

Naive Bayes Yöntemi ile Pnömoni Teşhisi

Seda POSTALCIOĞLU^{1*}, Alper KEŞLİ²

^{1,2}Bolu Abant İzzet Baysal Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, Bolu, Türkiye

Özet

Tıbbi teşhis, görüntü işlemenin etkin bir şekilde uygulandığı önemli prosedürlerden biridir. Pnömoni en yaygın hastalıklar arasındadır ve uzman eksikliği nedeniyle tespit edilmesi zordur. Çalışmada, Naive Bayes yöntemleri kullanılarak radyografi görüntüleri ile pnömoni teşhisi yapılmıştır. Sınıflandırmada Gaussian Naive Bayes (GNB), Multinomial Naive Bayes (MNB), Bernoulli Naive Bayes (BNB) kullanılarak performansları analiz edilmiştir. Kullanılan veri seti dengeli ve dengesiz olarak ele alınmıştır. Yapılan çalışma sonucu %81 oranında GNB'nin diğer yöntemlere göre daha iyi performans gösterdiği görülmüştür.

Anahtar Kelimeler: Gaussian Naive Bayes, Multinomial Naive Bayes, Bernoulli Naive Bayes, Pnömoni

Diagnosis of Pneumonia by Naive Bayes Method

Abstract

Medical diagnostics is one of the important procedures in which image processing is effectively implemented. Pneumonia is among the most common diseases and is difficult to detect due to lack of specialists. In the study, pneumonia was diagnosed by using radiographic images using Naive Bayes methods. In classification, their performances were analyzed using Gaussian Naive Bayes (GNB), Multinomial Naive Bayes (MNB) and Bernoulli Naive Bayes (BNB). The dataset is used as balanced and unbalanced. As a result of the study, it is seen that 81% of GNB performs better than other methods.

Keywords: Gaussian Naive Bayes, Multinomial Naive Bayes, Bernoulli Naive Bayes, Pneumonia

1 Giriş

Görüntü teşhisi tıpta önemli bir sorundur. Günümüzde, doktorların erken ve doğru bir şekilde hastalığı teşhis etmesini sağlayan, gereksiz tedavi prosedürlerinin ortadan kaldırılmasını sağlayan tıbbi tanı, görüntü işlemenin yararlı bir şekilde uygulandığı en önemli yöntemdir [1,2]. Artan hasta sayısı ile beraber doktorların fazla iş yükü nedeniyle teşhis ve tedavi gecikebilir. Teşhisin hızlı bir şekilde ortaya konulması ve tedaviye başlanması oldukça önemlidir. Çalışma, Naive Bayes yöntemine dayalı pnömoni tespitini içermektedir. Pnömoni hastalığı ağır seyreden ciddi bir hastalık olmakla birlikte ülkemizde ve dünyada görülme sıklığı da oldukça fazladır. Pandemi ile mücadele ettiğimiz şu günlerde yeni tip koronavirüsün enfekte olmuş kişilerde pnömoni hastalığına yol açabildiği ayrıca bilinmektedir.

Bu nedenle hastalığın teşhis safhasında kullanılabilecek bir teknoloji üretmek oldukça önemlidir. Ölümle sonuçlanan vakalara bakıldığında Tablo 1'de gösterildiği gibi pnömoni dünyada dördüncü, Türkiye'de dokuzuncu sıradadır [3,4]. Pnömoni(zatürre) akciğer iltihaplanması olarak bilinir. Özellikle 5 yaş altı çocuklarda görülen en ölümcül hastalık türüdür. Çocuk ölüm oranlarının %16'sını oluşturmaktadır [5].

