BIL545 Vize Ödevi

Teslim tarihi: 23.11.2024 23.59

**Önemli Not: Dosyanızı Numara\_ad\_soyad.docx olarak gönderiniz.**

**Örnek : 123456789\_Ersin\_Ersoy.docx**

**Ödeviniz iki sorudan oluşmaktadır. Seçimli değildir. İki soruya da cevap vermeniz beklenmektedir. Ödevinizi bu dokümanı kullanarak, yazım kurallarına dikkat ederek ve dosya ismini yukarıda belirttildiği şekilde değiştirerek gönderiniz.** Teslim edeceğiniz ödevinizin mevcutta bu şablonda olduğu gibi **toplam 4-5 sayfa olması gerekmektedir**.

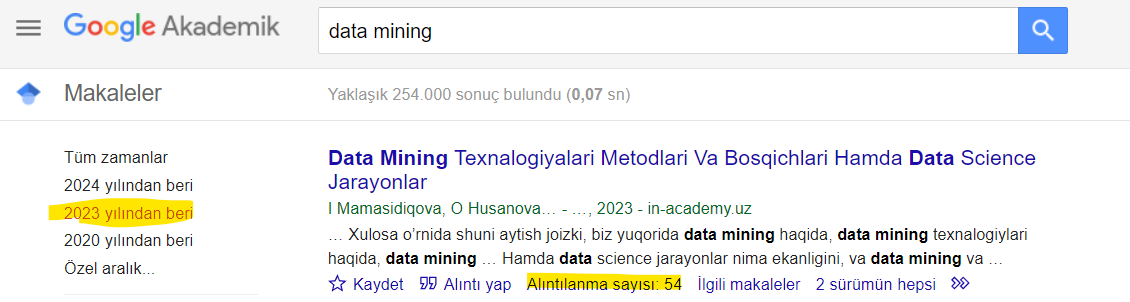
1. **Soru:** Bir çalışma alanı(Sigorta, Bankacılık, Finans, Yazılım mühendisliği, Güvenlik vb.) seçiniz. Seçtiğiniz alan ile ilgili aşağıdaki adımları uygulayarak ilk soruyu cevaplayınız.

1.1) <http://ccr.sigcomm.org/online/files/p83-keshavA.pdf> adresindeki “How to Read a Paper” makalesini okuyunuz. Gerekirse Türkçeye çevirmek için google translate ve türevlerini kullanabilirsiniz.

1.2) Google scholar, Dergipark ya da başka bir akademik\* veri tabanı üzerinden BELİRLEDİĞİNİZ alanda **Data mining ile ilgili en az 10 atıf almış**\*\* ve **en eski 2023\*\* yılında yazılmış** 2 yayın belirleyip her bir yayın için ayrı istenen bilgileri ayrı ayrı oluşturulmuş tabloları doldurarak cevaplayınız.

\*Bazı akademik veri tabanları: Google Scholar, ResearchGate, IEEE, ACM, ScienceDirect, Springer, Dergipark

