

SİGARA KULLANIMININ KOAH VE DEPRESYON ÜZERİNDEKİ ETKİLERİNİN

MAKİNE ÖĞRENMESİ VE WEB TABANLI KARAR DESTEK SİSTEMİ İLE ANALİZİ

1. PROJENİN AMACI VE KAPSAMI

Bu projenin temel amacı, **sigara kullanımının KOAH (Kronik Obstrüktif Akciğer Hastalığı) ve depresyon üzerindeki etkilerini**, büyük ölçekli sağlık verileri kullanarak **makine öğrenmesi yöntemleriyle analiz etmek** ve elde edilen sonuçları **web tabanlı bir karar destek sistemi** aracılığıyla kullanıcıya sunmaktır.

Günümüzde sağlık alanında veriye dayalı karar mekanizmaları büyük önem taşımaktadır. Ancak ham verilerin kullanıcıya doğrudan sunulması, hem anlaşılabilirlik hem de güvenlik açısından uygun değildir. Bu nedenle bu projede, veri seti doğrudan web arayüzünde kullanılmamış; bunun yerine **makine öğrenmesi modelleri eğitilmiş**, ardından bu modellerden elde edilen bilgi ve ilişkiler web uygulamasına entegre edilmiştir.

2. PROBLEM TANIMI

Sigara kullanımı, hem fiziksel hem de ruhsal sağlık üzerinde ciddi olumsuz etkilere sahiptir. Literatürde sigaranın KOAH ile güçlü bir ilişkisi olduğu bilinmekle birlikte, depresyon ile birlikte görülme durumu daha karmaşık bir yapı sergilemektedir.

Bu proje aşağıdaki sorulara yanıt aramaktadır:

- Sigara kullanımı, KOAH ve depresyon birlikteliğini ne ölçüde etkilemektedir?
- Hangi yaşam tarzı faktörleri bu risk üzerinde daha baskındır?
- Makine öğrenmesi modelleri ile riskli bireyler erken aşamada tespit edilebilir mi?
- Bu analizler, kullanıcıya anlaşılır bir web arayüzü ile sunulabilir mi?

3. VERİ SETİ VE VERİ ÖN İŞLEME

3.1 Veri Setinin Kaynağı

Projede kullanılan veri seti, **CDC (Centers for Disease Control and Prevention)** tarafından yayımlanan, toplum sağlığına yönelik kapsamlı bir ankete dayanmaktadır. Veri seti 370.000'den fazla bireye ait kayıt içermektedir.

Bu büyüklükteki bir veri seti, makine öğrenmesi için güçlü bir öğrenme ortamı sunmaktadır.

3.2 Kullanılan Değişkenler

Makine öğrenmesi sürecinde aşağıdaki değişkenler seçilmiştir:

- **Sigara kullanımı (smoking_status)**
- **Uyku süresi (sleep_hours)**
- **Vücut kitle indeksi (bmi)**
- **Fiziksel aktivite durumu (physical_activity)**
- **Yaş (age)**

Hedef değişken olarak:

- **KOAH ve depresyonun birlikte görülmesi (copd_and_depression)**

kullanılmıştır.

Bu değişkenler, hem literatürde yer alan risk faktörleriyle uyumlu olması hem de kullanıcıdan web formu aracılığıyla alınabilir olması nedeniyle tercih edilmiştir.

3.3 Veri Temizleme ve Küçültme

Veri setinin büyük olması nedeniyle hesaplama maliyetini düşürmek amacıyla **rastgele örnekleme yöntemiyle 10.000 satırlık bir alt veri seti oluşturulmuştur**. Bu işlem sırasında veri dağılımının korunmasına dikkat edilmiştir.

Bu yaklaşım, hem model performansını korumuş hem de analiz süresini önemli ölçüde azaltmıştır.

4. MAKİNE ÖĞRENMESİ YÖNTEMLERİ

4.1 Kullanılan Modeller

Projede üç farklı makine öğrenmesi modeli denenmiştir:

- Logistic Regression**
Basit ve yorumlanabilir bir model olmakla birlikte, dengesiz veri yapısında yetersiz kalmıştır.
- Random Forest**
Karar ağaçlarının birleşimiyle çalışan güçlü bir modeldir. Ancak hasta yakalama oranı sınırlı kalmıştır.
- XGBoost (Extreme Gradient Boosting)**
Yapılandırılmış veriler için günümüzde en yaygın kullanılan modellerden biridir. Hem hız hem de performans açısından üstünlük sağlamıştır.

