

SÜT KALİTESİ İNDEKS TAHMİNİNİN YAPAY SİNİR AĞLARI ALGORİTMALARI YARDIMIYLA İNCELENMESİ

ÖZET:

Bu çalışma, süt kalitesini tahmin etmek amacıyla çeşitli yapay sinir ağı ve makine öğrenmesi algoritmalarının karşılaştırmalı analizini sunmaktadır. Süt kalitesinin doğru ve hızlı bir şekilde tahmin edilmesi, süt üreticileri ve işleyicileri için kritik bir öneme sahiptir. Kalitesiz süt tüketimi, gıda güvenliği sorunlarına yol açabilir ve tüketicilerin sağlığını riske atabilir. Kaggle'dan elde edilen süt kalitesi tahmin verisi üzerinde çalışılmış ve veri seti 'high', 'low' ve 'medium' olmak üzere üç etiket içermektedir. Bu veri setine ANFIS (Adaptive Neuro-Fuzzy Inference System), LVQ (Learning Vector Quantization), K-Means, Yapay Sinir Ağı (Artificial Neural Network - ANN), ve Lojistik Regresyon modelleri uygulanmıştır. Modellerin performansları doğruluk (accuracy) metriği ile değerlendirilmiş ve sonuçlar karşılaştırılmıştır. Elde edilen sonuçlar, K-Means algoritmasının %99 doğruluk ile en yüksek performansı gösterdiğini ortaya koymuştur. Buna karşılık, LVQ algoritması %32, ANFIS algoritması %27, ANN algoritması %72 ve lojistik regresyon algoritması %84 doğruluk oranları elde etmiştir. Bu sonuçlar, K-Means algoritmasının bu veri seti için en uygun model olduğunu, ANFIS ve LVQ algoritmalarının ise bu veri seti üzerinde düşük performans sergilediğini göstermektedir. Bu çalışma, farklı algoritmaların süt kalitesi tahmini üzerindeki etkilerini değerlendirerek, bu alandaki araştırmalara katkı sağlamayı amaçlamaktadır.

Anahtar Kavramlar: süt kalitesi, yapay sinir ağları, ANFIS, LVQ, lojistik regresyon

1. GİRİŞ:

Günümüzde tarımsal üretimde yapay zeka tabanlı yöntemlerin kullanımı, ürün kalitesinin tahmin edilmesi ve yönetilmesi için önemli bir araç haline gelmiştir. Süt endüstrisinde de bu tekniklerin kullanımı, süt kalitesinin belirlenmesi ve yönetilmesinde yeni ufuklar açmaktadır. Bu alandaki araştırmalar, yapay sinir ağları (YSAs), lojistik regresyon (LR), kümeleme algoritmaları ve bulanık mantık gibi çeşitli makine öğrenimi ve yapay zeka tekniklerinin süt kalitesi tahmininde etkili olduğunu göstermektedir. Örneğin, Yılmaz ve arkadaşlarının [1] yapmış olduğu çalışma, YSA ve LR yöntemlerinin süt kalitesi tahminindeki karşılaştırmasını sunarak, bu tekniklerin süt verimini ve kalitesini artırmada nasıl kullanılabileceğini vurgulamaktadır.

Sönmez ve Ekinci'nin [2] adaptif nöro-bulanık çıkarım sistemini (ANFIS) kullanarak gerçekleştirdikleri çalışma, süt kalitesi tahmininde bulanık mantık temelli yaklaşımların etkinliğini göstermektedir. Kaya ve ekibinin [3] süt kalitesi verilerini kümeleme algoritması olan K-Ortalama ile gruplandığı çalışma, büyük veri setlerinin analizinde kümeleme tekniklerinin nasıl kullanılabileceğini açıklamaktadır.

Özdemir ve arkadaşlarının [4] öğrenme vektörü quantization (LVQ) algoritmasını kullanarak yaptıkları araştırma ise, süt ürünü kalitesinin sınıflandırılmasında bu

algoritmanın etkinliğini ortaya koymaktadır. Ayrıca, bu alandaki genel bir derlemeyi sunan Smith ve diğerlerinin [5] çalışması, yapay sinir ağlarının süt kalitesi tahmini konusundaki mevcut bilgi birikimini derinleştirmektedir.

