**[1] Büyük veride sağlık bilişimi ve analitiği**

Keywords: healtcare informatics, healtcare technology, data analytics, big data

Sağlık hizmetlerinde büyük veri bazı karmaşık sorunları da beraberinde getiriyor. Tekdüze olmayan veri dağıtımı ve paralel işleme gibi mevcut analitik yöntemlerle verimsiz bir şekilde ele alınan çok sayıda değişken mevcut, büyük verilerden bilgi elde etmek için bir sağlık sistemi için geleneksel olmayan ve olgun veri depolama, yönetim, analiz ve görselleştirme araçları ve teknikleri içerir.

Web 1.0’da hastaneler veya sağlık kuruluşları, hastalarla herhangi bir etkileşim olmaksızın bilgi üretir. Sağlık bağlamında web 1.0 amacı, çevrimiçi bir hizmet varlığı oluşturmaktı.

Sağlık bilişimi ve analitiği 1.0 farklı istatiksel analitikler içerir; sağlık verilerinin sınıflandırılması, segmentasyonu, kümelenmesi, analizi için araçlar ve veri madenciliği için teknikler sunar. Bu tekniklerle sorgulama, raporlama ve veri madenciliği görselleştirmesi sağlayabiliyoruz.

Web 2.0, sosyal etkileşimde ortak çıkarları olan büyük küresel kalabalıkların yönetilmesine ve bir araya getirilmesine izin verir. Ayrıca, Web 2.0 yapılandırılmamış veri kaynaklarıyla(sosyal medya beslemeleri) ilgilenebilir bir veri tabanında veriler saklanır. Böylece Web 2.0’da hastalar bilgilerini paylaşabilirler. Web 2.0 teknolojileri, sağlık /ilaç ve gelişmiş veri madenciliği teknolojileri, veri ve ağ odaklı sosyal medyanın açılış döneminde sağlık bilişimi ve analitiği araştırmasının yeni ve heyecan verici bir çağını başlatmıştır.

Sağlık bilişimi ve analitiği 2.0 uygulamaları, farklı türde sağlık hizmeti için hastalardan hem de profesyonellerden gelen büyük hacimli tepkileri ve duyguları sistematik olarak bir araya getirebilir.

Sağlık bilişimi ve analitiği 3.0 terimi daha geniş bir kavramı temsil eder ve ortaya çıkan tüm kavramları içeren bir şemsiye terimidir. Birden fazla kaynaktan gelen veriler ses çıkarmadan yeni bir çağın, büyük verinin ortaya çıkmasına neden olmuştur. Sağlık alanında teşhis ve cerrahi departmanlarının her bölümüne inmiş bulunmaktadır. Sağlık bilişimi ve analitiğinin üç ana teknik akışı belirlenmiştir. Bunlar yazılım aracısı, makine öğrenimi ve bulut bilişimdir. Daha farklı akışlar altındaki teknikler zamanla gelişmektedir. Akıllı ajanlar da büyük hacimli veri kaynaklarından sağlık bilgilerinin alınması gibi farklı görevlerde kullanılmak üzere geliştirilmiştir.

Gelişmiş sağlık sistemlerinde tanı ve hastalık seyriçözmeye yardımcı olabilecek makine öğrenimi yöntemlerinin, araçlarının ve tekniklerinin uygulanmasında da ciddi bir artış olmuştur. Makine öğrenimi ile klinik parametrelerin performansını ve bunların hastalık ilerlemesi tahmini, tedavi veya cerrahi için karar desteği, ortaya çıkan araştırma ve uygulamalardan bilgi çıkarma ve genel sağlık sistemi yönetimi gibi çeşitli kombinasyonların performansını artırmak için kullanılıyor. Makine öğrenimi yaklaşımı, heterojen sağlık verilerini analiz etmek için de kullanılmaktadır. Makine öğrenimi tekniklerden PredictSNP **ve** CanPredicthastalıkla ilgili mutasyon tahmini için iyi bilinen araçtır ve tahmin doğruluğu %90’dan fazladır. Kullanılan madencilik algoritmaları (K-Means, SVM, Naive Bayes gibi) ölçeklenebilir ve hesaplamalı bir şekilde kullanılmaktadır. Veri madenciliğinde temel olarak sınıflandırma, kümeleme, regresyon, ilişkilendirme analizi ve ağı kapsar. Tıp alanında veri kaynakları ikiyi ayrılır bunlar ana sınıflardan ilki olan dahili kaynak(klinik karar destek sistemleri, bilgisayarlı sağlayıcı sipariş girişi vs.) ve dış kaynaklardır(devlet kayıtları, sigorta kayıtları vs.). Bu makine öğrenimi teknikleriyle toplanan bilgilerdeki kalıpları sınıflandırma ve tahminleri ortaya çıkararak maliyetleri azaltabiliriz.

