Yapay Sinir Ağı (YSA) ve 10 Katlı Çapraz Doğrulama Kullanılarak Farklı Veri Tabanlarında Sınıflandırma

CHALLENGE 1

KIELEH NGONG IVOLINE-CLARISSE
178229001010

Veri Setleri

- **UCI Iris - 150 örnek, 4 Özellik, 3 Sınıf**
- **UCI Wisconsin Breast Cancer – 569 Örnek, 30 Özellik, 2 Sınıf**
- **UCI BUPA Liver Disease – 345 Örnek, 6 Özellik, 2 Sınıf**

Programlama dili

Python

Python Packetleri

- **Keras**
- **Numpy**
- **Panda**
- **MatplotLib**

Adımlar

- Ön işleme
- Yapay Sinir Ağı
- Sonuçlar

Ön işleme

- Boş değerleri kontrol edin ve boş değerlere sahip kolonları kaldırın
- Daha sonra herhangi bir kategorik değişken için kontrol ettik ve bunları sayısal değişkenlere dönüştürdük.
- Eksik değerler için kontrol ettiniz.
- Son olarak, ölçekleme özelliği.

Bu, veri kümesindeki değişkenlerin önemini eşitlemek için yapıldı.

Yapay Sinir Ağı

Algoritma

- Feedforward & backpropagation

- Gizli Katman & Transfer Fonksiyonu
  - Iris → 1 katman & relu- softmax
  - Breast Cancer →2 katman & relu- sigmoid
  - Liver Disease → 2 katman & relu- sigmoid

Yapay Sinir Ağı

10-Kat Çapraz Doğrulamalı YSA

Performing k-fold cross validation

In [49]:
```
kfold = KFold(n_splits=10, shuffle=True, random_state=seed)
```

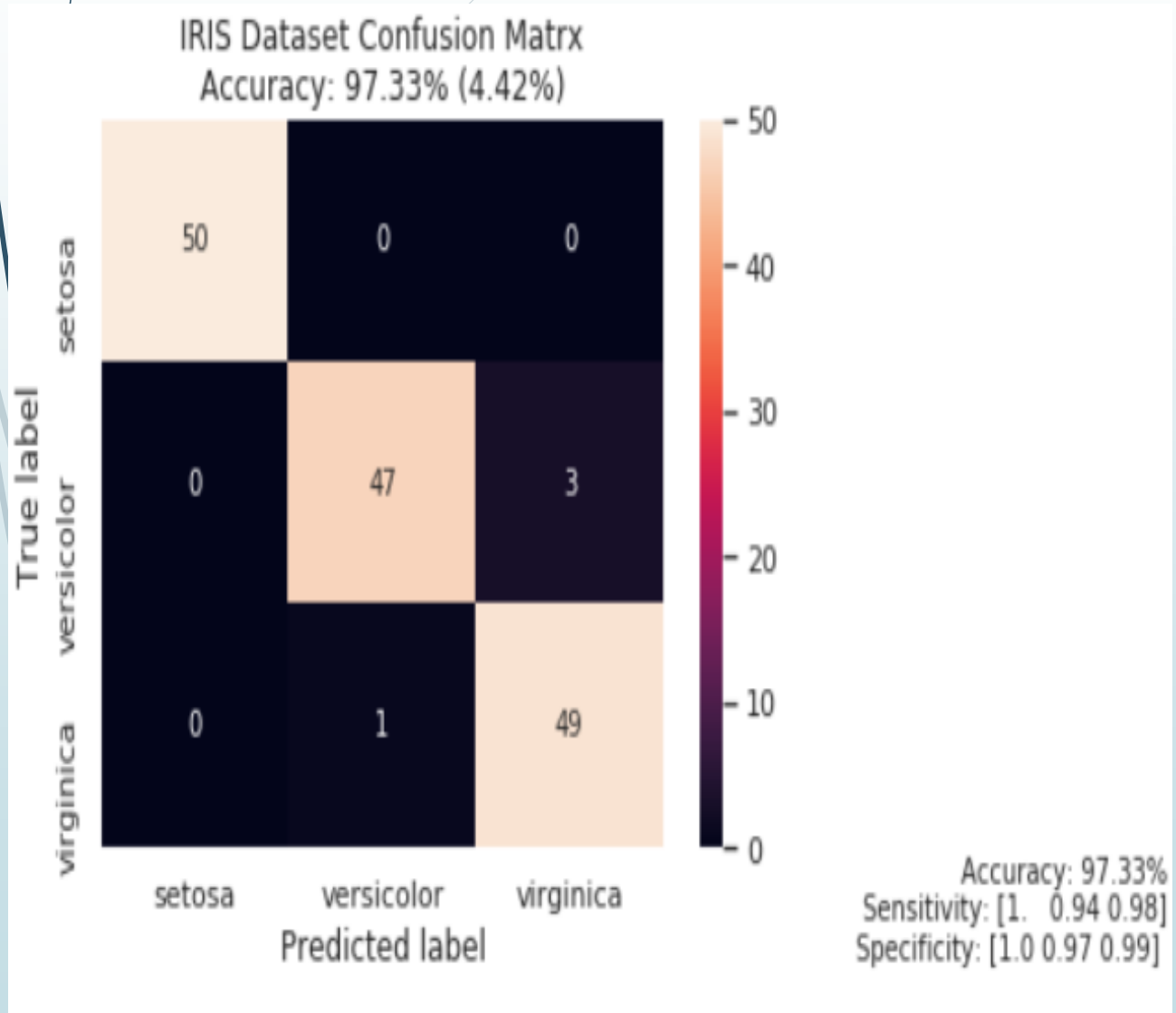
Evaluating the model with k-fold cross validation

