**Proje: Göğüs Kanseri Sınıflandırması**

**1. Proje Bilgileri:**

**Proje Amacı:**  
Göğüs kanseri hücrelerini **iyi huylu (benign)** ve **kötü huylu (malignant)** olarak sınıflandırmak. Amaç, hastalık teşhisinde doktorlara yardımcı olacak güvenilir bir sınıflandırma modeli geliştirmektir.

**Kullanılan Model:**  
Yapay Sinir Ağı (ANN - Artificial Neural Network).

Bu modelde, üç farklı yapay sinir ağı (ANN) mimarisi denendi ve her biri farklı hiperparametrelerle test edildi. Bu modeller, kanser teşhisi için ikili sınıflandırma yapmayı amaçladı. Kullanılan yapılar ve hiperparametreler şunlardı:

* **Model 1**: İki gizli katman (64 ve 32 nöron), ReLU aktivasyon fonksiyonu ve Adam optimizasyon algoritması.
* **Model 2**: İki gizli katman (128 ve 64 nöron), ReLU aktivasyon fonksiyonu ve SGD (Stokastik Gradyan İnişi) optimizasyon algoritması.
* **Model 3**: İki gizli katman (64 ve 64 nöron), Tanh aktivasyon fonksiyonu ve Adam optimizasyon algoritması.

**2. Veri Seti Bilgileri:**

**Veri Seti Kaynağı:**

UCI Machine Learning Repository - **Breast Cancer Wisconsin Dataset**.

Wolberg, W., Mangasarian, O., Street, N., & Street, W. (1993). Breast Cancer Wisconsin (Diagnostic) [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5DW2B.

Link: <https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic>

**Veri Seti Özellikleri:**

* **Toplam Örnek Sayısı:** 569
* **Sınıf Sayısı:** 2
  + Benign (İyi Huylu)- 357 örnek,
  + Malignant (Kötü Huylu)- 212 örnek

**Özellikler:** 30 adet sayısal özellik (örneğin, radius, texture, smoothness gibi kanser hücrelerinin fiziksel ve kimyasal özellikleri) boş özellik kolonu yok. İd ve unnamed kolonları hariç.

Örnek özellikler:

* Radius (yarıçap)
* Texture (doku)
* Perimeter (çevre)
* Area (alan)
* Smoothness (pürüzsüzlük)
* Compactness (sıkılık)
* Concavity (çukur derinliği)
* Symmetry (simetri)
* Fractal Dimension (fraktal boyut)
* **Sınıf Bilgisi:**
  + 0: Malignant (Kötü Huylu)
  + 1: Benign (İyi Huylu)
* **Veri Seti Tanıtım** **Makalesi:** Street, W. N., Wolberg, W. H., & Mangasarian, O. L. (1993, July). Nuclear feature extraction for breast tumor diagnosis. In Biomedical image processing and biomedical visualization (Vol. 1905, pp. 861-870). SPIE.

**3. Uygulama Detayları:**

**Model Geliştirme:**  
Yapay sinir ağı modeli şu aşamalarda eğitilecek:

1. **5-Fold Cross Validation**

**a. En iyi parametreler:** 'hidden\_layer\_sizes': (128, 64), 'activation': 'relu', 'optimizer': 'sgd'

**b. Ağ Topolojisi :**

Gizli Katmanlar (Hidden Layers): hidden\_layer\_sizes parametresi ile belirtilen boyutlara sahip iki gizli katman ekler. Örneğin, (128, 64) olarak ayarlandığında, ilk gizli katmanda 128 nöron, ikinci gizli katmanda ise 64 nöron bulunur.

Aktivasyon Fonksiyonu: activation='relu' ile her gizli katmanda ReLU aktivasyon fonksiyonu kullanılır.

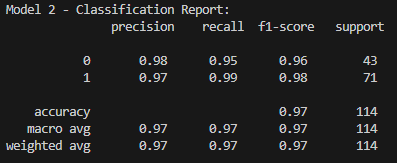
Çıktı Katmanı: Çıktı katmanı, ikili sınıflandırma için sigmoid aktivasyonu kullanır, böylece modelin çıktısı 0 ile 1 arasında bir değer olur.

Optimizer (Optimizasyon Algoritması): SGD (Stokastik Gradient Descent) kullanılır.