Sağlık alanında, tıbbi görüntüler önemli bir kaynaktır. Bu nedenle görüntü analizi sağlık bilişimi ve analitiği içinde aktif bir araştırma alanı haline geldi. Tıbbi görüntü analizinde segmentasyon, kümeleme, bilgi çıkarma ve entegrasyon yöntemleri gibi teknikler kullanılmaktadır.

**[2]** **Hastalar ve sağlık hizmetleri ile deneyime dayalı ortak tasarımı daha güvenli ilaç kullanımı için teoriye dayalı müdahaleler geliştirme**

Keywords: Co design, Intervention development, Medicines safety

Deneyime dayalı ortak tasarım, orijinal olarak geliştirilmiş ve halen öncelikli olarak bir sağlık hizmeti kalitesini iyileştirme aracı olarak kullanılan katılımcı bir tasarım yöntemidir. Geleneksel olarak deneyime dayalı ortak tasarım, hizmetler ve ortamlar içine yerleştirilmiştir.

Yöntem olarak bu makalede çok aşamalı araştırma çalışmasının parçası olarak sağlık sistemi içinde daha geniş bir şekilde test edilmek üzere, teoriyle desteklenen karmaşık müdahaleler geliştirmek için deneyime dayalı ortak tasarımın nasıl uyarlanabileceği sunulmuştur.

İlaç yönetiminin güvenliğini arttırmak için davranış değişikliği teorisiyle desteklenen karmaşık müdahaleler geliştirme örneğini kullanıyoruz. Bunu yaparken ilaç optimizasyonu araştırması bağlamında deneyime bağlı ortak tasarımın potansiyel kullanımı tartışılmış ve araştırmacıların müdahaleleri geliştirmek ve değerlendirmek için bu yöntemi nasıl uyarlayabileceklerinden bahsedilmiştir.

İlaçlar dünya çapında en yaygın kullanılan sağlık müdahalesidir ve sağlık koşullarının yönetilmesine esastır. Birleşik krallık ’ta, ulusal sağlık servisi için birinci ve ikinci basamak sağlık hizmetlerinde ilaçlara yaklaşık 16 milyar sterlin harcandığından bahsedilmiştir. İlaçlardan kaynaklanan zararı azaltmak için yapılan çok sayıda müdahale yönetim sürecinde farklı aşamalar olmuştur ve bunların çoğu ilaçları kendi kendine yönetme becerilerinin kazanılması da dahil olmak üzere hastaların ilaç kullanımına yöneliktir.

Sağlık hizmetlerinde kaliteyi iyileştirme uygulamalarında ortak tasarım kavramı benimsenmiştir. Burada, belirli bir hizmetle ilişkili personel ve hastalar farklı çalışma biçimlerini tasarlamak için birlikte çalışır. Sağlık hizmetleri için ortak tasarım, deneyime sahip bu bireylere etkileşimi ve bu bireylerin güçlendirilmesini içerir.

Geleneksel olarak deneyime dayalı ortak tasarım, sağlık hizmeti kullanıcılarının deneyimlerine ve duygularına odaklanan bir hizmet iyileştirme tekniği olarak kullanılmıştır. Deneyime dayalı ortak tasarımda yapılan araştırmalarda ilgilenilen sonuç ölçütlerinin tümü ilaçlara ilişkin bilgi ve ilaçlarla ilgili memnuniyet dahil olmak üzere ikincil sonuçlarla birlikte ölüm nedeni ve hastaneye yeniden yatış olmuştur.

MRC kılavuzu geliştirme, değerlendirme ve sağlıkta uygulama süreci beş aşamadan oluşmaktadır. Sırasıyla; geliştirme, pilot uygulama ve fizibilite, değerlendirme, raporlama ve uygulamadır. Deneyimlerimize göre tetik film çekilmektedir. Bu filmin amacı genel değişim hedefinin kaybolmasını ve gerekirse odağın kolayca yeniden sağlanabilmesini sağlamaktır.

Bu makalede müdahale tasarımını destekleyebilecek birçok teorik mercek olduğunu kabul ediyoruz, ancak dikkatimizi davranış değişikliği teorisi örneğine odaklıyoruz. Bunun nedeni, ilaç güvenliğini arttırmak için müdahalemizi tasarlarken değişime neden olacak olası davranışsal mekanizmaların anlaşılmasının araçlarla mümkün olmasıdır.

Sonuç olarak deneyime dayalı ortak tasarım, hastaları ve sağlık personelinin girişimlerini ve potansiyel değişikliklerin merkezine koyan sağlık hizmetlerini iyileştirmeye yönelik iş birlikçi yaklaşım olarak kabul edilmiştir. Deneyime dayalı ortak yaklaşımın bir araştırma projesine nasıl entegre edilebileceği ve mevcut araştırma yaklaşımlarının aşamalarına nasıl entegre edilebileceği gösterilmiştir. Müdahale değişikliği mekanizmalarını daha iyi anlamak için davranış değişikliği teorilerinin nerede kullanabileceği önerilmiştir.