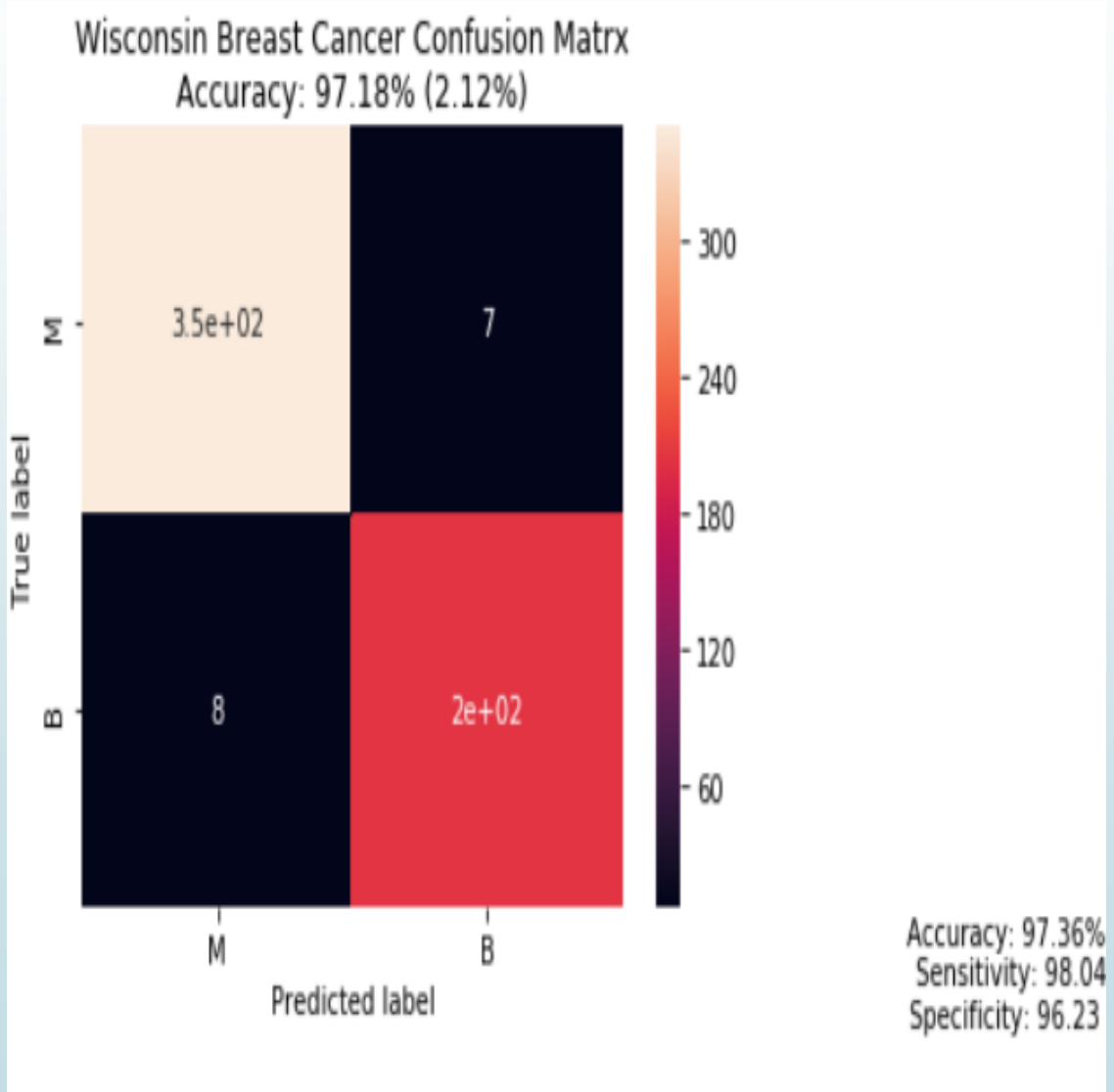
In [50]:
```
#Evaluating out estimator on our dataset using 10-fold cross valiation
results = cross_val_score(estimator, X, y, cv=kfold)
print("Accuracy: %.2f%% (%.2f%%)" % (results.mean()*100, results.std()*100))
```

