

Araştırma Yöntemleri Final Ödevi

Fatma Tuğba Yeşil - 21936459, Defne Erkoç - 21935908

Hacettepe Üniversitesi

15.06.2023

Makalenin Yayınlandığı Dergi Hakkında Genel Bilgiler

Dergi Adı: Medical Records (Med Records)

Dergi İndeksi: SCI-E

Dergi Hakkında: Medical Records (Med Records), Tıbbi Kayıtlar Derneği (Tıbbi Kayıtlar Derneği) tarafından yayınlanan uluslararası, bağımsız, tarafsız ve çift-kör hakemlik ilkelerine uygun olarak yayınlanan açık erişimli süreli yayındır. 2019 yılında yayın hayatına başlamış olup, 2019 yılından itibaren tüm sayıları elektronik ortamda erişime açılmıştır. Derginin resmi dili İngilizce olup yılda üç kez Ocak, Mayıs ve Eylül aylarında yayımlanır. Makaleler İngilizce yazılmalı ve derginin gerekliliklerini karşılamalıdır.

Dergide yayınlanacak tüm yazılar araştırma ve yayın etiğine uygun olmalıdır. Medical Records (Med Records), Tıbbın tüm yönleriyle ilgili hem klinik hem de deneysel kaliteli materyalleri yayınlamayı amaçlar. Araştırma bulguları, teknik değerlendirmeler ve incelemelerle ilgili makaleleri içerir. Dergide tıbbın tüm yönleriyle ilgili orijinal makaleler yayınlamanın yanı sıra, tıptaki son gelişmeler hakkında İncelemeler, Klinik Perspektifler ve Editoryal makaleler de yer almaktadır. Derginin hedef kitlesi, tüm tıp disiplinlerine ilgi duyan veya bu alanlarda çalışan araştırmacılar, hekimler ve sağlık profesyonelleridir. Medical Records tarafından yayınlanan tüm yazıların tam içeriği mevcuttur

Derginin Erişim Linki:

<https://dergipark.org.tr/tr/pub/medr>

Makale İncelemesi

Makale Hakkında Genel Bilgiler

- ▶ **Makale Adı:** Early Detection of Coronary Heart Disease Based on Machine Learning Methods
- ▶ **Yazarlar:** Rüstem Yılmaz, Fatma Hilal Yağın
- ▶ **Makale Dili:** İngilizce
- ▶ **Makalenin Konusu:** Makalenin konusu, kalp hastalığının tahmin edici sınıflandırması için makine öğrenimi yöntemlerinin performanslarının karşılaştırılmasıdır. Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM) yöntemlerini karşılaştırmıştır.
- ▶ **Makalenin Amacı:** Makalenin amacı, Kalp hastalıklarının dünya çapında sağlık sistemleri üzerinde bir yük olmaya devam etmesi nedeniyle, makine öğrenme yöntemlerini kullanarak kalp hastalığı tespiti olağanüstü bir araştırma konusu olmuştur. Bu nedenle, bu çalışmada, koroner kalp hastalığının tahmin edici sınıflandırması için makine öğrenme yöntemlerinin performansları karşılaştırılmıştır.

- **Makalenin Önemi:** Bu makale, kalp hastalığının erken tespitinin insan hayatının uzun vadede kurtarılmasına yardımcı olduğunu vurgulamaktadır. Kalp hastalığıyla ilgili sağlık verilerinin işlenmesi, bu prosedürün keşfedilmesini sağlamıştır. Makine öğrenimi algoritmaları, ham verilerin işlenmesinde kullanılarak kalp hastalığı için yeni ve orijinal öneriler sunabilir. Kalp hastalığı prognozu, tıpta en zor ve önemli konulardan biri olarak kabul edilmektedir. Durum erken teşhis edilirse, mortalite oranı kontrol altına alınabilir ve önleyici tedbirler mümkün olan en kısa sürede uygulanabilir. Bu çalışmada, makine öğrenimi yardımıyla kalp hastalığı tahmini için bir model önerilmektedir. Bu amaçla, RF, SVM ve LR olmak üzere üç makine öğrenimi algoritması kullanılmıştır.