**[3]** **Sağlık hizmeti analizine dayalı makine öğrenimi algoritmalarının değerlendirilmesi**

Keywords: Heart sickness, K-Means, MapReduce, Naive Bayes, Machine learning

Kalp hastalıkları uluslararası düzeyde insan öldüren hastalıklardan bir tanesidir. Araştırmacıların çoğu öncelikle tıbbi bilgilere dayanarak mağdurların tehlikelerini tespit ve tahmin etmek için kümeleme, kural oluşturma, karar ağacı ve makine öğrenme algoritmasında farkındalık yaratıyor. Bu nedenle denetimsiz öğrenme, denetimli öğrenme ve pekiştirmeli öğrenme olmak üzere üç önemli sınıflandırmaya ayrılan birçok farklı makine öğrenme algoritması türü vardır.

Koroner kalp hastalarının verilerini tahmin etmek için paralel K-Means kümeleme algoritması, MapReduce tabanlı Naive Bayes algoritması ve paralel yarı-Naive Bayes algoritması gibi üç makine öğrenme algoritmasının performansları analiz edilmiştir. Koroner kalp hastalığı hastalarını tahmin etmek için öncelikle tamamen kategori stratejilerine dayalı kümeleme önerilmiştir.

Bulut bilişim, arama sırasında etkilenen kişinin kayıtlarını toplamak ve almak için kullanılan en önemli sistemdir. Veri seti, eğitim için kullanılan kayıtların %20’si ve test için kullanılan kayıtların %80’i birlikte on bin bayt uzunluğundadır. Bu yaklaşımda öncelikle tamamen kategoriye dayalı kümeleme, koroner kalp hastalarının tehlike değerlendirmesini beklemek için uygulamaya konmuştur. Kayıtların eğitim ve test için bölümlenmesi ortam içindeki en önemli işlemlerdir.

Olasılıksal yaklaşımla geliştirilmiş paralel yarı-Naive Bayes algoritması, koroner kalp hastalarını tamamen vücut alan ağı aracılığıyla kaydedilen tıbbi belirtilerine dayalı olarak kategorize etmek için uygulanmaktadır. IMR algoritması ise MapReduce sistemini bitirmek için geliştirilmiştir.

Genel olarak Naive Bayes sınıflandırma algoritması verileri önceden belirlenmiş sınıf etiketlerine sınıflandırılır. Önerilen yöntem ayrıca kalp hastalarının risk analizini de sınıflandırmaktadır. Bu bilgi araştırmaların temel fikri değerli verileri çıkarmaktır. F1-skoru ise sistemin performanslarını analiz etmek için kullanılır.

Yapılan sonuç analizinde sırasıyla K-Means yöntemi kümeleme algoritması %88,43 doğruluk, Naive Bayes kategori algoritması %87,67 doğruluk, paralel yarı-Naive Bayes algoritması %90,53 doğruluk sağlamıştır.

Makine öğrenimi için tüm değerlendirme teknikleri MapReduce sistemi ile ölçeklenebilir ve genişletilebilir. Bu makalede üç değerlendirme makinesi öğrenme algoritmaları anlatılmıştır. Paralel yarı-Naive Bayes algoritması sonuçlarına göre kalp hastalıklarını tahmin etmek için kaliteli olanıdır. Önerilen model, çeşitli tıbbi uygulamalarda acil hasta verilerinin toplanmasını ve yorumlanması için kolayca kullanılabilir.

**[4]** **Sağlık Analitiğinde hassas ilacın karar vermeyi ne ölçüde etkilediği**

Keywords: Clinical decision making, Precision medicine, Big data analytics, Healthcare analytics

Sağlık verileri heterojen, çok boyutlu, parçalanmış, yarı yapılandırılmış ve çeşitli kaynaklardan gelmektedir. Ayrıca sağlık sektöründe nesnelerin interneti platformu aracılığıyla gelişen sensör teknolojiler elektronik tıbbi kaynakların oluşmasında öncü rol oynuyor. Bu verilerin karar vermede yararlı ve destekleyici olması için büyük veri analitiğine ihtiyaç vardır. Bu makale her bir büyük veri analitiğinin hassas ilaç ile ilgili olarak sağlık hizmetleri üzerindeki etkinlik düzeyini vurgulamayı amaçlamaktadır.

Herbert Simon tarafından tanıtılan “zekâ, tasarım, seçim”, bilim adamları tarafından problem çözme veya karar verme görevi için eksiksiz çerçeve olarak yaygın olarak kullanılmaktadır. Bu gelişen tekniğin her adımının bir geri besleme döngüsü vardır ve her adımın sonuçları nihai seçim seçilene kadar iletilir.