\*\*



1. **Soru Yanıtı:**

**Seçtiğiniz çalışma alanı: … (SAĞLIK)**

**1. Makale Yanıtı**

|  |  |
| --- | --- |
| **İstenen Bilgiler** | **Makale 1** |
| **Yayın Adı:** | Harnessing Big Data Analytics for Healthcare: A Comprehensive Review of Frameworks, Implications, Applications, and Impacts. |
| **Yazar(ları):** | Awais Ahmed, Rui Xi, Mengshu Hou, Syed Attique Shah, Sufian Hameed. |
| **Yazım Yılı:** | 2023 |
| **Atıf Sayısı:** | 19 Atıf |
| **Yayınlandığı Yer**(Dergi ise derginin ismi, konferans ise konferansın ismi vb.)**:** | IEEE Access Dergisi |
| **Makale Türü** (Category)**:**  (  <https://www.makaletercume.com/bilimsel-makale-turleri/>  linkindeki makale türlerinden uygun olanı belirtiniz) | Meta- Analiz Derleme |
| **Genel olarak makalenin ne ilgili olduğu, yapılan çalışma ile ilgilli detaylı bilgi veriniz** | Makale, büyük veri analitiğinin sağlık sektöründeki uygulama alanlarını, getirdiği zorlukları ve potansiyel faydalarını ele almaktadır. Çalışma, veri kaynaklarının entegrasyonu, işlenmesi ve analizi için kullanılan teknolojilere (örneğin, veri madenciliği, makine öğrenimi ve doğal dil işleme) odaklanır. Ayrıca, sağlık hizmetlerinde büyük veri teknolojilerinin nasıl kullanıldığını çeşitli örneklerle açıklamaktadır.  Özellikle veri madenciliği, büyük sağlık verilerinden anlamlı bilgiler çıkarma süreci olarak tanımlanır. Makale, data mining tekniklerinin teşhis modellerinin geliştirilmesinden, klinik karar destek sistemlerine kadar birçok uygulamada nasıl kullanıldığını detaylandırır. Örnek olarak, hasta sonuçlarını iyileştirmek için regresyon analizi, kümeleme ve sınıflandırma algoritmaları tartışılmıştır. |
| **Makale herhangi bir probleme çözüm sağlıyor mu? Sağlıyorla çözüm sağladığı problem ya da problemler nelerdir?** | Evet, makale sağlık sektöründeki şu temel sorunlara çözüm önerileri sunuyor:  Veri Yönetimi ve Entegrasyonu: Heterojen kaynaklardan gelen sağlık verilerinin entegrasyonu için standart protokoller önerilmektedir. Özellikle veri madenciliği algoritmalarının entegrasyonu kolaylaştırdığı belirtilmiştir.  Klinik Karar Destek Sistemleri: Data mining ile hasta teşhis modelleri geliştirilerek, yanlış teşhis oranlarının azaltılması ve daha doğru tedavi yöntemlerinin belirlenmesi hedeflenmiştir.  Tahminsel Analiz: Sağlık maliyetlerini düşürmek ve tedavi süreçlerini optimize etmek için tahminsel modelleme önerilmektedir. Bu modellerin temelinde veri madenciliği algoritmaları yatmaktadır.  Veri Güvenliği: Büyük veri analitiğinde kullanılan yöntemlerin, özellikle hasta bilgilerinin anonimleştirilmesi ve güvenliğinin sağlanmasında rol oynadığı ifade edilmiştir. |
| **Makaledeki varsayımlar nelerdir?** | Data mining teknikleri, büyük sağlık verilerinden anlamlı bilgiler çıkararak hasta sonuçlarını iyileştirebilir.  Farklı kaynaklardan gelen sağlık verileri entegre edilip analiz edildiğinde, sağlık hizmetlerinin verimliliği artırılabilir.  Veri madenciliği algoritmalarıyla oluşturulan modeller, hastalık teşhisinde ve önleyici tedavide etkin şekilde kullanılabilir.  Gizlilik ve güvenlik önlemleri sağlandığında büyük veri analitiği daha yaygın hale gelebilir. |
| **Makaledeki varsayımlar doğru mu?** (Correctness) | Varsayımlar genelde doğruluk payına sahip olsa da bazı durumlarda pratik zorluklarla karşılaşılabilir:  Doğruluğu: Data mining’in sağlık sektöründe hastalık teşhisinde başarılı sonuçlar verdiği, özellikle sınıflandırma algoritmalarının ve kümeleme yöntemlerinin etkili olduğu kanıtlanmıştır.  Zorluklar: Farklı veri kaynaklarından gelen sağlık verilerinin entegrasyonu sırasında yaşanan veri kalitesi problemleri, bazı varsayımların pratikte zorlayıcı olabileceğini göstermektedir. Ayrıca, data mining modellerinin yüksek doğruluk sağlaması için verilerin temizlenmesi ve uygun şekilde hazırlanması gereklidir. Veri entegrasyonu sırasında farklı kaynaklardan gelen veriler arasında uyumsuzluklar oluşabilir ve bu, analizlerin doğruluğunu etkileyebilir.  Veri güvenliği konusundaki önerilen teknolojiler, yasal düzenlemelere ve mevcut altyapılara bağlı olarak uygulanabilirlik zorluklarıyla karşılaşabilir.  Veri kalitesindeki eksiklikler, tahminsel modellerin başarısını sınırlayabilir. |
| **Makalenin ana katkıları nelerdir?** (Contributions) | Sağlıkta büyük veri ekosistemini detaylandırarak, veri madenciliği uygulamalarının potansiyelini açıklamıştır.  Data mining yöntemlerinin kullanıldığı başarılı sağlık projelerini analiz ederek, bu projelerin hasta sonuçlarını nasıl iyileştirdiğini göstermiştir.  Özellikle teşhis modelleri, hastalık tahmin sistemleri ve klinik veri analizinde veri madenciliğinin rolüne dikkat çekmiştir.  Gelecekteki araştırmalar için veri madenciliğiyle ilgili açık problemleri ve fırsat alanlarını tanımlamıştır.  Sağlık verilerinin gizliliğini ve güvenliğini sağlamak için veri madenciliği tekniklerinin nasıl kullanılabileceğine dair öneriler sunmuştur |
| **Makale iyi yazılmış mı?** (Clarity) | Evet, makale sistematik ve net bir şekilde yazılmıştır. Veri madenciliği ve diğer büyük veri analitiği tekniklerinin sağlık sektöründeki rolü detaylı bir şekilde açıklanmıştır. Özellikle uygulama örnekleri ve literatür incelemesi, makaleyi teorik olduğu kadar pratik açıdan da değerli hale getirmiştir. Ancak, veri madenciliği algoritmalarının gerçek hayatta uygulanmasında karşılaşılan zorluklar üzerine daha fazla detay verilmesi faydalı olabilirdi.  Bu analiz, veri madenciliği dahil olmak üzere makalenin temel konularını detaylı şekilde ele almıştır. Sağlık veri analitiği alanında çalışan öğrenciler ve araştırmacılar için rehber niteliğinde bir kaynak olarak değerlendirilebilir. |

