TEKNOFEST

HAVACILIK, UZAY VE TEKNOLOJİ FESTİVALİ

SAĞLIKTA YAPAY ZEKA YARIŞMASI

(Medikal Teknolojiler Kategorisi)

PROJE DETAY RAPORU

TAKIM ADI: ESIDRI'SALUTARY

TAKIM ID: 455780

PROJE ADI: ERKEN DİYABET TEŞHİSİ İÇİN KURAL TABANLI YAPAY ZEKA ALGORİTMASI

İçindekiler

1. **Problem Tanımı ve Mevcut Durum Değerlendirilmesi**
2. **Özgünlük**
3. **Yöntem**
   - **3.1. Verilerin Toplanması**
   - **3.2. Özellik Seçme ve Dengeleme**
   - **3.3. Makine Öğrenmesi ve Performans Değerlendirme**
   - **3.4. Sonuçlar**
4. **Uygulanabilirlik ve Ticarileşme Potansiyeli**
5. **Referanslar**

Proje Konusu

Proje konunuzu aşağıdaki listeden seçiniz.

☐ Hareket izleme ve destek sistemleri

☐ Kronik hastalık takibine yönelik giyilebilir teknolojiler

☐ Görme engelli bireyler için hastane içi yönlendirme sistemleri

☐ Gereksiz antibiyotik kullanımını azaltmaya yönelik teknolojiler

☐ Nefes bileşenlerinden hastalık teşhisi yapabilen sistemler

☐ Fiziksel tıp ve rehabilitasyona yönelik teknolojiler

☐ Ortez ve protez teknolojileri

☒ Hastalık teşhisine yönelik karar destek sistemleri

1. Problem Tanımı ve Mevcut Durum Değerlendirmesi

Diyabet; pankreasın insülin hormonunu yeterli miktarda üretemediği ya da insülin hormonunun yeterli düzeylerde üretildiği halde kullanımındaki bozukluk sonucu oluşan, kandaki şeker düzeyinin yükselmesiyle beraber ilerleyen ciddi olumsuz durumları beraberinde getiren kronik, bulaşıcı olmayan bir hastalıktır [1]. Diyabet hastalığı, dünya çapında gerçekleşen sağlık problemlerine bağlı ölüm nedenlerinin başında gelen rahatsızlıklardan biridir. Diyabet organ ve sistem sağlığını da olumsuz yönde etkilemektedir. Yapılan araştırmalara göre yaşam süresini 5-10 yıl arası kısaltmakta, kardiyovasküler rahatsızlıkların riskini 4 kata kadar arttırmakta, mikrovasküler ve makrovasküler hastalıklara sebebiyet vermekte ve böbrek sağlığını olumsuz yönde etkilemektedir [2]. Bu hastalığa dair yapılan çalışmalarda diyabetli olan bireylerin tüm koşul ve nedenlere bağlı ölüm riskleri değerlendirildiğinde sağlıklı olan hemcinslerine göre ölüm risklerinin 2 kat fazla olduğu sonucuna ulaşılmaktadır. Hastalığa bağlı azalmış kan akımına bağlı, ayak ülserleri, sinir hasarları ve enfeksiyon nedeniyle hasta bireylerde amputasyon ile sonuçlanan durumlar görülebilmektedir. Diyabet ve diyabete bağlı sağlık sorunları ile yaşayan insan sayısı her geçen yıl artış göstermekle birlikte diyabete bağlı oluşan sağlık sorunlarının sebeplerinden bir tanesi de erken teşhis konulamamasıdır. Diyabet ve beraberinde getirmiş olduğu sağlık sorunlarıyla mücadele eden insan sayısındaki artış kişi ve ülkelere ekonomik anlamda da yük oluşturmaktadır. Bu duruma bağlı olarak gelişen sağlık hizmetleri kullanımındaki artış, kardiyak problemler, uzun dönemde böbrek yetmezliği ve iş gücündeki kayıp gibi problemlerin ortaya çıkması ekonomik bir yük oluşumuna sebebiyet vermektedir [3]. Türkiye'de diyabetle ilişkili yapılan sağlı harcamalarının %74'ü diyabetin neden olduğu komplikasyonlar, %26'sı ise doğrudan hastalık üzerine bir maliyet oluşturmaktadır [3]. Bu tablolar değerlendirildiğinde, hastalığın sebebiyet verdiği bu komplikasyonların önüne geçebilmek için diyabeti erken teşhis edebilmek hem hayati hem de ekonomik bir önem taşımaktadır. Diyabet taraması ve hastalığın teşhisi kan şekeri seviyesi ölçülerek kandaki antikorlara bakılarak yapılmaktadır. Aynı zamanda hastalığın teşhisinde bazı tıbbi tahlillerin yanı sıra hastalığa sebebiyet verebilecek farklı semptomlar da bu tanı sürecinde önemli görev almaktadır. Diyabet hastalığının getirmiş olduğu komplike rahatsızlıklar ve hastalığın kendisi göz önüne alındığında erken tanının önemi bir kez daha vurgulanmaktadır. Diyabetli olan bireylerde bu durumu erken teşhis edip hastalığın tedavisine en ivedi şekilde başlanması bu noktada hayati önem taşımaktadır. Bu sayede hastalığın erken teşhisi ve hekimin doğru tedaviye yönlendirmesiyle beraber hem hastanın yaşam kalitesindeki düşüşün hem hastalığa bağlı olarak gerçekleşebilecek bazı komplikasyonların hem de ekonomik anlamda gerçekleşebilecek olan olumsuz durumların önüne geçilecektir. Tüm bu durumlar göz önüne alınıp değerlendirildiğinde kişinin kendi vücudu üzerinde herhangi bir tıbbi tahlile gereksinim olmadan kolaylıkla gözlemleyip teşhis edebileceği, hekimi doğru tanı ve tedaviye yönlendirecek, pratikte kullanıma uygun, ölçülebilir, hasta konforunu bozmayan ve aynı zamanda mali açıdan enjektör, kit, teknisyen ve laboratuvarda ekstra kullanılma olasılığı olan sarf gibi maliyetleri ortadan kaldıracak kural tabanlı yeni teknolojilere ihtiyaç vardır.