ABD Ulusal tıp kütüphanesi tarafından duyurulan hassas ilaç tanımına dayalı olarak, tıbbi karar vermede genetik profildeki bireysel farklılıkları, çevresel ve yaşam tarzı faktörlerini dikkate alan yeni bir yaklaşım olarak tanıtılmaktadır. Tıbbi karar vermedeki bu yeni yaklaşım 20 Ocak 2015’te eski Başbakan Barack Obama tarafından yüksek tedavi maliyetlerini azaltmak, tıbbi yan etkileri en aza indirmek ve sağlık hizmetlerinin kalitesini arttırarak halk sağlığını iyileştirmek için başlatılmıştır.

Veri analitiğinin genel sınıflandırması tanımlayıcı/teşhis amaçlı, tahmine dayalı ve kuralcıdır. İlk analitik sağlık alanında geniş çapta ortaya çıksa da daha çok klinik uygulamalarda yararlı olduğu görülmüştür. Makalede verilen örnekte bir laboratuvar muayenesinin sonucu, belirli bir hastalıkla ilgili olarak hastanın durumunu belirlemek için faydalı bilgiler içerir. Bu tür analizler etkili tedavi için belirli bir hastalığı erken bir aşamada tespit etmek için önemlidir. Ayrıca sağlık hizmeti sahtekarlığının verimli ve hızlı bir şekilde tespiti teşhis analitiğini kullanmanın bir başka avantajıdır. Özne ve nesnenin otomatik olarak sınıflandırılması, yalnızca hastalıkların doğru bir şekilde tanımlanmasını sağlamakla kalmadı aynı zamanda bu otomatik yapı, mevcut ilaçların çıkarılmasındaki performansı ve bireysel hastanın klinik verilerinde kaydedilen tedavi planlarını da iyileştirdi.

Bu makale, hassas tıbbın projeksiyonunu güçlendirmek için her bir sağlık hizmeti analitiğinin rolünü tartışmayı amaçlamıştır. Hassas ilaç, hedefe göre mümkün olan en iyi tedavi senaryosunu gerçekleştirmek için bireysel hasta verilerinin dikkate alındığı, tıbbi karar vermede yeni bir yaklaşım olarak tanımlanmaktadır. Tanımlayıcı ve tahmine dayalı analitik bilgi odaklıyken, karar vermeyi desteklemek için yorum gerektiren ve mümkün olan en iyi çözümü elde etmek için en yüksek etkiyle performans gösteren karar odaklı bir analitik olarak tanımlanır.

**[5]** **Makine öğrenme modelleri kullanarak verimli hastalık teşhisi**

Keywords: Machine learning, Risk estimation, Heart disease

Teknolojinin ilerlemesiyle daha iyi bilgi işlem gücü ve açık kaynak havuzlarındaki veri kümelerinin kullanılabilirliği, makine öğreniminin kullanımını daha da arttırdı. Makine öğrenimi, sağlık hizmetlerinde geniş alanlarda kullanılmaktadır. Sağlık sektörü, kalıpları belirlemeye ve tahminler yapmaya yardımcı olan görüntüler, hasta verileri açısından büyük miktarda veri üretir. Kalp hastalığı kişiye bağlıdır ve kalp hastalığının kapsamı kişiden kişiye değişebilir. Biz bir makine öğrenme modeli yapmak, onu veri kümesi üzerinde eğitmek ve bireysel hasta ayrıntılarını girmek tahminde yardımcı olabiliyor. Tahmin sonucu girilen verilere göre olacaktır ve dolayısıyla o kişiye özel olacaktır. Son zamanlarda birçok araştırmacı makine öğrenimi modellerine kıyasla önemli ölçüde daha iyi performans elde etmek için derin öğrenme modellerini kullanmaya başladılar.

Bir çalışmada kalp hastalığının öngörülmesi için topluluk tekniği kullanılmıştır. Araştırmalar, topluluk tekniğinin zayıf sınıflandırıcıların doğruluğunu arttırdığını göstermiştir. Makine öğrenimi modelleri, verileri kullanarak tıp alanındaki sorunları çözmek için koronavirüs hastalığı için de kullanılmıştır. Örneğin, enfekte insan sayısını ve Çin’de koronavirüsün muhtemelen ne zaman belirlemek için makine öğrenimi teknikleri ve matematiksel modeller kullanılmış. Diğer bir örnek vermek gerekirse makine öğrenme modelleri ile diyabet problemlerinde çözüm üretmek ve erken tahmin sağlamak için kullanılır.

Metodolojinin amacı makine kullanarak riskin tahmin edilmesidir. Kullanılan programlar Python 3, Java, Jupyter Notebook ve Android kullanılmıştır.

Ham verileri kullanılabilir hale dönüştürebilmek için temizlik yapılır. Veri temizliğinden sonra veri analizi yapılır. Veri analizinde, özellikler tanımlanır ve veriler üzerinde bir forma dönüştürülür. Daha sonra makine öğrenmesi modelleri uygulanabilir. Bu adımlar model tahminlerinin her biri için kullanılır. Yakın gelecekte önerilen model el yazısı tanıma, görüntü filtreleme, kanser sınıflandırılması ve tıbbi görüntü segmentasyonu gibi diğer bazı uygulamalarda uygulanabilir.