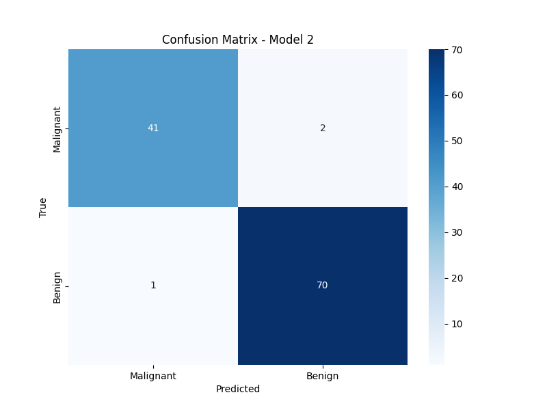
**c. Başarı Metrikleri:**

Cross-validation accuracy for Model 2: 0.9868

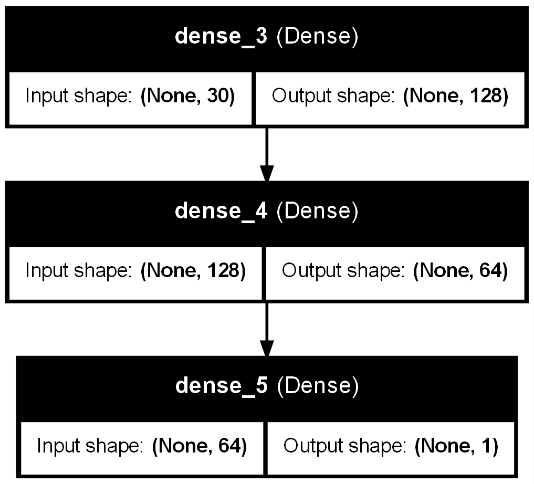
Model 2 - Accuracy: 0.9737, Precision: 0.9722, Recall: 0.9859, F1 Score: 0.9790

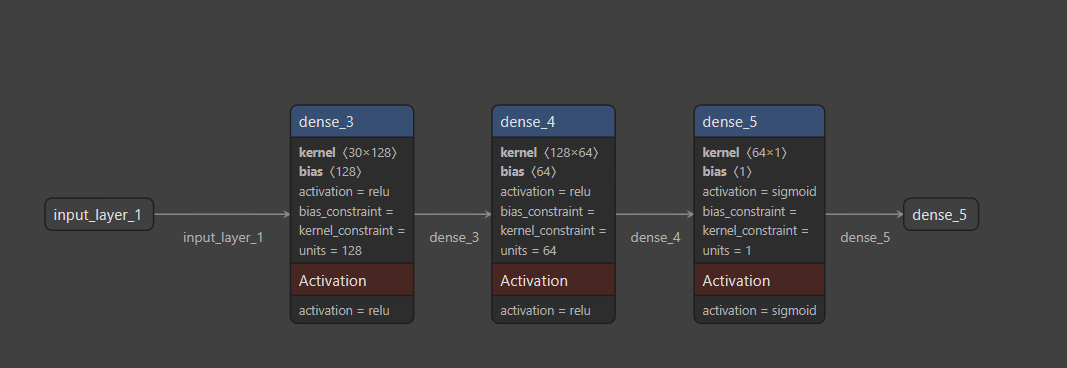


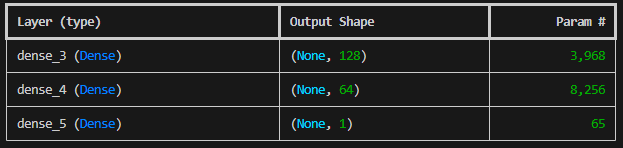
**d. Konfüzyon Matrisi**



**e. Gui ve Model Summary**







Total params: 12,289 (48.00 KB)

Trainable params: 12,289 (48.00 KB)

Non-trainable params: 0 (0.00 B)

Sonuçlar:

* **Model 1** (gizli katmanlar: 64, 32, aktivasyon: ReLU, optimizasyon: Adam): Çapraz doğrulama doğruluğu %99.56, test doğruluğu %96.49, hassasiyet, duyarlılık ve F1 skoru yaklaşık %97.18.
* **Model 2** (gizli katmanlar: 128, 64, aktivasyon: ReLU, optimizasyon: SGD): Çapraz doğrulama doğruluğu %98.68, test doğruluğu %97.37, hassasiyet %97.22, duyarlılık %98.59, F1 skoru %97.90.
* **Model 3** (gizli katmanlar: 64, 64, aktivasyon: Tanh, optimizasyon: Adam): Çapraz doğrulama doğruluğu %98.68, test doğruluğu %97.37, hassasiyet %97.22, duyarlılık %98.59, F1 skoru %97.90.