Yapılan literatür taramalarında, açık kaynaklı veri setlerindeki veriler ile bazı teşhis algoritmaları geliştirildiği görülmektedir. Literatürde yapılan bir çalışmada makine öğrenmesi temelli modeller oluşturulmuştur [4]. Bu çalışmanın başarı oranı yüksek olmasına rağmen klinikte pratikte kullanıma uygunluğu bulunmamaktadır. Bir teşhis algoritması geliştiriliyorsa bu çalışmanın hem yüksek doğruluk oranlarına sahip bir kurallar bütünü sunabilmesi hem de pratikte kullanıma uygun olması gerekmektedir. Bu yönden bakıldığında doğru metodlar ve analizler ile istatiksel sonuçlar değerlendirilmeli ve kullanacak kişinin konfor alanına hitap eden çalışmalar geliştirilmelidir. Yapmış olduğumuz çalışmada bu önemli noktalar dikkate alınarak elimizdeki veri seti incelenmiş ve bu yönde sonuç elde edilmiştir.

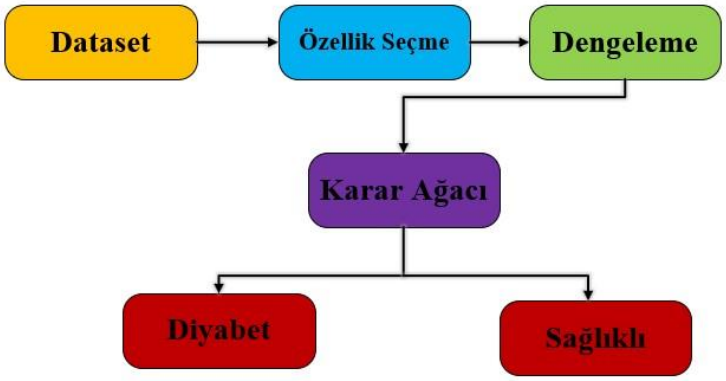
2. Özgünlük

Çalışmamızda kullanmış olduğumuz veri setinde yapmış olduğumuz bütün yapay zeka işlemlerinin yazılımları yerli yazılım ile yapılmıştır. Özellik Seçme ve Dengeleme işlemleri Matlab programı üzerinden var olan algoritmalar geliştirilerek çalışmamızın amacına uygun olacak şekilde en baştan yerli yazılım ile tekrardan yazılmıştır. Algoritmaların tarafımızca yazılması uygulama üzerinde bir yere bağlı kalmadan özgürce hareket ve verilerin şifrelenmesi imkanı sunmaktadır. Verilerin bilimsel olarak yorumlanabilmesi ve çalışmanın bilimsel çalışma niteliği taşıması için istatiksel performans değerlendirme kriterleri kullanılmıştır. Bu sayede kurallar bütününe sahip bir çalışma ortaya çıkmaktadır. Çalışma sonuçlarına göre web tabanlı bir ara yüz geliştirilmiştir. Çalışmanın ilerleyen aşamalarında Matlab üzerinden kullanıcının vermiş olduğu bütün özellikler veri tabanına kaydedilip bu şekilde veri depolama imkanı sunulması öngörülmektedir. Bilimsel çalışmaların her zaman için geliştirilebilir ve uygulanabilir olması büyük önem taşımaktadır. Bu doğrultuda yazmış olduğumuz algoritmanın uygun veri setleri ile başka hastalıkların tanı ve teşhisinde de kullanılabileceği öngörülmektedir.

Literatürde daha önce bu probleme yönelik çalışmalar bulunmaktadır [4]. Bu çalışmaların doğruluk oranları yüksek olmasına rağmen pratikte kullanıma uygunluğunun olmadığı tespit edilmiştir. Literatürdeki çalışmaya göre kişilerden elde edilen sonuçlar tekrar yapay zeka üzerinde konumlandırılıp tekrardan algoritma eğitimi yapılması gerekmektedir. Çalışmamızın amacının hitap ettiği kesim yalnızca mühendis veya bu yetkinliğe sahip kişiler olmadığı için pratikte kullanıma uygunluğu olmadığı sonucuna varılmıştır. Bu noktada elde edilen sonuç; ele almış olduğumuz problem için bilimsel dayanağı olan yeni teknolojilere ihtiyaç duyulduğu yönündedir. Yapmış olduğumuz çalışmadaki en temel amaç; diyabet hastalığında erken teşhise götürecek, pratikte kullanıma uygun, birey konforunu olumsuz yönde etkilemeyen, kural tabanlı, kolay anlaşılabilen, yorumlanabilen ve geliştirilmeye her zaman açık bir teşhis sistemi geliştirmektir. Bu amaç doğrultusunda literatürden farklı olarak hasta veya sağlıklı bireye, hekime hitap eden bir sistem geliştirilmiştir. Çalışmamızın sonuçlarından elde etmiş olduğumuz performans değerlendirme kriterleri göz önüne alındığında, doğruluk oranı yüksek bir çalışma sonucuna ulaşılması, çalışmamızı güvenilir ve amacına uygun sonuçlar verdiği yönünde değerlendirilmesine olanak vermektedir.