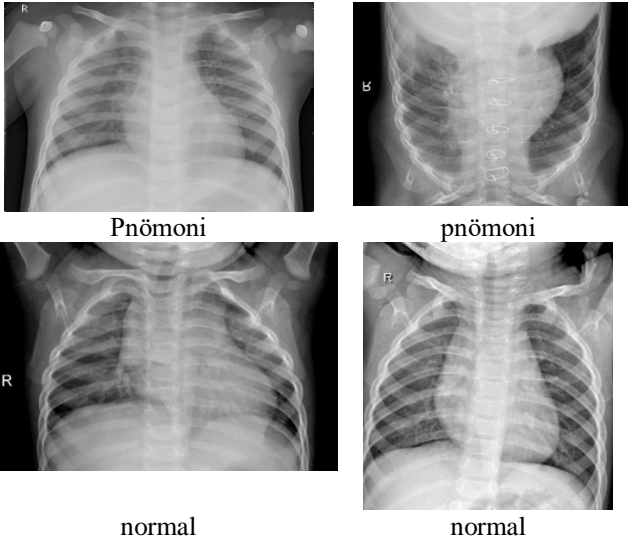
Radyografi görüntüleri, pnömoni teşhisi için kullanılan en yaygın yöntemdir. Radyografi görüntüleri ile hastalığın tespiti uzmanlar için de oldukça zordur [5].

Çalışmada Naive Bayes yöntemi kullanılmıştır. Radyografi görüntüleri Kaggle sitesinden elde edilmiştir [6]. Şekil 1'de bu veri setine ait örnek radyografi görüntüleri verilmiştir.

* İletişim e-posta: postalcioglu_s@ibu.edu.tr

Tablo 1: Ölümle sonuçlanan vakalar [3,4]

Dünya		Türkiye	
Sıra	Ölüm Nedeni	Sıra	Ölüm Nedeni
1	Kalp ve damar hastalıkları	1	kalp hastalığı
2	Kanser	2	Felç
3	Solunum yolu hastalıkları	3	Akciğer kanseri
4	Alt solunum yolu enfeksiyonu (Pnömoni)	4	KOAH
5	Bunama	5	Alzheimer hastalığı
6	Sindirim sistemi hastalıkları	6	Kronik böbrek hastalığı
7	Yenidoğan ölümleri	7	Diyabet
8	İshalli hastalıklar	8	Hipertansif kalp hastalığı
9	Diyabet	9	Alt solunum yolu enfeksiyonu (Pnömoni)
10	Karaciğer hastalıkları	10	Trafik kazaları



Şekil 1. Veri setine ait örnekler

2 Naive Bayes Yöntemi

Naive Bayes ile sınıflandırmada öğretilmiş veri sunulur. Öğretilmiş verilerden elde edilen olasılık değerleri ile sınıflandırma için verilen test verileri, daha önceden hesaplanmış olasılık değerlerine göre analiz edilir ve verilen test verisinin hangi sınıfa ait olduğu tespit edilir. Veri sayısı ne kadar çok ise, performansın artması o kadar mümkündür. Naive Bayes sınıflandırması, Bayes'in olasılık teoremini kullanan olasılıklı bir yaklaşıma dayanan makine öğreniminin denetimli bir sınıflandırmasıdır. Verilerin hesaplanan olasılık ile hangi sınıfa ait olduğunu tespit eder. Naive Bayes algoritması özellik oluşumlarının birbirinden bağımsız olduğu varsayımını yapar [7]. Naive Bayes sınıflandırıcı da iki temel kabul vardır. Bunlar; niteliklerin hepsi aynı derecede önemli olduğu ve niteliklerin birbirinden bağımsız olduğu varsayıdır. Bu yöntem koşullu olasılıklar arasında bağıntı kurar. Bir C sınıfının bilinmesi koşulunda x özneliliğinin var olma olasılığı ile x özneliliğinin bilinmesi koşulunda C sınıfının var olma olasılığı arasında Bayes teoremine göre eşitlik1'deki bağıntı vardır [8].

$$P(C|x) = \frac{P(C)P(x|C)}{P(x)} \quad (1)$$

Gaussian Naive Bayes, sürekli veriler için tercih edilmektedir. Her sınıfla ilişkili sürekli değerler Gauss dağılımına göre dağıtılır [9]. Eğitim verileri sınıfa göre bölümlere ayrılır ve her sınıfın ortalaması ve varyansı hesaplanır[9].