**2. Makale Yanıtı:**

|  |  |
| --- | --- |
| **İstenen Bilgiler** | **Makale 2** |
| **Yayın Adı:** | Big Data Analytics Cloud based Smart IoT Healthcare Network |
| **Yazar(ları):** | Joseph Bradley,Joel Barber,Doug Handler |
| **Yazım Yılı:** | 2023 |
| **Atıf Sayısı:** | 10 |
| **Yayınlandığı Yer**(Dergi ise derginin ismi, konferans ise konferansın ismi vb.)**:** | Proceedings of the 7th International Conference on Trends in Electronics and Informatics (ICOEI 2023) |
| **Makale Türü** (Category)**:**  (  <https://www.makaletercume.com/bilimsel-makale-turleri/>  linkindeki makale türlerinden uygun olanı belirtiniz) | Araştırma Makalesi (Orijinal Makale) |
| **Genel olarak makalenin ne ilgili olduğu, yapılan çalışma ile ilgilli detaylı bilgi veriniz** | Makale, sağlık sektöründe “büyük veri” ve “nesnelerin interneti” (IoT) uygulamalarını ele alır. “Smart Healthcare Network” olarak adlandırılan bir sistem önerilmektedir ve bu sistem, sağlık verilerinin büyük veriyi kullanarak analiz edilmesini amaçlar. Sistem, IoT cihazlarıyla toplanan biyomedikal verileri toplar, bulutta depolar ve analiz eder. Sonuç olarak, bu sayede hastalıkların daha doğru bir şekilde izlenmesi ve tedavi edilmesi sağlanır. Ayrıca, bu makale, veri analitiği ve makine öğrenmesi (machine learning) kullanarak uzaktan hasta izleme sistemleri geliştirilmesini hedefler.  Yapılan çalışma ise:  Sağlık sektörü için bir “Big Data” işleme ve IoT tabanlı sistem öneriyor. Bu sistem, çeşitli biyomedikal sensörlerden (kan şekeri ölçüm cihazı, kan basıncı sensörleri, hareket sensörleri vb.) veri toplar ve bu verileri büyük veri işleme platformlarında (özellikle Apache Hadoop) işler. Bu veriler, hastaların sağlık durumlarını izlemek için kullanılır ve verilerin analiz edilmesinin ardından sağlık profesyonellerine öneriler sunulur. Çalışma, özellikle uzaktan hasta izleme ve tedavi süreçlerine yönelik bir çözüm öneriyor. |
| **Makale herhangi bir probleme çözüm sağlıyor mu? Sağlıyorla çözüm sağladığı problem ya da problemler nelerdir?** | Evet, makale sağlık sektörü için önemli problemlere çözüm sağlıyor. Bu çözüm sağladığı temel problemler şunlardır:  Veri Toplama ve Depolama Problemi: IoT cihazları tarafından toplanan büyük miktarda biyomedikal verinin güvenli bir şekilde depolanması ve analiz edilmesi.  Uzaktan Hasta İzleme Problemi: Sağlık profesyonelleri, hasta verilerini uzaktan izleyerek erken teşhis ve daha doğru tedavi sağlayabilirler.  Yüksek Maliyetler ve Zaman Kaybı: Sağlık verilerinin bulut depolama ve büyük veri işleme yöntemleri kullanılarak işlenmesi, sağlık sektörü için maliyetleri azaltır ve tedavi sürelerini kısaltır.  Veri Hatalarının Azaltılması: Sağlık verilerinin işlenmesinde kullanılan yapay zeka (AI) ve makine öğrenmesi algoritmaları, yanlış verilerin tespit edilmesini sağlar. Veri madenciliği bu doğrultuda, hatalı verilerin bulunup düzeltilmesinde önemli bir rol oynar. |
