerinin ortalamasını temsil etmektedir.

$$R^2 = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2} \quad (5)$$

$$\text{ortalama kare hata} = \frac{1}{n} \sum (y - \hat{y})^2 \quad (6)$$

Doğrulama adımı için örgü aramasında başarılı görülen modellerin performansı, daha önce görülmemiş veriler ile birlikte değerlendirilmiştir. 2020 yılının ilk 7 ayı içerisindeki gerçekleştirme verileri ile seçilen modellerin doğruluk başarımı (accuracy) değeri takip edilmiştir.

$$\text{Doğrululuk başarımı} = \frac{y - \hat{y}}{y} \quad (7)$$

F. Canlıya Alma

Doğrulama adımı sonrasında doğruluk başarımı değerlerine göre sistem kullanılmak üzere devreye alınacaktır. Sistem, kullanım kolaylığı açısından bir kullanıcı ara yüzü tasarlanarak kullanıcılara teslim edilecektir. Model iyileştirme, öznetelik çıkarımı ve sürekli iyileştirme çalışmaları da sürekli olarak takip edilecektir.

III. MODEL SEÇİM ÇALIŞMALARI

İlgili üretim alanları için en yüksek tahmin başarımını gösterecek modelleri seçmek adına hazırlanan veri setleri ile örgü araması (grid search) yapılmıştır. Eldeki veri seti içerisinde de parametreler kendi aralarında kombine edilerek bazıları dışında bırakılacak şekilde kombinasyonları oluşturulmuştur. Modeller içerisindeki hiperparametreler, örneğin destek vektör makineleri için çekirdek fonksiyon seçimleri gibi, bu arama ile optimize edilmiştir. Çalışma içerisinde toplam 306000 model-veri seti kombinasyonu deneyerek en iyi başarıma sahip olanlar seçilmiştir.

Tablo 1’de her bir üretim alanı için denenen toplam model sayıları verilmiştir.

Alan Adı	Data Kombinasyon Sayısı	Her Data Kombinasyonu için Hiper Parametre Kombinasyonu	Toplam Denenen Model Sayısı
Mikser	16	3000	48000
Yardımcı İşletmeler	17	3000	51000
Ekstruder	12	3000	36000
Stok Hazırlama	13	3000	39000
Pişirme	14	3000	42000
Lastik İmalat	15	3000	45000
Bitirme	15	3000	45000

Tablo 1. Örgü Araması (Grid Search) Sürecinde Denenen Model Sayısı

Örgü araması sonucunda her bir alana özgü model performansları verilen metriklere göre karşılaştırılmıştır. Daha sonrasında her bir alandaki veri varyasyonunu en iyi açıklayabilecek ve en keskin tahminleri verecek yöntemler seçilmiştir. Kompleks ve yüksek dereceli model seçimlerinden kaçınılmıştır. Tablo 1’de üretim fabrikası içerisinde yer alan her bir alan için, o alana özgün olarak hazırlanan veri seti üzerinde en başarılı çapraz geçişleme sonuçlarını veren modeller sıralanmıştır. Pişirme ve lastik imalat alanları için modellerin R Kare skorları 0.800’ün altındadır. En başarılı modellerin seçilmesine karşın bu iki alan için elektrik enerji tüketiminde oluşacak varyasyonu modeller yeterli kadar açıklayamamaktadır. Dolayısıyla, tarihsel veri seti dağılımı dışında gelecek olan bir veri seti üzerinde hedeflenen başarımların %95 değerinin altında kalınabilir. Bu durumda, ilgili sahalar için modellerin daha gürbüz olması adına sahanın elektrik enerjisi tüketimini temsil edecek farklı öznitelikler çıkarmak üzere ilave çalışmalar yapmak gerekecektir.