Sonuçlar

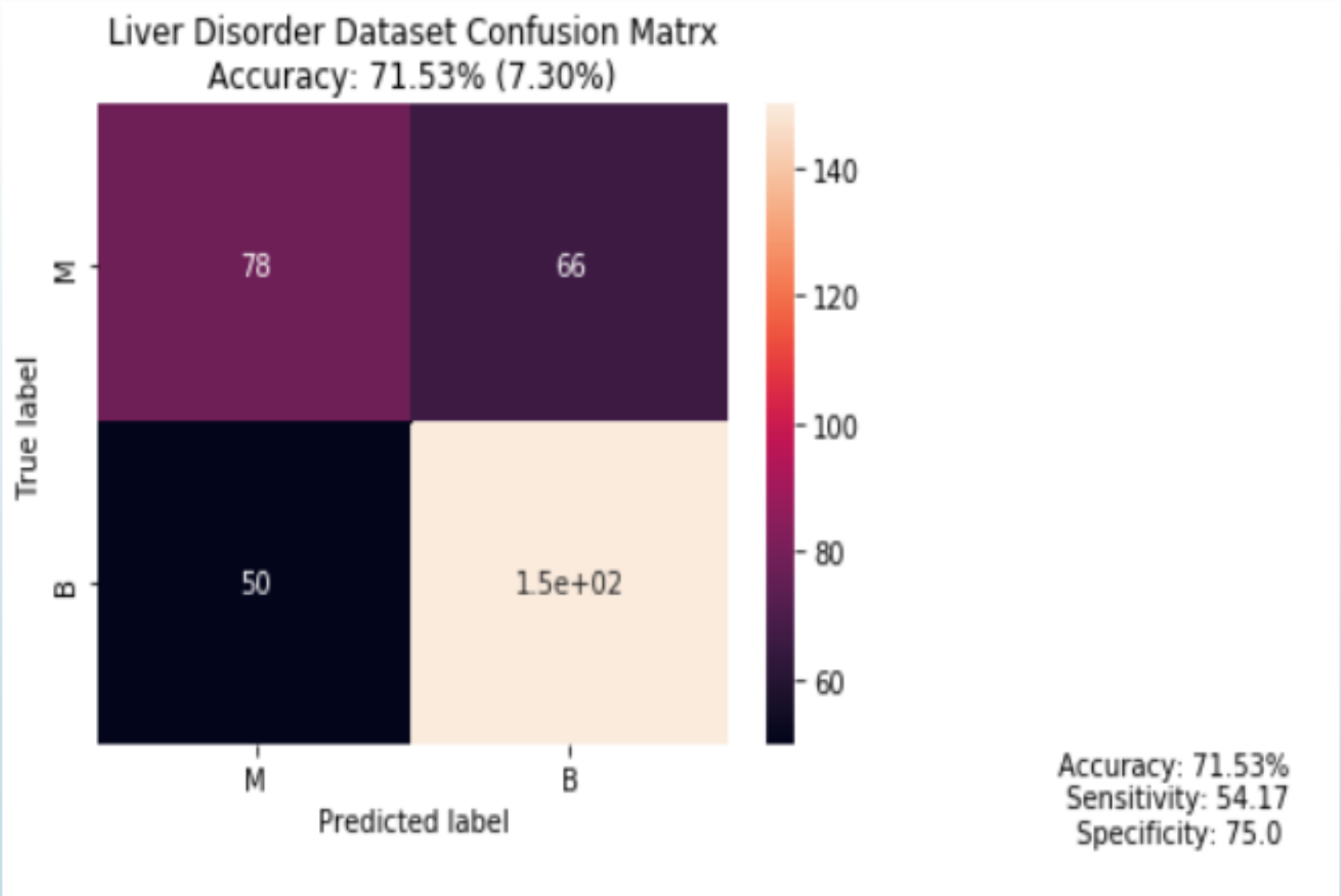
- $\boldsymbol{Doğruluk\ (Acc)} = (TP + TN)/(TP + TN + FP + FN)$
- $\boldsymbol{Duyarlılık\ (TPR)} = TP/(TP + FN)$
- $\boldsymbol{Özgüllük\ (TNR)} = TN/(TN + FP)$