Doğruluk değeri, modelde doğru tahmin ettiğimiz alanların toplam veri kümesine oranı ile hesaplanmaktadır. DP, DN, YP, YN sırası ile doğru pozitif, doğru negatif, yanlış pozitif, yanlış negatif ifade etmektedir. Performans değerlendirmesinde kullanılan ifadelerin eşitliği tablo 2'de verilmiştir [10].

Tablo 2. Performans hesapları

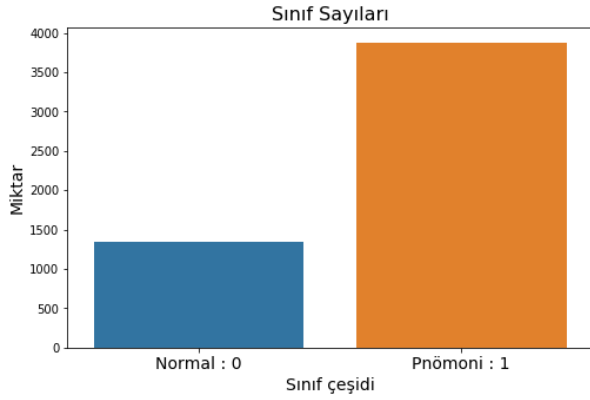
Doğruluk $\frac{DP + DN}{DP + YP + DN + YN}$	Kesinlik $\frac{DP}{DP + YP}$	Duyarlılık $\frac{DP}{DP + YN}$	F1-Skor $2 * \frac{Kesinlik * Duyarlılık}{Kesinlik + Duyarlılık}$
---	----------------------------------	------------------------------------	--

Kesinlik, pozitif olarak tahmin edilen değerlerin gerçekten kaç adedinin pozitif olduğunu ifade eder. Duyarlılık ise pozitif olarak tahmin edilenlerin, işlemlerin ne kadarını pozitif olarak tahmin ettiğimizi gösteren bir metriktir. F1 Skor değeri bize Kesinlik ve Duyarlılık değerlerinin harmonik ortalamasını göstermektedir.

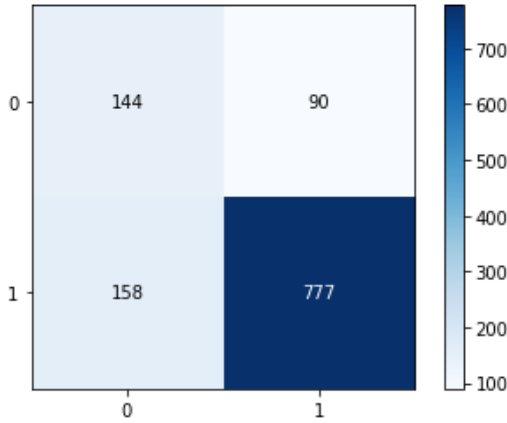
Çalışmada Naive Bayes yöntemi olarak Gaussian, Multinomial, Bernoulli teknikleri kullanılmıştır.

Veri dağılımı incelendiğinde şekil 2'deki durum görülmektedir. Dağılımın şekilden de görüldüğü gibi dengesiz olduğu görülmektedir. İlk olarak Gaussian Naive Bayes (GNB) yöntemi uygulanmıştır. GNB yönteminin hata matrisi şekil

3'de verilmiştir. Yöntemin doğruluk performansı %79 olarak hesaplanmıştır. Performans sonuçları tablo 3'de sunulmuştur.