3. Yöntem

Çalışma şekil 1'deki sürece göre tasarlanmıştır. 520 hasta-sağlıklı bireyin, hastalığa etkisi olduğu düşünülen semptomlarının kullanıldığı bir veri seti kullanılmıştır. İlk olarak bireyler kadın-erkek ve hasta-sağlıklı(pozitif-negatif) olarak ayrılarak gruplara bölünmüştür. Her gruptan etkili olacak özellikler seçilerek çıkarılmıştır. Özellik seçme algoritması kullanılarak ortaya çıkan özellikler azaltılarak ve dengelendikten sonra karar ağaçları yardımıyla diyabet teşhis edecek modeller oluşturulmaktadır.



ŞEKİL 1

3.1.Verilerin Toplanması

Çalışmada kullanılan veriler UCI Machine Learning Repository de açık kaynak paylaşılan veri setinden alınmıştır. Veri setinde bulunan 520 hasta-sağlıklı bireyin diyabet teşhisinde rol oynadığı, diyabete teşhisine etkisinin olabileceği düşünülen 14 adet semptom ve bireylerin diyabet negatif-diyabet pozitif sonuçları kullanılmış (tablo1) ve işlenmiştir.

Özellik		Erkek		Kadın		Toplam
No	Ad	Yes	No	Yes	No	
1	Polyuria	129	199	129	63	520
2	Polydipsia	108	220	125	67	520
3	Sudden Weight Loss	102	226	115	77	520
4	Weakness	177	151	128	64	520
5	Polyphagia	122	206	115	77	520
6	Genital Trush	95	233	21	171	520
7	Visual Blurring	121	207	112	80	520
8	Itching	153	175	100	92	520
9	Irritability	78	250	48	144	520
10	Delayed healing	138	190	101	91	520
11	Partial paresis	100	228	124	68	520
12	Muscle stiffness	112	216	83	109	520
13	Alopecia	152	176	27	165	520
14	Obesity	55	273	33	159	520
15	Class	147	181	173	19	520

Tablo 1

3.2.Özellik Seçme ve Dengeleme

Yaptığımız çalışmada özellik seçme algoritması olarak Fisher Özellik Sıralama Algoritması kullanılmıştır. Bu algoritma MATLAB programı üzerinden çalışmamızda kullanabileceğimiz şekilde yazılmış ve kullanılmıştır. Fisher Özellik Seçme Algoritmasında yapılacak olan sınıflandırma işlemi için veri setinde bulunan en mühim ve uygunluk derecesi en yüksek özellikler belirlenmektedir. Veri setindeki her bir özellik için bir F-score değeri hesaplanır. Her bir özellik için hesaplanan F-score değerlerinin ortalaması F-score eşik değerini vermektedir. Seçilecek olan özelliklerin ayırt edici olanının ve en uygun olanının belirlenebilmesi için F-score eşik değeri hesaplanmaktadır [5]. Veri setindeki her bir özelliğe ait hesaplanan F-score değeri, F-score eşik değerinden büyük ise o özellik ayırt edici, küçük ise o özellik ayırt ediciliği olmayan özellik olarak seçilmektedir. Özellikler F-score değerine göre büyükten küçüğe olacak şekilde sıralanmaktadır. Bu özellik seçme yöntemine göre en yüksek F-score değerine sahip olan özellikler çalışmada kullanılmak üzere çalışmanın amacına en uyumlu özellik olarak seçilmektedir [6]. Çalışmada kullanılan veri setinin bize sunmuş olduğu 15 tane hastalık belirteci olabileceği düşünülen özellik tablo 1 ‘de verilmiştir. Bu özelliklerin kadın-erkek bireylerde var olup olmaması durumuna göre ayrıştırılmıştır. 520 hasta-sağlıklı bireyden elde edilen veri setindeki özellikler Fisher Score değerlerine göre bu algoritma ile sıralanmıştır. Veri setinde erkek hasta-sağlıklı bireylerin semptomları incelendiğinde, bu bireylerin verilerinin seçiciliği ve ayırt ediciliğinin yüksek olduğu sonucuna ulaşılmaktadır. Bu sebeple çalışmada kullanılan veri seti hem veri setindeki tüm bireyler için hem de veri setinde bulunan erkek bireylerin ayrıştırılıp yalnız erkek bireyler için teşhis algoritması geliştirilmesine karar verilmiştir. Seçilen özellikler tüm bireyler ve yalnız erkek bireyler için tablo 2 ve tablo 3’de verilmektedir.

Tüm Veri Seti Özellik	
No	Ad
1	Gender
2	Polyuria
3	Polydipsia
4	Sudden Weight Loss
6	Polyphagia
10	Irritability
12	Partial Paresis
14	Alopecia

Tablo 2

Erkek Bireyler İçin Özellik	
No	Ad
2	Polyuria
3	Polydipsia
4	Sudden Weight Loss
5	Weakness
6	Polyphagia
7	Genital Trush
10	Irritability
12	Partial Paresis

Tablo3

Tüm bireyler ve yalnız erkek bireyler için seçilip F-score değerlerine göre sıralanan 8 tane özellik, veri setinde bulunan tüm özelliklerin %53,33333'ünü kapsamaktadır. Tablo 2'de yer alan veri setindeki bütün bireyler için; cinsiyet, poliüri, polidipsi, ani kilo kaybı, polifaji, sinirlilik, kısmi parezi, alopesi Fisher Score Özellik Seçme Algoritmasına göre en uyumlu özellikler olarak belirlenmektedir. Tablo 3'te yer alan yalnızca erkek bireyler için oluşturulan modelde; poliüri, polidipsi, ani kilo kaybı, dayanıksızlık güçsüzlük, polifaji, genital pamukçuk, sinirlilik ve kısmi parezi Fisher Score Özellik Seçme Algoritmasına göre en uyumlu özellikler olarak belirlenmektedir. Bu özellikler F-score değerleri hesaplanıp F-score eşik değerinden büyük olmalarına göre belirlenmiştir. Kendi aralarında F-score değerleri büyükten küçüğe olacak şekilde sıralanmıştır. Özellik seçme işlemi ile ilgili olmayan veriler tespit edilip ortadan kaldırıldıktan sonra kurulmuş olan algoritma hızı arttırılmıştır. Bu sayede çalışma verimiz daha anlaşılır, tanımlanabilir ve görselleştirilebilir hale gelmektedir.