Makalenin Metodolojisi

Çalışmada kullanılan kalp hastalığı veri setinin kaynağı ve bileşenleri hakkında bilgi verilmektedir. Veri seti, Cleveland, Hungarian, Switzerland, Statlog (Heart) Data Set ve Long Beach VA veri setlerinin 11 öz niteliği birleştirilerek oluşturulmuştur. Veri birleştirme işlemiyle, mevcut kalp hastalığı veri setlerine kıyasla oldukça büyük bir veri seti elde edilmiştir. Veri setindeki hastaların 281'i kadın ve 909'u erkektir. Kadınların ortalama yaşı 53 ± 10 iken erkeklerin ortalama yaşı 54 ± 9 'dur.

Makalede Kullanılan Materyal ve Metot

Çalışmada koroner kalp hastalığının sınıflandırılması için Rasgele Orman (RF), Lojistik Regresyon (LR) ve Destek Vektör Makinesi (SVM) algoritmaları ile üç farklı model oluşturulmuştur.

Hiperparametre optimizasyonu için 3 tekrarlı 10 katlı tekrarlı çapraz doğrulama yöntemi kullanıldı. Modellerin performansı Doğruluk, F1 Skoru, Seçicilik, Duyarlılık, Pozitif Tahmin Değeri, Negatif Tahmin Değeri ve Karışıklık Matrisi (Sınıflandırma matrisi) temel alınarak değerlendirilmiştir

Sonuç ve Tartışma

Bu makalede, kalp hastalığı tahmininde makine öğrenimi yöntemlerinin performansı karşılaştırılmıştır. RF, SVM ve LR olmak üzere üç farklı makine öğrenimi algoritması kullanılarak bir kalp hastalığı tahmin modeli oluşturulmuştur. Sonuçlar, RF algoritmasının diğer yöntemlere (LR, SVM) göre kalp hastalığı tahmininde daha iyi performans gösterdiğini göstermektedir. Makalede, literatürde kalp hastalıklarının tahminine yönelik birçok çalışma olduğu belirtilmektedir. Bir çalışmada J48, K En Yakın Komşu (KNN), Karar Ağacı ve Naive Bayes (NB) gibi algoritmalar kullanılarak kalp hastalığı tespiti yapılmış ve en yüksek doğruluk oranı (%83,732) J48 ile elde edilmiştir. Başka bir makalede, veri madenciliği sınıflandırma tekniklerini kullanarak bir kişinin kalp hastalığı olma olasılığı yüzdesi tahmin edilmiştir. Çalışmada, Karar Ağacı, KNN ve Naive Bayes (NB) algoritmaları kullanılmış ve NB, kalp hastalıklarını sınıflandırmada en yüksek doğruluk oranı %73.7 elde etmiştir.

2017 yılında, Hend Mansoor ve diğerleri, kardiyovasküler hastalığı olan hastaların risk maruziyetini değerlendirmek için LR ve RF sınıflandırma algoritmalarının performansını incelemişlerdir. LR Modelinin doğruluk oranı %89 iken RF Modelinin doğruluk oranı %88" olarak bulunmuştur. Başka bir makalede, Random Forest algoritması kullanılarak kalp hastalığı sınıflandırılmış ve test veri setinde yeni ve bilinmeyen hastalar %84,448 doğruluk oranıyla sınıflandırılmıştır. Yukarıda bahsedilen çalışmaların çoğunda elde edilen performans ölçütleri, bu çalışmadaki performanstan daha düşüktür. Bu çalışmada, kalp hastalığını sınıflandırmak için LR, SVM ve RF algoritmaları için hiperparametre optimizasyonu yapılarak en uygun modelin seçilmesine yardımcı olunmuştur. Diğer bir deyişle, hiperparametre optimizasyonu, makine öğrenimi modellerinde en uygun modeli oluşturmak için önemli bir adımdır.

Çalışmada birçok klasik makine öğrenimi algoritması denendikten sonra en yüksek performansa sahip üç algoritma üzerinde çalışmalar devam etmiştir. Bu çalışma, literatürden daha başarılı bir şekilde koroner arter kalp hastalığını tahmin etmiştir. RF modelinden elde edilen Doğruluk, Özgüllük, F1-skoru, Duyarlılık, Negatif tahmin edici değer ve Pozitif tahmin edici değer kriterleri sırasıyla 0.929, 0.929, 0.928, 0.928, 0.929 ve 0.928 olarak hesaplanmıştır. Sonuç olarak, bu çalışma, kabul gören ve uygulanması kolay olan ML algoritmaları arasında en iyi teknik olan önerilen RF algoritmasının, en azından bu veri kümesi için iyi performans gösterdiğini bulmayı amaçlamaktadır. Bu nedenle, RF algoritması, gelecekteki kalp ve farklı hastalıkların tahmin modellerinin geliştirilmesi için önerilebilir.