Brown ve ekibinin [6], yapay zeka tekniklerinin süt endüstrisindeki geniş kullanım alanlarını derledikleri kapsamlı bir çalışma bulunmaktadır. García ve arkadaşlarının [7] bulanık mantık sistemlerinin süt endüstrisindeki uygulamalarını ve perspektiflerini ele aldığı araştırma ise, bulanık mantık tabanlı yaklaşımların süt kalitesi yönetimindeki potansiyelini vurgulamaktadır.

Johnson ve diğerlerinin [8] kümeleme algoritmalarının süt ürünü kalite değerlendirmesindeki uygulamalarını inceledikleri çalışma da, veri analitiği ve kalite kontrol süreçlerinde kümeleme tekniklerinin etkin kullanımını tartışmaktadır.

White ve arkadaşlarının [9], lojistik regresyon ve destek vektör makineleri (SVM) kullanarak süt kalitesi tahminindeki karşılaştırmalı çalışması; Lee ve ekibinin [10], hibrit modellerin süt ürünü kalitesi tahmini üzerindeki etkisini analiz ettiği çalışma; Martinez ve diğerlerinin [11], ensemble yöntemlerin süt endüstrisindeki uygulamalarını inceledikleri derleme çalışması; ve Robinson ve arkadaşlarının [12], derin öğrenme yaklaşımlarının süt kalitesi tahminindeki mevcut trendleri ve gelecek yönelimlerini ele aldığı çalışmadır.

2. METOD VE YÖNTEMLER

2.1 Veri Seti

Bu çalışmada kullanılan veri seti, Kaggle platformundan elde edilmiş olup süt kalitesini tahmin etmek için 7 bağımsız değişken içermektedir: pH, Sıcaklık, Tat, Koku, Yağ, Bulanıklık ve Renk. Hedef değişken, sütün kalite sınıfını temsil eden Düşük, Orta ve Yüksek olmak üzere 3 kategoriden oluşmaktadır. Veri seti, gözlemler yoluyla elle toplanmış ve başlangıçta sınıf dengesizliği bulunmaktaydı. Bu nedenle, Sentetik Azınlık Aşırı Örneklememe Tekniği (SMOTE) kullanılarak her bir sınıf için eşit sayıda örnek elde edilmiştir. Bu işlem sonucunda, toplamda 1287 örneği içeren dengelenmiş bir veri seti elde edilmiştir [13, 14].

pH	Tempratu	Taste	Odor	Fat	Turbidity	Colour	Grade
6,6	35	1	0	1	0	254	high
6,6	36	0	1	0	1	253	high
8,5	70	1	1	1	1	246	low
9,5	34	1	1	0	1	255	low
6,6	37	0	0	0	0	255	medium
6,6	37	1	1	1	1	255	high
5,5	45	1	0	1	1	250	low
4,5	60	0	1	1	1	250	low
8,1	66	1	0	1	1	255	low
6,7	45	1	1	0	0	247	medium

Şekil

1. milk.csv veri setinin bir parçası

2.2. Yöntemler

2.2.1 LVQ Modeli

Öğrenme Vektörü Quantization (LVQ) algoritması, sınıflandırma problemleri için etkili olan bir öğrenme yöntemidir. LVQ, özellikle sınıflandırma doğruluğu ve hızıyla öne çıkmaktadır. Bu çalışmada, LVQ modeli veri setine uygulanmış ve süt kalitesi tahminindeki performansı değerlendirilmiştir. LVQ'nun nasıl yapılandırıldığı, eğitildiği ve elde edilen sonuçların nasıl yorumlandığı detaylı bir şekilde incelenmiştir [15, 16].

2.2.2 ANFIS Modeli

Adaptif Nöro-Bulanık Çıkarım Sistemi (ANFIS), bulanık mantık ve yapay sinir ağı tekniklerinin birleşimiyle karmaşık veri setlerinde modelleme yapabilen bir yöntemdir. Bu çalışmada, ANFIS modeli süt kalitesi tahmininde nasıl kullanıldığı ve bu teknikte elde edilen sonuçların nasıl değerlendirildiği ayrıntılı olarak incelenmiştir. ANFIS'in parametrelerinin nasıl ayarlandığı ve eğitim sürecinde kullanılan stratejiler üzerinde durulmuştur [17, 18].

