Günümüzde veriler üzerinde manuel yöntemler yerine makine öğrenme yöntemlerinin kullanımı artmaktadır. Bu artış uygulanan yöntemleri de iyileştirmektedir. Ayrıca endüstri alanında makine öğrenimi yöntemlerinin kullanım sıklığı artmıştır. Bu yöntemler aynı zamanda tahmini bakımda araç garanti verilerinde de kullanılır. Endüstride kullanılan garanti verileri şirketler için maliyet açısından önemlidir. Araçtaki bir parça arızaya neden olabilir ve arıza garanti süresi boyunca devam eden bir hataysa, bu maliyeti artıracaktır.

Bu çalışmada, otomotiv endüstrisinde kullanılan garanti verileri otomatik olarak etiketlenmiş ve etiketli kutu üzerinde anomali tespit süreci gerçekleştirilmiştir. 32360 araç verilerini garanti maliyetleri ile birlikte kullandık. Ayrıca, etiketleme için gama dağılımı ve hareketli ortalama yöntemleri kullanılmaktadır. Etiketleme durumları Kararlı, Sayı\_Yeni, Sayı\_Durum, Sayı\_Decrease ve Issue\_Increase olarak verilir. Bu vakalar anomali davranışını gösterir. Örneğin, Kararlı; gösterildiği gibi anomali no, Sayı\_Yeni; yeni bir anomali durumunu gösterir. Kararlı ve Sorun\_Yeni etiketler gama dağılımları kullanılarak belirlenir. Diğer etiketler hareketli ortalama ile belirlenir. Makine öğrenimi algoritmaları, etiketli veriler üzerinde anomali durum tespiti için kullanılır. İki farklı algoritma kullanıyoruz. Birincisi, karar ağacı ve degrade artırmaya dayalı bir xgboost algoritmasıdır. İkincisi, regresyona dayanan bir lojistik regresyon algoritmasıdır. Doğruluk bir değerlendirme kriteri olarak kullanılır ve sonuçlar karşılaştırılır. Lojistik regresyon algoritmasından daha iyi performans gösteren anomali tespit sürecinde xgboost algoritmasını gözlemledik.

Günümüzde toplumların en büyük ihtiyaçlarından birinin enerji olduğu bilinmektedir. Dünya genelinde sanayileşme ile beraber artan enerji tüketimi, karbon salınımı ve hava kirliliğinin de artmasına sebep olmaktadır. Bu nedenle, bu alanda yapılan çalışmaların; enerji tüketiminin planlanması, yönetilmesi, korunması ve sürdürülebilmesi için oldukça büyük önem arz ettiği bilinmektedir. Bunların yanı sıra arz ve talep yönetiminin verimli bir şekilde gerçekleştirilmesini sağlayacağı öngörülmektedir.

Bu çalışmada, 2013 ve 2018 yılları arasındaki doğalgaz tüketim verileri eğitim veri kümesi olarak belirlenmiş ve bu veri kümesi kullanılarak farklı makine öğrenmesi modelleri eğitilmiştir. Ardından 2019 yılı doğalgaz tüketimi için günlük bazda bir tahminleme yapılmıştır. Modelleme teknikleri ve gerekliliklerinden önce gaz tüketiminde dikkat edilmesi gereken iklim ve bina metrekaresi gibi farklı faktörler bulunmaktadır. Bu çalışmanın amacı, iklim özellikleri, bina metrekaresi, dini ve milli tatiller gibi parametreler ele alınarak bir senelik gaz tüketimini günlük bazda tahmin edebilmek ve seçilen özelliklerin yapılan tahmin üzerindeki etkisini belirleyebilmektir. İlk olarak özellik çıkarım adımı gerçekleştirilmiş; iklim, dini ve milli tatiller ile ortalama gaz tüketimi arasında yüksek bir korelasyon değeri bulunmuştur. Gaz tüketimi ile en yüksek korelasyona sahip olan özelliğin, iklim özelliği olduğu ve doğru tahmin edilen iklim verileri alındığında modellerin yüksek doğruluk oranına sahip olduğu gözlemlenmiştir. Özellik çıkarma işleminin ardından 2013 ve 2018 yılları arasındaki doğalgaz tüketim verileri kullanılarak üç farklı model eğitilmiş ve 2019 yılı için bir doğalgaz tüketim tahmininde bulunulmuştur. Modellerin doğruluk oranı test edilirken değerlendirme ölçütü olarak R2 puanlaması kullanılmıştır. İlk olarak Rastgele Orman modeli seçilmiş ve bütün veri kümesi kullanılarak tahmin yapılmıştır. Modelin yüksek doğruluk oranına sahip olduğu gözlemlense de özellik çıkarımının ardından beklenen başarıyı gerçekleştirememiştir. Ardından Xgboost modeli tercih edilmiş ve sonuç olarak rastgele orman modeline yakın sonuçlara ulaşılmıştır. Son olarak 10 farklı makine öğrenmesi modelinin en iyi sonuçlarını birleştiren Superlearner modeli kullanılmış ve en iyi doğruluk oranı bu model ile elde edilmiştir.