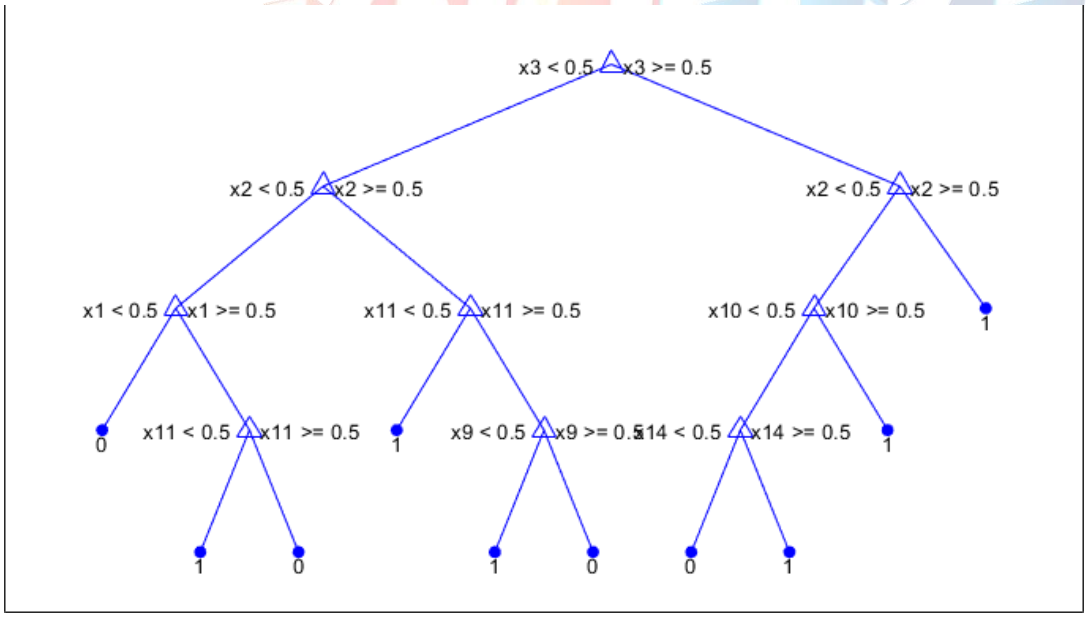
3.3. Makine Öğrenmesi ve Performans Değerlendirme

Çalışmamızda kullandığımız açık kaynaklı veri seti Excel programı kullanarak Matlab programının kullanabileceği matrisler formatına getirilmiştir. Bu format ile birlikte her bir sütuna özellikler her satıra ise bir bireyi temsil edecek şekilde atama yapılmıştır. Düzenlediğimiz matrisin son sütunu ise etiket; hasta olup olmama durumuna göre 1 veya 0'ların atandığı parça olarak belirlenmektedir. Bu yaptığımız ön işlemden sonra veri matrisimizi Matlab kullanarak yazdığımız dengeleme, özellik çıkartma ve özellik seçme algoritmalarından geçirilmiştir. Hazırlamış olduğumuz verinin, ön veri işleme adımlarından sonra Matlab'ın sunmuş olduğu yapay zeka geliştirme kütüphanesi olan 'Classification Linear' için veri matrisimiz uygun hale getirilmiştir. Classification Linear kütüphanesi kullanarak ilk olarak veri setimizin hangi sütunlarının data, hangi kısımlarının etiket olduğunu tanımlanmaktadır. Veri dağıtıcı olarak Cross-Validation seçilmiştir ve yapay zeka modelinin doğruluk oranın belirlenebilmesi için veri setinin %15'lik kısmı test veri kümesi olarak belirlenmiştir. Sınıflandırmada cTree (Karar ağaçları) algoritması seçilmiştir. İterasyon olarak 100 adım belirlenmiştir. Yapılan model eğitiminden sonra algoritmanın önceden bildiği ve bilmediği birey özelliklerine karşı verdiği tahmin sonuçları Confusion Matrix olarak alınmış, gelen verilerden istatiksel olarak net bir sonuç elde edebilmek için istatiksel performans kriterleri hesaplanmıştır.

Çalışmamızda kullanmış olduğumuz karar ağaçları sınıflandırma ve regresyon modeli olarak veri tabanı sistemleriyle kolayca entegre edilebilmeleri ve güvenilirlik düzeyinin yüksek olması kolay yorumlanabilmeleri sebebiyle kullanılmaktadır.

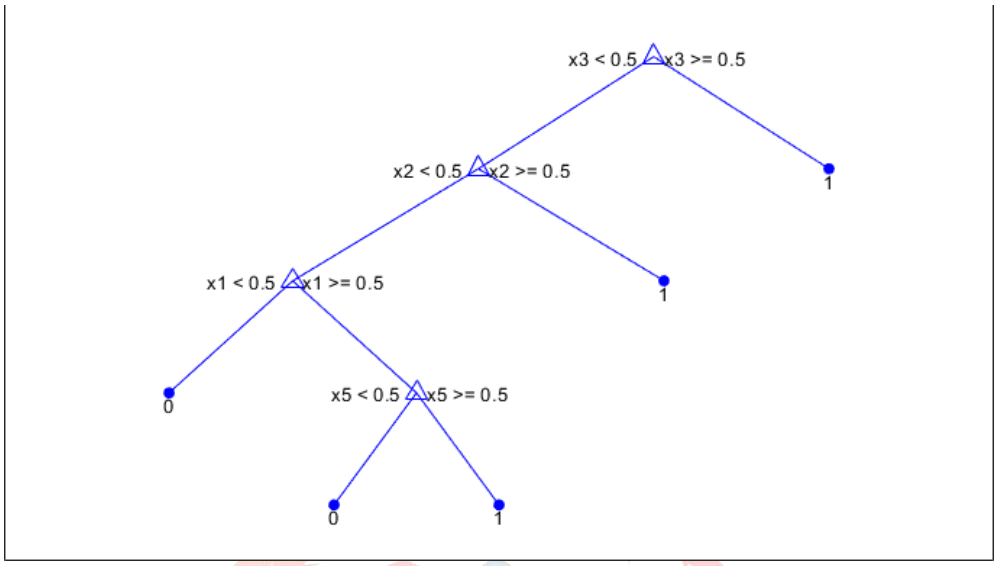
Sınıflandırma için bir ağaç oluşturulduktan sonra veri setinde bulunan her bir özellik bu ağaç sisteminin içine gömülmektedir. Yapmış olduğumuz araştırmalara göre; veri setinin oldukça karmaşık olduğu durumlarda bile yapının bağımlı değişkeni etkileyen diğer değişkenleri basit