**[6] Veri madenciliğinde CART ve Lojistik regresyon analizinin yeri: İlaç provizyon sistemi verileri üzerine araştırma**

Keywords: Veri madenciliği, CART, Lojistik Regresyon

Veri madenciliği, büyük veri tabanlarından faydalı bilgileri ortaya çıkararak hizmet kalitesinin artırılması bakımından büyük katkılar sağlamaktadır. Genellikle araştırmalarda büyük veri kümelerini sınıflandırarak önemli veri sınıflarını ortaya koyan veya gelecek veri eğilimlerini tahmin etmede faydalanılan yöntemlerden, veri madenciliği teknikleri içerisinde en yaygın kullanıma sahip olanlarından bir tanesi sınıflama ve regresyon modelleridir. Bu çalışmada veri madenciliği metotları içerisinde, sınıflama ve regresyon modellerinden en çok kullanılan karar ağacı algoritmalarından biri olan sınıflama ve regresyon ağaçları (CART) algoritması ile lojistik regresyonun sınıflama özellikleri karşılaştırılarak gerçek bir veri seti üzerinde uygulama yapılmış ve söz konusu iki yöntemin başarısını göstermek amaçlanmıştır.

Bilgisayar sistemleri ile üretilen bu veriler kendi başına değersizdir, çünkü tek başlarına herhangi bir anlam ifade etmemektedir. Veriler belirli bir amaca yönelik olarak işlenerek bilgiye dönüştürüldüğünde bir anlam ifade etmeye başlamaktadır. 1990’lı yılların başından itibaren kullanılmaya başlanan, büyük veri kümeleri içinde saklı durumda bulunan ve işlenmemiş bilgiyi anlaşılabilir ve yorumlanabilir hale getiren işlemlerden biri veri madenciliğidir. İstatistiksel uygulamalarda sınıflama ve regresyon yöntemleri, bağımlı ve bağımsız değişken arasındaki ilişkiyi tanımlamaya yönelik veri analizlerinin önemli bir parçası olmaya başlamıştır. Bağımsız değişkenler için herhangi bir varsayım olmaksızın kategorik bağımlı değişkeni tahmin etmek için sadece lojistik regresyon değil aynı zamanda karar ağaçları da kullanılmaktadır. Çeşitli şekillerde elde edilmiş veriyi analiz ederek anlaşılır ve faydalı bir yapıya dönüştürmeyi hedefleyen veri madenciliği metotlarından biri olan karar ağaçları; kolay anlaşılır olması, görsel sunumunun ön planda olması gibi nedenlerle sıklıkla tercih edilmektedir. Sınıflama tahmin edici bir model olup, havanın bir sonraki gün nasıl olacağı veya bir kutuda kaç tane mavi top olduğunun tahmin edilmesi bir sınıflama işlemidir.

Ele alınan bağımlı değişken kategorik ise yöntem sınıflama ağaçları (Classification Tree), sürekli ise regresyon ağaçları (Regression Tree) olarak adlandırılmaktadır. Lojistik regresyon analizini, doğrusal regresyon analizinden ayıran en belirgin özellik de lojistik regresyon analizinde bağımlı değişkenin iki ya da çok sınıflı olmasıdır. Bir modeldeki bağımsız değişkenler ile bağımlı değişken arasındaki lineer ilişkiyi veren fonksiyona “link fonksiyonu” adı verilmektedir.

Makalede belirtilen çalışma, çok geniş bir çalışma alanı olan veri madenciliğinin sınıflama ve regresyon modellerine ilişkin teknikleri ile sınırlandırılan, Sosyal Güvenlik Kurumu (SGK) İlaç Provizyon Sistemi veri tabanından elde edilen veri kümesi üzerinde bir uygulama yapılmıştır. Uygulamaya konu olan veri kümesi, İlaç Provizyon Sisteminden solunum sistemi hastalıkları için antibiyotik kullanan yaklaşık 50 milyon hasta içerisinden örnekleme yoluyla seçilen 18.931.000 hastanın 12 farklı değişkene ilişkin değerlerini içermektedir. Uygulamada söz konusu veri kümesi üzerinde Clementine 12.0 yazılımı kullanılarak veri madenciliğinin sınıflama ve regresyon problemine ilişkin CART ile lojistik regresyon teknikleri için örnek bir uygulama ortaya konmuştur. Çalışmanın amacı bu veri kümesi için, veri madenciliği uygulaması ile penisilin grubu antibiyotik kullanan hastaların profilini belirleyen önemli faktörlerin araştırılarak ortaya çıkarılmasıdır. Penisilin grubu antibiyotik kullanmayan 4.287.961 hastanın ise 4.118.361 tanesi lojistik regresyon ile yapılan sınıflandırma işleminde doğru, 169.600 hasta ise hatalı olmak üzere % 96,04’lük doğruluk yüzdesiyle sınıflandırılmıştır. Lojistik regresyon ile yapılan sınıflandırma işleminde genel doğruluk değeri ise 6.772.313 hastanın 6.188.070 tanesi doğru sınıflandırılarak % 91,37 olarak hesaplanmıştır.