Model 2 ve Model 3, benzer test doğruluğu ve yüksek duyarlılıkla (yaklaşık 0.986) en iyi performansı gösterdi, bu da malign vakaların tespitinde daha başarılı olduklarını gösteriyor. Ancak, Model 2, hassasiyet ve F1 skoru açısından biraz daha iyi performans sergiledi. Model 1, yüksek çapraz doğrulama doğruluğuna rağmen, test doğruluğu ve F1 skoru bakımından Model 2 ve 3'e kıyasla biraz daha düşük sonuçlar verdi.

Sonuç olarak, Model 2, tüm metrikler açısından sağlam bir performans sergileyerek en iyi model olarak öne çıkıyor ve özellikle hassasiyet ve duyarlılık bakımından üstünlük sağlıyor.

1. **10-Fold Cross Validation**

**a. En iyi parametreler:** 'hidden\_layer\_sizes': (64, 32), 'activation': 'relu', 'optimizer': 'adam'

**b. Ağ Topolojisi :**

Gizli Katmanlar (Hidden Layers): hidden\_layer\_sizes parametresi ile belirtilen boyutlara sahip iki gizli katman ekler. Örneğin, (64, 32 olarak ayarlandığında, ilk gizli katmanda 64 nöron, ikinci gizli katmanda ise 32 nöron bulunur.

Aktivasyon Fonksiyonu: activation='relu' ile her gizli katmanda ReLU aktivasyon fonksiyonu kullanılır.

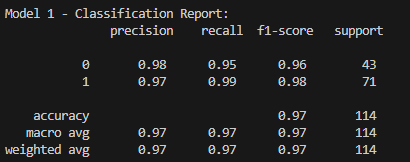
Çıktı Katmanı: Çıktı katmanı, ikili sınıflandırma için sigmoid aktivasyonu kullanır, böylece modelin çıktısı 0 ile 1 arasında bir değer olur.

Optimizer (Optimizasyon Algoritması): Adam optimizasyon algoritması kullanılır

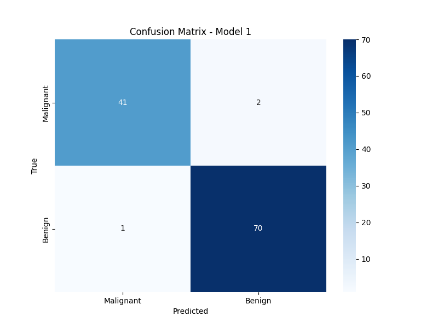
**c. Başarı Metrikleri:**

Cross-validation accuracy for Model 1: 0.9957

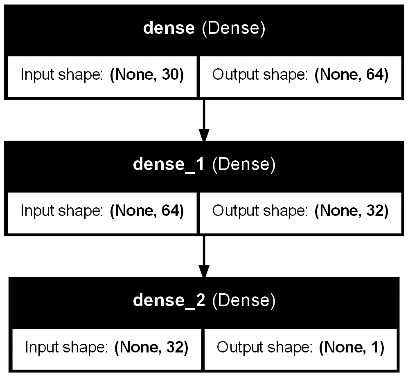
Model 1 - Accuracy: 0.9737, Precision: 0.9722, Recall: 0.9859, F1 Score: 0.9790

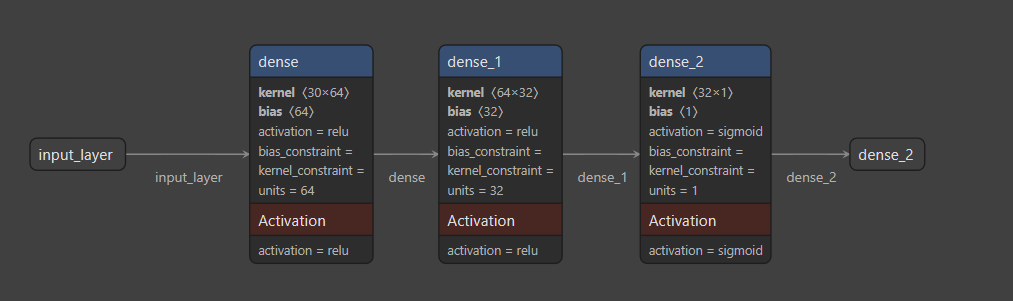


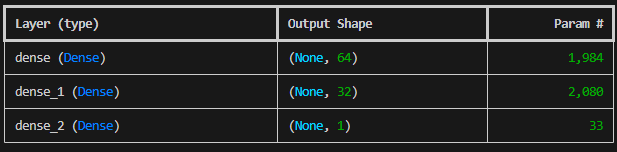
**d. Konfüzyon Matrisi**



**e. Gui ve Model Summary**







Total params: 4,097 (16.00 KB)

Trainable params: 4,097 (16.00 KB)

Non-trainable params: 0 (0.00 B)

Yaptığınız deneylerde üç farklı modelin performansını karşılaştırmışsınız. Her modelin farklı yapılandırmaları ve optimizasyon teknikleri bulunuyor. İşte her bir modelin kısa özeti ve karşılaştırması:

1. **Model 1**: Bu modelde, iki katmanlı bir yapıda 64 ve 32 nöron kullanılıyor ve ReLU aktivasyon fonksiyonu ile Adam optimizasyonu tercih ediliyor. Modelin çapraz doğrulama (k-fold cross-validation) doğruluğu %99.57, eğitim doğruluğu ise %97.37. Precision, recall ve F1 skorları sırasıyla %97.22, %98.59 ve %97.90. Bu modelin genel olarak yüksek doğruluk oranlarına sahip olduğunu söyleyebiliriz.
2. **Model 2**: Model 2'de 128 ve 64 nöronlu iki katman var, ayrıca ReLU aktivasyonu ve SGD optimizasyonu kullanılıyor. Çapraz doğrulama doğruluğu %98.69, eğitim doğruluğu %96.49. Precision, recall ve F1 skorları sırasıyla %97.18, %97.18 ve %97.18. Bu model de oldukça iyi sonuçlar elde etmiş, ancak Model 1'e göre biraz daha düşük doğruluk oranlarına sahip.
3. **Model 3**: Bu modelde de iki katman bulunuyor, fakat her katmanda 64 nöron kullanılıyor ve aktivasyon fonksiyonu olarak tanh seçilmiş. Adam optimizasyonu ile çalışıyor. Çapraz doğrulama doğruluğu %99.35, eğitim doğruluğu ise %97.37. Precision, recall ve F1 skorları sırasıyla %98.57, %97.18 ve %97.87. Model 3, Model 1'e oldukça yakın bir performans sergiliyor, ancak recall açısından biraz daha düşük sonuçlar alıyor.

**K-Fold Cross-Validation**: Her üç modelde de çapraz doğrulama kullanılmış ve her modelin performansı birkaç farklı veri kümesi üzerinde test edilmiş. Bu, modellerin genelleme yeteneklerini daha doğru bir şekilde değerlendirmek için çok önemli bir adım. K-Fold, modelin overfitting yapmasını engellemek için güçlü bir yöntemdir.

**Sonuç ve En İyi Model**: Modellerin performansına bakıldığında, **Model 1** en yüksek çapraz doğrulama doğruluğuna sahip ve yüksek precision ve recall skorları sunuyor. Ancak, Model 3 de benzer sonuçlar veriyor ve daha iyi bir precision skoruna sahip. Model 2 ise doğruluk açısından biraz daha düşük kalıyor. Bu bağlamda, **Model 1** en iyi performansı gösteren model olarak öne çıkıyor.

1. **%66-%34 Eğitim-Test Ayrımı (5 farklı rastgele ayırma ile)**

**a. En iyi parametreler:** 'hidden\_layer\_sizes': (64, 64), 'activation': 'tanh', 'optimizer': 'adam'

**b. Ağ Topolojisi :**

Gizli Katmanlar (Hidden Layers): hidden\_layer\_sizes parametresi ile belirtilen boyutlara sahip iki gizli katman ekler. Örneğin, (64, 64 olarak ayarlandığında, ilk gizli katmanda 64 nöron, ikinci gizli katmanda ise 64 nöron bulunur.

Aktivasyon Fonksiyonu: activation='tanh'.

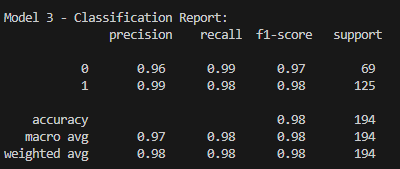
Çıktı Katmanı: Çıktı katmanı, ikili sınıflandırma için sigmoid aktivasyonu kullanır, böylece modelin çıktısı 0 ile 1 arasında bir değer olur.

Optimizer (Optimizasyon Algoritması): Adam optimizasyon algoritması kullanılır

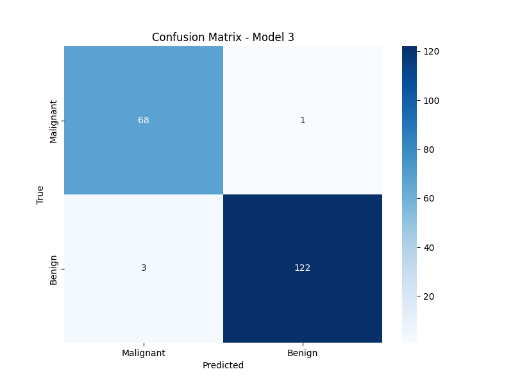
**c. Başarı Metrikleri:**

Cross-validation accuracy for Model 3: 0.9840

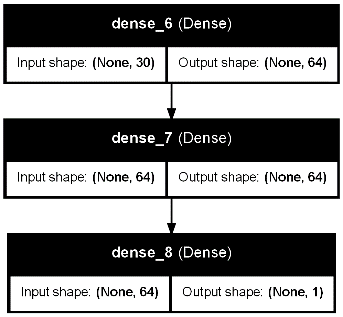
Model 3 - Accuracy: 0.9794, Precision: 0.9919, Recall: 0.9760, F1 Score: 0.9839

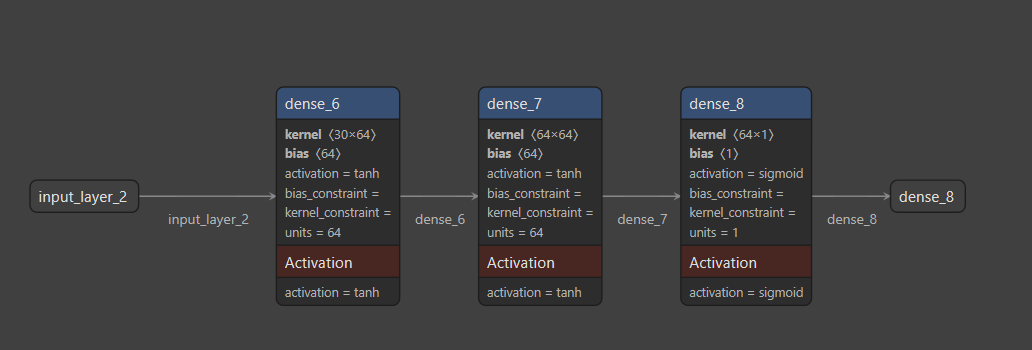


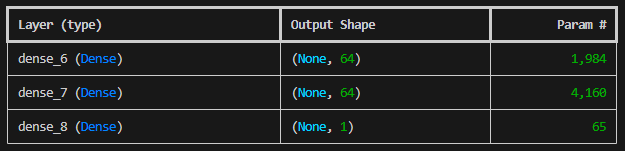
**d. Konfüzyon Matrisi**



**e. Gui ve Model Summary**







Total params: 6,209 (24.25 KB)

Trainable params: 6,209 (24.25 KB)

Non-trainable params: 0 (0.00 B)

Üç farklı modelin karşılaştırıldığı bu çalışmada, her biri farklı parametreler ve optimizasyon yöntemleriyle eğitilmiştir. Sonuçlar aşağıdaki şekilde özetlenmiştir:

1. **Model 1 (64, 32, ReLU, Adam):** Bu modelin çapraz doğrulama doğruluğu %98.67 olarak hesaplanmış ve test setinde %97.42 doğruluk, %98.39 hassasiyet, %97.60 geri çağırma ve %97.99 F1 skoru elde edilmiştir.
2. **Model 2 (128, 64, ReLU, SGD):** Bu modelin çapraz doğrulama doğruluğu %98.13, test setindeki sonuçları ise Model 1 ile benzer şekilde %97.42 doğruluk, %98.39 hassasiyet, %97.60 geri çağırma ve %97.99 F1 skoru sağlamıştır.
3. **Model 3 (64, 64, Tanh, Adam):** Çapraz doğrulama doğruluğu %98.40 olan bu model, test setinde %97.94 doğruluk, %99.19 hassasiyet, %97.60 geri çağırma ve %98.39 F1 skoru sunmuştur.

**Sonuç:** Tüm modeller yüksek performans sergilemiştir, ancak **Model 3** en yüksek çapraz doğrulama doğruluğu ve F1 skoru ile en iyi sonuçları elde etmiştir. Eğitim-test ayrımı %66-%34 ile yapılmış ve her model 5 farklı rastgele eğitim ve test seti ile test edilmiştir.