bir dallanma modeli ile sunulduğu görülmektedir [7]. Veri setindeki verilerimiz değişkenlere göre gruplandırılarak ağaç şeklinde modellenmiştir. Karar ağaçları sınıflandırma modelinde temel olarak elde edilmek istenen sonuç ve öğrenilmek istenen veriler ağaç şeklinde modellenmektedir. Akış diyagramına benzeyen bu yapıda diyagramda akış kök kısımdan başlayıp yapraklara yani uçlara doğru dallanma yapmaktadır. Ağaçta yaprak olarak adlandırılan yerler veri setindeki her bir veriyi temsil etmektedir. Bu verilerin her birinin bir etiketi bulunmaktadır. Ağaçtaki dallanma durumlarının olduğu yerlere düğüm noktaları adı verilmektedir [7]. Çalışmamızdaki düğüm noktaları eğitim algoritması ile belirlenmiştir. Ara düğümlere, kullanmış olduğumuz ve özellik seçerek dengelediğimiz veri setindeki en iyi ayırt etme yeteneği bulunan değişkenler atanmaktadır. Değişkenlerin değerlerine göre düğüm noktalara ayrılmaktadır. Sonuçta elde etmiş olduğumuz veriler yapraklara atanmaktadır. Çalışmada bu işlem eğitilmiş olan makine ile defalarca tekrarlanır. Bu işlem boyunca değişkenler verileri en uygun şekilde ayırarak tekrarlanır. Bu işlemler sonucunda gruplar elde edilmiştir. İnceleme yapıldığında oluşan bu grupların varyanslarının minimum olduğu tespit edilmektedir. Çalışmada tüm bu süreç tamamlanıp karar ağaçları ortaya çıktığında veri setinini ayırabilmemiz için bir kurallar bütünü ortaya çıkmaktadır. Bu sayede hem kategorik sınıflara hem de sürekli değişkenlere ait kural tabanlı modeller elde edilmiştir. Yapmış olduğumuz çalışmada veriler sınıflandırılmış ve karar ağaçları oluşturulduktan sonra performans kriterleri veri setindeki hem yalnız erkek hem de tüm veri setinde bireyler için ayrı olarak hesaplanmaktadır. Performans değerlendirme kriteri olarak seçmiş olduğumuz özellik sayısına göre; sınıflandırma doğruluk oranı (DO), özgüllük, duyarlılık, F-ölçümü, Kappa Kat Sayısı ve AUC değerleri hesaplanmıştır. Doğruluk oranının yüksek olduğu ağaç modelleri teşhis algoritmasında kullanılmak üzere baz alınmıştır.
Çalışma sonucu MATLAB üzerinden hesaplanmış performans değerlendirme kriterlerinden elde edilmiş olan verilere göre karar ağaçları ile modellenmiştir. Tüm veri seti için veri setindeki 15 semptom- özelliğin kullanılması durumunda en yüksek doğruluk oranı elde edilmektedir. Oluşturulan karar ağaçları durumlarından doğruluk oranını yüksek veren karar ağacı modeli seçilmiş ve şekil 2'de gösterilmiştir.



Şekil 2

Veri setindeki yalnız erkek bireyleri alarak yapılan çalışmada 5 semptom- özelliğin kullanılması durumunda en yüksek doğruluk oranının elde edildiği görülmüştür. Bu değerlendirme sonucunda, yalnız erkek bireyler için oluşturulan karar ağaçları modellerinden, en yüksek doğruluk oranını veren model seçilmiş ve şekil 3'te gösterilmiştir.



Şekil 3

Elde edilen karar ağaçları modellemeleri ile birlikte kolay yorumlanabilen, kural tabanlı, karmaşıklığı az teşhis algoritmaları üretilmiştir.

3.4. Sonuçlar

Çalışmamızdaki amaç diyabeti erken teşhis eden, anlaşılır ve pratikte kullanıma uygun kural tabanlı bir model geliştirmektir. Diyabet her geçen gün risk ve önem sürecini taşımaktadır. Hastalığın erken teşhisi ve erken tedavi süreci büyük önem taşımaktadır. Yapmış olduğumuz çalışmada diyabetin herhangi bir tıbbi sarf malzemesine ihtiyaç duyulmadan erken teşhis edilebilmesi ile kişiyi doğru tedaviye yönlendirerek geç kalınmış teşhis ve bununla beraber gelişen diğer riskli komplikasyon risklerinden korumak, konforlu bir yaşam sürebilmesine yardımcı olabilme gayesi hedef alınmıştır. Diyabeti erken teşhis edebilecek yöntemler ve modeller geliştirildikçe mali açıdan da sağlık sektörü başta olmak üzere beraberinde toplum ekonomisine de fayda sağlayacaktır. Bu yönde yapmış olduğumuz çalışma karar ağacı ile performans kriterleri veri setindeki hem yalnız erkek bireyler hem de tüm bireyler için hesaplanmaktadır. Tüm veri seti erkek bireyleri ayırmadan performans değerlendirme kriterlerine bakılarak incelendiğinde, kullanılan özellik sayısı 15 iken başarı oranı %93.75 olmaktadır. Veri setindeki erkek bireyler ayrılarak performans değerlendirme kriterlerine göre incelendiğinde kullanılan özellik sayısı 5 iken başarı oranı %97.22 olmaktadır. Çalışmamızda elde etmiş olduğumuz doğruluk oranları değerlendirildiğinde çalışmanın umut vaad eden bir sonuç verdiği, oluşturulan sisteminin güvenilirlik düzeyinin yüksek olduğu ve yazılımsal olarak da kullanılabilir olduğu değerlendirilmesi yapılmaktadır.