Alan Adı	Seçilen Model	Çekirdek Fonksiyon	R Kare Skoru
Mikser	Destek Vektör Makinesi	Polinom 2. Derece	0.986
Yardımcı İşletmeler	Destek Vektör Makinesi	Doğrusal	0.905
Ekstruder	Çoklu Doğrusal Bağlanım	-	0.881
Stok Hazırlama	Destek Vektör Makinesi	Doğrusal	0.818
Pişirme	Destek Vektör Makinesi	Polinom 3. Derece	0.749
Lastik İmalat	Destek Vektör Makinesi	Doğrusal	0.786
Bitirme	Çoklu Doğrusal Bağlanım	-	0.910

Tablo 2. Çapraz Geçişleme Sonucu Her Alan İçin Seçilen Modeller

Modeller 2020 yılı ilk 7 ay verisi ile doğrulama adımında test edileceklerdir. Modelin gerçek başarısı hiç görmediği veri seti üzerinden değerlendirilerek, sonuçların başarımına göre devreye alma işlemi gerçekleştirilecektir

IV. SONUÇLAR

Hazırlanan modeller 2020 yılı içerisinde her ay sonunda gerçekleşme verisi ile doğrulama adımına sokulmuştur. 7 aylık deneme süreci modellerde hiçbir değişiklik yapılmadan sürdürülmüş ve doğruluk

başarımları kayıt edilmiştir. Tablo 3’de verilen sonuçlara göre mikser, yardımcı işletmeler, bitirme, pişirme, stok hazırlama üretim alanlarının ortalama doğruluk başarımları, Nisan 2020 ayı haricinde proje öncesinde belirlenen hedefle paralel olarak %95’in üzerinde bulunmuştur. Lastik imalat ve ekstruder üretim sahalarında ise Nisan 2020 ayı haricinde minimum %89 başarımlarına ulaşılmıştır. Tüm alanlar için doğruluk adımı tamamlanmış olup devreye alma süreci başlatılmıştır.

Alan Adı	Doğruluk Başarımı						
	Ocak 2020	Şubat 2020	Mart 2020	Nisan 2020	Mayıs 2020	Haziran 2020	Temmuz 2020
Mikser	98.2%	98.9%	98.2%	75.5%	96.6%	98.2%	98.0%
Yardımcı İşletmeler	97.7%	99.8%	96.3%	90.2%	98.2%	96.2%	95.0%
Ekstruder	96.6%	95.4%	95.4%	65.4%	93.1%	89.8%	96.2%
Stok Hazırlama	91.4%	98.5%	95.2%	75.7%	97.7%	96.2%	92.8%
Pişirme	98.0%	93.1%	95.8%	75.4%	99.3%	94.0%	99.6%
Lastik İmalat	98.7%	96.9%	96.3%	50.7%	92.1%	95.7%	89.8%
Bitirme	96.9%	97.2%	91.5%	3.5%	91.0%	96.3%	94.8%

Tablo 3. Doğrulama Adımı Aylara Göre Doğruluk Başarımı

Sonuçlara ilişkin bir diğer çıkarım, tüm alanlar için Nisan 2020 ayı doğruluk başarımlarının diğer 6 aya nazaran oldukça düşük olduğu yönündedir. Bu durum, COVID-19 döneminde fabrika üretiminin alışılmışın dışında uzunca bir süre durmasından kaynaklanmaktadır. Tarihsel veri havuzu içerisinde çalışma günü ve çalışma tonajının hiç bu kadar az olmadığı gözlemlenmiştir. Bağlanım modelleri, veri havuzu içerisindeki aralığa bağımlı olarak modelleri inşa ettiğinden dolayı, eğitim verilerinden çok farklı veri setleri için yeterli gürbüzlüğü gösterememektedir. Bununla birlikte, normal üretim dönemlerinde bu denli uç değerler (outlier) ile karşılaşılacağı için Nisan 2020 ayı değerlendirmelere dahil edilmemiş olup, yalnızca literatüre katkıda bulunmak amacıyla raporlanmıştır.

Özet olarak, geliştirilen elektrik enerjisi tahmin modeli, ilgili sahaların çoğunluğunda hedeflenen doğruluk başarımlarını karşılamıştır. Doğruluk başarımları istenilen seviyede olmayan sahalar başta olmak üzere, tüm sahaların başarımları ilave veri ve öznitelik çıkarımı yapılarak geliştirilebilecek düzeydedir. Bir sonraki aşamada, tarihe bağlı olarak artacak veri seti modele dahil edilerek modellerin daimi olarak geliştirilmesi hedeflenmektedir.

Yazar Katkıları

Birinci yazar çalışmanın sorumlu yazarı olup çalışma tasarımını yapmış, verilerin hazırlanması ve modellerin kurulması adımlarını yerine getirmiştir. *İkinci yazar* model

sonuçlarının değerlendirilmesi adımını yürütmüş ve veri analizi sürecine katkıda bulunmuştur. Üçüncü yazar verileri toplamış ve devreye alma işlemini gerçekleştirmiştir.