Yapılan bu çalışmada, penisilin grubu antibiyotik kullanan hastaların profilini belirlemek amacıyla bir uygulama yapılmış ve çalışmaya alınan veri seti için CART analizinin lojistik regresyon analizine göre daha iyi bir doğru sınıflandırma oranına sahip olduğu görülmüştür.

**[7] Denetimli makine öğrenimi sınıflandırıcıları ile sağlık hizmetleri kararlarını tahmin etme**

Keywords: Machine learning cllassifiers, Healthcare decision making, K nearest neighbor, Random forest

Günümüzde sağlık hizmetleri hayat kurtarmak, sağlık hizmetlerinin maliyetini azaltmak ve bulaşıcı hastalıkları erken keşfedilmesi için etkili yöntemlere ve araştırma metodolojilerine ihtiyaç duymaktadır. Makine öğrenimi teknikleri, sağlık kuruluşlarının hasta koşullarındaki ve davranışlarındaki eğilimleri tahmin etmelerini sağlayabilir. Bu şekilde elde edilen bilgiler, hastaların durumunun eğilimlerini mümkün olan en kısa sürede tahmin etmek ve sağlık hizmetlerinin maliyetini azaltmak için kullanılabilir. Ayrıca, özellikle sonuçların tahmin edilmesinin zor olduğu ve en iyi ameliyat yöntemini seçmenin zor olduğu durumlarda hekimlerin hastaları teşhis etmesine yardımcı olmak için makine öğrenme teknikleri de kullanılabilir.

Bu makale sağlık hizmetleri operasyonları sırasında doktorun verdiği gerçek kararlara kıyasla kararları tahmin etmek için makine öğrenimi sınıflandırıcılarına dayalı bir sağlık bakımı operasyonel karar verme sistemini açıklamaktadır. Denetimli makine öğrenimi sınıflandırma ve optimizasyon tekniklerinin çoğu, bu tür karar verme tekniklerinde kullanılır. Bu sistemi anne ve bebeği kurtarmaya yardımcı olmak için dünyada en sık yapılan doğum operasyonu olan sezaryen üzerinde araştırma yapılmıştır. Kullanılan bu sistemi ameliyatı ne zaman kullanmamız gerektiğini tahmin etmemize yardımcı oluyor. Kullanılan veri setinde bilinen etiketler varsa makine öğrenimi denetimli olarak adlandırılır bu sistemde eğitim aşamasında örneklerle birlikte verilir, oysa denetimsiz makinede örnekler etiketlenmez. Bu konuda makine öğrenme tekniklerinden olan destek vektör makinesi(SVM) üzerine odaklanılmıştır.

Destek vektör makinesi(SVM), hem sınıflandırma hem de regresyon problemleri için kullanılabilen ayırt edici bir sınıflandırıcıdır. SVM amacı, etiketli eğitim verileri verildiğinde algoritma yeni örnekleri kategorize eden bir hiper düzlem üretmesidir. Karaları tahmin etmek için gerçekleştirilmesi gereken üç adım gereklidir: veri seti toplama, makine öğrenimi sınıflandırıcılarının eğitimi ve makine öğrenimi sınıflandırıcılarının değerlendirilmesidir.

Sezaryen veri setinde 80 örnek toplanmış ve sağlık hizmetleri operasyon kararlarını tahmin etmek için kullanılan denetimli makine öğrenimi sınıflandırıcılarını değerlendirmek ve eğitmek için kullanılmış. Bu araştırma WEKA yazılımı kullanılarak denetimli makine öğrenmesi sınıflandırıcılarının performansları hesaplanmıştır.

Yapılan araştırmada K- En yakın komşu ve Random Forest teknikleri %95 vaka için doğru sonucu tahmin etmiştir. Bu nedenle makine öğrenme algoritmalarının kullanımıyla elde edilen bilgiler, sağlık hizmeti karar verme başarısını ve hastaların sağlığını iyileştirecek başarılı kararlar almak için faydalı olduğu saptanmıştır.

**[8]** **Büyük veri çağında işletmelerde veri bilimi**

Keywords: Büyük veri, Veri bilimi, Veri analitiği

Günümüzde erişime açık mevcut veri kaynaklarının ve miktarının artması ile, hemen her sektör rekabet avantajı elde etmek için iş süreçlerini veri odaklı güncellemeye büyük önem vermektedir. Öyle ki, yapılan bir araştırmaya göre şirketlerin üst düzey yöneticilerinin %70’e yakını, veri analizlerinin satış pazarlama süreçlerine dâhil edilmesinin şirketlerinin rekabet avantajı elde edebilmesi için “çok önemli” veya “önemli” olduğu vurgulanmıştır. 330 şirket üzerinde yapılan diğer bir araştırma ise kendilerini veri odaklı şirket olarak tanımlayan şirketlerin finansal ve operasyon başarılarının diğer şirketlere oranla daha yüksek olduğunu gözlemlemiştir.