Sonuçlar


IRIS Dataset Confusion Matrx
Accuracy: 97.33% (4.42%)
True label
setosa
versicolor
virginica
50
0
0
0
47
3
0
1
49
50
40
30
20
10
0
setosa
versicolor
virginica
Predicted label
Accuracy: 97.33%
Sensitivity: [1. 0.94 0.98]
Specificity: [1.0 0.97 0.99]


Wisconsin Breast Cancer Confusion Matrx
Accuracy: 97.18% (2.12%)
True label
M
B
3.5e+02
7
8
2e+02
300
240
180
120
60
M
B
Predicted label
Accuracy: 97.36%
Sensitivity: 98.04
Specificity: 96.23

Sonuçlar


Liver Disorder Dataset Confusion Matrx
Accuracy: 71.53% (7.30%)
78
66
50
1.5e+02
M
B
True label
M
B
Predicted label
140
120
100
80
60
Accuracy: 71.53%
Sensitivity: 54.17
Specificity: 75.0

Sonuçlar

UCI DataSets	Iris	Wisconsin Breast Cancer	Liver Disorder
Accuracy	97.33	97.38	71.53
Sensitivity	1.0	0.98	0.54
	0.94		
	0.98		
Specificity	1.0	0.96	0.75
	0.97		
	0.99		

UCI DataSets	Iris	Wisconsin Breast Cancer	Liver Disorder
[1]	93.8		
[7]	97.8	94.5	68.3
[8]			95.73
[9]			88.38
[10]	96.6		
[11]	96.6		
[12]		95	
[13]		99.68	
[14]			73.3
[15]			90
[16]	96		
[17]	97.3		
[18]		94.74	
This Study	97.33	97.38	71.53