4.2 Model Değerlendirme Kriterleri

Veri seti dengesiz olduğu için, **accuracy tek başına yeterli bir ölçüt olarak kabul edilmemiştir**. Bu nedenle aşağıdaki metrikler esas alınmıştır:

- Recall (Hasta Yakalama Oranı)**
- F1-Score**
- Confusion Matrix**

Sağlık alanında yanlış negatiflerin (hastayı kaçırma) sonuçları daha ciddi olduğu için **recall değeri öncelikli metrik olarak değerlendirilmiştir**.

4.3 Model Karşılaştırması ve Seçimi

Model	Recall	F1-Score	Accuracy
Logistic Regression	Çok düşük	Çok düşük	Yanıltıcı yüksek
Random Forest	%62	%16	%80
XGBoost	%72	%14	%74

Bu sonuçlara göre, **XGBoost modeli**, riskli bireyleri daha yüksek oranda tespit ettiği için **final model** olarak seçilmiştir.

5. MODEL GÖRSELLEŞTİRMELERİ

Model performansını daha iyi yorumlayabilmek için aşağıdaki görseller oluşturulmuştur:

- **Feature Importance Grafiği:**
Sigara kullanımının en güçlü belirleyici faktör olduğu açıkça görülmüştür.
- **Confusion Matrix Heatmap:**
Random Forest ve XGBoost modellerinin hasta yakalama performansları karşılaştırılmıştır.
- **ROC Eğrisi:**
Modelin sınıfları ayırt etme yeteneği değerlendirilmiştir.

Bu görseller, analiz sayfasında kullanıcıya sunulmuştur.

6. WEB TABANLI KARAR DESTEK SİSTEMİ

6.1 Mimari Yaklaşım

Web uygulaması, **makine öğrenmesi modellerinin doğrudan web sunucusunda çalıştırılması yerine**, bu modellerden elde edilen bilgilerin kullanılacağı şekilde tasarlanmıştır.

Bu mimari:

- Performans
- Güvenlik
- Kullanılabilirlik

açılarından doğru bir yaklaşımdır.

6.2 Web Uygulamasında Veri Setinin Rolü

Veri seti:

- Web arayüzünde doğrudan gösterilmemektedir.
- Makine öğrenmesi modellerini eğitmek için kullanılmıştır.

- Modellerden elde edilen ilişkiler web uygulamasına aktarılmıştır.

Web uygulaması, kullanıcıdan alınan verileri bu öğrenilmiş ilişkilere göre değerlendirerek **kişiselleştirilmiş risk analizi** sunmaktadır.

6.3 Kullanıcı Etkileşimi

Kullanıcı:

1. Web sitesine giriş yapar.
2. Risk hesaplama formunu doldurur.
3. Sistem:
 - a. Risk seviyesini
 - b. Açıklayıcı nedenleri
 - c. Görsel ilerleme çubuğunu
 - d. Bilgilendirici kaynakları kullanıcıya sunar.

7. GÜVENLİK, OTURUM VE SEO

- Kullanıcı girdileri XSS saldırılarına karşı temizlenmiştir.
- Session kullanılarak kullanıcı etkileşimleri takip edilmiştir.
- Cookie kullanılarak son değerlendirme bilgileri saklanmıştır.
- SEO uyumlu meta etiketler eklenmiştir.

8. NESNE TABANLI PROGRAMLAMA (OOP)

PHP tarafında:

- Abstract sınıf kullanılmıştır.
- Kalıtım (inheritance) uygulanmıştır.
- Encapsulation prensipleri gözetilmiştir.
- Risk hesaplama işlemleri sınıf yapısına taşınmıştır.

9. SONUÇ VE GENEL DEĞERLENDİRME

Bu projede, sigara kullanımının KOAH ve depresyon üzerindeki etkileri hem analitik hem de uygulamalı olarak ele alınmıştır. XGBoost modeli, sağlık alanı için kritik olan hasta yakalama başarısı nedeniyle final model olarak seçilmiştir.

Proje, web tasarımı, veri analizi ve makine öğrenmesini bir araya getiren **çok disiplinli ve özgün bir çalışma** sunmaktadır.