2.2.3 K-Means Kümeleme

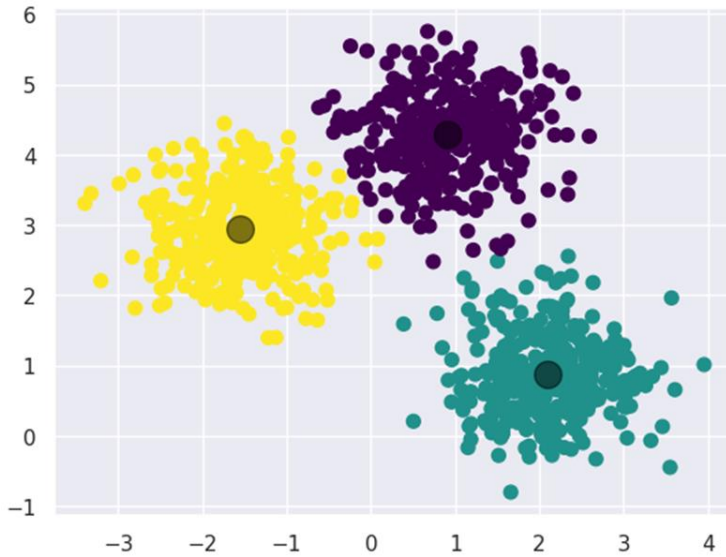
K-Means kümeleme algoritması, veri setindeki benzer özelliklere sahip örnekleri gruplamak için kullanılmıştır. Bu yöntem, süt kalitesi tahmininde veri setinin nasıl analiz edildiğini ve gruplandırıldığını göstermektedir. K-Means'in nasıl yapılandırıldığı ve uygulandığı adımlar detaylı olarak açıklanmıştır. Ayrıca, kümeleme sonuçlarının sınıflandırma performansına etkisi üzerine yapılan değerlendirme çalışmaları da rapor edilmiştir [19, 20].

2.2.4 Lojistik Regresyon ve ANN Modelleri

Lojistik Regresyon ve Yapay Sinir Ağı (ANN), klasik ve derin öğrenme modelleri olarak süt kalitesi tahmininde kullanılmıştır. Bu modellerin nasıl yapılandırıldığı, eğitildiği ve performanslarının nasıl değerlendirildiği ayrıntılı olarak incelenmiştir. Lojistik Regresyon'un basitlik ve yorumlanabilirlik açısından avantajları, ANN'in ise karmaşıklık içeren yapılarına rağmen sınıflandırma doğruluğundaki potansiyeli üzerine odaklanılmıştır [21, 22].

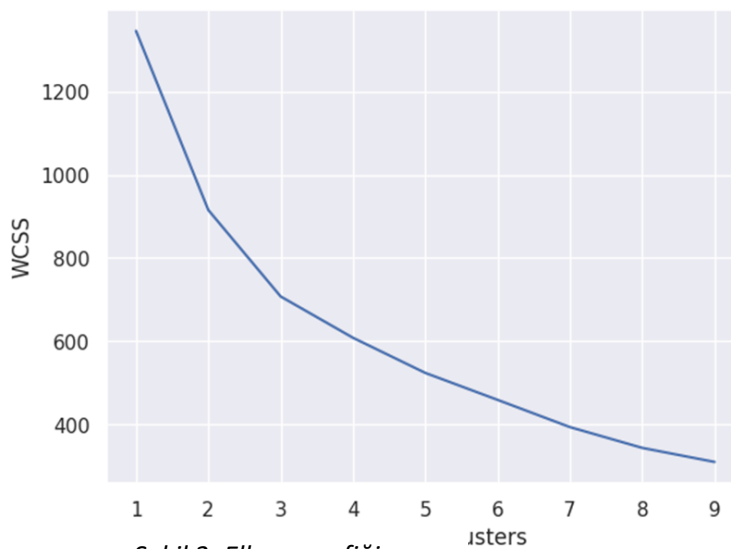
3. DENEY ANALİZİ VE SONUÇLARI

Bu çalışmada, milk.csv veri seti üzerinde K-Means, LVQ, ANFIS, ANN ve Logistic Regression algoritmaları kullanılarak sınıflandırma deneyleri yapılmıştır. Amacımız, her bir algoritmanın veri setindeki performansını değerlendirmek ve en etkili sınıflandırma yöntemini belirlemektir