Çalışmanın tüm veri seti kullanılarak ve yalnız erkek bireyler kullanılarak elde edilmiş sonuçları tablo4 ve tablo 5'te verilmektedir.

Tüm Veriler için
Performans Değerlendirme Kriterleri

Özellikler Sayısı	Doğruluk Oranı	Duyarlılık	Özgüllük	F-Ölçümü	Kappa Katsayısı	AUC
2	82.50	0.85	0.80	0.82	0.65	0.83
3	76.25	0.90	0.63	0.74	0.53	0.76
5	85.00	0.83	0.88	0.85	0.70	0.85
6	78.75	0.85	0.73	0.78	0.58	0.79
8	90.00	0.90	0.90	0.90	0.80	0.90
15	93.75	0.98	0.90	0.94	0.88	0.94

Tablo 4

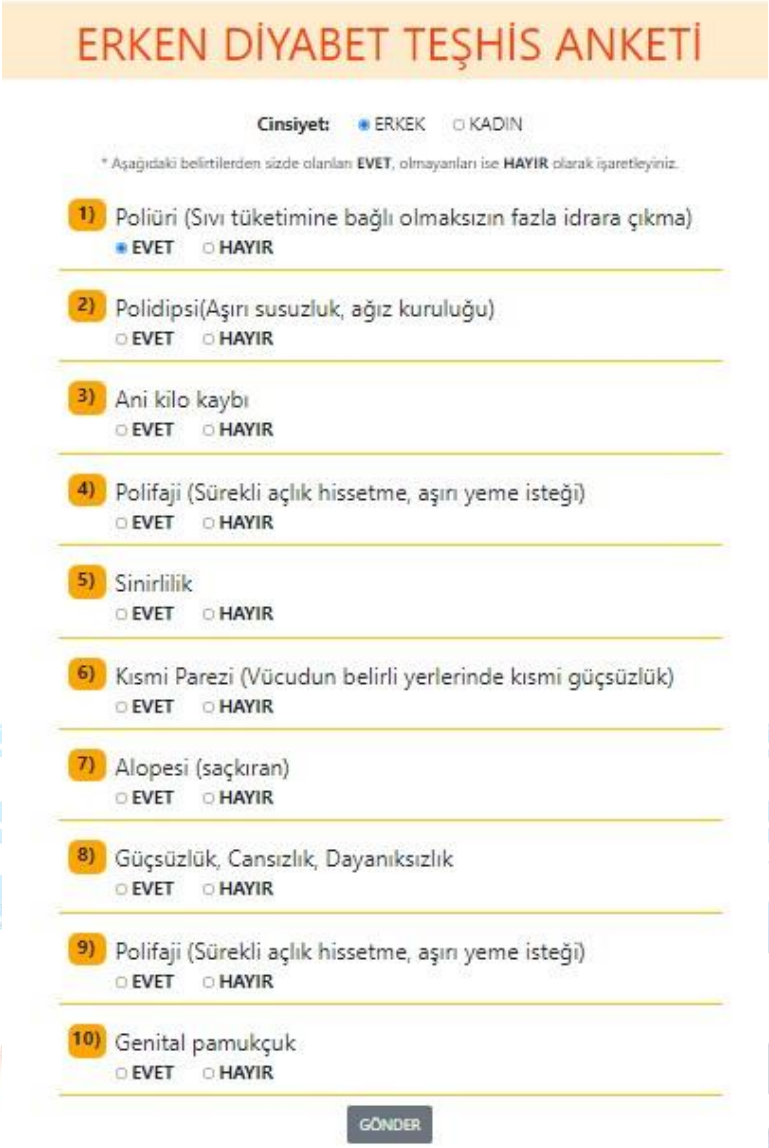
Erkek Bireyler için
Performans Değerlendirme Kriterleri

Özellikler Sayısı	Doğruluk Oranı	Duyarlılık	Özgüllük	F-Ölçümü	Kappa Katsayısı	AUC
2	94.44	0.95	0.94	0.94	0.89	0.94
3	94.44	0.95	0.94	0.94	0.89	0.94
5	**97.22**	**0.97**	**0.97**	**0.97**	**0.94**	**0.97**
6	95.83	0.97	0.94	0.96	0.92	0.96
8	94.44	0.97	0.91	0.94	0.89	0.94
15	94.44	1.00	0.89	0.94	0.89	0.94