Makalenin Artı Yönleri

- ▶ **Makine öğrenimi algoritmalarının kullanımı:** RF, SVM ve LR gibi üç farklı makine öğrenimi algoritması kullanılarak kalp hastalığı tahmini için bir model önerildi. Farklı algoritmaların performansını karşılaştırmak ve en iyi sonuçları elde etmek için çeşitli yöntemlerin kullanılması.
- ▶ **Performans değerlendirmesi:** Accuracy, Specificity, F1-score, Sensitivity, Negative predictive value ve Positive predictive value gibi performans ölçütleri kullanılarak modellerin değerlendirilmesi. RF modelinin diğer yöntemlere göre daha iyi performans gösterdiğinin ortaya konulması.

- ▶ **Hyperparameter optimizasyonu:** LR, SVM ve RF algoritmaları için hyperparameter optimizasyonu yapılması. En uygun parametrelerle modellerin oluşturulması ve daha yüksek performans elde edilmesi.
- ▶ **Yüksek performans:** Elde edilen sonuçların literatürdeki diğer çalışmalardan daha yüksek performans göstermesi. RF modelinden elde edilen Accuracy, Specificity, F1-score, Sensitivity, Negative predictive value ve Positive predictive value değerlerinin yüksek olması.

Makalenin Eksi Yönleri

- ▶ **Diğer algoritmalara karşı sınırlı karşılaştırma:** RF, SVM ve LR algoritmaları kullanılarak bir model önerilmesi ve RF algoritmasının daha iyi performans gösterdiğinin belirtilmesi. Diğer popüler makine öğrenimi algoritmalarının (J48, KNN, Naive Bayes gibi) performansıyla karşılaştırma yapılmaması.
- ▶ **Veri setinin karakteristikleri hakkında yetersiz bilgi:** Veri setinin cinsiyet dağılımı, yaş ortalamaları gibi bazı temel bilgilerin verilmesi. Diğer önemli demografik veya klinik özellikler hakkında yeterli bilgi sağlanmaması.

- ▶ **Dışarıdan doğrulama eksikliği:** Sunulan sonuçların sadece kullanılan veri seti üzerinde elde edilen sonuçlara dayanması. Önerilen modelin başka veri setleri veya gerçek dünya koşullarında ne kadar etkili olduğu hakkında daha fazla dışarıdan doğrulama yapılması gerekliliği.
- ▶ **Makalenin genel sınırlamaları:** Veri setinin temsiliyeti, yöntemlerin optimizasyon süreci veya kullanılan performans ölçütleri gibi faktörlerdeki sınırlamalar.

Sonuç

Makale, kalp hastalığı tahmini için RF, SVM ve LR gibi makine öğrenimi algoritmalarının kullanıldığı bir model önermektedir. Makalenin artı yönleri arasında farklı algoritmaların kullanılması, performans değerlendirmesi, hyperparameter optimizasyonu ve yüksek performans yer almaktadır. Bununla birlikte, diğer algoritmalara karşı sınırlı karşılaştırma, veri setinin karakteristikleri hakkında yetersiz bilgi, dışarıdan doğrulama eksikliği ve makalenin genel sınırlamaları gibi eksi yönler de bulunmaktadır. Daha kapsamlı bir karşılaştırma, daha fazla dışarıdan doğrulama ve sınırlamaların dikkate alınması, gelecekteki araştırmaların odak noktaları olabilir.

Kaynaklar

1. İST480-ARAŞTIRMA YÖNTEMLERİ-LATEX DERS NOTLARI
2. <https://dergipark.org.tr/tr/pub/medr/issue/67333/1011924> (2022) *Medical Records*.
3. T. Watanabe, K. Ando, H. Daidoji, Y. Otaki, S. Sugawara, M. Matsui, et al., "A randomized controlled trial of eicosapentaenoic acid in patients with coronary heart disease on statins," *Journal of cardiology*, vol. 70, pp. 537-544, 2017.
4. N. P. Paynter, R. Balasubramanian, F. Giulianini, D. D. Wang, L. F. Tinker, S. Gopal, et al., "Metabolic predictors of incident coronary heart disease in women," *Circulation*, vol. 137, pp. 841-853, 2018.