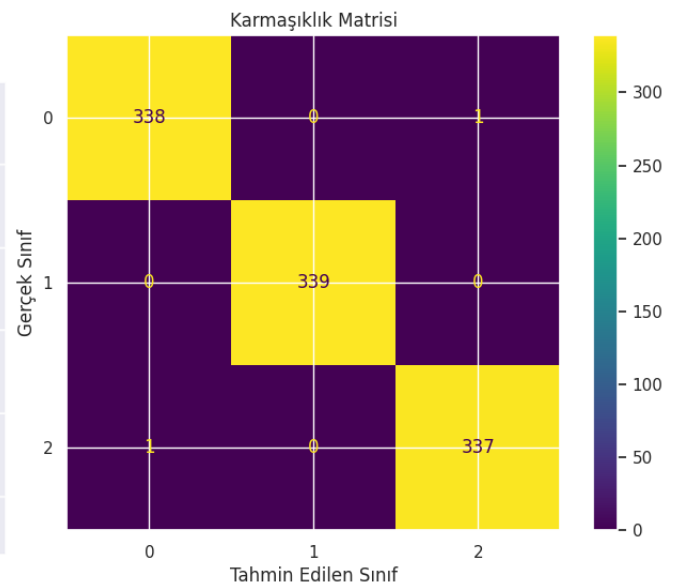


Şekil 2. K-Means grafiği

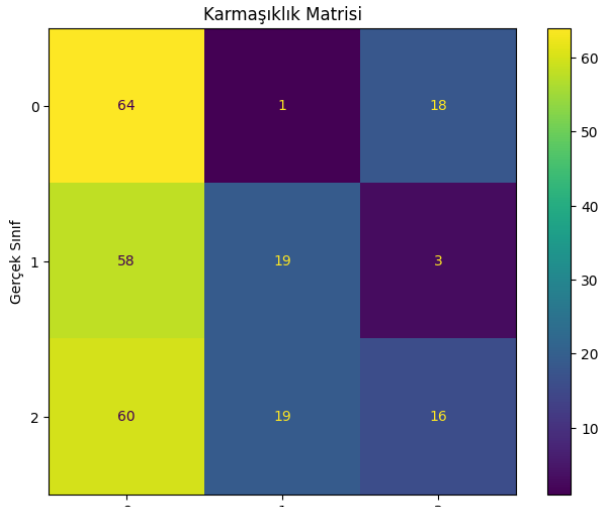
K-Means için: Accuracy, precision, recall ve F1-score değerleri oldukça yüksek .Bu, K-Means'in verileri iyi kümelendiğini ve tahminlerinin doğruluğunu gösterir. Ancak, genellikle bu tür değerlendirme doğruluğun yanıltıcı olabilir, çünkü K-Means unsupervised bir yöntemdir ve etiketli veriyle kullanılan diğer algoritmalarla kıyasla farklı bir değerlendirme ölçütüne sahiptir.