Bu araştırma kurgulanırken cevap aradığımız sorular şu şekildedir: “Veri bilimi ve büyük veri alanları hangi yönleriyle birbirine benzemektedir ve hangi yönleriyle birbirlerinden ayrılmaktadır?”, “Veri odaklı bir organizasyon kurmak için veri bilimi ve büyük veri alanlarından kaynaklanan sorunlar nelerdir?” ve “Büyük veri çağında veri biliminde yaşanan gelişmeler nelerdir?”. Bu bildiri, verilen sorulara cevap ararken yaptığımız keşfedici araştırmaların sonuçlarını raporlamaktadır.

Büyük veri kavramı Nesnelerin İnterneti, web kaynakları ve benzer bilgi teknoloji servisleri ile üretilen yapılandırılmış, yarı yapılandırılmış ve/veya yapılandırılmamış yapıda, hızlı ve hacimli verileri tanımlamak için kullanılmaktadır. Büyük veri uygulamalarının ortak sahip oldukları üç temel karakteristik vardır: Hız, Hacim ve Çeşitlilik. Bu üç boyut; veriyi analiz edecek modelleri, verinin saklanma yöntemlerini ve veriyi işleyecek araçları doğrudan etkilemektedir. Büyük veri uygulamalarındaki en büyük zorluklar; verilerin verimli bir şekilde işlenmesi, saklanması ve sorgulanmasıdır. veri işleme ve saklama yönteminin belirlenmesi gerekmektedir. Örneğin, anlık olarak üretilen veriler üzerinde hızlı kararlar verebilmek için eş-zamanlı akan veri işleme yaklaşımı uygun iken, hali hazırda veri tabanında saklanmış verilerin toplu olarak işlenmesi için yığın veri işleme yöntemi daha uygundur.

Veri bilimi, işletmede ortaya çıkan gerçek iş problemlerini verilerle anlamlandıran ve bu problemlere çözüm sunacak veri uygulamaları geliştiren, içinde bilgisayar bilimleri, istatistik, matematik ve bilişim bilimleri dahil olmak üzere birçok bilim dalını içeren disiplinler arası bir çalışma alanıdır. Veri bilimi ham veriden bir soruna çözüm ortaya koyabilecek bilgiyi üretmek ve atılması gereken adımları bulmak için verilerin toplanması, hazırlanması, analizi ve analizlerin sonuçlarının anlamlandırılarak aksiyon alınması gibi temel süreçleri içinde barındırmaktadır. Veri bilimi, büyük veri, makine öğrenimi, derin öğrenme ve veri madenciliği teknolojilerini ve yöntemlerini kullanarak dört farklı analitik türü kullanarak veri seti üzerinde farklı sorulara cevap bulmayı amaçlamaktadır.

Büyük veri ve veri bilimi her ne kadar ortak amaçları taşısa da kendi içlerinde tamamen ayrışan iki alandır. Yapılan son araştırmalar gösteriyor ki; kurumların veri kullanılabilirliğini %10 oranında arttırması şirket verimliliğini %17 ile %49 arasında arttırırken, şirketlerin büyük veriyi, veri bilimi süreçlerine başarı ile entegre etmesinin yatırım kârlılığını %241 oranında arttırdığı gözlemlenmiştir. Büyük veri bilimi projelerinin sadece %4’ lük bir kısmı istenen başarı ile sonuçlanırken, %43’lük kısmı elde edilen tecrübelerden ve bilgiden başarı elde ettiğini düşünürken, %23’lük bir kısmı hiçbir fayda elde edilmeden sonuçlanmıştır. Örgütsel sorunlardan biri de organizasyon yapısının verinin iş süreçlerinde etkili ve verimli şekilde kullanılabilmesini destekleyecek şekilde düzenlemesidir. Örneğin büyük ölçekli bir işletmede tek başına bilgi teknolojileri koordinatörlüğü altında veri bilimi bölümü kurulması yeterli olmayabilir. Bu durumda saha çalışanları ve uzmanlar veri bilimcilerle direk çalışmak durumunda kalırlar ki bu pek çok soruna yol açar. Yöneticiler eğer veri analizi sonuçlarındaki değeri kavrayamazsa veya sonuçların uygulanmasını sağlayamazsa veri analizi sonuçları anlam ifade etmez. Veri toplamak için otomatik veri toplayıcı araçlar ve algılayıcılar kullanılabilir. İş süreçleri ve operasyonların ihtiyaçlara göre değişkenlik gösterdiği durumlarda veya ilave kaynaktan veri toplanması gerektiği zaman otomatik veri toplayıcı araçların kodlarının hızlıca uyarlanması gerekebilir. Bu zaman alıcı bir süreçtir ancak bazı yaklaşımlar bu sürecin kısaltılmasına yardımcı olabilir.

