

**T.C.**

**FENERBAHÇE ÜNİVERSİTESİ**

**MÜHENDİSLİK VE MİMARLIK FAKÜLTESİ**

**BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ**

**MAKİNE ÖĞRENMESİ MAKALE RAPORU**

**Hazırlayan**

**Ogün Berat Gürses**

**190301005**

**Makine öğrenimi tekniklerini kullanarak önemli özelliklerin seçimi ve şarap kalitesinin tahmin edilmesi**

Seçmiş olduğum makale ürün kalitesi için doğrusal regresyon, Neural Network ve SVM (Support Vector Machine) gibi makine öğrenimi tekniklerinin kullanımını iki şekilde incelemektedir. Makalede ilk olarak hedef değişkenin bağımsız değişkenlere göre regresyon analizi yapılarak bağımlılığı belirlenmiştir. İkinci hedef olarak bu değişkenlerin regresyonuna göre kırmızı şarap için 7 beyaz şarap için 8 en önemli sekiz adet değişken belirlenmiş olup hem toplam değişkenlere göre hem de belirlenen değişkenlere göre Neural Network ve SVM gibi makine öğrenmesi metotları ile bu veri seti üzerinde tahmin ettirme işlemleri yapılmıştır. Bu makalede hedef değişkenin bağımsız değişkenlere bağımlılığını belirlemek için doğrusal regresyon kullanılmıştır. Tüm deneyler, Kırmızı Şarap ve Beyaz Şarap veri kümeleri üzerinde gerçekleştirilmiştir. Bu makale, tüm özellikleri dikkate almak yerine seçilen featureların (özelliklerin) dikkate alınması durumunda daha iyi tahmin yapılabileceğini kanıtlamaktadır.

Bu makalede makine öğrenmesi tekniklerinin kullanılmasının en önemli sebebi Neural Network ve Support Vector Machine (SVM) gibi farklı makine öğrenimi tekniklerini araştırıp bu teknikler ile ürünün mevcut özellikleri yardımıyla kalite güvence sürecini gerçekleştirerek insan müdahalesini en aza indirip süreci otomatikleştirmektir. Çalışma ayrıca bağımlı değişkenlerin değerlerini tahmin etmek için önemli özellikleri de tanımlamaktadır.

Şarap, yoğunluk, pH değeri, alkol ve diğer asitler gibi çeşitli özelliklere sahiptir. Şarap kalitesi iki tür testle değerlendirilebilir; birincisi fizikokimyasal test, ikincisi ise duyusal testtir. Fizikokimyasal test laboratuvar testleri ile belirlenebilir ve insan uzmana gerek yoktur, ancak duyusal test için insan uzman gerekir. Ayrıca, fizikokimyasal ve duyusal analiz arasındaki ilişkiler karmaşık olduğundan ve hala tam olarak anlaşılmadığından, Şarap kalitesi değerlendirmesi çok zordur. Bu makalede, şarap kalitesinin farklı fizikokimyasal özelliklere bağımlılığını belirlemek için doğrusal regresyon, NN ve SVM uygulanmaktadır. Ayrıca, bağımlılıklarına göre seçilen önemli özellikler temelinde şarap kalitesi için tahminler de yapılmıştır [1].

Bu makale 4898 adet Beyaz Şarap verisi, 1599 adet de Kırmızı Şarap verisi ile işlemler yapılmıştır. Her iki şarap türünün her numunesi 12 fizyokimyasal değişkenden oluşmuştur: fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol, and quality rating. Kalite derecelendirilmesi ise 0-çok kötü 11-mükemmel olacak şekilde duyusal testler ile derecelendirilmiştir.

**Deneysel sonuçlar ve analizler:**

Önceki sayfada söylediğim üzere veri setinde 12 adet değişken bulunmaktadır. Bu değişkenlerden quality rating bağımlı değişken olarak kabul edilir ve diğer 11 adet değişken bağımsız değişken olarak kullanılmıştır. Analizlerde öncelikle regresyon ile wine quality değişkenine hangi değişkenlerin daha çok etki ettiği bulunmuştur. Ardından tahmim ediciler ile şarap kalitesinin değeri tahmin edilir.

Kırmızı Şarap için regresyon analizi yapıldığında wine quality değişkenine en çok etki eden değişkenler volatile acidity, chlorides, free sulfur dioxide, total sulfur dioxide, sulphates, alcohol olmak üzere 7 değişken şeklinde çıkmıştır.

Beyaz Şarap için regresyon analizi yapıldığında ise fixed acidity, volatile acidity, residual sugar, free sulfur dioxide, density, pH, sulphates, alcohol olmak üzere 8 adet değişken çıkmıştır.

Bu çıkan değişkenler doğrultusunda kırmızı şarap üzerine öncelikle tüm değişkenler için (11-5-1) öğrenme sonucu kalite tahminleri çıkartılmıştır. Ardından seçilen 7 adet değişken için kalite tahminlerin çıktıları değerlendirilmiştir. Ardından beyaz şarap için aynı iş yapılmıştır.

Neural Network için çıktılar tamamlandıktan sonra aynı testler SVM üzerinde tekrarlanmıştır.

Bu testler sonucunda asıl bakılması gereken şey öncelikle tüm özellikler ile seçili özellikler arasındaki farkın görülmesidir. Beklenen durum seçili özelliklerdeki sonuçların genel özelliklerdeki sonuçlara nazaran gerçeğe daha yakın çıkmasıdır.

Neural Network ile Support Vector Machine için bu karşılaştırmalar yapıldığında Neural Network içerisinde birkaç test verisi haricinde beklenene nazaran daha az yaklaşan sonuçlar ortaya çıktığı gözlenmiştir. Suppor Vector Machine’de ise gerçek değerlere daha fazla yaklaşma gözlenmiştir.

Kısacası bu iki makine öğrenmesi tekniği karşılaştırıldığında bu veri seti ve seçilen özellikler nezdinde SVM’in tahmin edici olarak daha iyi bir sonuç verdiği görülmektedir.

Referanslar:

[1] Portuguese Wine- Vinho Verde. Comissa ̃o de Viticultura da Regia ̃o dos Vinhos Verdes (CVRVV), http://www.vinhoverde.pt, July 2008.

Kanyak:

Selection of important features and predicting wine quality using machine learning techniques, Yogesh Gupta, <https://www.sciencedirect.com/science/article/pii/S1877050917328053> December 2017.