| **Makaledeki varsayımlar nelerdir?** | Makaledeki varsayımlar şunlardır:  IoT Cihazları ve Sensörler Sağlıklı ve Güvenilir Çalışır: Makale, IoT cihazlarının doğru veri toplaması gerektiğini varsayar.  Bulut Depolama Yöntemleri Veriyi Güvenli Bir Şekilde Saklayabilir: Verilerin bulut tabanlı sistemlerde güvenli bir şekilde depolanabileceği varsayılır.  Büyük Veri Analitiği Doğru Sonuçlar Verecek: Apache Hadoop ve MapReduce gibi büyük veri işleme yöntemlerinin doğru analiz yapabileceği varsayılır.  Yapay Zeka ve Makine Öğrenmesi Algoritmaları Veriyi İyi Yorumlayacak: Verilerin analizi için kullanılan yapay zeka ve makine öğrenmesi algoritmalarının doğru sonuçlar üretmesi beklenir. |
| **Makaledeki varsayımlar doğru mu?** (Correctness) | Bu varsayımlar genellikle doğrudur, ancak uygulamada bazı zorluklar olabilir:  IoT Cihazları ve Sensörlerin Güvenilirliği: IoT cihazlarının güvenilirliğini sağlamak, özellikle hasta verilerinin doğru şekilde toplanması açısından kritik bir konu olabilir. Sensörler bazen hatalı ölçümler yapabilir, bu da yanlış analizlere yol açabilir.  Bulut Depolama ve Güvenlik: Verilerin bulutta depolanması güvenli olsa da, sağlık verileri hassas bilgiler içerdiği için bu verilerin güvenliği büyük bir risk teşkil eder. Bu nedenle şifreleme ve veri koruma önlemleri çok önemlidir.  Veri Analizinin Doğruluğu: Büyük veri işleme yöntemleri genellikle doğru sonuçlar verir, ancak çok büyük veri setlerinde bazı istisnalar olabilir. Yapay zeka ve makine öğrenmesi algoritmalarının doğruluğu, verinin kalitesine ve algoritmanın eğitim sürecine bağlıdır. |
| **Makalenin ana katkıları nelerdir?** (Contributions) | Makalenin ana katkıları şunlardır:  Yeni bir Akıllı Sağlık Sistemi Önerisi: IoT cihazları, büyük veri ve yapay zeka yöntemlerini birleştirerek akıllı sağlık ağları için bir çözüm önerisi sunar.  Uzaktan İzleme ve Tedavi: Sağlık profesyonellerine hastaların sağlık verilerini uzaktan izleme ve daha hızlı karar alma imkanı tanır.  Büyük Veri ve Bulut Depolama Kullanımı: Sağlık verilerinin büyük veri işleme yöntemleri ve bulut depolama sistemleriyle analiz edilmesi, maliyetleri azaltma ve verimliliği artırma potansiyeline sahiptir.  Makine Öğrenmesi ile İyileştirilmiş Analizler: Yapay zeka ve makine öğrenmesi algoritmaları ile hasta verilerinin daha doğru ve verimli bir şekilde analiz edilmesini sağlar. |
| **Makale iyi yazılmış mı?** (Clarity) | Makale genel olarak iyi yazılmıştır ve veri madenciliği (data mining) yöntemleri, sağlık IoT sistemlerinin verimliliğini artırmak amacıyla etkili bir şekilde entegre edilmiştir. Ancak, veri madenciliği süreçlerinin detayları konusunda daha fazla örnek ve açıklama sağlanması, özellikle veri hazırlığı, model seçimi ve sonuçların doğruluğu konularında, bu alanda daha az deneyime sahip okuyucuların konuyu daha iyi anlamasına yardımcı olabilir. Ayrıca, kullanılan algoritmaların nasıl optimize edildiği veya hangi parametrelerin önemli olduğu gibi ek bilgiler, uygulama açısından daha derin bir bakış açısı kazandırabilir |