Şekil 2. Veri dağılımı

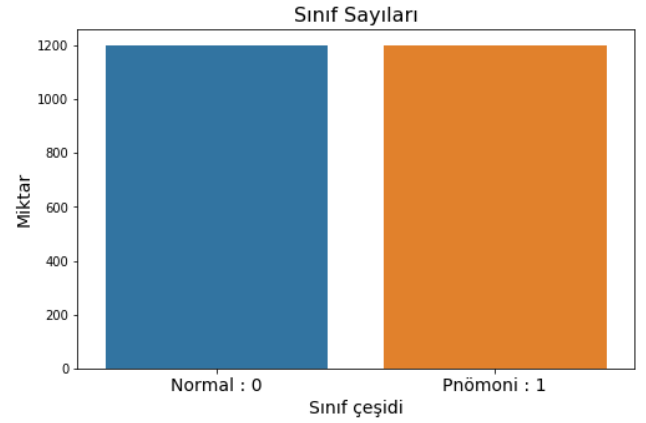


Şekil 3. Hata matrisi

Tablo 3. Performans sonuçları

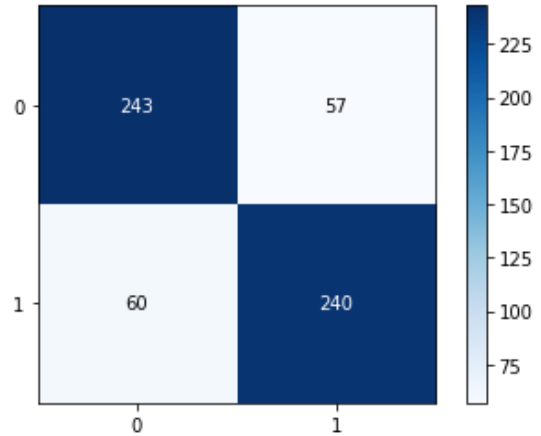
	Kesinlik	Duyarlılık	F1 Skor
Normal	%48	%62	%54
Pnömoni	%90	%83	%86

Verilerin düzgün dağılımı gerçekleştirilip, GNB yöntemi tekrar test edilmiştir. Şekil 4'de düzgün dağılımlı veri seti gösterilmektedir. Veri setinin 2400'ü eğitim ve 600'ü test verisi olacak biçimde %80'e %20 oranında ayarlanmıştır.



Şekil 4. Düzgün dağılımlı veri seti

Hata matrisi ve performans değerleri sırası ile şekil 5 ve tablo 4'de verilmiştir. Doğruluk değeri %81 olarak elde edilmiştir. Veri setinin düzgün dağılımı olumlu yönde etki etmiştir.

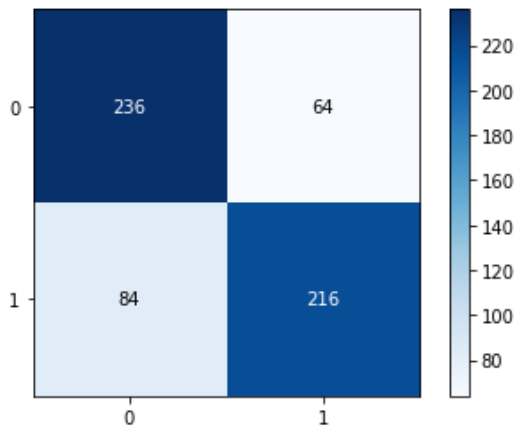


Şekil 5. GNB için hata matrisi

Tablo 4. GNB için performans sonuçları

	Kesinlik	Duyarlılık	F1 Skor
Normal	%80	%81	%81
Pnömoni	%81	%80	%80

GNB yönteminin performansı düzgün dağılımlı veri setinde artmış olduğundan Multinomial NB (MNB) ve Bernoulli NB (BNB) için de aynı veri seti ile çalışılmıştır. MNB ile elde edilen hata matrisi ve performans değerleri şekil 6 ve tablo 5 de verilmiştir. Doğruluk performans değeri %75 olarak hesaplanmıştır.