Tablo 5

Performans değerlendirme kriteri olarak kullanılmış olan AUC değeri geliştirmiş olduğumuz yöntemin doğru teşhis edebilme yeteneğini belirleyen istatiksel bir yöntemdir ve bu değerin 1'e yakın olması beklenmektedir. Tablo 4' de tüm veriler için geliştirilen modelde doğruluk oranının en yüksek olduğu şartta yani 15 özelliğin hepsinin teşhis algoritmasında kullanıldığı durumda AUC değeri 0.94 olarak hesaplanmaktadır. Tablo 5'de yalnız erkek bireyler için geliştirilen teşhis algoritmasında yine en yüksek doğruluk oranını veren özellik sayısı dikkate alındığında AUC değeri 0.97 olarak hesaplanmıştır. Performans değerlendirme kriteri olarak kullanılan Kappa Kat Sayısı çalışmada geliştirerek kullandığımız referans yöntem ve çalışmada geliştirilen yöntem arasındaki güveninirliği uyum derecesini ölçen istatiksel bir yöntemdir. Uyum derecesinin belirlenmesi adına bazı sayısal aralıklar tanımlanmıştır. Yapmış olduğumuz çalışmada en yüksek doğruluk oranına sahip özellik sayıları baz alınarak Tablo 4 ve Tablo 5 değerlendirildiğinde tüm veri seti için Kappa Kat Sayısı 0.88, erkek bireyler için oluşturulmuş modelde 0.94 olarak hesaplanmaktadır. Bu değerler baz alındığında sistemin uyum derecesinin verilen aralıklarda her veri durumu içinde "çok iyi düzeyde uyum" sonucuna götürdüğü görülmektedir. Duyarlılık ve özgüllük, hasta bireyi teşhis edebilmekteki başarı ve hassasiyeti göstermektedir. Bu değerlerin 1'e yakın olması beklenmektedir. Yapılan çalışmada Tablo 4 ve Tablo 5'te bulunan sonuçlardan en yüksek doğruluk oranlarını veren özellik sayıları referans alındığında tüm veri setindeki bireyler için duyarlılık ve özgüllük değerleri sırası ile 0.98 ve 0.90, yalnız erkek bireyler için yapmış olduğumuz çalışmada duyarlılık ve özgüllük değerleri sırası ile 0.97 ve 0.97 olarak hesaplanmıştır. Performans değerlendirme kriteri olarak kullanılan F-ölçüt değerinin yine 1'e yakın olması istatiksel olarak beklenmektedir. Bu değer çalışmamızdaki en yüksek doğruluk oranını veren özellik sayısı referans alındığında veri setindeki tüm bireyler için 0.94, yalnız erkek bireyler için oluşturulmuş durumda 0.97 olarak 1'e oldukça yakın değerler olarak hesaplanmıştır.

Geliştirilmiş olan teşhis sistemlerinin pratikte kullanılabilmesi için en az %80 doğruluk oranına sahip olması gerekmektedir [8]. Çalışmanın amacına uygun olacak şekilde doğruluk oranları yüksek, kural tabanlı, pratikte kullanıma uygun bir teşhis algoritması ortaya konmuştur. Bu çalışmanın baştaki nihai amaçlarının bir tanesi de hekime yardımcı bir sistem geliştirmektir. Hekimin bilgisayarında bulunan çalışmamızda geliştirmiş olduğumuz web tabanlı arayüz hastanelerde hekimlerin kullandığı sistemlerle entegre edilinecektir. Bu sayede hekimin basit bir anket mantığıyla kişinin diyabet riskini öngörmesi durumunda hasta olduğu düşünülen bireyden laboratuvar tahlili istemesi ve bu sayede ekonomik anlamda da büyük faydalar sağlanması öngörülmektedir. Çalışmanın farklı bir kullanım alanının olabileceğini düşündüğümüz bir diğer senaryo; E-Nabız ve bu tarz sağlık uygulamalarında kullanılabilecek olan geliştirdiğimiz web tabanlı anket mantığında olan anlaşılması ve kullanımı kolay arayüze sahip uygulama sayesinde hasta olduğundan şüphelenen veya şüphelenmeyen kişi kolaylıkla bu test ile bir hekim görüşü alıp almayacağının kararını verebilmektedir.



ERKEN DİYABET TEŞHİS ANKETİ

Cinsiyet: ◉ ERKEK ○ KADIN

* Aşağıdaki belirtilerden sizde olanlan **EVET**, olmayanları ise **HAYIR** olarak işaretleyiniz.