1. **Soru:** Aktif olarak çalıştığınız veya ilgi duyduğunuz bir alanda ne gibi bir veri madenciliği uygulamasına ihtiyaç bulunduğunu, bu tür bir uygulama olup olmadığını (yoksa bir proje önerisi şeklinde detaylandırabilirsiniz, varsa mevcut durum hakkında bilgi vererek ne gibi eklemeler yapılabileceğini açıklayabilirsiniz) açıklayınız.
2. **Soru Yanıtı (Toplam 2 sayfa):**

Sağlık sektöründe öncü olan Acıbadem hastanelerine bağlı Acıbadem Teknoloji şirketinin AR-GE departmanında Junior Veri bilimcisi olarak görev yaptım. Yaptığımız ve üzerinde çalıştığımız çeşitli projeler oldu. Bu projelerin genel amacı hasta verilerini kullanarak doktorlara teşhis ve tedavi planlanmasında yardımcı sistemler tasarlamaktı. Hastane ziyaretlerimiz ve doktorlardan yapmış olduğumuz istişareler sonucunda projesini tasarladığım şu an güncel olarak geliştirmeleri Acıbadem hastanelerinde devam eden projeyi anlatacağım.  
  
“Yoğun bakım ünitelerinde ventilasyon cihazına bağlı olan hastaların ventilasyon cihazından ayrılma” tahmini üzerine. Peki neden ventilasyon cihazına ihtiyaç duyarız? Ventilasyon cihazına bağlı hastalar, kendi başlarına yeterli solunum yapamadıkları için bu cihaza bağlanır. Başlıca nedenler arasında solunum yetmezliği, ağır zatürre, KOAH gibi akciğer hastalıkları, travma, ameliyat sonrası durumlar ve COVID-19 gibi akut solunum yolu hastalıkları bulunur. Bu hastaların gözlem altında tutulması ve takibi hastanın gidişatını etkileyen hayati önem taşımaktadır. Bu hastaların takibini yapabilmek adına hastane monitörizasyon cihazları ile çeşitli hastalar üzerinden veri toplamak mümkün. Peki bu veriler bize hasta hakkında detaylı bilgi verebiliyor mu? Tıbbi değerlendirmeler sonucunda hastalar hakkında yeterli bilgileri toplayabildiğimizi kararlaştırdık. Gerekli olan parametreler şu şekilde; **oksijen satürasyonu**, **kalp hızı**, **kan basıncı**, **solunum hızı**, **end-tidal CO₂** bu parametreli monitörizasyon cihazları ile elde edebiliyoruz. Ancak bu parametreler bize hatanın ventilasyon durumunu yorumlamada yardımcı olsa da ventilasyondan ayrılma kararını vermekte yardımcı olmuyor. Ek cihaz ve manuel testler ile **arteriyel kan gazları**, **laktat seviyesi** ve **glukoz düzeyi** bizlere daha anlamlı sonuçlar çıkarmak için imkân verebiliyor. Veri setinin hazırlanması belki de işin en zor kısmı çünkü bu veri seti zaman serisi formatlı, ventilasyon cihazına bağlı kaldığı süre boyunca 3 saniyede ya da 5 saniyede bir veri ataması yapılıyor. Bazı durumlarda cihazdan veri akışı sağlanmıyor. Bu sancılı sürecin sonunda hazırlanmış veri setimiz ile analizlerimize başlayabilmemiz mümkün. Veri seti çeşitli hastalardan toplanan; kimi hastadan toplamda 30 günlük kimi hastalardan 4-5 gün kimi hastalardan ise sadece 20 saatlik verilerden oluşmaktadır. Ek olarak sorulması gerek en kritik sorulardan biri de veri setinde çok farklı gruplara ait hastalar bulunmakta eğitilecek bir modelin genellenebilir olması en önemli adımlardan peki elimizdeki bu veri seti ile eğiteceğimiz model ne kadar genellenebilir olacak?