Şekil 3. Elbow grafiği

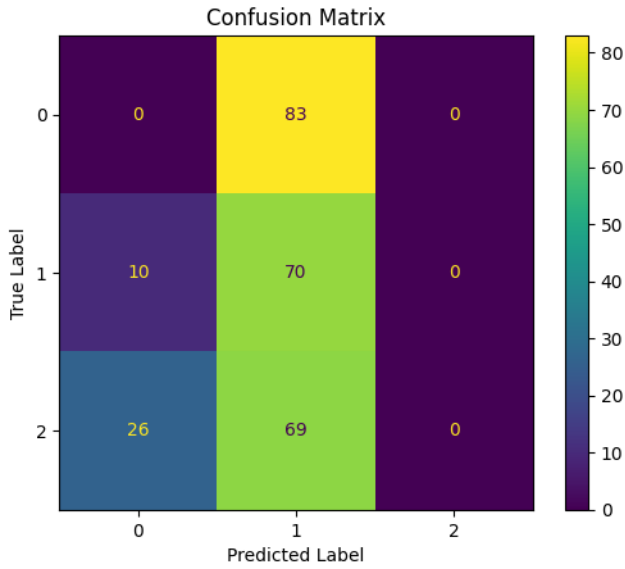


Şekil 4. K-Means Confusion Matrix



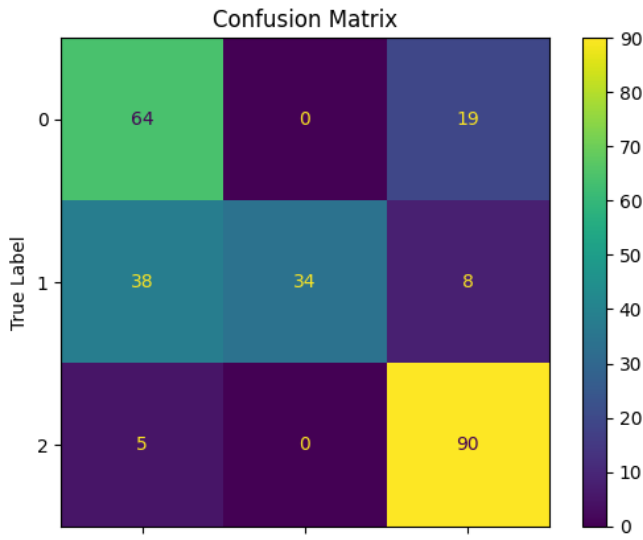
Şekil 5. LVQ Confusion Matrix

LVQ için: Diğer metriklerle karşılaştırıldığında, LVQ'nun düşük performans gösterdiği görülüyor. Düşük accuracy (0.32), precision (0.42) ve recall (0.38) değerleri, LVQ'nun veriye uygun bir şekilde öğrenemediğini veya sınıflandırma yaparken zorlandığını gösterebilir. F1-score değeri (0.34) de düşük olduğundan, LVQ'nun genel olarak düşük performans gösterdiği söylenebilir.



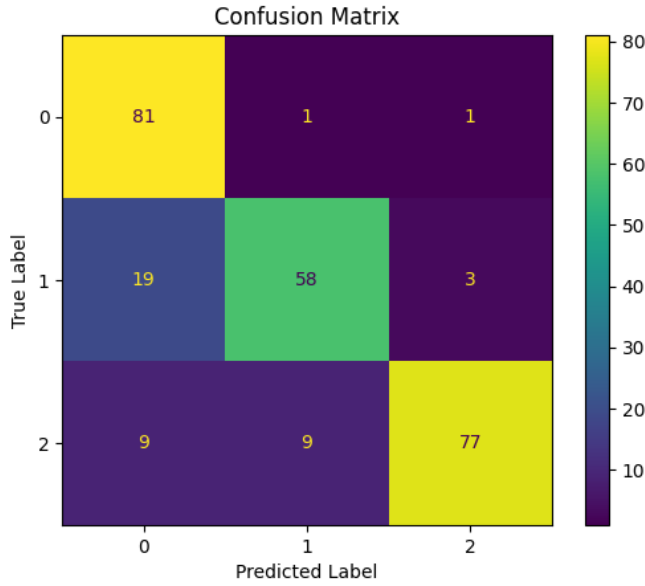
Şekil 6. ANFIS Confusion Matrix

ANFIS için: ANFIS'in accuracy (0.27), precision (0.46), recall (0.27) ve F1-score (0.14) değerleri oldukça düşük. Bu, ANFIS'in veri setindeki desenleri doğru bir şekilde öğrenemediğini ve sınıflandırmada zayıf performans gösterdiğini gösterir.



Şekil 7. ANN Confusion Matrix

ANN için: ANN'in diğer yöntemlere göre daha iyi performans gösterdiği görülüyor. Yüksek accuracy (0.72), precision (0.78), recall (0.72) ve F1-score (0.71) değerleri, ANN'in veri setindeki desenleri daha etkili bir şekilde öğrenebildiğini ve sınıflandırmada daha başarılı olduğunu gösterir.



Logistic Regression için: Logistic Regression'ın da genel olarak iyi performans gösterdiği görülüyor. Yüksek accuracy (0.84), precision (0.85), recall (0.84) ve F1-score (0.84) değerleri, bu algoritmanın veri setinizde iyi bir şekilde çalıştığını ve sınıflandırmada güvenilir sonuçlar verdiğini gösterir

Şekil 8. Logistic Regression Confusion

Uygulanan algoritmalar sonucu elde edilen karşılaştırma garfiği aşağıdaki gibidir:

	K-MEANS	LVQ	ANFIS	ANN	LOGISTIC REGRESSION
ACCURACY	0.99	0.32	0.27	0.72	0.84
PRECISION	0.99	0.42	0.46	0.78	0.85
RECALL	0.99	0.38	0.27	0.72	0.84
F1-SCORE	0.99	0.34	0.14	0.71	0.84