KAYNAKÇA

- [1] R. Wirth, Jochen. Hipp, "CRISP-DM: Towards a standard process model for data mining. Proceedings of the 4th International conference on the Practical Applications of Knowledge Discovery and Data Mining", 2000.
- [2] S. Gorenstein, "Planning Tire Production. Management Science," 17(2), B-72-B-82. doi:10.1287/mnsc.17.2.b72, 1970.
- [3] International Organization for Standardization, "Energy management systems-Requirements with guidance for use (ISO/DIS Standard No. 50001)," Retrieved from iso.org/publication/PUB100400, 2018.
- [4] T.-Y. Chiu, S.-L. Lo, Y.-Y. Tsai, "Establishing an Integration-Energy-Practice Model for Improving Energy Performance Indicators in ISO 50001 Energy Management Systems," *Energies* 2012, 5, 5324-5339, 2012.
- [5] A. Mckane, D. Desai, M. Matteini, W. Meffert, R. Williams, R. Risser, "Thinking Globally: How ISO 50001- Energy Management can make industrial energy efficiency standard practice," 10.2172/983191, 2009.
- [6] B.C. Rocha, R.T. Júnior, "IDENTIFYING BANK FRAUDS USING CRISP -DM AND DECISION TREES," *International Journal of Computer Science and Information Technology*, 2, 162-169, 2010.
- [7] E. Rivo, J. de la Fuente, Á. Rivo, E. García-Fontán, M.-Á. Cañizares, P. Gil, "Cross-Industry Standard Process for data mining is applicable to the lung cancer surgery domain, improving decision making as well as knowledge and quality management," *Clinical and Translational Oncology*, 14(1), 73-79, 2012.
- [8] S. Huber, H. Wiemer, D. Schneider, S. Ihlenfeldt, "DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP-DM model," *Procedia CIRP*, 79, 403-408. doi:10.1016/j.procir.2019.02.106, 2019.
- [9] C. M. Musil, C. B. Warner, P. K. Yobas, S. L. Jones, "A Comparison of Imputation Techniques for Handling Missing Data," *Western Journal of Nursing Research*, 24(7), 815-829. doi:10.1177/019394502762477004, 2002.
- [10] K. Lakshminarayan, S. A. Harp, T. Samad, "Imputation of Missing Data in Industrial Databases," *Applied Intelligence*, 11(3), 259-275. doi:10.1023/a:1008334909089, 1999.
- [11] S. Sinharay, H. S. Stern, D. Russell, "The use of multiple imputation for the analysis of missing data," *Psychological Methods*, 6(4), 317-329. doi:10.1037/1082-989x.6.4.317, 2001.
- [12] G. Hawthorne, P. Elliott, "Imputing Cross-Sectional Missing Data: Comparison of Common Techniques," *Australian & New Zealand Journal of Psychiatry*, 39(7), 583-590. doi:10.1080/j.1440-1614.2005.01630, 2005.
- [13] B. Rubin. Donald, "Inference and missing data, Biometrika," Volume 63, Issue 3, December 1976, Pages 581-592, https://doi.org/10.1093/biomet/63.3.581, 1976.
- [14] D. A. Belsley, "Conditioning Diagnostics: Collinearity and Weak Data in Regression," Wiley, New York, 1991.
- [15] D. E. Farrar, R. R. Glauber, "Multi-collinearity in regression analysis: the problem revisited," *Rev. Econ. Statist.*, 49, 92-107, 1967.
- [16] A. Alin, "Multicollinearity," *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(3), 370-374. doi:10.1002/wics.84, 2010.
- [17] E. R. Mansfield, B. P. Helms, "Detecting Multicollinearity", *The American Statistician*, 36(3a), 158-160. doi:10.1080/00031305.1982.10482818, 1982.
- [18] H. Abdi, L. J. Williams, "Principal components analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol.2, pp.433-450, 2010.
- [19] E. Alpaydin, "Introduction to Machine Learning," *The MIT Press*, Pages 105-130, ISBN:978-0-262-02818-9, 2014.
- [20] J. Bergstra, Y. Bengio, "Random search for hyperparameter optimization," *J. Mach. Learn. Res.* 13, null (3/1/2012), 281-305, 2012.