Buradaki ilk adım, hastaların ventilasyon cihazından ayrılma süreçlerini anlamak adına Keşifsel Veri Analizi (EDA) ile verilerimizi incelemek olacak. Bu analiz sırasında, her hastanın iyileşme ve kötüleşme dönemlerine ilişkin örüntüler çıkarabiliriz. Elde edilen bu örüntüler, modelin hangi parametrelerin ayrılma sürecinde belirleyici rol oynadığını anlamamızda bize rehberlik eder. Ventilasyondan ayrılma tahmin modelimizi geliştirmek için ilk adım, elde edilen verilerdeki trend ve desenleri anlamak olacak. Zaman serisi analizi yaparak her bir hastanın ventilatör cihazına bağlı kaldığı süredeki oksijen satürasyonu, kalp hızı, kan basıncı, solunum hızı ve end-tidal CO₂ gibi parametrelerin nasıl değiştiğini gözlemleyeceğiz. Örneğin, oksijen satürasyonunda düzenli bir artış ya da solunum hızında belirgin bir iyileşme eğilimi, hastanın kendi başına nefes alabilecek duruma yaklaştığını işaret edebilir. Aynı şekilde, kan basıncında istikrar veya end-tidal CO₂ seviyesindeki düşüş gibi veriler de iyileşme süreci hakkında bize kritik ipuçları verebilir.

Bu tür verilere dayanarak, zaman serisi analizine yönelik özel algoritmalar kullanmayı planlıyoruz. Zaman serisi analizinde, veri noktalarının ardışık düzenleri arasındaki ilişkileri keşfetmek çok önemli. Örneğin, her bir parametredeki değişim hızını (bir önceki zamana göre nasıl arttığı ya da azaldığı) anlamak, ventilasyondan ayrılma kararı için önemli bir gösterge olabilir. Bu durumda, LSTM (Long Short-Term Memory) veya GRU (Gated Recurrent Unit) gibi tekrarlayan sinir ağları (RNN - Recurrent Neural Networks) bu tür ardışık veri analizleri için oldukça uygun. LSTM ve GRU modelleri, zaman içindeki değişimlere duyarlı olduğundan, hastanın ventilasyon cihazına bağlı kalma sürecindeki karmaşık örüntüleri ve hastanın genel iyileşme eğilimini tahmin edebilir. Eğitilecek olan multi-variate lstm model her kolon için 1 saat sonraki değeri tahmin edebilecek.

|  |  |
| --- | --- |
| TIME | O2\_SATURASYON |
| 01.01.2024 08.00 | 95 |
| 01.01.2024 09.00 | 94 |
| 01.01.2024 10.00 | 93 |
| 01.01.2024 11.00 | 92 |
| 01.01.2024 12.00 | 91 |

Tablo 1

Tabloda örnek veri akışını gösteren, her saat başında değişen oksijen satürasyonu değeri görülmektedir. LSTM modelinin mantığı tahmin edilecek saat dilimini kendinden önce gelen saatlik aralıklardan eğitim yapar ve öğrenir. Model sonucunda tüm kolonlar için saatlik tahminler yapılabilir. Zaman serisi analizine ek olarak, modelimizi manuel ölçümlerle desteklemek de doğru tahmin yapmamıza yardımcı olacaktır. Örneğin, arteriyel kan gazı, laktat seviyesi ve glukoz düzeyi gibi veriler, doktorların manuel olarak ölçümlediği ve hastanın solunum sisteminin işlevselliğine dair değerli bilgiler sunan parametrelerdir. Bu veriler, modelin karar verme sürecini daha sağlam hale getirerek hastanın ventilasyon cihazından ayrılmaya uygun olup olmadığını belirlemede doğruluğu artıracaktır.