Referanslar

- M. Navin JR and Balaji K, "Performance Analysis of Neural Networks and Support Vector Machines using Confusion Matrix", "International Journal of Advanced Research in Science,Engineering and Technology", vol. 3, Issue 5, 2016. (Introduction too)
- S. Haykin,"Neural Networks and Learning Machines","Pearson Prentice Hall", 3rd Edition,2008. (haykin)
- R. Andrews, J. Diederich and A. B. Tickle, "Survey and Critique of Techniques For Extractng rules from Trianed Artificial Neural Networks", "Knowledge-Based Systems", vol. 8, 1995 (info)
- S.Sathish Kumar, Dr N Duraipandian, "Artificial Neural Network Based Method for Classification of Gene Expression Data of Human Diseases along with Privacy Preserving", "International Journal of Computers & Technology",vol. 4,2013. (ANN)
- M. Zakaria, M. AL-Shebany, S. Sarhan, "Artificial Neural Network : A Brief Overview", "Int. Journal of Engineering Research and Applications",vol. 4,2014.
- A. Krenker, J. Bester and A. Kos, Introduction to the Artificial Neural Networks, Artificial Neural Networks - Methodological Advances and Biomedical Applications, Prof. Kenji Suzuki (Ed.), ISBN: 978- 953-307-243-2, InTech, Available from: http://www.intechopen.com/books/artificial-neural-networksmethodological-advances-and-biomedical-applications/introduction-to-the-artificial-neural-networks , 2011
- L. Zheng, H. Huo, Y. Guo and T. Fang," Supervised Adaptive Incremental Clustering for data stream of chunks", "Neurocomputing", http://dx.doi.org/10.1016/j.neucom.2016.09.054, 2016.
- P. H. Kassani, A. B. J. Teoh and E. Kim, "Evolutionary-modified fuzzy nearest-neighbor rule for pattern classification", "Elsevier Ltd.", 2017.
- B. V. Ramana, M.S. P. Babu and N. B. Venkateswarlu, "A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis", "International Journal of Database Management Systems" (IJDMS), Vol.3, No.2, 2011.
- A. S. Aneeshkumar and C. J. Venkateswaran, " Estimating the Surveillance of Liver Disorder using Classification Algorithms", "International Journal of Computer Applications (0975 – 8887) " Volume 57– No.6,2012.
- M. Swain, S. K. Dash, S. Dash and A. Mohapatra, "An Approach For Iris Plant Classification Using Neural Network", "International Journal on Soft Computing (IJSC)", Vol.3, No.1,2012.
- S. T. Halakatti and S. T. Halakatti , "Identification Of Iris Flower Species Using Machine Learning" , "IPASJ International Journal of Computer Science (IIJCS)", Volume 5, Issue 8,2017.
- R. Setiono, "Extracting Rules from Pruned Neural Networks for Breast Cancer Diagnosis", "Appears in Artificial Intelligence in Medicine", Vol. 8, No. 1, 1996.
- A. M. Abdel-Zaher and A. M. Eldeib, "Breast Cancer Classification Using Deep Belief Networks", "Expert Systems With Applications",2016.
- B. V. Ramana and M.S. P. Babu, " Liver Classification Using Modified Rotation Forest"," International Journal of Engineering Research and Development",vol.1,2012.
- R. Lin, "An intelligent model for liver disease diagnosis", "Artificial Intelligence in Medicine",2009.
- K. Patel, J. Vala and J. Pandya, "Comparison of various classification algorithms on iris datasets using WEKA", "International journal of Advance Engineering and Research Development (IJAERD)", Vol. 1 Issue 1,2014.
- W. Yue,Z. Wang, H. Chen, A. Payne and X. Liu, "Machine Learning with Applications in Breast Cancer Diagnosis and Prognosis", "MDPI",2018
- A. Roshanpoor, M. Ghazisaeidi, S. Niakan, K. Maghooli and R. Safdari, "The Performance of K-Nearest Neighbors on Malignant and Benign Classes: Sensitivity, Specificity, and Accuracy Analysis for Breast Cancer Diagnosis"," International Journal of Computer Applications (0975 – 8887)", Volume 180 – No.8, 2017
- W. Zhu, N. Zeng, N. Wang, "Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS®