****

**KARAR DESTEK SİSTEMLERİ**

**MEME KANSERİ KARAR DESTEK SİSTEMİ**

**ÖĞRETİM GÖREVLİSİ**

MUHAMMET HANEFİ CALP

SEDANUR GEBİÇ 368702

ÖZET

Veri ambarı ve veri madenciliğinin birlikte kullanımı sağlık alanına yeni bir soluk getirmiştir. Sağlık sektörü bilgiye dönüşmeyi ve değerlendirilmeyi bekleyen çok fazla miktarda veri barındırmaktadır. Hastanelerin son zamanlardaki hedefleri arasında olan dijital hastaneler için önemli uygulamalardan biri, karar destek sistemleridir. Doktorların hasta ile ilgili karar verme sürecine standart bir uygulama getirerek yol gösterici olması ve hastaya ayrılan zamanın daha aktif kullanılması amaçlanmıştır. Hastane bilgi yönetim sistemi ve karar destek sistemi yapısının incelenmesi ve etkin bir karar destek modülünün geliştirilmesi için çalışmalar yapılmıştır. Çalışma kapsamında, karar destek sistemi algoritmalarının oluşturulması, sisteme entegre edilmesi, uygulaması ve uygulama sonuçları yer almaktadır. Çalışmanın sonraki aşamasında kurulan bu veri ambarı kullanılarak veri madenciliği teknikleri ile doktorlara karar verme aşamalarında destek sağlayacak bir örnek sistem geliştirilmiştir. Kanser verilerinin büyüklüğü ve karmaşıklığı düşünüldüğünde veri ambarı ve veri madenciliğinin birlikte kullanılmasıyla kurulan bir karar destek sistemi, doktorların tedavilerini şekillendirmesinde onlara ışık tutabilecektir. Bu çalışmada çağımızın en ciddi hastalıklarından olan meme kanseri üzerinde durulmuştur. Çalışmada Wisconsin Meme Kanseri (Teşhis) veri seti (data.csv) isimli bir veri seti kullanılmıştır. Bu veri seti bir göğüs kitlesinin ince iğne aspiratının (FNA) sayısallaştırılmış bir görüntüsünden özellikleri alınarak iyi huylu veya kötü huylu olarak tanımlamak için oluşturulmuştur. Veri setindeki verileri en doğru şekilde tahmin edilmesi için Karar Ağacı sınıflandırma yöntemleri kullanılmıştır. Bu yöntemde “diagnosis, radius\_mean, texture\_mean, perimeter\_mean, area\_mean, smoothness\_mean, concavity\_mean, symmetry\_mean, radius\_se, texture\_se, premeter\_se, area\_se, smoothness\_se, concavity\_se, smemetry\_se, radius\_se, radius\_worst, texture\_worst, perimeter\_worst, area\_worst, smoothness\_worst, concavity\_worst, symmetry\_worst” parametreleri temel alınarak ilerlenmiştir. Yöntemlerde makineye veri setinin sonucu bilinen değerlerin %65 eğitim %35 i test kümesi alınarak model oluşturulmuştur. Yöntem için gerekli olan hesaplamalar yapılmış, sonuçlar ve grafikler kullanılmıştır. Göğüs kanseri verileri belirli formatlara getirilerek, hatalardan ve tekrarlardan arındırılarak, oluşturulan veri ambarı yapısında tutulmuştur. Bu veriler hastaların değerlerden oluşmaktadır. Veri ambarı kullanımıyla büyük miktardaki bu veriler arasında karmaşık sorgulamaları bile yapmamıza olanak sağlayan bir yapı kurulmuştur. Çalışmanın ilerleyen aşamalarında değerli bilgilerin açığa çıkarılması ve meme kanseri hastalığının tedavisiyle uğraşan uzmanlara karar aşamasında fikir sağlayan bir karar destek sistemi geliştirilmesi amaçlanmaktadır. Ayrıca meme kanseri hakkında yapılan bilimsel araştırmalar için de kullanıma hazır hale getirilmiş bu verilerden daha birçok şekilde yararlanılabilir. Uygulamanın detaylı açıklaması 3. Bölümde bulunmaktadır.

1.GİRİŞ

Dünyadaki teknolojik gelişmelere bağlı olarak ülkemizde sağlık alanında bilgisayar kullanımı artmıştır. Sağlık kuruluşlarında, diğer kuruluşlarda olduğu gibi verimliliği artırırken maliyet ve karmaşayı azaltmak amacıyla bilgi yönetim sistemleri kullanımına ihtiyaç duyulmuştur[1]. Hastalıklara kısa sürede teşhis konulması, mevcut verilerin daha kolay yönetilmesi ve hızlı bir şekilde tedavi yolunun seçilip uygulanması gerekmektedir. Bu gereklilik tıp alanında bilişim sektöründen yararlanılmasını kaçınılmaz kılmıştır. Bilgi teknolojilerinden ve sağlık bilişim sistemlerinden faydalanılarak hastalık süreçleri için çok kıymetli olan zamanın kısaltılması, daha kaliteli hizmet verilmesi ve pozitif yönde sonuçlar alınması sağlanabilmektedir.

GLOBOCAN 2012 [2] verilerine göre 2012 yılında Dünya’da toplam 14,1 milyon yeni kanser vakası gelişmiş ve 8,2 milyon kansere bağlı ölüm olmuştur. En çok tanı konulan kanserler akciğer (%13,0), meme (%11,9) ve kolon (%9,7) iken kanserden ölümlerin ise en çok akciğer (%19,4), karaciğer (%9,1) ve mideden (%8,8) gerçekleştiği belirtilmiştir. Meme kanseri, hem gelişmiş hem de gelişmekte olan ülkelerde akciğer kanserinden sonra kadınlarda kanser ölümlerinin önde gelen nedenlerinden biri olduğu için önemli bir sağlık sorunu olarak kabul edilmektedir[3].ABD'de tüm kanser teşhislerinin üçte birini meme kanseri oluşturmaktadır ve kadınlar arasındaki en yaygın kanser türüdür. Araştırmalar, 50 yaşın üstünde mamografi çektiren kadınlar için ölüm oranının %30 oranında azalabileceğini göstermiştir[4].

Eğer teknoloji daha etkin kullanılırsa bu hastalıkların tedavi süreci daha olumlu sonuçlanacak, hatta hastalıklar belki de tamamen önlenebilecektir. Bunun için sağlık kuruluşlarının, karar verme aşamasında verileri daha bilinçli şekilde kullanarak değerlendirmeye, verimliliği artırarak maliyetleri düşürmeye, kaynakların daha iyi yönetilmesini sağlamaya ihtiyaçları vardır. Sağlık sektörü tüm bu amaçları gerçekleştirmek için çeşitli bölümlerden, laboratuvarlardan ve ilgili idari süreçlerden gelen, bilgiye dönüşmeyi bekleyen çok fazla miktarda veri barındırır. Bu verilerin dağınıklığı, karmaşıklığı, heterojenliği, birbirinden ayrı yerlerde olması güvenilir şekilde bütünleştirilmelerini zorlaştırır.

Bütünleşmemiş veriler, doktorların birbirlerinin kararlarından haberdar olmasını engellediği için tutarlı sonuçlar elde etmek çok zaman alıcı ve zahmetli bir iş haline gelir. Bu sorunun çözümü için hatalardan ve belirsizliklerden arındırılmış, belirli formatlara getirilmiş verilerin veri ambarı denilen yapılarda tek bir çatı altında toplanması gerekir.

Veri ambarları farklı yerlerden toplanmış verileri kendi yapısında bulundurduğu gibi, bu veriler üzerinde sorgulama ve analiz yapılmasına olanak sağlar[5]. Bununla beraber sağlık alanında bir veri ambarı oluşturmak diğer sektörlere oranla çok daha fazla çaba gerektiren bir iştir. Sağlık alanında veriler zamana bağlı olarak sürekli değişen bir yapıdadır. Bu verilerin daima güncel tutulması doktorların daha kesin ve tutarlı kararlar verebilmesi için son derece önemlidir. Tıp alanındaki röntgen, laboratuar sonuçları gibi verilerin çok karmaşık ve çeşitli tiplerde olması ise diğer bir sorundur. Bu verileri analiz çalışmaları için bütünleştirmek ayrı bir uğraş ve uzmanlık istemektedir.

Doktorların karar verme süreci incelendiğinde geçmiş bilgilerin ve deneyimlerin etkili olduğu görülmektedir. Dolayısı ile deneyimsizlik, insani durumlar ve benzeri anlık veya kalıcı problemler nedeniyle kararlar gerektiği yönde verilemeyebilir. Bunun sonucunda da hata toleransı çok düşük olan tıp alanında ölüme kadar varabilen istenmeyen sonuçlar ile karşılaşılabilmektedir [5].

Veri ambarı ve veri madenciliği tekniklerinden yararlanarak yapılan karar destek sistemleri doktorların hastalarla ilgili en iyi kararı verebilmesi için destek sağlayan bilişim sistemi uygulamalarıdır.Hastanelerde işlerin etkin ve verimli yürütülmesini sağlayan ve karar mekanizması ile yönlendirebilecek bir sistem olan hastane bilgi yönetim sistemi (HBYS) kullanılmaktadır[2]. Sağlık hizmetlerinde hastane bilgi yönetim sistemlerine yapılan kayıtlarda önemli ve büyük miktarda veri üretilmektedir[6]. Hastane bilgi yönetim sistemleri, etkinlik ve verimliliği en üst düzeye çıkarmak, elde edilen bilgileri doğru şekilde saklamak ve bu bilgilerin bölümler arası paylaşımını gerçekleştirmek amacıyla insan sağlığı için önemli hususlarda riski minimum seviyede tutmak amacıyla geliştirilmiş sistemlerdir[7]. Bu sistemler en güncel bilgileri doktorların kullanımına sunarak karar vermelerine yardımcı olmak amacıyla tasarlanmaktadır. Bilginin çok çabuk eskidiği ve artan bilgi miktarına bağlı olarak doktorların karar vermesinin daha da güç bir hale geldiği durumlarda karar almaya yardımcı olmaktadır [8].

Ülkemizde hastane bilgi yönetim sistemleri tamamlanmaya çalışılırken gelişmiş ülkelerde dijital hastane uygulamalarına geçilmiştir. Dijital hastane; yüksek kalitede sağlık hizmeti sunabilmek için klinik ve idari iş akış süreçlerine bilgi ve iletişim teknolojilerini entegre ederek, her türlü tıbbi cihazın bilgi yönetim sistemine ağlar ve sensörler aracılığıyla veri/bilgi gönderebildiği; çalışanların ve hastaların yetki ve onayları dâhilinde bu sistemdeki veri/bilgiye hastaneden veya uzaktan erişebildiği bir hastane modelidir[9]. Hastanelerde teknolojik gelişmelere uyum ve yenilik anlamında en çok dikkat çeken konu dijital hastane olma yolundaki ilerlemelerdir [10].

Dijital hastane uygulamaları için önemli uygulamalardan biri de karar destek sistemleridir (KDS). Karar verme işlemi, karar vericinin değişik seçeneklerle karşı karşıya kaldığı durumlarda, amaçlarına uygun, kendisince belirlenmiş ölçütlere en uygun olanı seçebilmesidir. Çeşitli alternatifler arasından birini seçme işlemi olan karar verme, aynı zamanda problem çözme işlemini içeren faaliyetleri düşünme ve sonuca varma sürecidir[11]. Klinik karar verme ise, topluma kaliteli ve güvenli bakım sunma ve hasta bakım sonuçlarını geliştirme açısından önemlidir.

Karar destek sistemleri sayesinde etkili, standart, hızlı ve doğru karar verilmekte, zaman uygun kullanılmaktadır. 1950’lerin ortalarından bu yana sağlık bakım alanında geliştirilen KDS’ler Klinik Karar Destek Sistemleri (KKDS) olarak adlandırılmıştır. KKDS; hastaya özgü klinik verilerin analiz edilerek elde edilen yeni bilginin hasta bakımına uygulanmasına yardım etmektedir ve hastanelerde uygulanmaktadır[12]. Klinik karar destek sistemleri, teşhis, tedavi ve bakımları iyileştirirken, hastanelerin maliyet ve tanılama hatalarını azaltmak için kullanılan bir sistemdir. Klinik Karar Destek Sistemleri müdahalelerinin performansları hasta sonuçlarını iyileştirirken, bu uygulamaların raporlanmasını kolaylaştırmaktadır[13].

Institute of Medicine tarafından 2000 yılında yapılan bir çalışmaya göre her yıl Amerika’da tıbbi hatalardan dolayı 98.000 hasta ölmektedir. Enstitünün amacı ise geliştirilen klinik karar destek sistemleri ile bu sayıyı minimize etmektir [14]. Klinik karar destek sistemleri kullanıldıkları klinik bölümlerde oluşan beklenmedik durumlarda erken uyarı, hızlı cevap, anında analiz, maliyet düşürme, doğru karar, etkin takım çalışması, zaman tasarrufu ve veri kaynaklarını iyi kullanabilme gibi etkin özellikleri ile sınanan alternatif sayısında artış sağlamaktadır [15].

2.LİTERATÜR TARAMASI

Amacı, kurumsal yapılarda karar destek sistemlerinde yapay zeka kullanımının uygulanabilirliğini araştırmak ve bu araştırmanın sonucunda bir model ve uygulama geliştirerek KDS simülasyonlarının uzman sistemler desteği ile modele uygun çıkarımlarda bulunmasını sağlamaktır. Bu sayede taktiksel düzeyde karar alan orta düzey yöneticilerinin karar süreçlerinin ne ölçüde hızlanabileceği ve doğru kararlar verebileceği ortaya konulacaktır. Çalışma üretim yapan ve tüm Türkiye’de satış faaliyetlerini yürütmekte olan bir A Firmasının geçmişe dönük 5 yıllık verilerinin aylık seriler halinde Veri ambarına aktarılması ve bu veri ambarı üzerinden KDS’nin What-If modellemeleriyle manipüle edilerek ilgili senaryoların görselleştirilmesi yöntemi ile oluşturulmuştur. Gerçekleştirilen araştırma ile, yapay zekâ uzman sistemlerin organizasyonlara uygun modellenmiş KDS‟ler ile orta düzey yöneticilere taktiksel kararlar alırken hız ve verimlilik kattığını ölçümlemelerle ortaya koymuştur [16].

Bu çalışmanın amacı, Türkiye’de üç büyükşehirde görev yapan kadın hastalıkları ve doğum uzmanlarının klinik karar destek sistemi (KKDS) kullanım düzeyinin belirlenmesidir. Kesitsel ve analitik tipteki çalışma, İzmir Sağlık Bilimleri Üniversitesi Tepecik Eğitim ve Araştırma Hastanesi Girişimsel Olmayan Klinik Araştırmalar Etik Kurulu’ndan etik onay alındıktan sonra gerçekleştirildi. Veriler, 17 sorudan oluşan bir soru formu kullanılarak yüz yüze görüşme yöntemi ile toplandı. Bu çalışmada KHDU arasında KKDS kullanma düzeyi %71,9 olarak belirlenmiştir. Her gün düzenli olarak KKDS kullanan hekimlerin düzeyi düşüktür (%14,8). Uygulamaları genç hekimlerin ve akademik unvana sahip olanların daha fazla kullanması nedeniyle 45 yaş üzeri ve akademik olmayan gruplara uzmanlık dernekleri tarafından bu konuda tanıtım ve eğitim yapılması sonucuna varılmıştır[17].

Çalışmada amaç, çalışanların değerlendirilmesinde kullanılması gereken görev performansı ölçütlerini oluşturmak değildir. Bir çalışanın mevcut işini 7 iş değerlendirme faktörü açısından nasıl yerine getirdiğini yansıtan bilgi ve beceriler, görev performans kriterleri olarak tasarlanmıştır. Bu çalışmada, hem Aritmetik Artış ve hem de Geometrik Artış yöntemleri kullanılmış, faktörlerin derece tanımları dikkate alınarak, her bir faktör için hangi yönteme göre derece puanlarının belirleneceği kararlaştırılmış, ancak derece tanımları dikkate alınarak puanlarda kısmi düzenlemeler yapılarak faktör-puan çizelgesi oluşturulmuştur. Karar destek sistemi karar vericinin yaptığı en güncel değişikliklere göre çalışmakta ve elde edilen tüm sonuçları veri tabanında, veri güvenliği sağlayacak biçimde saklı tutulmaktadır[18].

Bu çalışmada müşteri memnuniyetini ve firmanın önceliklerini dengeleyebilmek için teslim tarihi belirleme problemini çok kriterli bir karar yapısı olarak ele alan bir karar destek modeli için geliştirilmiştir. Yapılan çalışmada, Türkiye’de kalkınmada öncelikli TRC1 bölgesindeki tekstil sektörünün ihtiyaçları baz alınarak uygulama geliştirilmiştir. Bölgede hızla büyüyen ve yüksek teknolojik bir altyapıya sahip olan dokumasız kumaş üretici firmalarının satış pazarlama birimi yöneticileri ile yapılan yüz yüze görüşmelerde uzman görüşleri sorularak sipariş sıralamasında etkili faktörler olarak aşağıdaki kriterler belirlenmiştir. Uygulama Veritabanına tanımlanan tabloların veri giriş çıkış işlemleri için Müşteri, İşletme, Bakım, Sipariş ve Parametre formları tasarlanmıştır. Müşteri ve sipariş gibi çoklu kayıtların hazır bir Excel dosyasından içe aktarımının yapılabilmesi ve elde edilen sonuçların dışa aktarılabilmesi için Excel içe aktar/ dışa aktar fonksiyonları kodlanmıştır. Yapılan çalışmada önerilen karar modeli, hiyerarşik olarak müşteri ve sipariş kriterlerinden oluşmaktadır. Uygulama sektörü olarak dokumasız kumaş üreten işletmeler seçilmiştir[19].

Yöneticilerin verdiği stratejik kararlarda önemli toplantı sonuçları çok etkilidir. Araştırmanın en genel amacı toplantıların kalitesini artırmaktır. Araştırmada, toplantı tutanaklarından kelime çıkarımı yapılarak, toplantılara ait genel konu başlıkları metin madenciliği ile elde edilecektir. Yöntem olarak, dökümanlara metin madenciliği süreci içerisinde, otomatik doğal dil işleme teknikleri kullanılmıştır. Otomatik analizlerin yanı sıra elde edilen yapısal veri tarafımızdan elle kontrol edilerek yeniden gözden geçirilmiştir. K-ortalamalar ve benzerlik temelli modelleme araçları, sosyal ağ grafikleri verilere uygulanmıştır. Sonuç olarak makine öğrenimi ve yapay zeka teknikleri ile güçlü modeller geliştirilebilir. Elde edilen temalar, ilerde tasarlanacak bilgi sistemlerinde girdi olarak kullanılabilir[20].

Bu çalışmada yemek tüketim miktarını tahmin etmek için ileri beslemeli çok katmanlı yapay sinir ağları, destek vektör makineleri ve lineer regresyon analizi kullanılmıştır. Bu metotlar için veri madenciliği amacıyla geliştirilmiş açık kaynak kodlu WEKA yazılım aracından faydalanılmıştır. Yemek tüketim tahminini yapmak için Yapay Sinir ağları, Destek Vektör Makineleri ve Lineer Regresyon analizi yöntemleri kullanılmış olup, açık kaynak kodlu WEKA yazılım aracından faydalanılmıştır. Yapılan bu araştırmada makine öğrenme sistemlerinin yemek tahmini üzerinde sonuçlar elde edilebileceği gösterilmiş. Farklı profil ve ölçekteki yerleşkeler için metotların performansları sunulmuştur. Sonraki çalışmalar için bir karşılaştırma verisi olup daha büyük ölçekli işletmeler için farklı metotlar geliştirilerek bu çalışmanın genişletilebilir olduğuna varılmıştır[21].

Çalışma; kronik böbrek yetmezliği hastalarının sosyal destek algısını geliştirmeye yönelik yürütülen grup çalışmasını içermektedir. Atatürk Eğitim Araştırma Hastanesi hemodiyaliz ünitesinde tedavi gören bireylerle 2018 yılı Ekim-Kasım ayları arasında 8 haftalık grup çalışması yoluyla gerçekleştirilen bu uygulamada; katılımcıların yaşam deneyimlerini, düşünce, duygu, istek ve gereksinimlerini rahatça paylaşabilecekleri bir ortam yaratarak hemodiyaliz uygulaması sırasında kendilerini daha iyi hissetmeleri, hastalar arasındaki etkileşimin ve hastaların sosyal destek algılarının değişim ve gelişimi amaçlanmıştır. Yürütülen grup çalışmasının etkisini ortaya koymak amacıyla, nicel araştırma desenlerinden deneysel araştırma olarak tasarlanan çalışmada kontrol grupsuz ön test-son test modeli kullanılmıştır. Bu çalışma sonrasında yapılacak araştırmalar ve uygulamalarda, grup çalışmasından yararlanılmasının sürdürülmesi başta hastalar olmak üzere hastaların aileleri, sosyal çevreleri ve sağlık alanında çalışan profesyonelleri olumlu etkileyecektir. Bu doğrultuda yalnızca hastalar ile değil hastaların aileleri ve hizmet veren personellerle de grup çalışmaları planlanabilir [22].

Telekomünikasyon sektöründe reklam ve tanıtım faaliyetlerinde geleneksel ve dijital pazarlama kanalları yoğun olarak kullanılmaktadır. Bu çalışmada Telekomünikasyon sektöründe yer alan Türkiye’nin önde gelen bir işletmesinin ürünlerini ve kampanyalarını pazarlamak için sosyal medya platformunun seçiminde, seçimi etkileyen kriterlerin belirlenmesi ve kriter ağırlıklarının hesaplanması amaçlanmıştır. Araştırma kapsamında hazırlanan anketler işletme bünyesinde görev yapan on sekiz pazarlama yöneticisine uygulanmıştır. Ağırlıkların hesaplanmasında Bulanık DEMATEL yönteminden, durulaştırma aşamasında ise CFCS yönteminden faydalanılmıştır. Araştırma kapsamında görünüm ve algı, içerik, maliyet, üye profili, izlenim oranı ve teknik destek olmak üzere 6 kriter belirlenmiştir. Analiz sonuçlarına göre izlenme oranı (0,185), görünüm ve algı (0,181) ve üye profili (0,170) kriterleri sosyal medya platformu seçiminde en fazla ağırlığa sahip ilk üç kriter olarak karşımıza çıkmıştır [23].

Çalışmada hastanelerde HBYS ‘de kullanılacak yazılım paket program seçimi problemi ele alınmıştır. Yapılan çalışmada belli başlı kriterler, belirli uzman kişiler tarafından ve literatür araştırması sonucu belirlenmiş ve buna karşılık en uygun yazılım paket programının bulunması amaçlanmıştır. Bu çalışma altı kısımdan oluşmaktadır. Çalışmada, birinci kısımda HBYS ile ilgili bilgi, ikinci kısımda ise kullanılan yöntemler ile ilgili olarak literatür araştırması yapılmıştır. Üçüncü kısımda ise çalışmanın çözümünde kullanılan ÇKKV tekniklerinden AHP, TOPSIS ve PROMETHEE yöntemlerinden bahsedilmiştir. Dördüncü kısımda problem tanımlanarak AHP, TOPSIS ve PROMETHEE yöntemleri ile problem çözümü yapılmıştır. Beşinci kısımda ise değerlendirme ve sonuçlara yer verilmiştir. Bu çalışmada elde edilen sonuçlara göre; öncelikle belirlenen kriterler doğrultusunda kriterler arası karşılaştırma matrisinin AHP yöntemiyle çözülmesinin ardından “güvenlik ve firma” en önemli önceliğe sahip kriter olarak belirlenmiştir. Diğer adımda ise TOPSIS ve PROMETHEE yöntemleri sonucu en iyi alternatif olarak “VII Pusula” belirlenmiştir. Araştırmada elde edilen sonuçlara göre sağlık kuruluşları, yöneticiler, çalışanlar ve hastalar yazılım paket programının güvenirliliğine ve gizliliğine önem vermektedir [24].

Bu çalışma da yaşlı bireylerin sosyal destek algılarının yaşam doyumları üzerindeki etkisini belirlemek amaçlanmıştır. Bununla birlikte çalışmada yaşlı bireylerin birlikte yaşadıkları bireyler, çocuklarını görme sıklıkları ve akrabalarını görme sıklıkları değişkenleri ile sosyal destek algısı ve yaşam doyumu arasındaki ilişkilerde analiz edilmiştir. Araştırmaya alınacak birey sayısının belirlenmesinde örneklem genişliği formülünden yararlanılmıştır. %95 güvenirlik ve %5 hata payı ile evrenden alınması gereken örneklem büyüklüğü 384 olarak belirlenmiştir. Örnekleme alınacak birey sayısının belirlenmesinde sistematik örnekleme yöntemi kullanılmıştır. Sosyal destek olduğunu söylemek mümkündür. Sonuç olarak yaşlılıkta yaşam doyumu aile üyeleri, akraba ve arkadaşların sağlamış olduğu sosyal destekten önemli ölçüde etkilenmektedir [25].

**3.YÖNTEM VE TEKNİKLER**

Meme kanseri veri seti yapay zeka sınıflandırma yöntemi olan Karar Ağacı algoritması Spyder ve Weka ile analiz edilmiştir. Analiz sonucunda elde edilen karar ağacı oranları temel alınarak PHP, CSS, HTML ve MySQL ile kodlanarak karar destek sistemi oluşturulmuştur.

**Karar Ağacı**

Karar ağaçları, sınıflandırma ve tahmin için sıkça kullanılan bir veri madenciliği yaklaşımıdır. Karar ağaçlarının sunduğu mantıksal modelin yansıttığı karar kuralları, insanlar tarafından kolayca anlaşılabilecek kadar açıktır. Yüksek sınıflandırma doğruluk oranı ve üretilen basit kurallar gibi özelliklere sahip olduğundan dolayı bu yöntem geniş bir uygulama yelpazesine sahiptir. Karar ağaçları; kişilerin kredi geçmişlerini kullanarak kredi tercihinde bulunması, geçmişte işletmeye en faydalı olan bireylerin özelliklerini kullanarak işe alma süreçlerinin tespit edilmesi, tıbbi gözlem verilerinden hareketle en etkin kararların verilmesi, satışı etkileyen değişkenlerin saptanması, üretim verilerini inceleyerek ürün hatalarına yol açan değişkenlerin belirlenmesi gibi uygulamalarda kullanılmaktadır [26].

Karar ağaçları;

* Düşük maliyetli olması,
* Anlaşılmasının, yorumlanmasının ve veri tabanları ile entegrasyonun kolaylığı,
* Güvenilirliklerinin iyi olması gibi nedenlerden ötürü en yaygın kullanılan sınıflandırma tekniklerinden biridir.

**3.1.ÇALIŞMANIN AMACI**

Meme kanseri veri setinin yapay zeka tekniği olan Karar Ağacı sınıflandırma yöntemi ile incelenmesi ve elde edilen en yüksek başarı oranı veren karar ağacı modelinin temel alınarak karar destek sisteminin kodlanması. Kodlanan karar destek sistemi laboratuvardan meme kanserine ilişkin verileri kayıt ve analiz etmesi. Karar destek sistemi bu veriler ışığında test sonucunun iyi huylu ya da kötü huylu olduğuna ilişkin çıktı vermesi amaçlanmıştır.

**3.2.UYGULAMA**

**3.2.1VERİ SETİNİN TANIMI**

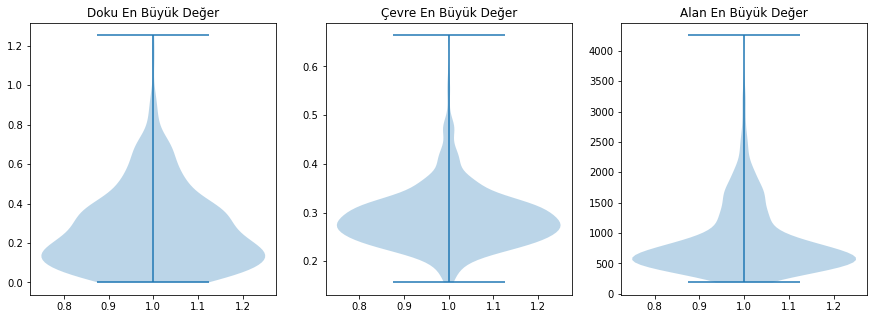
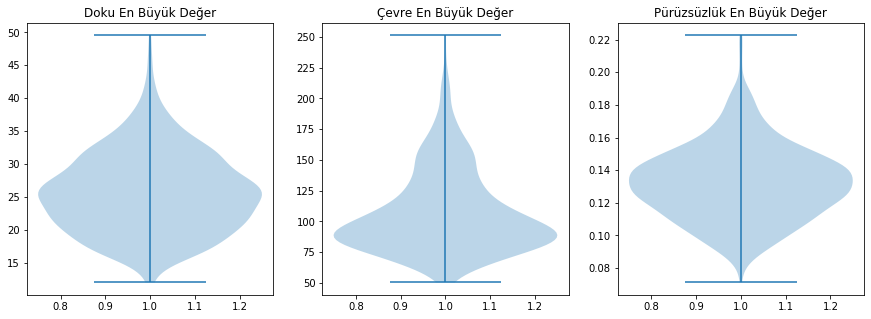
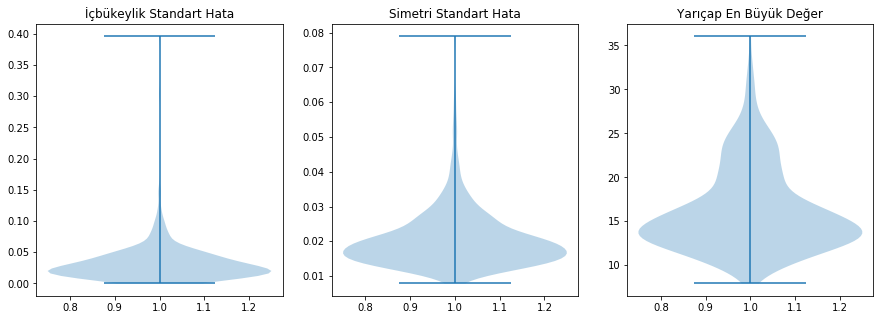
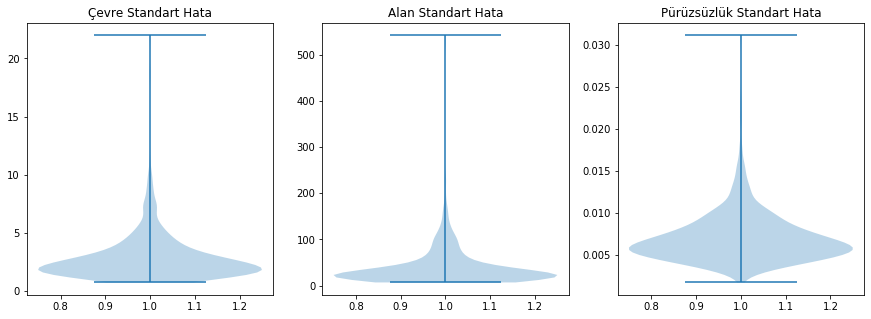
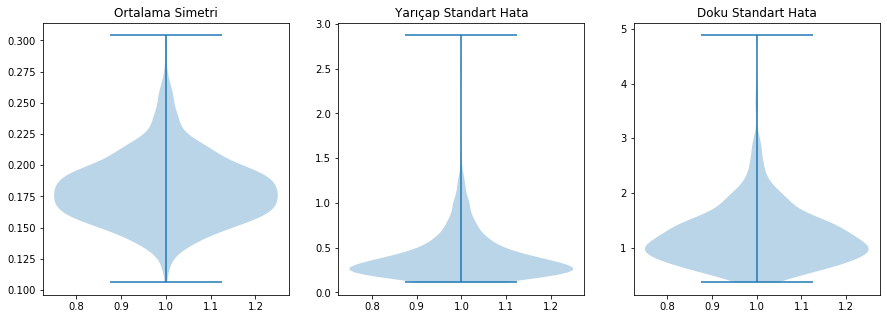
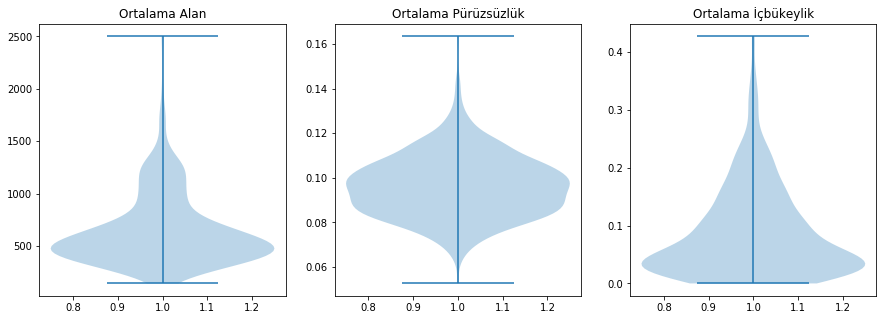
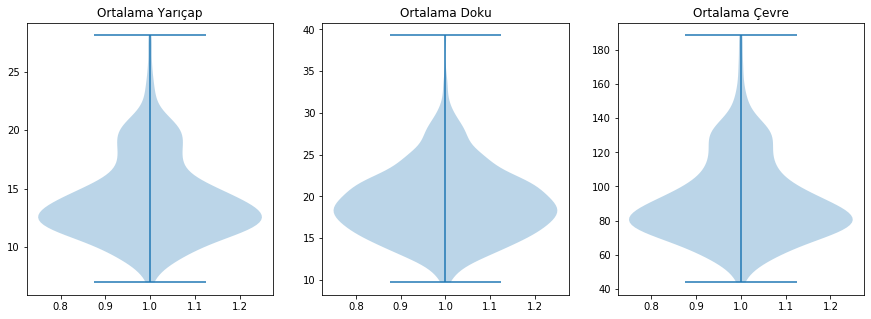
Wisconsin Meme Kanseri (Teşhis) Veri Seti

Veri seti göğüs kanseriyle ilgilidir. Özellikler, bir göğüs kütlesinin ince iğne aspiratının (FNA) sayısallaştırılmış bir görüntüsünden hesaplanmıştır. Veri setinde görüntüde bulunan hücre çekirdeklerinin özellikleri verilmiştir. Verilen alanlar; ortalama, standart hata ve en kötü olmak üzere 3’e ayrılır. Örneğin yarıçap (Radius) ile ilgili olan 3 alan vardır. Bunlar yarıçap ortalaması, yarıçap standart hatası ve en kötü yarıçaptır. Verilen özellikler iyi huylu ve kötü huylu olmak üzere 2 kategoriye ayrılır. 22 sütundan oluşan veri setinde sınıf dağılımı; 357 iyi huylu, 212 kötü huylu olmak üzere toplamda 569 adet veri vardır.

**Veri Kümesi**

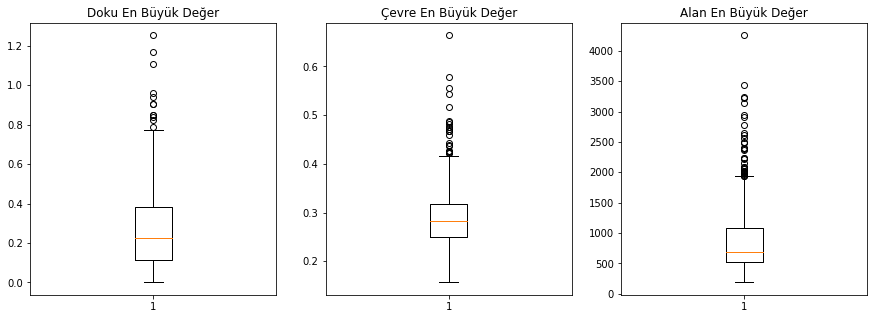
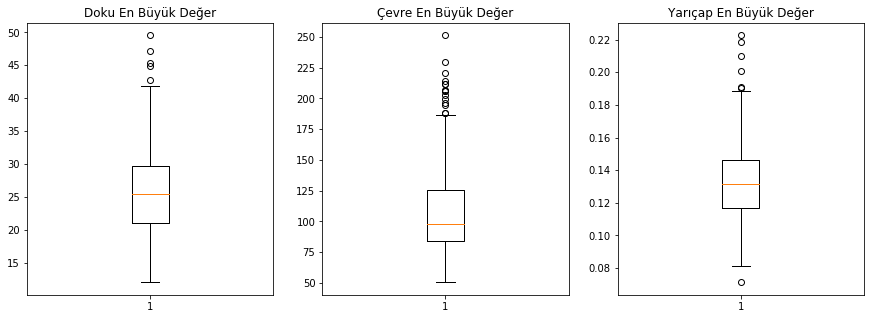
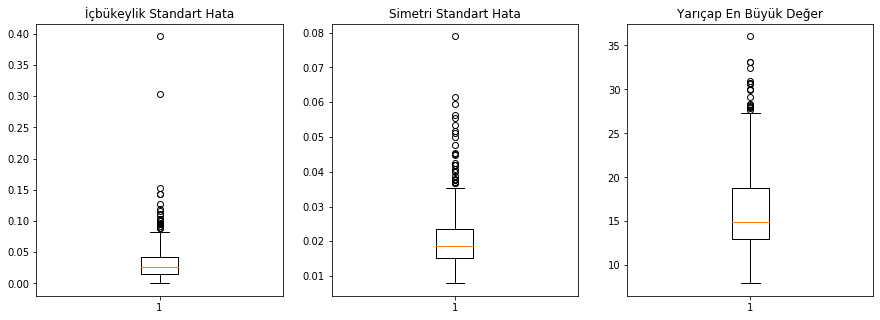
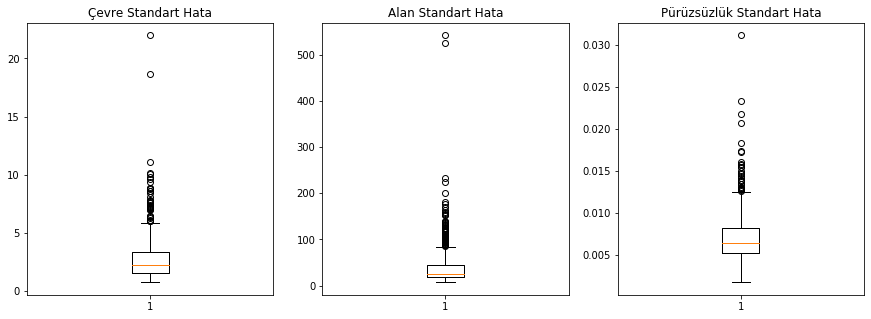
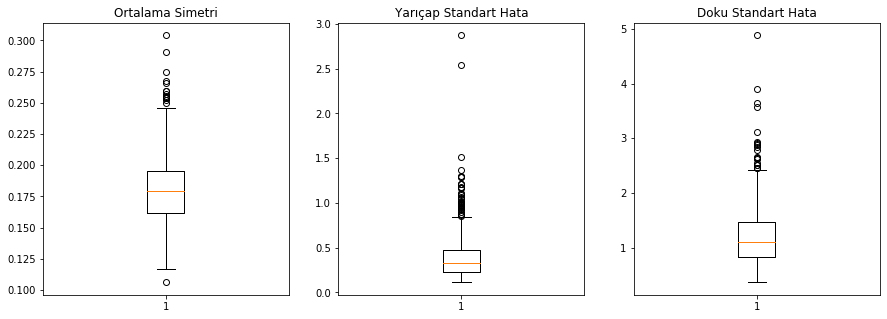
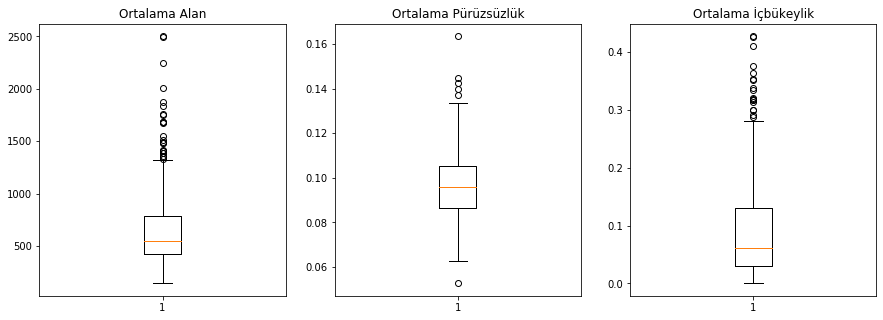
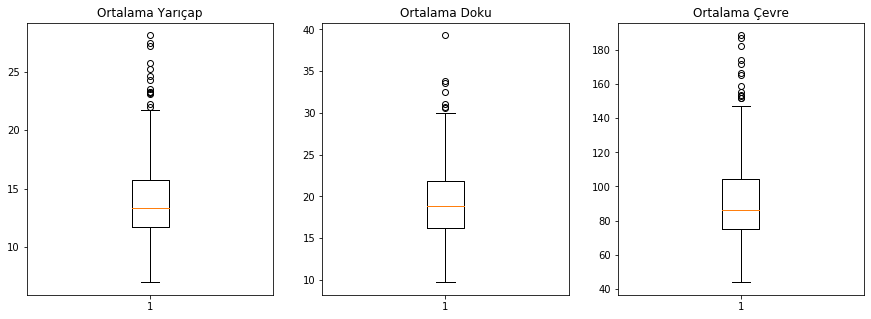
|  |  |
| --- | --- |
| SÜTUNLAR | AÇIKLAMA |
| Diagnosis | M = Kötü Huylu, B = İyi huylu |
| Radius\_mean | Yarıçap Ortalaması (merkezden noktaya olan uzaklıkların ortalaması) |
| Texture\_mean | Doku Ortalaması (gri tonlama değerlerinin standart sapması) |
| Perimeter\_mean | Çevre Ortalaması (çekirdek tümörün ortalama büyüklüğü) |
| Area\_mean | Alan Ortalaması |
| Smoothness\_mean | Pürüzsüzlük Ortalaması (yarıçap uzunluklarında yerel varyasyon) |
| Concavity\_mean | İçbükeylik Ortalaması (konturun içbükey kısımlarının şiddeti) |
| Symmetry\_mean | Simetri Ortalaması |
| Radius\_se | Yarıçap Standart Hata |
| Texture\_se | Doku Standart Hata |
| Perimeter\_se | Çevre Standart Hata |
| Area\_se | Alan Standart Hata |
| Smoothness\_se | Pürüzsüzlük Standart Hata |
| Concavity\_se | İçbükeylik Standart Hata |
| Symmetry\_se | Simetri Standart Hata |
| Radius\_worst | Yarıçap En Kötü |
| Texture\_worst | Doku En Kötü |
| Perimeter\_worst | Çevre En Kötü |
| Area\_worst | Alan En Kötü |
| Smoothness\_worst | Pürüzsüzlük En Kötü |
| Concavity\_worst | İçbükeylik En Kötü |
| Symmetry\_worst | Simetri En Kötü |

**Parametrelerin Aldıkları Değerlerin Dağılımı (Violin Grafik)**



Yukarıdaki violin grafikleri incelediğimizde zaman veri setinde temel aldığımız parametrelerin hangi değerler arasında yığılım yaptığı gösterilmiştir. Sığ olan dağılımlar dengeli olduğunu gösterirken sığ olmayan dağılımlar dengesizlik olduğunu gösterir. Sığ olmayan dağılımlar veri seti için olumsuz bir anlam ifade etmez çünkü her parametre kendi içerisinde farklı bir ölçüte göre hesaplanır. Sığ dağılımlarda dalgalanmanın birden azaldığı yerlerde dengeli dağılımın azaldığı gözlemlenir. Örneğin “İçbükeylik standart hata” parametresinde 2.5 değerinden sonra keskin bir azalış görülmüştür. Bu durum “İçbükeylik standart hata” parametresinin ölçümü sırasında aykırı sonuçlar elde edildiğini gösterir

**Parametrelerin Aldıkları Değerlerin Aykırılık Durumu (Kutu Grafik)**



Yukarıdaki grafikler incelendiğinde parametreler ölçülürken ele alınan değerlerin aykırılık oranları gösterilmiştir. Grafiklerde bulunan yuvarlaklar aykırı verileri temsil etmektedir. Örneğin “Alan En Büyük Değer” sütunu 2000-4000 arasında aykırı veriye sahiptir.

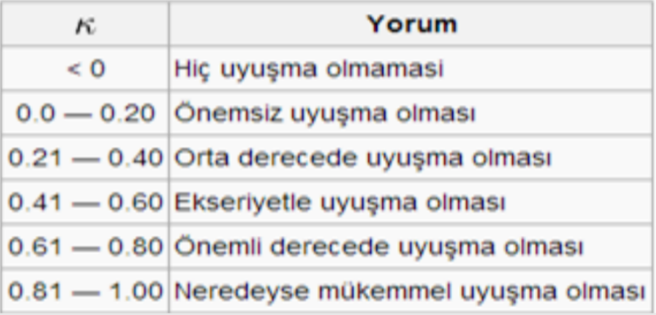
**3.2.2.DEĞERLENDİRME KRİTERLERİNİN AÇIKLAMASI**

**Correctly Classified Instances**

Veri setinde sınıflandırma yönteminde tahmin edilen doğruluk sayısıdır. Aynı zamanda yüzde olarakta sınıflandırma yönteminin başarı oranını göstermektedir.

**Kappa Statistic**

Cohen'in kappa katsayısı iki değerleyici arasındaki karşılaştırmalı uyuşmanın güvenirliğini ölçen bir [istatistik](https://tr.wikipedia.org/wiki/%C4%B0statistik) yöntemidir. Cohen'in kappa ölçüsü her biri *N* tane maddeyi *C* tane birbirinden karşılıklı hariç olan kategoriye ayıran iki değerleyicinin arasında bulunan uyuşmayı ölçer. Cohen'in kappa ölçüsü bu uyuşmanın bir şans eseri olabileceğini de ele aldığı için basit yüzde orantı olarak bulunan uyuşmadan daha güçlü bir sonuç verdiği kabul edilir. Cohen’in kappa istatistiklerinin sonucunun yorumlamasının tablosunu alt kısımda görmekteyiz.



**Mean Absolute Error**

Ortalama mutlak hata iki sürekli değişken arasındaki farkın ölçüsüdür. MAE, yönlerini dikkate almadan bir dizi tahmindeki hataların ortalama büyüklüğünü ölçen, tüm tekil hataların ortalamada eşit olarak ağırlıklandırıldığı doğrusal bir skordur. MAE değeri 0’dan ∞’a kadar değişebilir. Negatif yönelimli puanlar yani daha düşük değerlere sahip tahminleyiciler daha iyi performans gösterir.

**Root Mean Squared Error**

Bir makine öğrenmesi modelinin, tahminleyicinin tahmin ettiği değerler ile gerçek değerleri arasındaki uzaklığın bulunmasında sıklıkla kullanılan, hatanın büyüklüğünü ölçen bir metriktir. Verilere en iyi uyan çizgi etrafında o verilerin ne kadar yoğun olduğunu söyler. RMSE değeri 0’dan ∞’a kadar değişebilir. Negatif yönelimli puanlar yani daha düşük değerlere sahip tahminleyiciler daha iyi performans gösterir. RMSE değerinin sıfır olması modelin hiç hata yapmadığı anlamına gelir.

**Relative Absolute Error**

Bir tahminde hesaplanan değeri ve gerçek değeri biliyorsak bağıl mutlak hata hesaplaması yapılır. Gerçek değer ile hesaplanan değer arasındaki farkın gerçek değere oranlanması ile elde edilir. Göreli mutlak hata 0’a yakın olduğunda başarı elde edilir ve negatif değer almaz.

**Root Relative Squared Error**

Gerçek değerlerin yalnızca ortalamasıdır. Böylece, göreceli kare hata toplamın karesi alınmış hatayı alır ve basit tahmin edicinin toplam kare hatası ile bölerek normalleştirir. Göreceli karesel hatanın karekökü alınarak, hata tahmin edilen miktarla aynı boyutlara indirgenir. RRSE değeri 0’dan ∞’a kadar değişebilir. Göreli hata karekök 0’a yakın olduğunda başarı elde edilir ve negatif değer almaz.

**Total Number of Instances**

Veri setinde sınıflandırma yönteminde kullanılan verilerin toplam sayısıdır.

**Gerçek Pozitif Değerlerin Oranı (True Positive Rate)**

Sınıflayıcının ne kadar gerçek pozitif değeri doğru tahmin ettiğinin bir ölçüsüdür. Hassasiyet, İsabet Oranı veya Hatırlama olarak da bilinir. TP Rate kısaca doğruya doğru demektir

**Yanlış Pozitif Değerlerin Oranı (False Positive Rate)**

Gerçek değeri 0 olmasına karşın 1 olarak tahmin edilenlerin oranıdır. FP Rate kısaca doğruya yanlış demektir.

**Hassasiyet (Precision)**

Tüm sınıflardan, doğru olarak ne kadar tahmin edildiğinin bir ölçüsüdür. Doğru olarak tahmin edilenlerin, toplama oranıdır. Doğru ne kadar tahmin edildiğinin bir ölçüsüdür. 0 ile 1 arasında değer alır, mümkün olduğunca yüksek olmalıdır.

**Geri Çağırma (Recall)**

Pozitif durumların ne kadar başarılı tahmin edildiğini gösterir. En iyi değer 1, en kötü değer 0’dır. Örnekle açıklarsak kullandığım veri setinde true pozitifler sesin kadın olduğu tahmin edilen ve gerçekte sesin kadın olduğu insanlar, false negativeler yani sesin erkek olduğu tahmin edilen ve gerçekte sesin erkek olmadığı insanlar. Recall’e bakılma sebebi tamamiyle paydadaki false negativeler, yani sesin erkek olduğu tahmin edilen ve gerçekte sesin kadın olduğu insanlar. Bu verisetiyle ilgili bir çalışma yapıldığında kadın olan bir sesin erkek olarak algılanmasının maliyeti büyük olacağı için recall false negativelerin göz ardı edilemez durumlarda kullanılan önemli bir metriktir.

**Roc-Area**

ROC bize modelin true positive rate’iyle false positive rate’i cinsinden ne kadar iyi ayrım yapabildiğini açıklar. AREA ise ROC eğrisinin altında kalan alanı verir, 0'la 1 arasındadır, 0'sa bütün tahminler yanlıştır. True positive rate kısaca gerçekte durum pozitifse bunların kaçını pozitif tahmin ettiğimizi gösterir, false positive rate de gerçekte durum negatifken bunların kaçını pozitif olarak tahmin ettiğimizi (yanlış alarm da denir) gösterir.

**F-Measure**

Gerçek olumlu oranın (recall) ve hassasiyetin ağırlıklı ortalamasıdır. F-Measure 0 ile 1 arasında değer alır. Çıkan değer 1’e ne kadar yakınsa model o kadar başarılı olur.

**MCC**

İkili (iki sınıflı) sınıflandırmaların kalitesinin bir ölçüsü olarak kullanılır. Doğru ve yanlış pozitif ve negatifleri hesaba katar ve genellikle sınıflar çok farklı boyutlarda olsa bile kullanılabilecek dengeli bir ölçü olarak kabul edilir. 0 ile 1 arasında değer alır. Çıkan değer 1’e ne kadar yakınsa model o kadar başarılı olur.

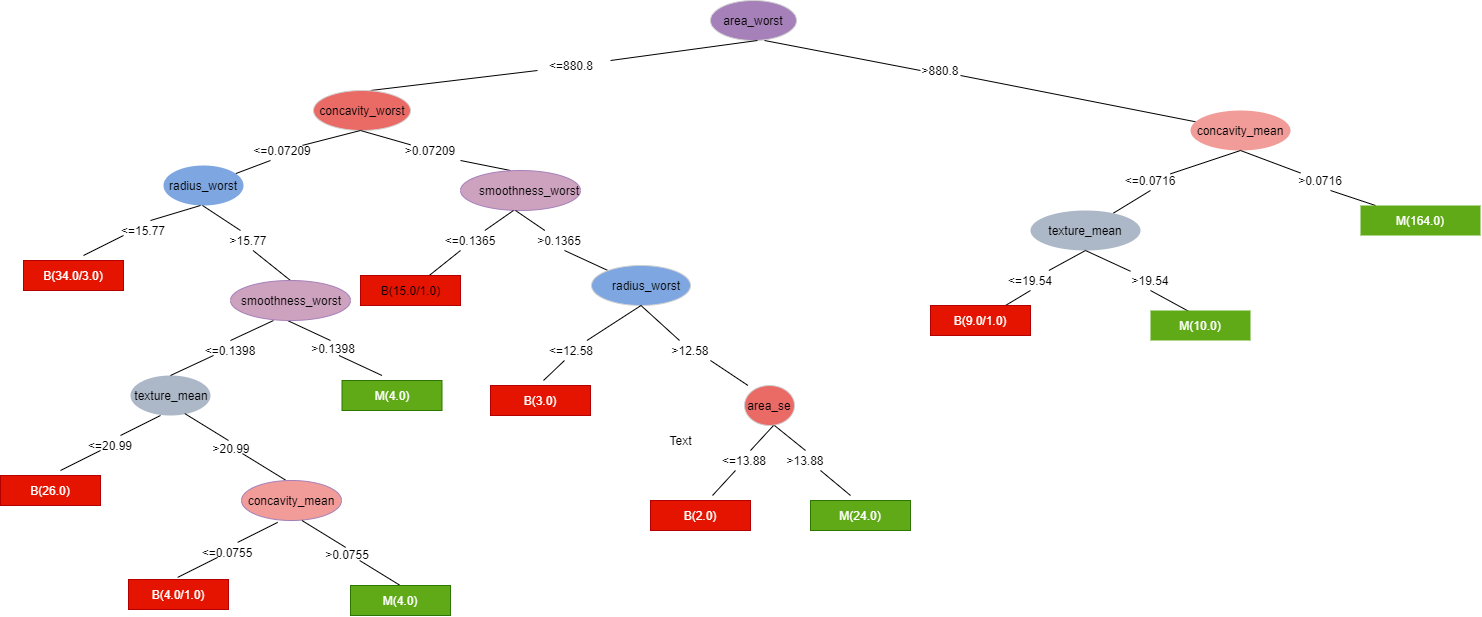
**PRC Area**

Bununla birlikte, dengesiz veri setlerine dayalı olarak ROC eğrilerinin görsel yorumu ve karşılaştırmaları yanıltıcı olabilir. ROC eğrisine bir alternatif, hassaslık-geri çağırma eğrisidir (PRC). ROC eğrilerinden daha az kullanılır, ancak göreceğimiz gibi PRC, dengesiz veri kümeleri için daha iyi bir seçim olabilir. 0 ile 1 arasında değer alır. Çıkan değer 1’e ne kadar yakınsa model o kadar başarılı olur.

**3.2.3. WİSCONSİN MEME KANSERİ (TEŞHİS) VERİ SETİNİN KARAR AĞACI SINIFLANDIRMA YÖNTEMİ**

Veri setimizin içerisinden diagnosis, radius\_mean, texture\_mean, perimeter\_mean, area\_mean, smoothness\_mean, concavity\_mean, symmetry\_mean, radius\_se, texture\_se, premeter\_se, area\_se, smoothness\_se, concavity\_se, smemetry\_se, radius\_se, radius\_worst, texture\_ worst, perimeter\_worst, area\_worst, smoothness\_worst, concavity\_worst, symmetry\_worst parametrelerini temel alarak ilerledik. Karar Ağacı algoritmasından makineye veri setinin sonucu bildiğimiz değerlerin %65’ini eğitim vererek %35’ini tahmin etmesini istedik. diagnosis parametresi üzerinden de sınıflandırmasını yaptık. Yaprak başına düşen minimum obje (MinNumObj) sayısı 2 (varsayılan değer) olarak belirlenmiştir.

ÖRNEK 1



Veri setimizin içerisinden diagnosis, radius\_mean, texture\_mean, perimeter\_mean, area\_mean, smoothness\_mean, concavity\_mean, symmetry\_mean, Radius\_se, texture\_se, perimeter\_se, area\_se, smoothness\_se, concavity\_se, symmetry\_se, Radius\_worst,texture\_worst, perimeter\_worst,, area\_worst, smothness\_worst, concavity\_worst, symmetry\_worst parametrelerini temel alarak ilerledik. Karar Ağacı algoritmasından makineye veri setinin sonucu bildiğimiz değerlerin %65’ini eğitim vererek %35 ini tahmin etmesini istedik. Diagnosis parametresi üzerinden de sınıflandırmasını yaptık. Yaprak başına düşen minimum obje (MinNumObj) sayısı 2 (varsayılan değer) olarak belirlenmiştir.Karar ağacını incelediğimizde dallanmanın ilk olarak “area\_wost” parametresinden başlandığı görülmektedir. İlk koşulun “area\_wost” parametresi olma sebebi entropi değerinin yüksek olmasıdır. Modellediğimiz karar ağacı 12 yaprak, 23 daldan oluşmaktadır.

SONUÇ 1

|  |  |
| --- | --- |
| **Başarı Oranı** | **95.4774 %** |
| Doğru Sınıflandırılmış Örnekler | 190 |
| Kappa İstatistiği | 0.9044 |
| Ortalama Mutlak Hata | 0.0557 |
| Kök Ortalama Kare Hatası | 0.1986 |
| Göreli Mutlak Hata | 11.8548 % |
| Kök Göreli Kare Hatası | 40.7368 % |
| Toplam Örnek Sayısı | 199 |

Modeli oluşturmak için geçen süre: 0,00 saniyedir.

**Correctly Classified Instances**

Veri setinin %35’i test kümesi olarak ele alındığı için 569 satır veriden 199’u üzerinde tahmin yapılmıştır. Bunun sonucunda 199 adet veriden 190’ını doğru tahmin ederek %95.4774oranında başarı elde edilmiştir. Modelimiz tahmin yaparken 9 adet veriyi yanlış bulmuştur.

**Kappa Statistic**

İyi huylu ve kötü huylu sonuçlar arasındaki karşılaştırmalı uyuşmanın güvenilirlik oranı 0.9044’tür. Elde edilen değer 0.81-1. 00 aralığında olduğu için iyi huylu ve kötü huylu değerleri arasında neredeyse mükemmel bir uyuşma olduğu görülmektedir.

**Mean Absolute Error**

Tahmin sonucunda elde edilen ortalama mutlak hata oranı 0.0557’dir. Bu değer 0’a yakın olduğu için tahmin edilen modelde başarı elde edilmiştir.

**Root Mean Squared Error**

Tahmin sonucunda elde edilen karekök ortalama hata oranı 0.1986’dır. Bu değer 0’a yakın olduğu için tahmin edilen modelde başarı elde edilmiştir.

**Relative Absolute Error**

Tahmin sonucunda gerçek değer ile hesaplanan değer arasındaki farkın gerçek değere oranlanması sonucunda 11.8548 değeri elde edilmiştir. Bu değer 0’a yakın olduğu için tahmin edilen modelde başarı elde edilmiştir.

**Root Relative Squared Error**

Tahmin sonucunda kök göreceli hata oranı 40.7368’dir. Bu değer 0’a çok yakın olmasa da diğer metriklerin 0’a yakın olmasından dolayı modelin başarı oranını çok etkilememiştir.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **TP Rate** | **FP Rate** | **Presicion** | **Recall** | **F-Measure** | **MCC** | **ROC Area** | **PRC Area** | **Class** |
|  | 0,935 | 0.033 | 0.947 | 0.935 | 0.941 | 0.904 | 0.953 | 0.918 | M |
|  | 0.967 | 0.065 | 0.959 | 0.967 | 0.963 | 0.904 | 0.953 | 0.961 | B |
| **Ağırlıklı Ortalama** | 0.955 | 0.052 | 0.955 | 0.955 | 0.955 | 0.904 | 0.953 | 0.944 |  |

**TP Rate**

Tahmin sonucunda sınıfı “M” olan verilerden kötü huylu olanı kötü huylu olarak tahmin oranı 0,935’tir. Sınıfı “B” olan verilerden iyi huylu olanı iyi huylu olarak tahmin oranı 0,967’dir. Bu değerler 1’e yakın olduğu için iyi bir isabet oranı elde edildiği görülmektedir. Yapılan karar ağacı sınıflandırma yöntemi sonucunda verilerin tamamına yakın bir kısmını doğru tahmin ettiği görülmektedir. Aynı zamanda “M” ve “B” sınıflarında en çok doğru tahmini yapan B sınıfıdır.

**FP Rate**

Tahmin sonucunda sınıfı “M” olan verilerden “M” sınıfını “B” sınıfı olarak tahmin etme oranı 0.033’dür. Sınıfı “B” olan verilerden “B” sınıfını “M” sınıfı olarak tahmin etme oranı 0,065’dir. Bu değerler 0’a yakın olduğu için yapılan hatalı tahminin çok az olduğu görülmektedir. Aynı zamanda “M” ve “B” sınıfları arasında en çok hata yapan “B” sınıfıdır.

**Precision**

Tahmin sonucunda “M” sınıfı hassasiyet oranı 0,947 iken “B” sınıfı hassasiyet oranı 0,959’dur. Hassasiyet oranları 1’e yakın olduğu için sınıflandırma modelinin doğru tahmin yaptığı görülmektedir. Sınıfı “M” olan verilerin hassasiyet oranı daha yüksektir.

**Recall**

Tahmin sonucunda “M” sınıfı geri çağırma oranı 0,935 iken “B” sınıfı geri çağırma oranı 0,967’dir. Örneğin gerçekte sonucun “B” olduğu durumda tahminin “M” olarak yapılmasıdır. Bu hata 0’a yaklaştıkça artar 1’e yaklaştıkça azalır. Bu durumda sınıflandırma modelinin doğru tahmin yaptığı görülmektedir. “B” sınıfına ait veriler daha doğru tahmin edilmiştir.

**F-Measure**

Tahmin sonucunda “M” sınıfı F-Measure oranı 0.941 iken “B” sınıfı F-Measure oranı 0.963 çıkmıştır. Bu değerin 1’e yakın olması yapılan tahmin modelinin başarılı olduğunu göstermektedir.

**MCC**

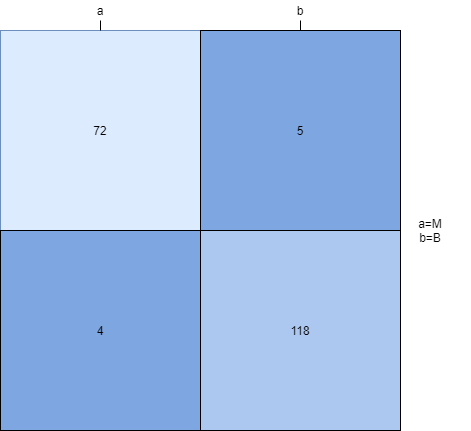
Tahmin sonucunda “M” ve “B” sınıfı MCC oranı 0,904 çıkmıştır. Bu değerin 1’e yakın olması yapılan tahmin modelinin başarılı olduğunu göstermektedir.

**ROC Area**

Tahmin sonucunda “M” ve “B” sınıfı ROC Area oranı 0.953 çıkmıştır. Bu değerin 1’e yakın olması yapılan tahmin modelinin başarılı olduğunu göstermektedir.

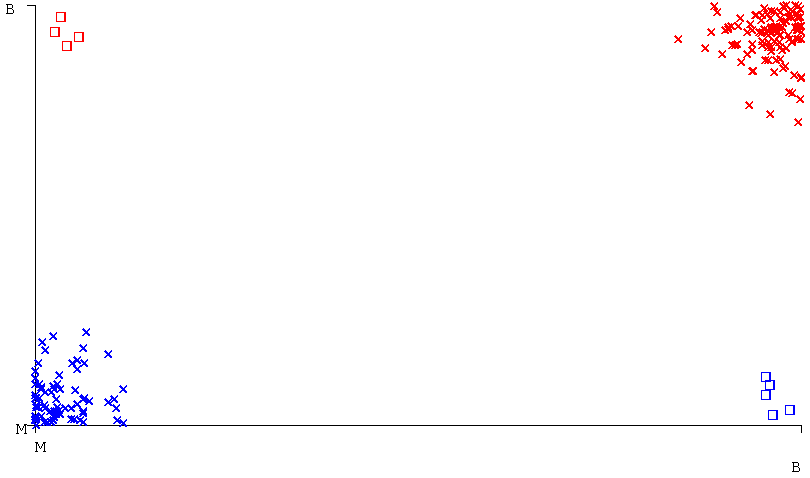
**PRC Area**

Tahmin sonucunda “M” sınıfı PRC Area oranı 0,918 iken “B” sınıfı oranı 0,961 çıkmıştır. Bu değerin 1’e yakın olması yapılan tahmin modelinin başarılı olduğunu göstermektedir. Aynı zamanda “B” sınıfının PRC Area oranı kadın sınıfından daha başarılıdır.

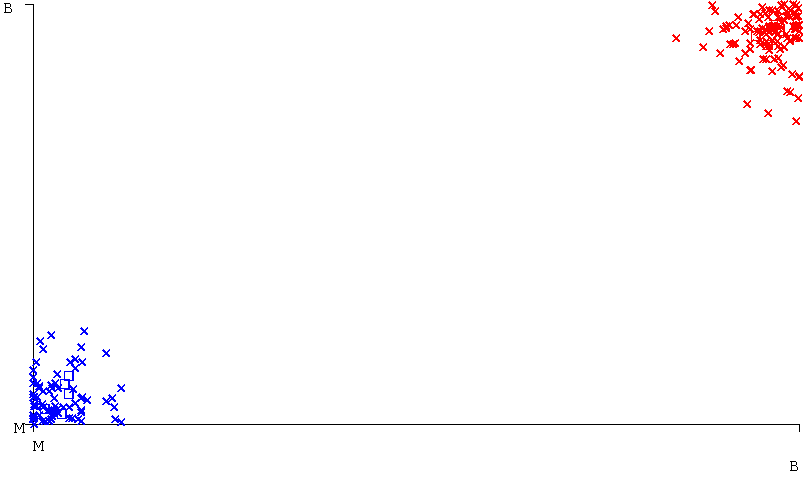


Confusion matrix kısmında oluşturduğumuz test ve eğitim kümeleri sonucunda label parametremizde bulunan “M” ve “B” tahminlerindeki başarımızı görmekteyiz. Confusion matrixde male ifadesini a olarak, female ifadesini b olarak ele almaktayız. A olarak ele aldığımız “M” olan 77 verinin 72 tanesini “M” olarak doğru tahmin ederken 5 tanesini “B” olarak yanlış tahmin etmiştir. B olarak ele aldığımız “B” olan 122 verinin 118 tanesini “B” olarak doğru tahmin ederken 4 tanesini “M” olarak yanlış tahmin etmiştir. Tüm veri setine baktığımız zamanda 199 veriden 190 veriyi doğru tahmin ederken 9 veriyi yanlış tahmin etmiştir.

**Yanlış Tahmin Edilen Verilerin Grafik Olarak Gösterimi**

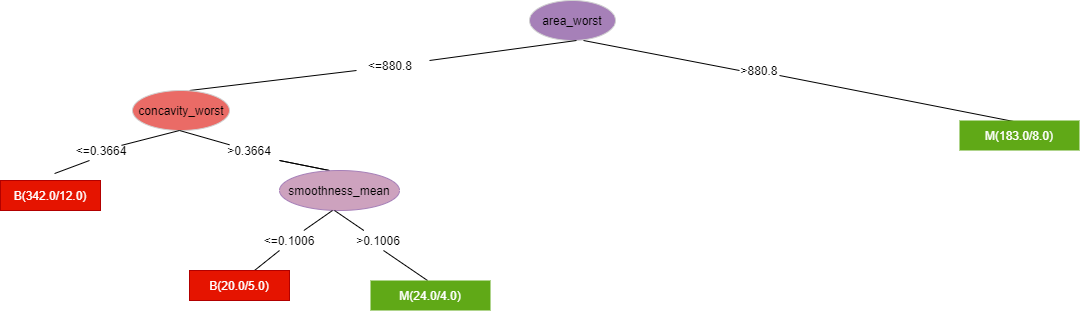


Yukarıdaki grafikte yapılan yanlış ve doğru tahminlerin birleşimi gösterilmiştir. Sınıflandırıcı hata grafiğinde “B” sınıfı kırmızı, “M” sınıfı mavi renktedir. Grafikte ise “**x”** ile gösterilenler doğru sınıflandırılmış verileri “**kare”** şeklinde gösterilen yerler ise yanlış sınıflandırılmış verileri ifade ediyor. Yukarıdaki grafikte yapılan yanlış ve doğru tahminlerin birleşimi gösterilmiştir.



Örneğin **“M”** kısmında bulunan **mavi** bir kare bu değerin “B” sınıfına ait olduğunu ama yanlış şekilde sınıflandırılarak “M” sınıfına dâhil edildiğini gösteriyor. Tahmin modelimizin sınıflandırıcı hata grafiğine baktığımızda 4 “M” 5 “B” olmak üzere 9 hata yapıldığı görülmektedir.

ÖRNEK 2



Veri setimizin içerisinden diagnosis, radius\_mean, texture\_mean, perimeter\_mean, area\_mean, smoothness\_mean, concavity\_mean, symmetry\_mean, Radius\_se, texture\_se, perimeter\_se, area\_se, smoothness\_se, concavity\_se, symmetry\_se, Radius\_worst,texture\_worst, perimeter\_worst,, area\_worst, smothness\_worst, concavity\_worst, symmetry\_worst parametrelerini temel alarak ilerledik. Karar Ağacı algoritmasından makineye veri setinin sonucu bildiğimiz değerlerin %61’ini eğitim vererek %39 unu tahmin etmesini istedik. Diagnosis parametresi üzerinden de sınıflandırmasını yaptık. Yaprak başına düşen minimum obje (MinNumObj) sayısı 20 (varsayılan değer) olarak belirlenmiştir. Karar ağacını incelediğimizde dallanmanın ilk olarak “area\_wost” parametresinden başlandığı görülmektedir. İlk koşulun “area\_wost” parametresi olma sebebi entropi değerinin yüksek olmasıdır. Modellediğimiz karar ağacı 4 yaprak, 7 daldan oluşmaktadır.

SONUÇ 2

|  |  |
| --- | --- |
| **Başarı Oranı** | **89.1892%** |
| Doğru Sınıflandırılmış Örnekler | 198 |
| Kappa İstatistiği | 0.764 |
| Ortalama Mutlak Hata | 0.1593 |
| Kök Ortalama Kare Hatası | 0.3074 |
| Göreli Mutlak Hata | 33.8503 % |
| Kök Göreli Kare Hatası | 62.6616 % |
| Toplam Örnek Sayısı | 222 |

Modeli oluşturmak için geçen süre: 0,01 saniyedir.

**Correctly Classified Instances**

Veri setinin %49’u test kümesi olarak ele alındığı için 569 satır veriden 222’si üzerinde tahmin yapılmıştır. Bunun sonucunda 222 adet veriden 198’ini doğru tahmin ederek %89.1892 oranında başarı elde edilmiştir. Modelimiz tahmin yaparken 24 adet veriyi yanlış bulmuştur.

**Kappa Statistic**

İyi huylu ve kötü huylu sonuçlar arasındaki karşılaştırmalı uyuşmanın güvenilirlik oranı 0.764’tür. Elde edilen değer 0.61-0. 80 aralığında olduğu için iyi huylu ve kötü huylu değerleri arasında önemli derecede uyuşma olduğu görülmektedir.

**Mean Absolute Error**

Tahmin sonucunda elde edilen ortalama mutlak hata oranı 0.1593’dur. Bu değer 0’a yakın olduğu için tahmin edilen modelde başarı elde edilmiştir.

**Root Mean Squared Error**

Tahmin sonucunda elde edilen karekök ortalama hata oranı 0.3074’dur. Bu değer 0’a yakın olduğu için tahmin edilen modelde başarı elde edilmiştir.

**Relative Absolute Error**

Tahmin sonucunda gerçek değer ile hesaplanan değer arasındaki farkın gerçek değere oranlanması sonucunda 33.8503 değeri elde edilmiştir. Bu değer 0’a yakın olduğu için tahmin edilen modelde başarı elde edilmiştir.

**Root Relative Squared Error**

Tahmin sonucunda kök göreceli hata oranı 61.6616’dır. Bu değer 0’a çok yakın olmasa da diğer metriklerin 0’a yakın olmasından dolayı modelin başarı oranını çok etkilememiştir.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **TP Rate** | **FP Rate** | **Presicion** | **Recall** | **F-Measure** | **MCC** | **ROC Area** | **PRC Area** | **Class** |
|  | 0,739 | 0.007 | 0.985 | 0.739 | 0.844 | 0.782 | 0.866 | 0.831 | M |
|  | 0.993 | 0.261 | 0.853 | 0.993 | 0.917 | 0.782 | 0.866 | 0.851 | B |
| **Ağırlıklı Ortalama** | 0.892 | 0.161 | 0.905 | 0.892 | 0.888 | 0.782 | 0.866 | 0.843 |  |

**TP Rate**

Tahmin sonucunda sınıfı “M” olan verilerden kötü huylu olanı kötü huylu olarak tahmin oranı 0,739’dur. Sınıfı “B” olan verilerden iyi huylu olanı iyi huylu olarak tahmin oranı 0,993’dür. Bu değerler 1’e yakın olduğu için iyi bir isabet oranı elde edildiği görülmektedir. Yapılan karar ağacı sınıflandırma yöntemi sonucunda verilerin tamamına yakın bir kısmını doğru tahmin ettiği görülmektedir. Aynı zamanda “M” ve “B” sınıflarında en çok doğru tahmini yapan B sınıfıdır.

**FP Rate**

Tahmin sonucunda sınıfı “M” olan verilerden “M” sınıfını “B” sınıfı olarak tahmin etme oranı 0.007’dir. Sınıfı “B” olan verilerden “B” sınıfını “M” sınıfı olarak tahmin etme oranı 0,261’dir. Bu değerler 0’a yakın olduğu için yapılan hatalı tahminin çok az olduğu görülmektedir. Aynı zamanda “M” ve “B” sınıfları arasında en çok hata yapan “B” sınıfıdır.

**Precision**

Tahmin sonucunda “M” sınıfı hassasiyet oranı 0,985 iken “B” sınıfı hassasiyet oranı 0,853’dür. Hassasiyet oranları 1’e yakın olduğu için sınıflandırma modelinin doğru tahmin yaptığı görülmektedir. Sınıfı “M” olan verilerin hassasiyet oranı daha yüksektir.

**Recall**

Tahmin sonucunda “M” sınıfı geri çağırma oranı 0,739 iken “B” sınıfı geri çağırma oranı 0,993’dür. Örneğin gerçekte sonucun “B” olduğu durumda tahminin “M” olarak yapılmasıdır. Bu hata 0’a yaklaştıkça artar 1’e yaklaştıkça azalır. Bu durumda sınıflandırma modelinin doğru tahmin yaptığı görülmektedir. “B” sınıfına ait veriler daha doğru tahmin edilmiştir.

**F-Measure**

Tahmin sonucunda “M” sınıfı F-Measure oranı 0.844 iken “B” sınıfı F-Measure oranı 0.917 çıkmıştır. Bu değerin 1’e yakın olması yapılan tahmin modelinin başarılı olduğunu göstermektedir.

**MCC**

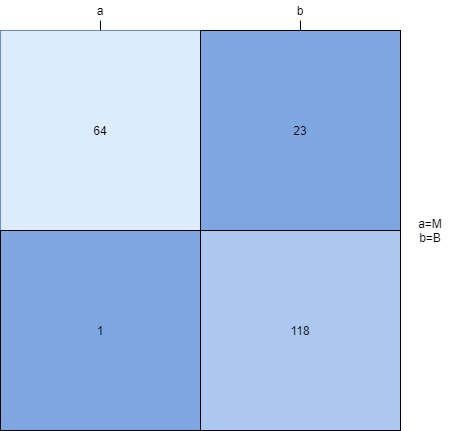
Tahmin sonucunda “M” ve “B” sınıfı MCC oranı 0,782 çıkmıştır. Bu değerin 1’e yakın olması yapılan tahmin modelinin başarılı olduğunu göstermektedir.

**ROC Area**

Tahmin sonucunda “M” ve “B” sınıfı ROC Area oranı 0.866 çıkmıştır. Bu değerin 1’e yakın olması yapılan tahmin modelinin başarılı olduğunu göstermektedir.

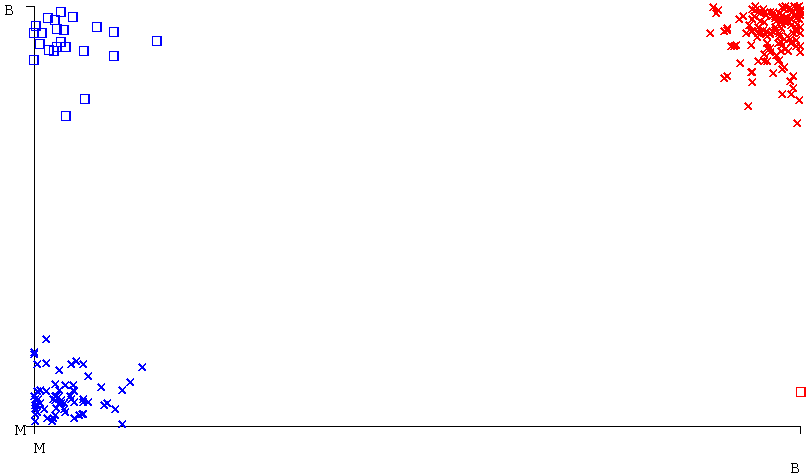
**PRC Area**

Tahmin sonucunda “M” sınıfı PRC Area oranı 0,831 iken “B” sınıfı oranı 0,851 çıkmıştır. Bu değerin 1’e yakın olması yapılan tahmin modelinin başarılı olduğunu göstermektedir. Aynı zamanda “B” sınıfının PRC Area oranı kadın sınıfından daha başarılıdır.

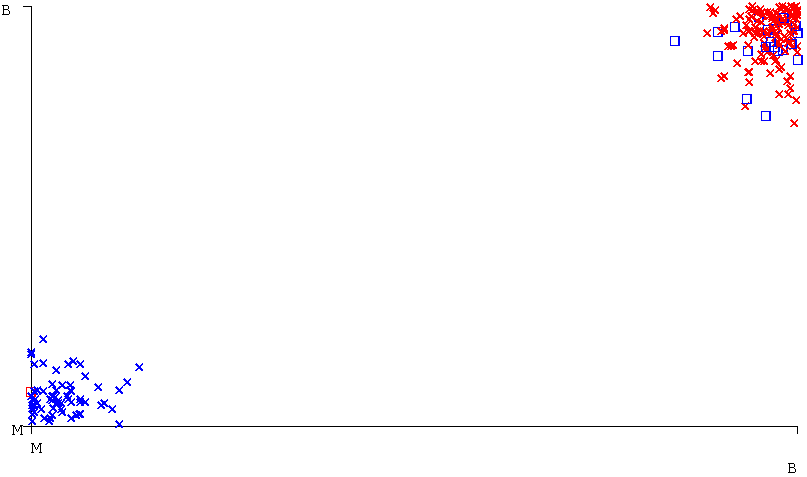


Confusion matrix kısmında oluşturduğumuz test ve eğitim kümeleri sonucunda label parametremizde bulunan “M” ve “B” tahminlerindeki başarımızı görmekteyiz. Confusion matrixde male ifadesini a olarak, female ifadesini b olarak ele almaktayız. A olarak ele aldığımız “M” olan 88 verinin 65 tanesini “M” olarak doğru tahmin ederken 23 tanesini “B” olarak yanlış tahmin etmiştir. B olarak ele aldığımız “B” olan 134 verinin 133 tanesini “B” olarak doğru tahmin ederken 1 tanesini “M” olarak yanlış tahmin etmiştir. Tüm veri setine baktığımız zamanda 222 veriden 198 veriyi doğru tahmin ederken 24 veriyi yanlış tahmin etmiştir.

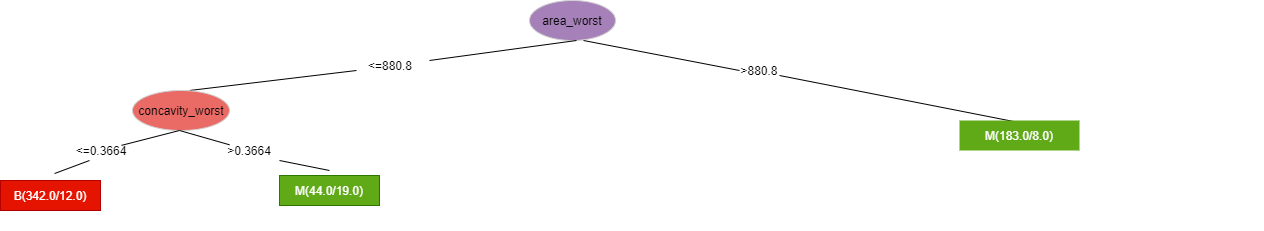
**Yanlış Tahmin Edilen Verilerin Grafik Olarak Gösterimi**



Yukarıdaki grafikte yapılan yanlış ve doğru tahminlerin birleşimi gösterilmiştir. Sınıflandırıcı hata grafiğinde “B” sınıfı kırmızı, “M” sınıfı mavi renktedir. Grafikte ise “**x”** ile gösterilenler doğru sınıflandırılmış verileri “**kare”** şeklinde gösterilen yerler ise yanlış sınıflandırılmış verileri ifade ediyor. Yukarıdaki grafikte yapılan yanlış ve doğru tahminlerin birleşimi gösterilmiştir.

Örneğin **“M”** kısmında bulunan **mavi** bir kare bu değerin “B” sınıfına ait olduğunu ama yanlış şekilde sınıflandırılarak “M” sınıfına dâhil edildiğini gösteriyor. Tahmin modelimizin sınıflandırıcı hata grafiğine baktığımızda 23“M” 1 “B” olmak üzere 24 hata yapıldığı görülmektedir.

ÖRNEK 3



Veri setimizin içerisinden diagnosis, radius\_mean, texture\_mean, perimeter\_mean, area\_mean, smoothness\_mean, concavity\_mean, symmetry\_mean, Radius\_se, texture\_se, perimeter\_se, area\_se, smoothness\_se, concavity\_se, symmetry\_se, Radius\_worst,texture\_worst, perimeter\_worst,, area\_worst, smothness\_worst, concavity\_worst, symmetry\_worst parametrelerini temel alarak ilerledik. Karar Ağacı algoritmasından makineye veri setinin sonucu bildiğimiz değerlerin %80’ini eğitim vererek %20 sini tahmin etmesini istedik. Diagnosis parametresi üzerinden de sınıflandırmasını yaptık. Yaprak başına düşen minimum obje (MinNumObj) sayısı 25 (varsayılan değer) olarak belirlenmiştir. Karar ağacını incelediğimizde dallanmanın ilk olarak “area\_wost” parametresinden başlandığı görülmektedir. İlk koşulun “area\_wost” parametresi olma sebebi entropi değerinin yüksek olmasıdır. Modellediğimiz karar ağacı 3 yaprak, 5 daldan oluşmaktadır

SONUÇ 3

|  |  |
| --- | --- |
| **Başarı Oranı** | **88.5965%** |
| Doğru Sınıflandırılmış Örnekler | 101 |
| Kappa İstatistiği | 0.7714 |
| Ortalama Mutlak Hata | 0.1258 |
| Kök Ortalama Kare Hatası | 0.2513 |
| Göreli Mutlak Hata | 26.1951 % |
| Kök Göreli Kare Hatası | 50.2481 % |
| Toplam Örnek Sayısı | 114 |

Modeli oluşturmak için geçen süre: 0,02 saniyedir.

**Correctly Classified Instances**

Veri setinin %20’si test kümesi olarak ele alındığı için 569 satır veriden 114’ü üzerinde tahmin yapılmıştır. Bunun sonucunda 114 adet veriden 101’ini doğru tahmin ederek %89.8876oranında başarı elde edilmiştir. Modelimiz tahmin yaparken 13 adet veriyi yanlış bulmuştur.

**Kappa Statistic**

İyi huylu ve kötü huylu sonuçlar arasındaki karşılaştırmalı uyuşmanın güvenilirlik oranı 0.7714’ dür. Elde edilen değer 0.61-0. 80 aralığında olduğu için iyi huylu ve kötü huylu değerleri arasında önemli derecede uyuşma olduğu görülmektedir.

**Mean Absolute Error**

Tahmin sonucunda elde edilen ortalama mutlak hata oranı 0.1258’dir. Bu değer 0’a yakın olduğu için tahmin edilen modelde başarı elde edilmiştir.

**Root Mean Squared Error**

Tahmin sonucunda elde edilen karekök ortalama hata oranı 0.2513’dür. Bu değer 0’a yakın olduğu için tahmin edilen modelde başarı elde edilmiştir.

**Relative Absolute Error**

Tahmin sonucunda gerçek değer ile hesaplanan değer arasındaki farkın gerçek değere oranlanması sonucunda 26.1951değeri elde edilmiştir. Bu değer 0’a yakın olduğu için tahmin edilen modelde başarı elde edilmiştir.

**Root Relative Squared Error**

Tahmin sonucunda kök göreceli hata oranı 50.2481’dir. Bu değer 0’a çok yakın olmasa da diğer metriklerin 0’a yakın olmasından dolayı modelin başarı oranını çok etkilememiştir.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **TP Rate** | **FP Rate** | **Presicion** | **Recall** | **F-Measure** | **MCC** | **ROC Area** | **PRC Area** | **Class** |
|  | 0,939 | 0.154 | 0.821 | 0.939 | 0.876 | 0.777 | 0.954 | 0.940 | M |
|  | 0.846 | 0.061 | 0.948 | 0.846 | 0.894 | 0.777 | 0.954 | 0.936 | B |
| **Ağırlıklı Ortalama** | 0.886 | 0.101 | 0.894 | 0.886 | 0.887 | 0.777 | 0.954 | 0.937 |  |

**TP Rate**

Tahmin sonucunda sınıfı “M” olan verilerden kötü huylu olanı kötü huylu olarak tahmin oranı 0,939’dur. Sınıfı “B” olan verilerden iyi huylu olanı iyi huylu olarak tahmin oranı 0,846’dır. Bu değerler 1’e yakın olduğu için iyi bir isabet oranı elde edildiği görülmektedir. Yapılan karar ağacı sınıflandırma yöntemi sonucunda verilerin tamamına yakın bir kısmını doğru tahmin ettiği görülmektedir. Aynı zamanda “M” ve “B” sınıflarında en çok doğru tahmini yapan B sınıfıdır.

**FP Rate**

Tahmin sonucunda sınıfı “M” olan verilerden “M” sınıfını “B” sınıfı olarak tahmin etme oranı 0.154’ dür. Sınıfı “B” olan verilerden “B” sınıfını “M” sınıfı olarak tahmin etme oranı 0,161’dir. Bu değerler 0’a yakın olduğu için yapılan hatalı tahminin çok az olduğu görülmektedir. Aynı zamanda “M” ve “B” sınıfları arasında en çok hata yapan “B” sınıfıdır.

**Precision**

Tahmin sonucunda “M” sınıfı hassasiyet oranı 0,987 iken “B” sınıfı hassasiyet oranı 0,862’dir. Hassasiyet oranları 1’e yakın olduğu için sınıflandırma modelinin doğru tahmin yaptığı görülmektedir. Sınıfı “M” olan verilerin hassasiyet oranı daha yüksektir.

**Recall**

Tahmin sonucunda “M” sınıfı geri çağırma oranı 0,939 iken “B” sınıfı geri çağırma oranı 0,846’ dır. Örneğin gerçekte sonucun “B” olduğu durumda tahminin “M” olarak yapılmasıdır. Bu hata 0’a yaklaştıkça artar 1’e yaklaştıkça azalır. Bu durumda sınıflandırma modelinin doğru tahmin yaptığı görülmektedir. “B” sınıfına ait veriler daha doğru tahmin edilmiştir.

**F-Measure**

Tahmin sonucunda “M” sınıfı F-Measure oranı 0.876 iken “B” sınıfı F-Measure oranı 0.894 çıkmıştır. Bu değerin 1’e yakın olması yapılan tahmin modelinin başarılı olduğunu göstermektedir.

**MCC**

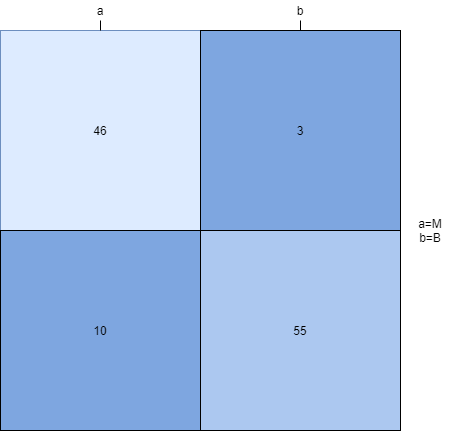
Tahmin sonucunda “M” ve “B” sınıfı MCC oranı 0,777 çıkmıştır. Bu değerin 1’e yakın olması yapılan tahmin modelinin başarılı olduğunu göstermektedir.

**ROC Area**

Tahmin sonucunda “M” ve “B” sınıfı ROC Area oranı 0.954 çıkmıştır. Bu değerin 1’e yakın olması yapılan tahmin modelinin başarılı olduğunu göstermektedir.

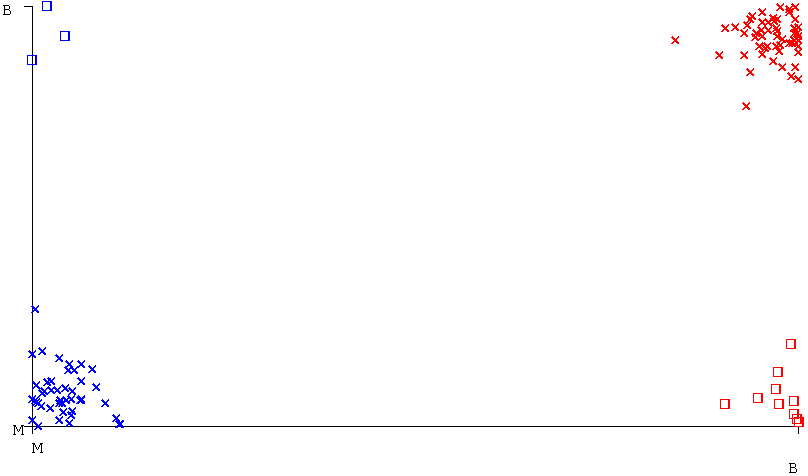
**PRC Area**

Tahmin sonucunda “M” sınıfı PRC Area oranı 0,940 iken “B” sınıfı oranı 0,936 çıkmıştır. Bu değerin 1’e yakın olması yapılan tahmin modelinin başarılı olduğunu göstermektedir. Aynı zamanda “B” sınıfının PRC Area oranı kadın sınıfından daha başarılıdır.

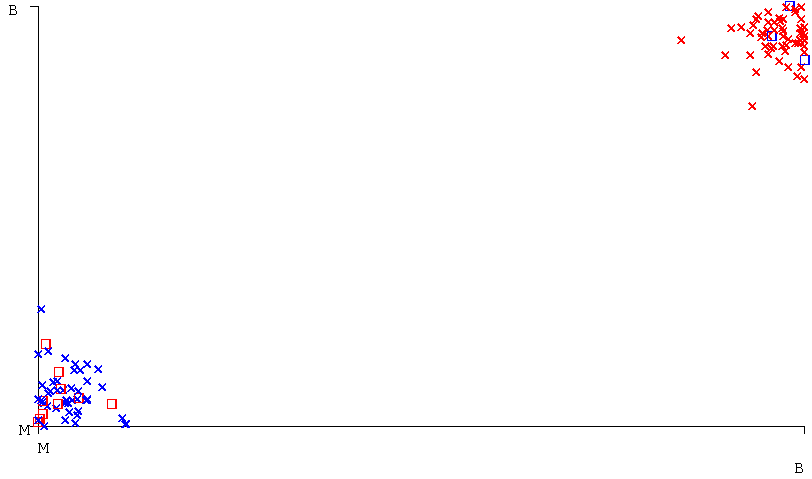


Confusion matrix kısmında oluşturduğumuz test ve eğitim kümeleri sonucunda label parametremizde bulunan “M” ve “B” tahminlerindeki başarımızı görmekteyiz. Confusion matrixde male ifadesini a olarak, female ifadesini b olarak ele almaktayız. A olarak ele aldığımız “M” olan 49 verinin 46 tanesini “M” olarak doğru tahmin ederken 3 tanesini “B” olarak yanlış tahmin etmiştir. B olarak ele aldığımız “B” olan 65 verinin 55 tanesini “B” olarak doğru tahmin ederken 10 tanesini “M” olarak yanlış tahmin etmiştir. Tüm veri setine baktığımız zamanda 114veriden 101 veriyi doğru tahmin ederken 13 veriyi yanlış tahmin etmiştir.

**Yanlış Tahmin Edilen Verilerin Grafik Olarak Gösterimi**



Yukarıdaki grafikte yapılan yanlış ve doğru tahminlerin birleşimi gösterilmiştir. Sınıflandırıcı hata grafiğinde “B” sınıfı kırmızı, “M” sınıfı mavi renktedir. Grafikte ise “**x”** ile gösterilenler doğru sınıflandırılmış verileri “**kare”** şeklinde gösterilen yerler ise yanlış sınıflandırılmış verileri ifade ediyor. Yukarıdaki grafikte yapılan yanlış ve doğru tahminlerin birleşimi gösterilmiştir.



Örneğin **“M”** kısmında bulunan **mavi** bir kare bu değerin “B” sınıfına ait olduğunu ama yanlış şekilde sınıflandırılarak “M” sınıfına dâhil edildiğini gösteriyor. Tahmin modelimizin sınıflandırıcı hata grafiğine baktığımızda 3 “M” 10 “B” olmak üzere 13 hata yapıldığı görülmektedir.

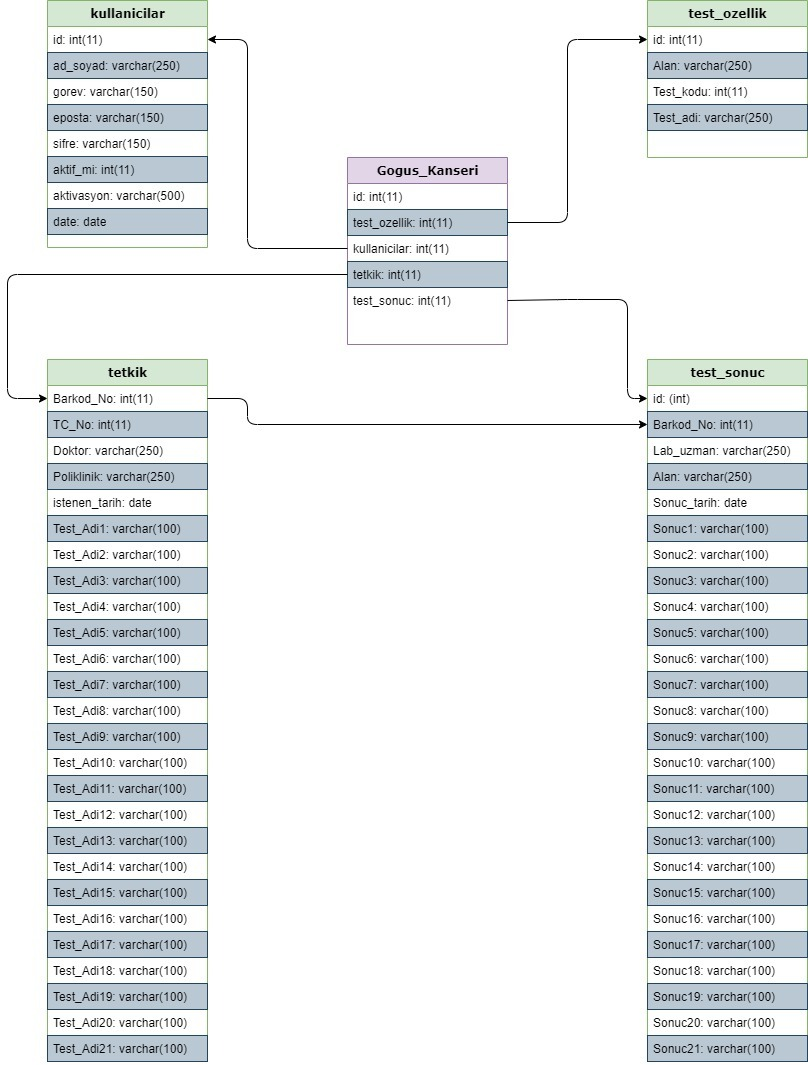
KARAR AĞACI SINIFLANDIRMA YÖNTEMİ SONUÇLARIN KARŞILAŞTIRILMASI

|  |  |  |  |
| --- | --- | --- | --- |
|  | *SONUÇ 1* | *SONUÇ 2* | *SONUÇ 3* |
| Eğitim ve Test Oranı | *%65-%35* | *%61-%39* | *%80-%20* |
| Doğru Tahmin Sayısı | 190 | 198 | 101 |
| Yanlış Tahmin Sayısı | 9 | 24 | 13 |
| Toplam Veri Sayısı | 199 | 222 | 114 |
| Yaprak ve Dal Sayısı | 12-23 | 4-7 | 3-5 |
| Model Oluşturmak İçin Harcanan Zaman | 0.00 saniye | 0.01 saniye | 0.02 Saniye |
| Yaprak başına düşen minimum obje sayısı | 2 | 20 | 25 |
| Başarı Oranı | %95.4774 | %89.1892 | %88.5965 |

Veri seti karar ağacı sınıflandırma algoritması ile 3 farklı test ve eğitim seti ile incelenmiştir. Örnek1’de %65 eğitim seti vererek %35’lik tahminde minimum obje sayısı 2 alınarak %95.4774’lük başarı elde edilmiştir. Örnek2’de %61 eğitim seti vererek %39’luk tahminde minimum obje sayısı 20 alınarak %89.1892’lik başarı elde edilmiştir. Örnek3 ’de %80 eğitim seti vererek %20’lik tahminde minimum obje sayısı 25 alınarak %88.5965’lik başarı elde edilmiştir. En çok başarı Örnek 1 de elde edildiği için uygulamanın yapımında Örnek1 sonucundaki karar ağacı algoritması temel alınmıştır.

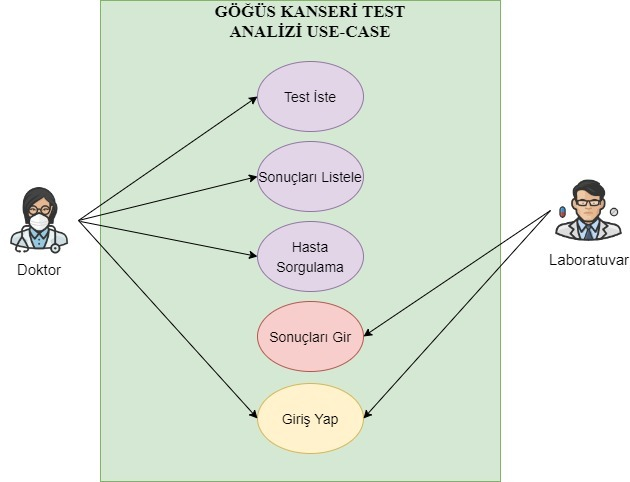
**3.2.4 İSTERLERİN TANIMLANMASI**

3.2.4.1 VERİTABANI



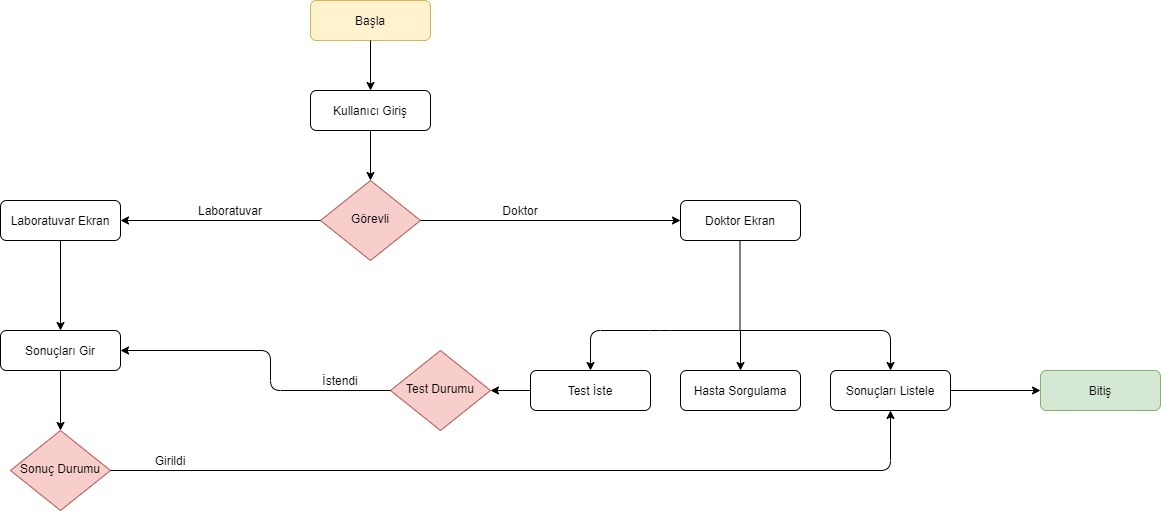
Veritabanında 4 adet tablo vardır. Kullanıcılar tablosunda laboratuvar ve doktora ait kullanıcı ve şifre bilgileri bulunmaktadır. Giriş sırasında girilen bilgiler veritabanındaki bilgiler ile eşleşiyorsa giriş işlemi gerçekleşir eğer eşleşmiyorsa sistem uyarı verir. Test özellik tablosunda test bilgileri yer almaktadır. Bu bilgileri doktorun kullandığı test isteme sayfasındaki comboboxlara çekilir. Tetkik tablosunda hasta, doktor, poliklinik, test ve tarih bilgileri yer almaktadır. Doktor test isteme sayfasından test talebinde bulunduğunda sayfadaki veriler tetkik tablosuna kaydolur. Bu tetkikler barkod numarası ile laboratuvarın erişim sağladığı sonuç girme sayfasına çekilir. Aynı zamanda doktorun kullandığı hasta sorgulama ekranına TC kimlik numarası ile veri çeker. Test sonucu tablosunda laboratuvarın sonuçları girdiği sayfadaki verileri kaydeder. Bu tabloda kaydedilen veriler barkod numarası ile doktorun kullandığı sonuç listeleme sayfasına çekilir. Tetkik ve Test Sonuç tabloları Barkod\_No ile ilişkisel olarak birbirine bağlıdır.

3.2.4.2 USE CASE



Laboratuvar ve doktorun aynı yaptığı tek işlem giriş yapmasıdır. Doktor sistemde test isteme, sonuçları listeleme ve hasta sorgulama işlemlerini yapabilmektedir. Laboratuvar sistemde sadece sonuçları girme işlemini gerçekleştirmektedir.

3.2.4.3 AKIŞ DİYAGRAMI



Kullanıcı ilk olarak sisteme giriş yapar. Kullanıcı Doktor ise test isteme, hasta sorgulama ve sonuçları listeleme menülerini görüntüleyebilir. Kullanıcı Laboratuvar ise sonuçları girme menüsünü görüntüleyebilir. Sistemde ilk olarak doktor laboratuvardan test ister. Her teste ait bir barkod numarası vardır sistem barkod numarasını üretir. Üretilen barkod numarası hastaya verildiği varsayılır. Test istendikten sonra laboratuvar istenen testleri barkod numarası ile görüntüler. Laboratuvar görüntülediği testleri elde ettiği değerlere göre doldurur ve onaylar. Sonuçlar karar yapısında analiz edilir. Elde edilen sonuçlar doktorun ekranına gönderilir. Doktor sonuçları barkod numarası ile sorgular ve test sonucuna erişir.

3.2.5. UYGULAMANIN ANLATIMI

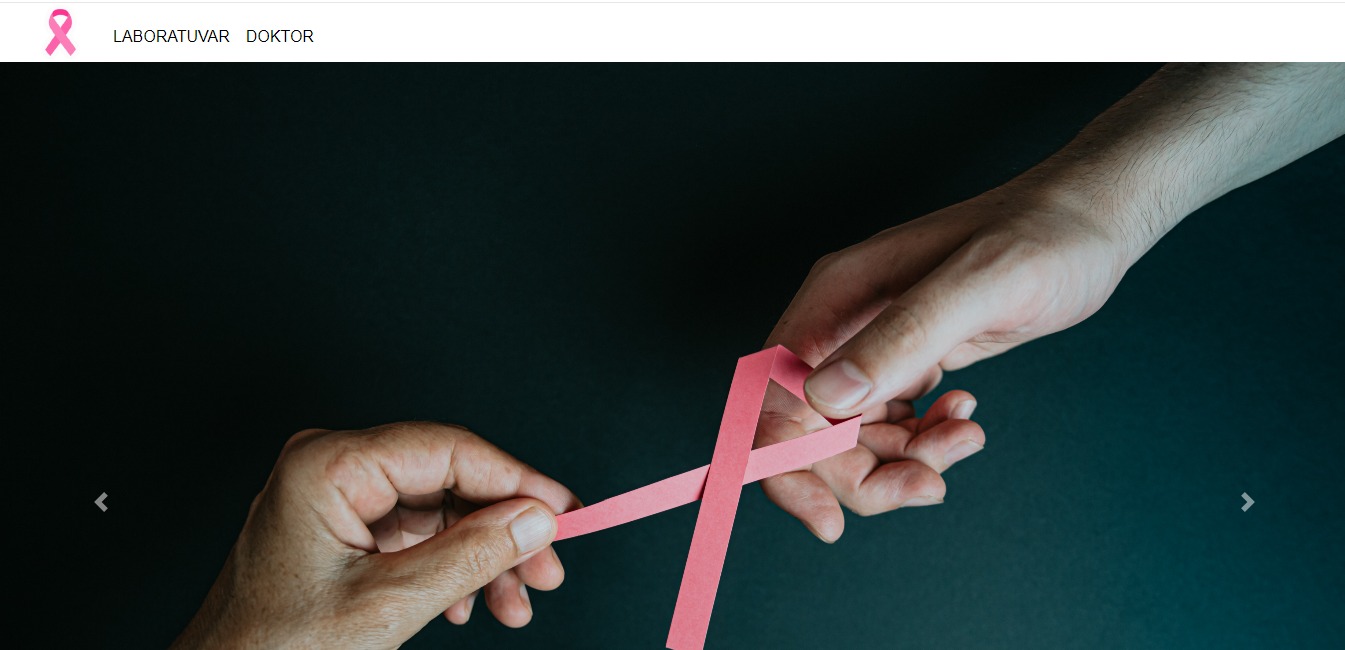
Şekil 1’ de üst kısımda “Laboratuvar” ve “Doktor” isimli ayrı iki sayfanın bulunduğu bir navbar vardır. Aynı zamanda logoya tıklanıldığı zaman kullanıcı tekrar ana sayfaya yönlendirilmektedir.

Şekil1: Ana Ekran

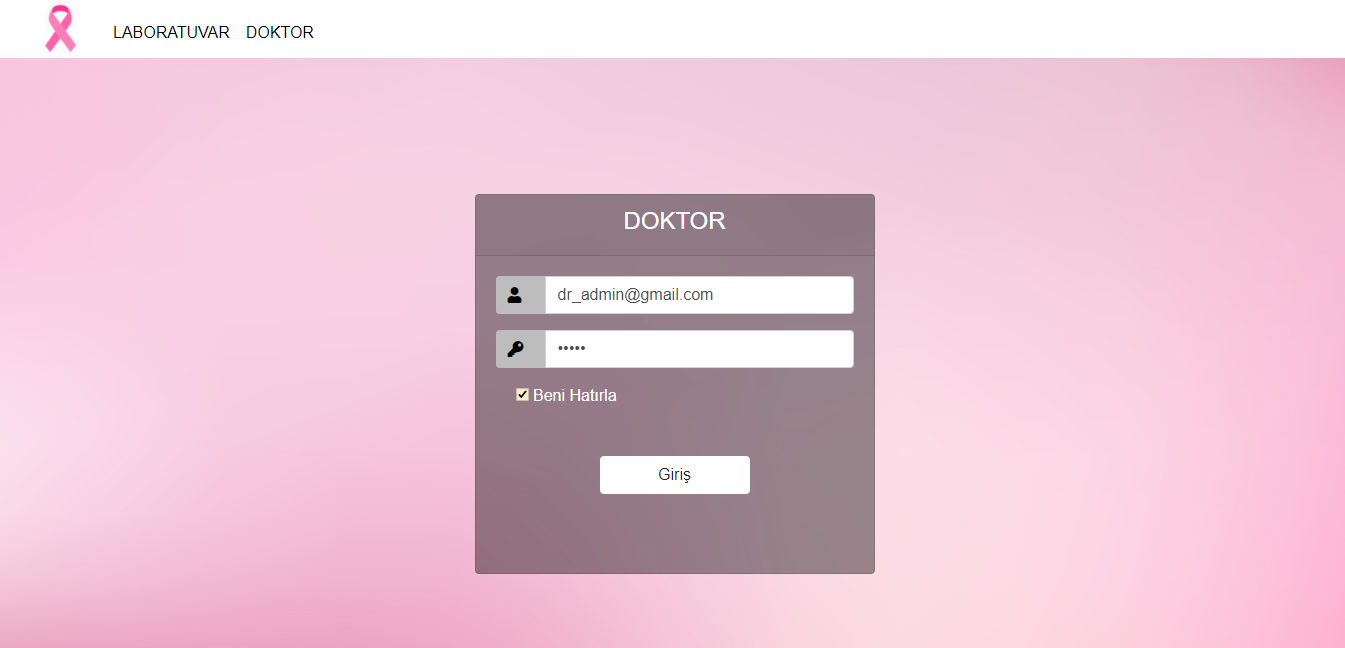


Şekil 1’ de ve Şekil2’de üst kısımda “Laboratuvar” ve “Doktor” isimli ayrı iki sayfanın bulunduğu bir navbar vardır. Aynı zamanda logoya tıklanıldığı zaman kullanıcı tekrar ana sayfaya yönlendirilmektedir.

Şekil2: Ana Ekran 2

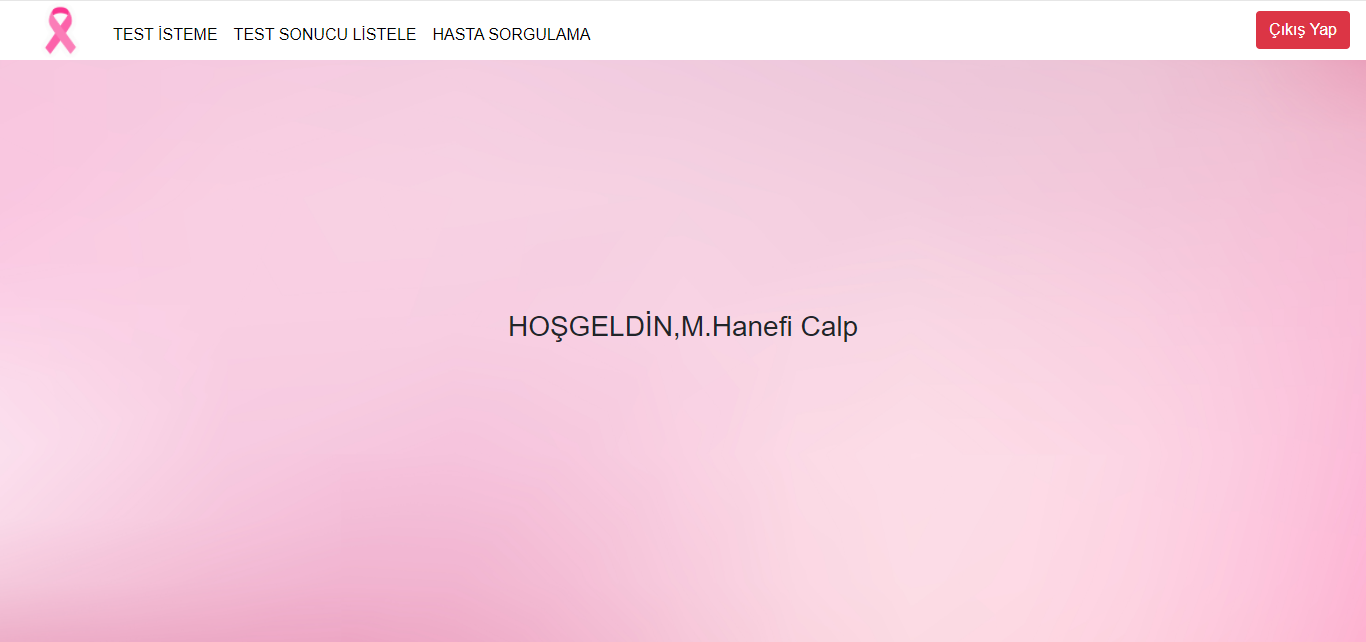


Şekil3: Doktor Giriş Ekranı



Navbar kısmında bulunan “Doktor” butonuna tıklanıldığı zaman Şekil3’ de yer alan ekran açılmaktadır. Doktor kendi e-posta adresi ve şifresini girerek sisteme giriş yapacaktır. Daha sonraki girişlerde e-posta adresi ve şifresinin hatırlanması için “Beni Hatırla” butonuna tıklayabilir. Doktorun e posta adresi [dr\_admin@gmail.com](mailto:dr_admin@gmail.com) şifresi de “dr123” olarak veri tabanına eklenmiştir.

Şekil4: Doktor Giriş Ana Ekran



E-Posta ve şifre yazılarak giriş yapıldıktan sonra doktor giriş ana ekranına yönlendirilir. Navbarda sadece doktorun erişebileceği “Test İsteme”, “Test Sonucu Listeleme”, “Hasta Sorgulama” butonları bulunmaktadır. Doktor gelen hastadan gerekli testleri istemek için “Test İsteme” butonuna tıklamalıdır.

Şekil5: Test İsteme Ekranı



Doktor “Hasta Bilgileri” kısmına hastanın tc kimlik numarasını, “Doktor Bilgileri” kısmına kendi adını, soyadını, polikliniği, hastadan test istediği tarihi, “Test Bilgileri” kısmına meme kanserinin iyi huylu veya kötü huylu olup olmadığını anlayabilmek için gerekli olan tüm testleri seçerek ekler. Seçilecek olan testlerde en önemli kısım testler sırasıyla seçilmeli ve boş bırakılmamalıdır. Aksi taktirde sonuç yanlış çıkabilir. Tüm bunları tamamladıktan sonra “Kaydet” butonuna basılır.

Şekil6: Hasta Sorgulama Ekranı



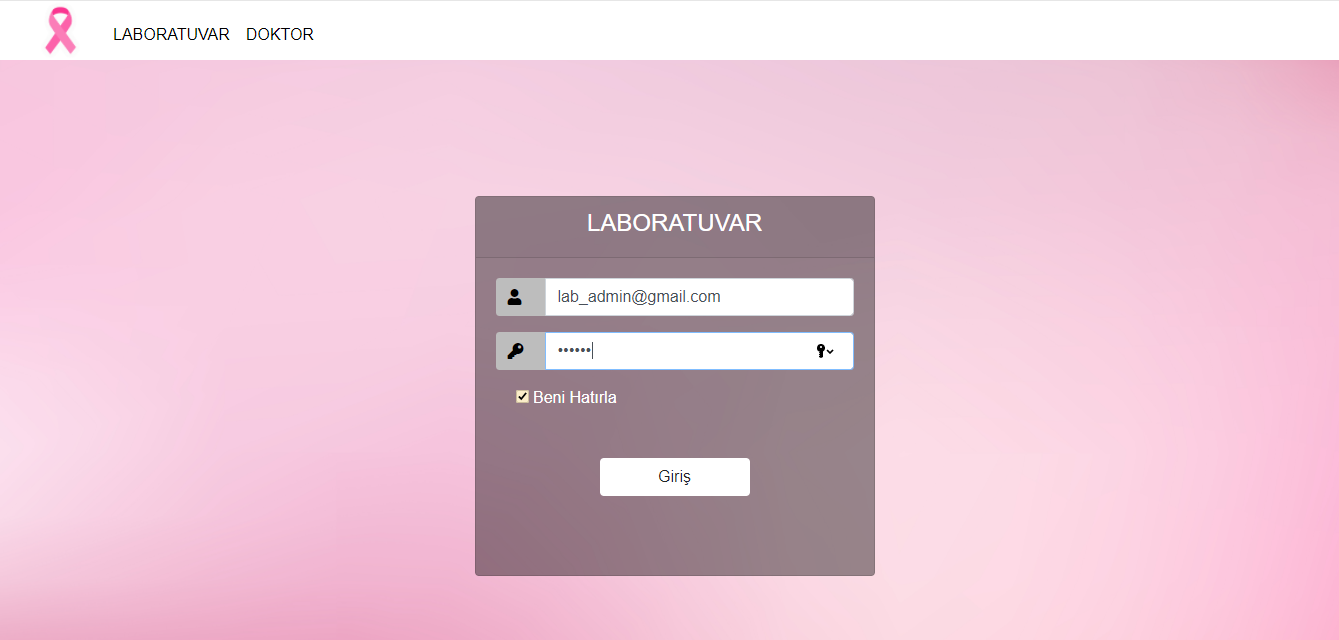
Doktor navbarda bulunan “Hasta Sorgulama” butonuna tıkladığında Şekil6’ da ki ekran açılacaktır.

Şekil7: Hasta Sorgulama Ekranı 2



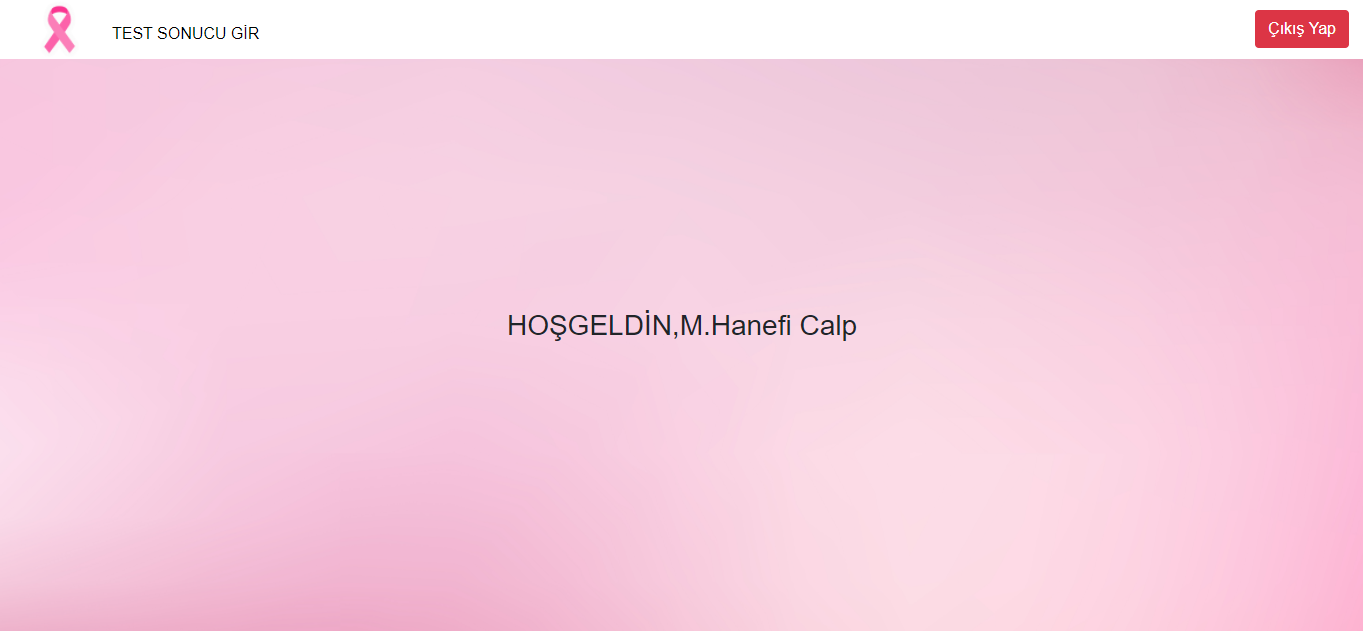
Doktor “Hasta Sorgulama” butonuna tıklayarak “Arama” kısmından hastanın tc kimlik numarasını aratarak hastanın “Barkod No, Tc Kimlik No, Poliklinik, İlgili Doktor, ve son olarak İstenen Tarih” bilgilerini görüntüleyebilir. Sistemde yer alan barkod numarası sonuç aratırken gerekli olacaktır ve barkod numarasını kaybeden hastaların tc kimlik numarasıyla barkod numarasına ulaşılacaktır. Doktor sistemden çıkış yapmak için navbarın sağ köşesinde bulunan “Çıkış Yap” butonuna tıklamalıdır.

Şekil8: Laboratuvar Görevlisi Giriş Ekranı

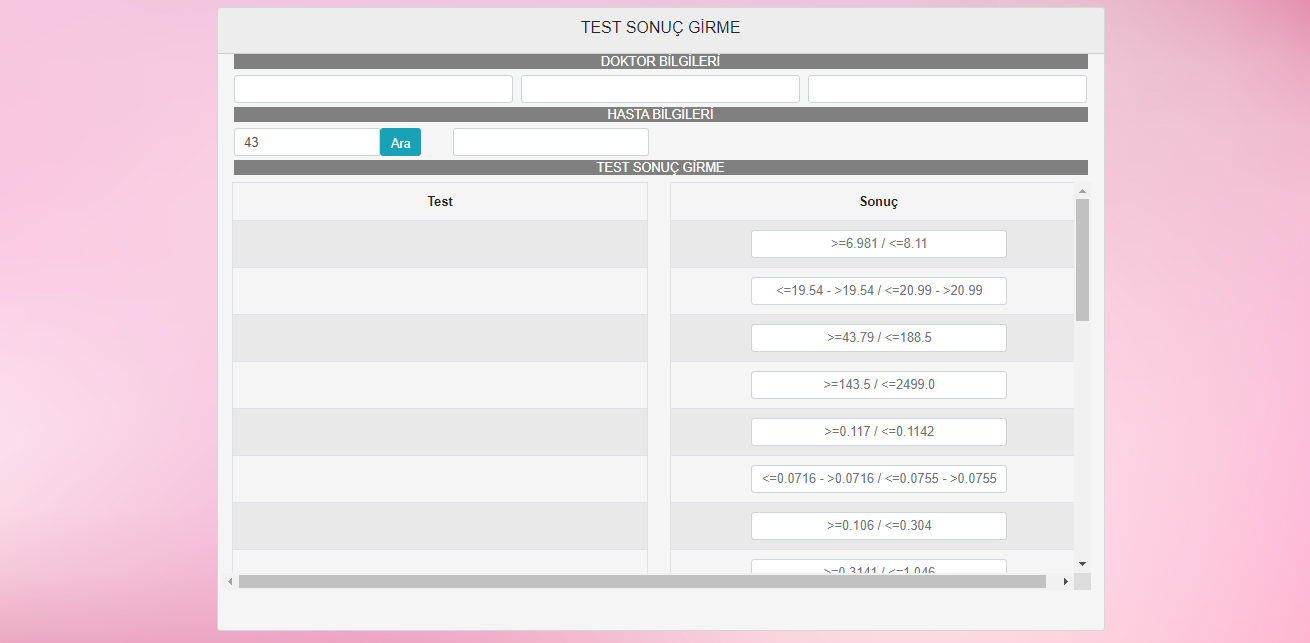


Navbar kısmında bulunan “Laboratuvar” butonuna tıklanıldığı zaman Şekil8’ de yer alan ekran açılmaktadır. Laboratuvar görevlisi kendi mail adresi ve şifresini girerek sisteme giriş yapacaktır. Daha sonraki girişlerde mail adresi ve şifresinin hatırlanması için “Beni Hatırla” butonuna tıklayabilir. Laboratuvar görevlisinin e –posta adresi [lab\_admin@gmail.com](mailto:lab_admin@gmail.com), şifresi de “lab123” olarak veri tabanına eklenmiştir.

Şekil9: Laboratuvar Görevlisi Giriş Ana Ekran

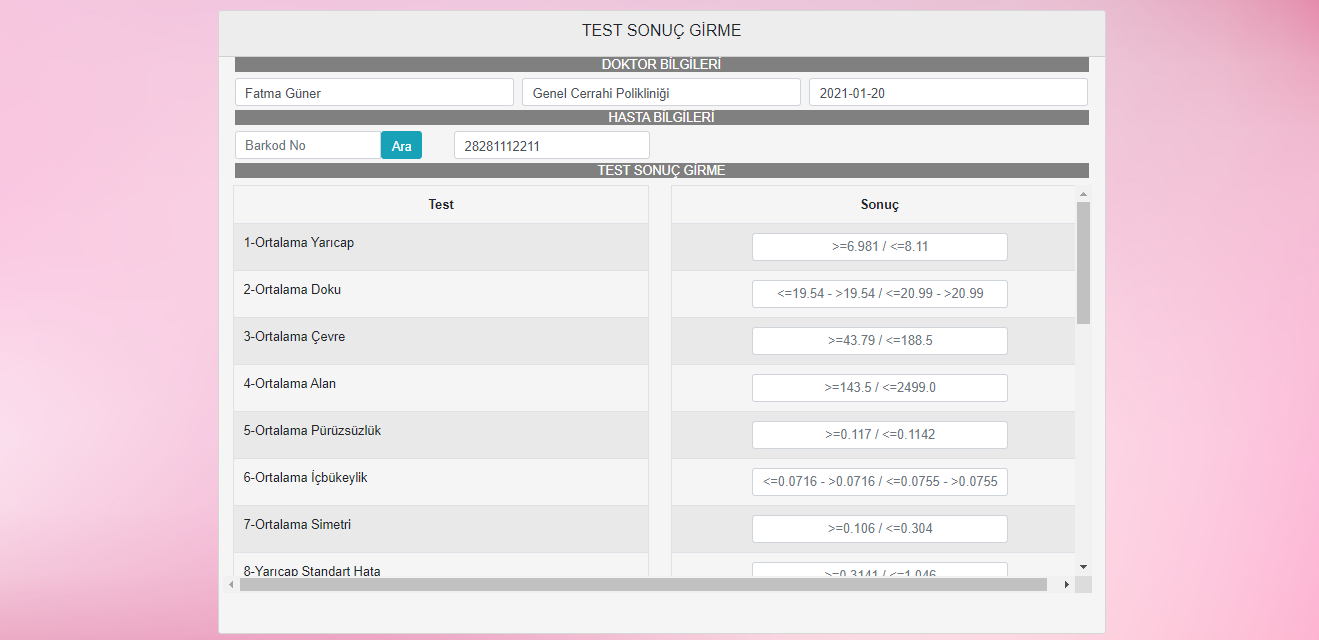


E-Posta ve şifre yazılarak giriş yapıldıktan sonra laboratuvar giriş ana ekranına yönlendirilir. Navbarda sadece laboratuvar görevlisinin erişebileceği “Test Sonucu Gir” butonu bulunmaktadır. Laboratuvar görevlisi, doktorların hastalardan istemiş olduğu testlerin sonuçlarını sisteme girecektir.

Şekil10: Test Sonucu Girme Ekranı

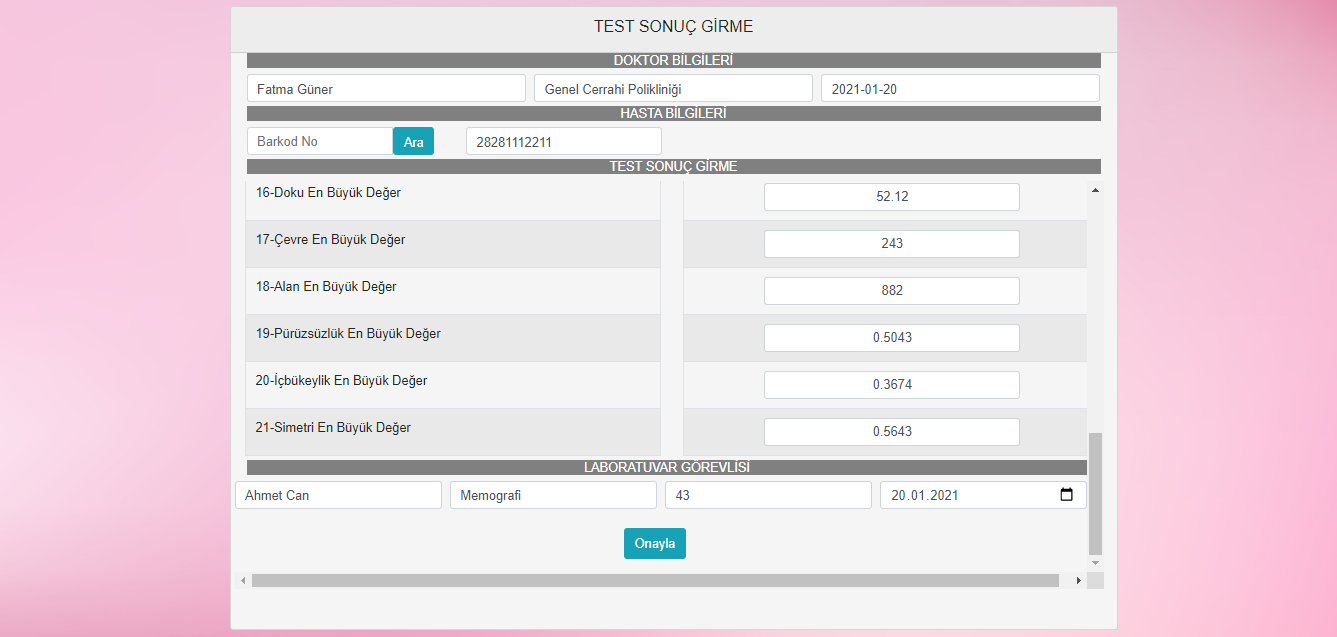
Laboratuvar görevlisi hastanın barkod numarasını sistem üzerinden aratarak, doktor bilgilerine, hastanın tc kimlik numarasına ve doktorun hastadan istemiş olduğu testlere ulaşabilir.

Şekil11: Test Sonucu Girme Ekranı 2



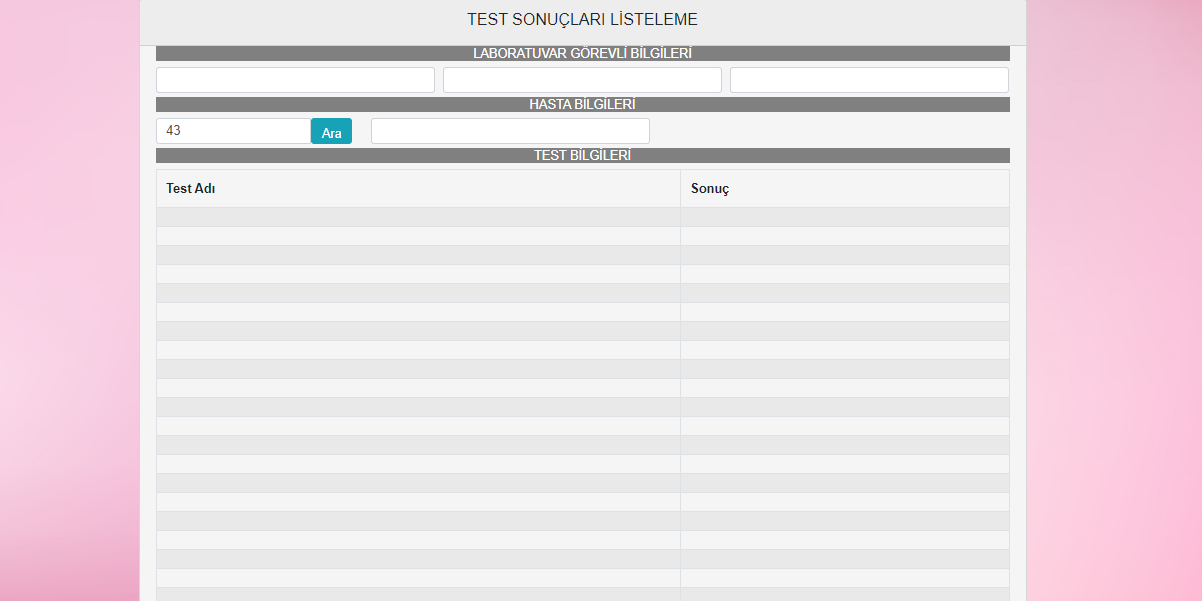
Barkod numarası aratıldıktan sonra, hastanın doktorunun adı, hangi poliklinikten olduğu, testleri istediği tarih, hastanın tc kimlik numarası ve doktorun hastadan istemiş olduğu testler gözükür. Görevli kişi testlerin sonuçlarını doğru bir şekilde “Sonuç” kısmına yazmalıdır. Hangi testtin sonucu hangi aralıklar içinde olması gerektiği her testin karşısında olan sonuç kısmında yazmaktadır.

Şekil12: Test Sonucu Girme Ekranı 3



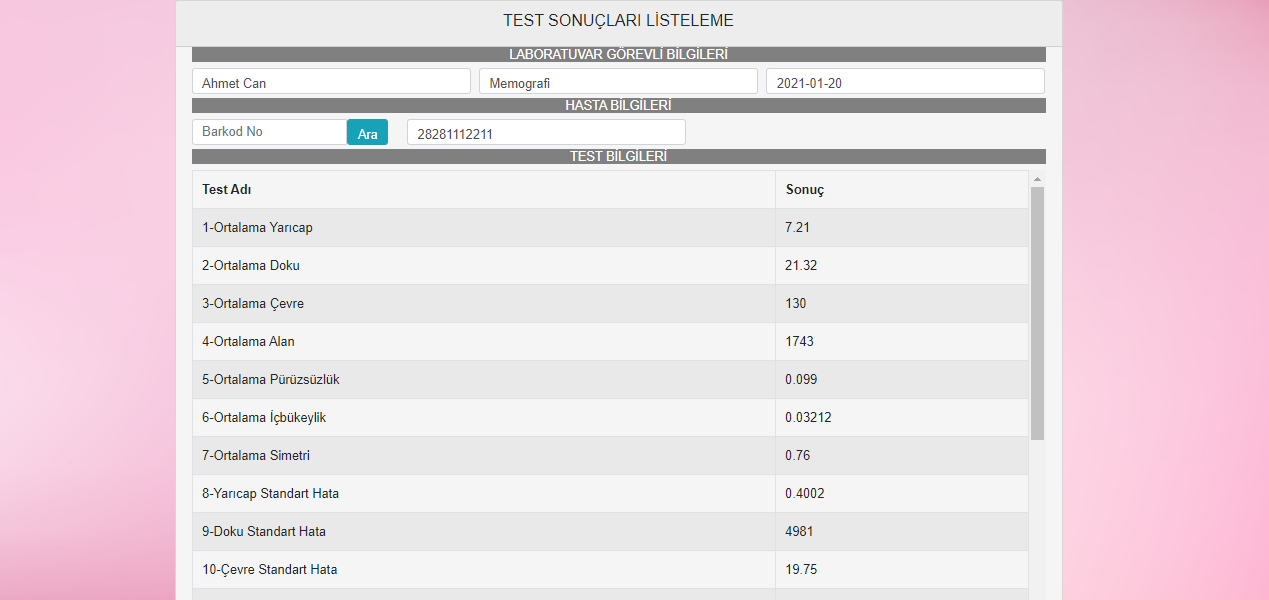
Laboratuvar görevlisi testlerin sonuçlarını gerekli yerlere girdikten sonra en alt kısımda bulunan “Laboratuvar Görevlisi” kısmından adını soyadını, laboratuvar alanını, hastanın barkod numarasını ve sonuçların girildiği tarihi yazıp onaylaması gerekmektedir.

Şekil13: Test Sonuçları Listeleme Ekranı



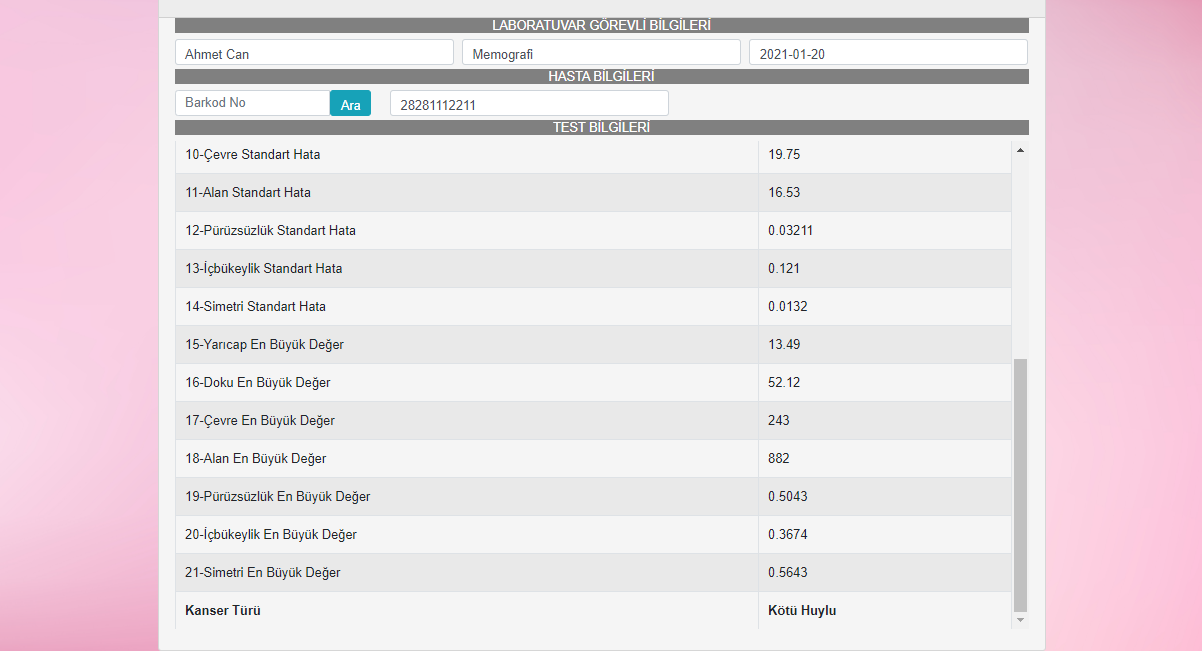
Doktor e-postasını ve şifresini girdikten sonra Şekil4’ deki ekrana yönlendirilecektir. Bu ekranda navbar kısmında buluna “Test Sonuçları Listele” butonuna tıkladığında Şekil13’ de bulunan ekran açılacaktır. Burada doktor “Hasta Bilgileri” kısmında yer alan “Barkod No” kısmına hastanın barkod numarasını yazıp “Ara” butonuna basarak sistem üzerinden aratmalıdır.

Şekil14: Test Sonuçları Listeleme Ekranı 2



Barkod numarası sayesinde doktor “Laboratuvar Görevli Bilgileri” kısmından test sonuçlarını giren kişinin adına, soyadına, laboratuvar alanına ve sonuçları girdiği tarihi görebilecektir. Aynı zamanda “Hasta Bilgileri” kısmından hastanın tc kimlik numarasına ve son olarak “Test Bilgileri” kısmından testlerin isimlerini ve sonuçlarını görebilecektir.

Şekil15: Test Sonuçları Listeleme Ekranı 3

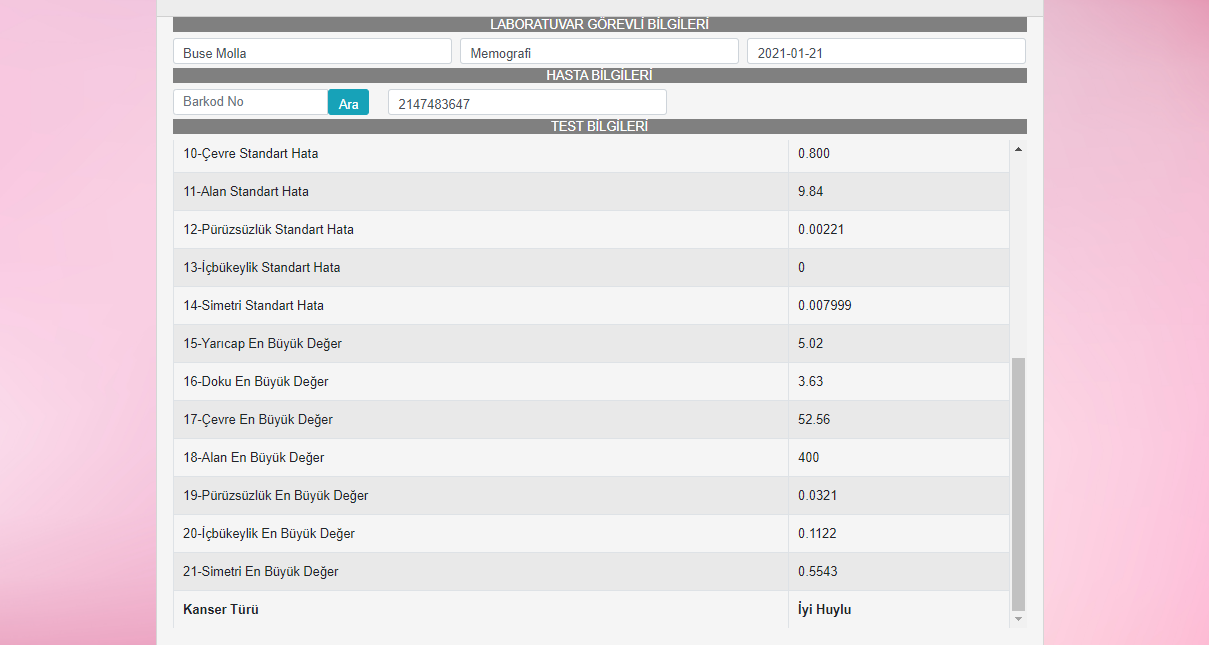


Tüm test sonuçlarının değerlerine göre kanser türünün ne olduğunu “Kanser Türü” kısmından görebilirsiniz. Buradaki değerler sonucunda “2828112211” tc numarasına sahip olan hastanın kanserinin “Kötü Huylu” olduğu ortaya çıkmıştır. Doktor bu değeri baz alarak tedavisini ona göre şekillendirecektir.

Şekil16: Test Sonuçları Listeleme Ekranı 4

Başka bir hastanın barkod numarası aratıldığında, hastanın test sonuçlarının farklılık göstermesi durumunda kanser türünün değiştiği görülmektedir.

Şekil16: Test Sonuçları Listeleme Ekranı 5



Elde edilen test sonuçlarına göre Tc kimlik numarası “2147483647” olan hastanın kanser türü “İyi Huylu” olarak belirtilmiştir

4. SONUÇ VE ÖNERİLER

Sağlık sektörü bilgiye dönüşmeyi bekleyen çok fazla miktarda veri barındırır. Sağlık kuruluşlarının, karar verme aşamasında bu verileri daha bilinçli şekilde kullanarak değerlendirmeye, daha etkin tedavi yöntemleri geliştirmeye, kaynakların daha iyi yönetilmesini sağlayarak maliyetleri düşürmeye ihtiyaçları vardır[27]. Ülkemizde sağlık hizmetleri hastane odaklı sunulmakta ve sağlık sektörüne ayrılan payın en büyük bölümünü hastanelere tüketmektedir. Kaynakların büyük bir bölümünü tüketmelerine rağmen sunulan hizmet toplum ihtiyacına cevap vermemekte ve hastane kapılarında uzun hasta bekleyişleri oluşmaktadır. Bu tür çalışmalar hem hastanın bekleyiş süresini kısaltmak hem de daha nitelikli bir sağlık hizmeti sunabilmek için çağın gerektirdiği teşhis ve tedavi sistemlerini, hastanelere kazandırmak için gayret göstermektedir. Bu sistemlerin kullanılması ile daha fazla sayıda hastaya, çabuk, kaliteli ve en doğru şekilde sağlık hizmeti sunulabilmektedir.

Meme kanseri, hücrelerde kötü huylu lezyon gelişimini göstermektedir. Akciğer kanseri sonrasında dünyadaki en yaygın kanser türüdür. Her 8 kadından birinin hayatlarının belirli bir zamanında göğüs kanseri geçirdiği bildirilmektedir. Erkeklerde görülmesine karşın, kadın vakaları erkek vakalardan 100 kat fazladır. Göğüs kanserine karşı en iyi önleyici yöntem erken teşhis yöntemidir. Bu nedenle erken tespiti artırmak için çeşitli yöntemler geliştirilmektedir. Bilgisayar destekli tanı sistemleri, klinik karar destek sistemleri doktorlara tıbbi görüntüleri yorumlamada yardımcı olmak için kullanılmaktadır. Bilgisayar destekli tanı sistemleri yardımcı bir teşhis sistemidir; nihai karar doktora aittir. Bu çalışmada, veritabanında yer alan parametler sonucunda otomatik olarak kanserin iyi huylu veya kötü huylu olduğunun ortaya çıkmasını sağlayan, derin öğrenme tabanlı bir karar destek sistemi geliştirilmiştir. Deneysel sonuçlar, önerilen sistemin etkililiğini ve gerçek zamanlı klinik uygulamalardaki kullanılabilirliğini göstermektedir.

Farklı sistemlerden gelen, tekrarlardan, hatalardan ve belirsizliklerden arındırılan veriler seçim, ön işleme, dönüştürme işlemlerinden geçirilerek veri ambarında tutulması bu amaçları gerçekleştirmenin ilk adımıdır. Bütünleşmiş bir şekilde veri ambarında bulunan bu veriler analiz, sorgulama, raporlama ve veri madenciliği için kullanılır. Veri ambarı ve veri madenciliği tekniklerinin karar destek amacıyla kullanılması sağlık alanında yeni bir yön olarak karşımıza çıkmıştır. Veri ambarları farklı sistemlerden gelen büyük miktarda ve çok boyutlu verileri saklayabildiği, karmaşık sorgulara yanıt verebildiği için karar verme sürecini kolaylaştırdığı gibi aynı zamanda verileri veri madenciliği algoritmaları içinde hazır hale getirmiş olurlar. Bu çalışmada meme kanseri ile ilgili veriler içeren bir veri seti elde edilmiş ve veri madenciliği tekniklerinin uygulanması için hazır duruma getirilmiştir. Tekniklerden Karar Ağacı tekniği uygulanmıştır. Veri madenciliği tekniği kullanılarak meme kanseri hastalığıyla ilgilenen doktorların tedavide daha çabuk karar vermelerini sağlayacak bir sistem geliştirmek amaçlanmıştır. Veri madenciliği sonucu elde edilen sonuçlar doktorların kanseri daha erken seviyelerde teşhis etme ya da daha etkin tedavi yollarının seçilmesi gibi kararları almalarında onlara yardımcı olabilmektedir.

Klinik çalışma, tıbbi tedavilerin insanların sağlığını iyileştirip iyileştiremeyeceğini öğrenmek için yapılan bir araştırma çalışmasıdır. Tıbbi tedavi bir ilaç, tıbbi cihaz veya tıbbi prosedür olabileceği gibi beslenme şekli veya egzersiz gibi kişinin hayatındaki bir değişiklik de olabilir[28]. Bu çalışmada bir nevi klinik çalışmadır ve %95.4774oranında başarı elde edilmiştir. Bu nedenle çalışmanın sağlık sektöründe meme kanseri araştırma ve sonuçlarında faydalı olacağı düşünülmektedir. Veri setini karar ağacı sınıflandırma algoritması ile değerlendirdik ama farklı sınıflandırma algoritmaları ile incelemelerde yapılabilir.

5. KAYNAKLAR

1. Akkoç L. (2009) Hastane Bilgi Yönetim Sistemi (HBYS)’nin Isparta’da Bulunan Sağlık Kuruluşları Üzerindeki Etkililiğinin Araştırılması. Süleyman Demirel Üniversitesi Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, Isparta.
2. http://globocan.iarc.fr/ (Erişim Tarihi: 18.01.2021)
3. http://kanser.gov.tr/ (Erişim Tarihi: 18.01.2021)
4. Siegel, Rebecca L, Kimberly D Miller, and Ahmedin Jemal. 2015. “Cancer Statistics, 2015.” CA: A Cancer Journal for Clinicians 65 (1). American Cancer Society: 5–29. <http://dx.doi.org/10.3322/caac.21254>.
5. Kallmeyer V,Venkat K., “Beyond eHealth: Health and IT Converge”, Siliconindia, Vol. 6, Issue 4, April 2002, p. 42.
6. Altındiş S. ve Kıran Morkoç İ. (2018) Sağlık Hizmetlerinde Büyük Veri. Ömer Halis Demir Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi 11(2): 257-271.
7. Uslu D., Toygar Ş. A. ve Mansur F. (2016) Hastane Bilgi Yönetim Sisteminin Kullanılabilirliğini Belirlemeye Yönelik Bir Araştırma. Uluslararası Sağlık Yönetimi ve Stratejileri Araştırma Dergisi 2(3): 45-57.
8. Koç E, Şengül A. Y, Özkaya U. A, Klinik Karar Destek Sistemlerinin Sağlık Hizmetleri Verimliliğine Etkileri, 6.Sağlık ve Hastane İdaresi Kongresi, 2012
9. Yelmen A. (2016) Klinik Bilgi Sistemlerine İlişkin Lisans Sözleşmeleri Hakkında Değerlendirmeler. İnönü Üniversitesi Hukuk Fakültesi Dergisi 7(2): 365-382.
10. Tüfekçi N., Yorulmaz R. ve Cansever İ. H.(2017) Dijital Hastane. Journal of Current Researches on Healt Sector 2(7): 143-156.
11. Özata M. ve Aslan Ş. (2012) Klinik Karar Destek Sistemleri ve Örnek Uygulamalar. Kocatepe Tıp Dergisi 5(11): 11-17.
12. Koç E., Şengül Atılgan Y., Uyar Özkaya A. ve Gökçe B. (2012) Klinik Karar Destek Sistemleri Kullanımına Yönelik Bir Araştırma: Acıbadem Hastanesi Örneği. IX. Ulusal Tıp Bilişimi Kongresi, 64-74.
13. Madden G. R., Mesner G., Cox H. L., Mathers A. J., Lyman J. A., Sifri C. D. And Enfield K. B. (2018) Reduced Clostridium Difficile Tests and Laboratory-Identified Events with a Computerized Clinical Decision Support Tool and Financial Incentive. Infection Control & Hospital Epidemiology 39(6): 737-740.
14. Özata M., Aslan Ş., “Klinik Karar Destek Sistemleri ve Örnek Uygulamalar”, Afyon Kocatepe Üniversitesi Tıp Dergisi, Cilt: 5, Sayı:1, Ocak 2004.
15. Savaş S., Topaloğlu N.,Yılmaz M., Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri, İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, Sayı:21 ,Sayfa:1-23
16. Mert Dönerçark, Vahap Tecim “Kurumsal Karar Destek Sistemlerinde Yapay Zekâ Kullanımı: Tasarım Ve Uygulama” Yönetim Bilişim Sistemleri Dergisi Cilt 6, Sayı 2, Yıl 2020, Sayfalar 77-103
17. Raziye Özdemir, İbrahim Gülhan, Gökhan Tosun, İbrahim Uyar, Aşkın Doğan, Yüksel Kurban, Suna Yıldırım Karaca, Bülent Kılıç. “Kadın Hastalıkları Ve Doğum Uzmanlarının Klinik Karar Destek Sistemlerine Yaklaşımı” Sted / Sürekli Tıp Eğitimi Dergisi Yıl 2020, Cilt 29, Sayı 3, Sayfalar 153-160
18. Emin Kahya, Merve Türkoğlu. “Personel Değerleme Sistemi İçin Karar Destek Sistemi Tasarımı” Süleyman Demirel Üniversitesi İktisadi Ve İdari Bilimler Fakültesi Dergisi Yıl 2018, Cilt 23, Sayı 3, Sayfalar 937 – 960
19. Cemal Aktürks “Sipariş Teslim Tarihi Problemi İçin Çok Kriterli Ve Çok Yöntemli Karar Destek Sistemi Önerisi” İstanbul Üniversitesi Yayınevi Yıl 2018, Cilt 29 , Sayı 84, Sayfalar 65 – 78
20. Feyza Gürbüz, Esra Kahya Özyirmidokuz” Toplantı Tutanaklarının Analizi İle Bir Karar Destek Sistemi” Sakarya Üniversitesi Fen Bilimleri Enstitüsü Dergisi Yıl 2018, Cilt 22, Sayı 2, Sayfalar 257 – 268
21. Fatih Kılıç, Murat Reis Akkaya, Nuran Memili “Yemekhane İçin Yapay Zekâ Teknikleri Kullanımı İle Günlük Talep Tahmini”. Avrupa Bilim Ve Teknoloji Dergisi Yıl 2018, Sayı 13, Sayfalar 65 - 71
22. Betül Kübra Doğan, Aslıhan Burcu Öztürk. “Kronik Böbrek Yetmezliği Olan Hastalar İle Sosyal Destek Algısını Geliştirmeye Yönelik Grup Çalışması” Türkiye Sosyal Hizmet Araştırmaları Dergisi Yıl 2020, Cilt 4, Sayı 3, Sayfalar 28-36
23. Burcu Oralhan “Sosyal Medya Platformu Seçimini Etkileyen Kriter Ağırlıklarının Bulanık Dematel Yöntemiyle Belirlenmesi” Ibad Sosyal Bilimler Dergisi Yıl 2019, Cilt , Sayı , Sayfalar 408 – 420
24. Burak Yeşilyurt, Kübra Karakuş, Şeyda Gür, Tamer Eren “Çok Ölçütlü Karar Verme Yöntemleri İle Hastane Bilgi Yönetim Sistemleri İçin Paket Programı Seçimi” Başkent Üniversitesi Ticari Bilimler Fakültesi Dergisi Yıl 2019, Cilt 3, Sayı 1, Sayfalar 1-21
25. Ayfer Aydıner Boylu, Gülay Günay “Yaşlı Bireylerde Algılanan Sosyal Desteğin Yaşam Doyumu Üzerine Etkisi” İnsan Ve Toplum Bilimleri Araştırmaları Dergisi Yıl 2018, Cilt 7, Sayı 2, Sayfalar 135 -1363
26. Akpınar, H. (2000) “Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği”, İstanbul Üniversitesi İşletme Fakültesi Dergisi, 29(1): 1-22.
27. Poyraz Ö., Tıpta Veri Madenciliği Uygulamaları: Meme Kanseri Veri Seti Analizi, 2012
28. Yılmaz A, Aloğlu E. Hastane Bilgi Sistemleri. Eskişehir. 5. Sağlık Kuruluşları ve Hastane Yönetimi Sempozyum Kitabı,Osmangazi Üniversitesi, p.338, 2002.