1) Poliüri (Sıvı tüketimine bağlı olmaksızın fazla idrara çıkma)
◉ EVET ○ HAYIR

2) Polidipsi(Aşırı susuzluk, ağız kuruluğu)
○ EVET ○ HAYIR

3) Ani kilo kaybı
○ EVET ○ HAYIR

4) Polifaji (Sürekli açlık hissetme, aşırı yeme isteği)
○ EVET ○ HAYIR

5) Sinirlilik
○ EVET ○ HAYIR

6) Kısmi Parezi (Vücudun belirli yerlerinde kısmi güçsüzlük)
○ EVET ○ HAYIR

7) Alopesi (saçkıran)
○ EVET ○ HAYIR

8) Güçsüzlük, Cansızlık, Dayanıksızlık
○ EVET ○ HAYIR

9) Polifaji (Sürekli açlık hissetme, aşırı yeme isteği)
○ EVET ○ HAYIR

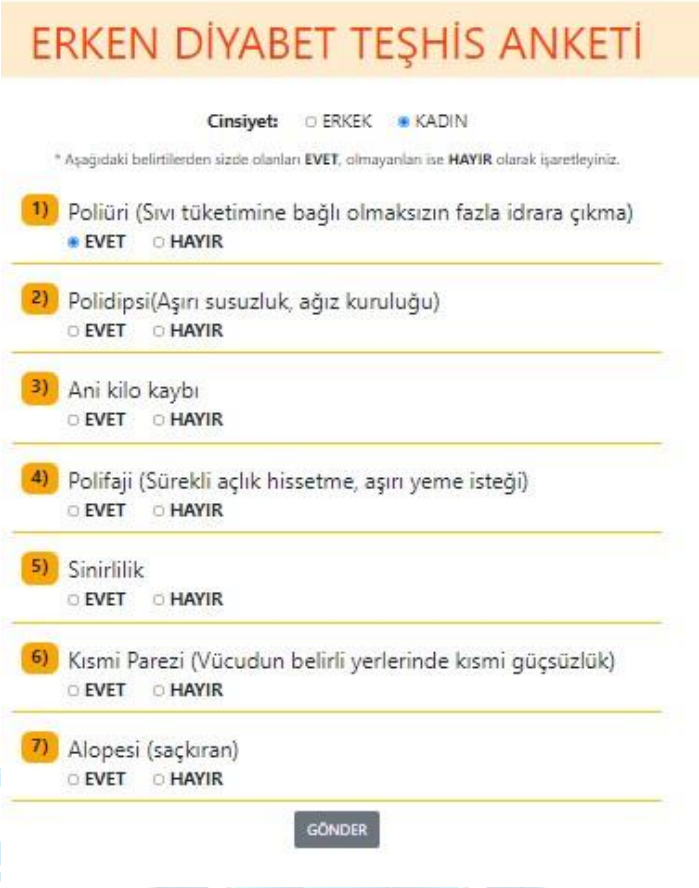
10) Genital pamukçuk
○ EVET ○ HAYIR

GÖNDER

Şekil 4

HAVACILIK, UZAY VE TEKNOLOJİ FESTİVALİ

Geliştirmiş olduğumuz web tabanlı arayüzler şekil 4 ve şekil 5'te verilmiştir. Çalışmanın sonuçlarına değerlendirmek ve sonucunu görmek adına basit bir arayüz tasarımı yapılmıştır. Bu çalışmanın doğruluğunu test edebilmek adına çevremizde bulunan kişilere bu test uygulanmış ve doğru sonuçlar elde edildiği sonucuna mutlulukla ulaşılmıştır.



Şekil 5

4. Uygulanabilirlik ve Ticarileşme Potansiyeli

Çalışmanın pratikte kullanıma uygunluğunun sağlanabilmesi için Web tabanlı bir arayüz geliştirilmiştir. Geliştirilen arayüz çok temel seviyede anket mantığında ilerlemektedir. Veri setinde bulunan hastalığa etkisi olduğu düşünülen semptomlara kişinin kendinde olup olmama durumuna göre vereceği evet- hayır cevapları ile oransal tanı belirlenmektedir. Çalışmanın mevcut şartlar altında pratik kullanım imkanı sağlayarak çalıştığı görülmektedir. Her bilimsel çalışmada olduğu gibi bizim çalışmamızda da bazı risk faktörleri bulunmaktadır. Çalışmamızda kullanıcıya sunduğumuz sistemin web kısmında bazı değişik entegrasyonlar yapılmak istediğinde sıkıntı çıkması halinde Matlab' deki algoritma düzenini bozulmadan aynı işlemlerin Python'da yapılması risk ve çözüm analizi olarak belirlenmiştir. Ayrıca veri setindeki verilerden kaynaklı oluşabilecek ve yanlış tanıya sebebiyet verecek hatalar da risk planları içerisindedir. Bu durumda eğitilmiş olan makinenin baştan eğitilmesine karar verilmiştir. Projemiz yazılım tabanlı bir proje olması sebebiyle doğrudan malzeme sarf ihtiyacı gibi bir maliyet oluşturmamaktadır. Yapılan saha çalışmalarında, projeyi sağlık sektöründe bulunan uzman kişiler ile paylaştığımızda bu projenin arayüz tasarımı ile beraber ticarileştirilip özel hastanelerde kullanılabileceği sonucuna varılmıştır. Saha verileri ve uzman görüşleri değerlendirildiğinde çalışmanın ticarileştirme potansiyelinin yüksek olduğu görülmektedir. Çalışmamıza en başta topluma fayda sağlamak ve gelişmekte olana katkımız olması amacıyla başladığımız için bu çalışmanın arayüz tasarımlarını geliştirerek halka açık E-nabız gibi uygulamalarda kişi kullanımına ve devlet hastanelerinde hekimlerin kullanımına açmanın isteği içerisinde bulunmaktayız.

5. Referanslar

[1] Ekin Koç and Selim Güler. Diyabet Şeker Hastalığı. pages 1–4, 2015.
[2] Hatice Öntürk and Hanefi Özbek. Deneysel diyabet oluşturulması ve kan şeker seviyesinin ölçülmesi. Genel Tıp Dergisi, 17(4):231–236, 2007.
[3] Ece Öcal and Muhammed Önsüz. Diyabet hastalığının ekonomik yükü. ESTÜDAM Halk Sağlığı Dergisi, 3(1):24–41, 2018.
[4] M. M. Faniqul Islam, Rahatara Ferdousi, Sadikur Rahman, and Humayra Yasmin Bushra. Likelihood Prediction of Diabetes at Early Stage Using Data Mining Techniques. Advances in Intelligent Systems and Computing, 992:113–125, 2020.
[5] Hüseyin Budak. Özellik Seçim Yöntemleri ve Yeni Bir Yaklaşım. Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 22(Özel Sayı):21–31, 2018.
[6] Mehmet Recep Bozkurt, Muhammed Kürşad Uçar, Ferda Bozkurt, and Cahit Bilgin. Development of hybrid artificial intelligence based automatic sleep/awake detection. IET Science, Measurement and Technology, 14(3):353–366, may 2020.
[7] Zeynep Burcu Kıran. Lojistik Regresyon ve Cart Analizi Teknikleriyle Sosyal Güvenlik Kurumu İlaç Provizyon Sistemi Verileri Üzerinde Bir Uygulama. PhD thesis, Gazi Üniversitesi, 2010.
[8] Reha Alpar. Spor, Sağlık ve Eğitim Bilimlerinden Örneklerle Uygulamalı İstatistik ve Geçerlik- Güvenirlik. Detay Yayıncılık, 2016.

TEKNOFEST
HAVACILIK, UZAY VE TEKNOLOJİ FESTİVALİ