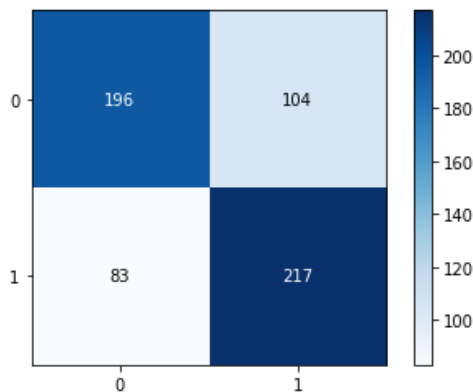


Şekil 6. MNB için Hata matrisi

Tablo 5. MNB için performans sonuçları

	Kesinlik	Duyarlılık	F1 Skor
Normal	%74	%79	%76
Pnömoni	%77	%72	%74

Son olarak Bernoulli NB kullanılmıştır. Hata matrisi ve performans değerleri şekil 7 ve tablo 6 da verilmiştir. Doğruluk performans değeri %69 olarak bulunmuştur.



Şekil 7. BNB için hata matrisi

Tablo 6. BNB için performans sonuçları

	Kesinlik	Duyarlılık	F1 Skor
Normal	%70	%65	%68
Pnömoni	%68	%72	%70

Doğruluk performansları açısından sonuçlar tablo 7'de verilmiştir. Sonuçlardan da görüldüğü gibi Gaussian NB yöntemi daha başarılı bir performans göstermiştir.

Tablo 7. Doğruluk performansları açısından sonuçlar

	Gaussian NB	Multinomial NB	Bernoulli NB
Doğruluk	%81	%75	%69

3 Sonuçlar

Çalışmada, Naive Bayes yöntemleri kullanılarak radyografi görüntüleri, pnömoni açısından sınıflandırılmıştır. Pnömoni en yaygın hastalıklar arasındadır ve uzman eksikliği nedeniyle tespit edilmesi zordur. Bu çalışma ile Gaussian NB, Multinomial NB ve Bernoulli NB kullanılarak sınıflandırma performansları karşılaştırılmıştır. Yapılan çalışma sonucu Gaussian NB %81 oranında performans göstermiştir.

Kaynaklar

- [1] P. N. Kieu, H. S. Tran, T. H. Le, T. Le and T. T. Nguyen, "Applying Multi-CNNs model for detecting abnormal problem on chest x-ray images," 2018 10th International Conference on Knowledge and Systems Engineering (KSE), Ho Chi Minh City, 2018, pp. 300-305, doi: 10.1109/KSE.2018.8573404
- [2] A. A. Abdullah, Norafifah Md Posdzi and Y. Nishio, "Preliminary study of pneumonia symptoms detection method using Cellular Neural Network," International Conference on Electrical, Control and Computer Engineering 2011 (InECCE), Pahang, 2011, pp. 497-500, doi: 10.1109/INECCE.2011.5953933.
- [3] <http://www.healthdata.org/turkey>, erişim tarihi : 14.06.2020
- [4] <https://ourworldindata.org/causes-of-death>, erişim tarihi : 14.06.2020
- [5] Talo, Muhammed. (2019). Konvolüsyonel Sinir Ağları ile Radyografi Görüntülerinden Pnömoni Tespiti. 1-4. 10.1109/SIU.2019.8806614.
- [6] <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>, 14.06.2020
- [7] H. T. Zaw, N. Maneerat and K. Y. Win, "Brain tumor detection based on Naïve Bayes Classification," 2019 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST), Luang Prabang, Laos, 2019, pp. 1-4, doi: 10.1109/ICEAST.2019.8802562.
- [8] Solmaz, R., Günay, M., Alkan A., 2014. Fonksiyonel Tiroit hastalığı tanısında Naive Bayes Sınıflandırıcının kullanılması, Akademik Bilişim Konferansı, Mersin, 891-897.
- [9] Gayathri, B.M., & Sumathi, C.P. (2016). An Automated Technique using Gaussian Naïve Bayes Classifier to Classify Breast Cancer. International Journal of Computer Applications, 148, 16-21.
- [10] Powers, David & Ailab,. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. J. Mach. Learn. Technol. 2. 2229-3981. 10.9735/2229-3981.