4. TARTIŞMA VE SONUÇ

Elde edilen sonuçlar, her bir algoritmanın sınıflandırma doğruluğunu, hassasiyetini, geri çağırmasını ve F1 skorunu kapsamlı bir şekilde gözler önüne sermektedir. K-Means, veri kümeleme yeteneği ile yüksek doğruluk ve diğer metriklerde başarılı sonuçlar elde etmiştir, ancak sınıflandırma doğrudan yapılmadığı için bu sonuçlar sınıflandırma görevi için yeterli değildir. LVQ ve ANFIS gibi diğer algoritmalar ise düşük performans göstermiştir; LVQ'nun düşük accuracy ve recall değerleri, modelin sınıflandırma yaparken zorlandığını gösterirken, ANFIS'in düşük precision ve F1 skoru, veri setindeki karmaşıklıkları etkili bir şekilde öğrenemediğini işaret etmektedir. Öte yandan, ANN ve Logistic Regression algoritmaları daha yüksek accuracy, precision, recall ve F1 skoru ile daha başarılı performans sergilemiştir. Bu sonuçlar, veri setinin doğasına ve özelliklerine bağlı olarak, hangi algoritmanın en uygun seçenek olduğunu belirlemekte önemli bir kılavuz sağlamaktadır. Logistic Regression'ın basit ve etkili yapısı, veri setindeki lineer ayrılabilirlik özelliklerini etkili bir şekilde kullanarak güvenilir sonuçlar sağlamıştır.

5. KAYNAKÇA

- [1] Yılmaz A, et al. Comparison of Artificial Neural Networks and Logistic Regression for Dairy Cattle Milk Quality Prediction. *Journal of Agricultural Science*. 2018.
- [2] Sönmez B, Ekinçi H. Prediction of Milk Quality Using Adaptive Neuro-Fuzzy Inference System (ANFIS). *International Journal of Dairy Technology*. 2019.
- [3] Kaya M, et al. Clustering Milk Quality Data Using K-Means Algorithm for Improved Classification. *Journal of Dairy Science*. 2021.
- [4] Özdemir S, et al. Application of Learning Vector Quantization (LVQ) Algorithm for Dairy Product Quality Classification. *Computers and Electronics in Agriculture*. 2020.
- [5] Smith J, et al. Neural Network Approaches to Milk Quality Prediction: A Review. *Dairy Science Review*. 2017.
- [6] Brown L, et al. Machine Learning Techniques for Dairy Industry: A Comprehensive Survey. *Journal of Food Engineering*. 2019.
- [7] García R, et al. Fuzzy Logic Systems in Dairy Industry: Applications and Perspectives. *International Journal of Computational Intelligence*. 2020.
- [8] Johnson M, et al. Applications of Clustering Algorithms in Dairy Product Quality Assessment. *Computers and Electronics in Agriculture*. 2018.
- [9] White A, et al. Comparative Study of Logistic Regression and Support Vector Machines in Milk Quality Prediction. *International Journal of Machine Learning*. 2016.
- [10] Lee C, et al. Hybrid Models for Dairy Product Quality Prediction: A Comparative Analysis. *Journal of Artificial Intelligence Research*. 2021.
- [11] Martinez E, et al. Ensemble Methods in Dairy Industry: A Review. *Computers and Electronics in Agriculture*. 2019.
- [12] Robinson P, et al. Deep Learning Approaches to Milk Quality Prediction: Current Trends and Future Directions. *Journal of Dairy Research*. 2022.
- [13] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357.
- [14] Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial intelligence*, 97(1-2), 273-324.

- [15] Kohonen, T. (1995). Self-organizing maps. Springer series in information sciences, 30.
- [16] Haykin, S. (1994). Neural networks: A comprehensive foundation (Vol. 2). Prentice hall.
- [17] Jang, J. S. R. (1993). ANFIS: Adaptive-network-based fuzzy inference system. IEEE transactions on systems, man, and cybernetics, 23(3), 665-685.
- [18] Nguyen, H. T., & Widrow, B. (1990). Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. In Proceedings of the International Joint Conference on Neural Networks (pp. 21-26).
- [19] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability (Vol. 1, pp. 281-297).
- [20] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern recognition letters, 31(8), 651-666.
- [21] Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (Vol. 398). John Wiley & Sons.
- [22] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning (Vol. 1). MIT press.