Büyük veri ve veri bilimi ile ilgili yapılmış araştırmaların literatürdeki durumunu analiz etmek amacıyla, Kitchenham sistematik literatür taraması metodu kullanılmıştır. Bu tarama Temmuz 2018’de gerçekleştirilmiş olup, arama tarih aralığı Ocak 2010-Ocak 2018 olarak belirlenmiştir ve elde edilen sonuçlar konferans bildirisi, makale, kitap ve kitap bölümü seçenekleri ile sınırlandırılmıştır.

Sonuç olarak günümüzde büyük veri bilimi özellikle finans, yazılım ve internet tabanlı şirketler gibi hali hazırda birçok kaynaktan veri toplama imkânı olan sektörler için büyük bir avantaj sağlamaktadır. Ancak gelişen ağ bilimi, yazılım ve donanım teknolojileri sayesinde yakın gelecekte hemen hemen her sektörün sundukları hizmetlerin üretim aşamasından son kullanıcı deneyimine kadar geçen her süreçte veri toplama imkânı olacaktır.

**[9]** **Kronik hastalık tanısında makine öğrenimi öngörü modellerinin uygulamaları**

Keywords: chronic diseases, prediction models, accuracy, disease classification

Bu makale de kronik hastalıkları tanılamada makine öğrenimi öngörücü modellerinin uygulamalarını kullanılmaktadır. Bu hastalıklardan muzdarip hastaların ömür boyu tedaviye ihtiyacı vardır. Günümüzde, tahmine dayalı modeller, bu hastalıkların tanısında ve öngörülmesinde sıklıkla başvurulan bir yöntemdir. Bu analiz 2015-2019 yılları arasında yayınlanmış 453 makaleyi kapsamaktadır ve doküman taramamız yapılmıştır.

Yapay zekâ, makineler tarafından gerçekleştirilen insan zekâsı olarak tanımlanabilir. Bilgisayarda bilim, makinenin hiçbir şey kullanmadan kendi başına akıllı davranışı taklit etme kapasitesi olarak tanımlanır. Makine öğrenmesi ise karmaşık modelleri belirlemek ve tıbbi bilgileri ortaya çıkarmak için kullanılır. Yapılan bu makine öğrenmesi modelleriyle ilaç reçetesine dahil edilmesi doktorları işini hafifletir ve yeni tıbbi fırsatlar ortaya çıkarabilir. Makine öğrenmesi modelleri ile tıbbi verilerin kalitesini artırmak, veri akışındaki dalgalanmaları azaltmak da mümkün olabilir. Ölüm oranlarını azaltmak için kronik hastalıkların neden olduğu, erken teşhis ve etkili tedaviler tek çözümdür. Bu nedenle, çoğu tıp bilimci hastalıkta öngörü modellerinin yeni teknolojilerine ilgi duyuyor. En yüksek doğruluğa sahip modellerin tıbbi teşhiste büyük önem kazanmaktadır.

Sistematik literatür taraması PubMed (Medline) ve Sağlık Literatürü Kümülatif İndeksi kütüphaneleri aracılığıyla yapıldı. Verilerin değerlendirilmesi iki aşamada gerçekleştirilmiştir. İlk aşamada, katılıma bağlı olarak kriterleri, genişletilmiş revizyonlar için 55 belge belirlendi. İkinci aşamada iki kişi kalite kontrolü için eşit olarak dağıtıldı. Her makalenin analizi ayrı ayrı gözden geçirildi ve kaydedildi. Kalite için yayınlanan her makalenin zayıf (0–4), orta (5–6) veya güçlü (7–9) olarak puanlanarak değerlendirildi. Seçilen her çalışma tek bir kâğıdın uygun olup olmadığını hesaplamak için puan ayrı Excel sayfalarına kaydedildi. Sonuç olarak, kronik hastalıktaki tahmin modelleriyle uyumlu 22 çalışma seçildi.

Regresyona dayalı makine öğrenmesi modelleri büyük ölçüde karaciğer, gaz kromatografisi ve patolojik değişiklikleri tahmin etmek için kullanılır. Yapılan araştırma da sonuçlar, RF modellerinin tanımlamada daha iyi olduğunu vurgulamaktadır. Bir diğer yapılan araştırmada bilim adamları, böbreğin glomerüler filtrasyon hızının topluluk modelleri aracılığıyla yapılabileceğini tahmin ettiler. KNN modeller hastalık paternlerini belirlemek için daha iyi kullanılabilir. Araştırmada yapılan bir diğer sonuç demans yaşlılarda görülen kronik hastalıklardan biridir ve özellikle Alzheimer, bunama vakalarının %60-70'i ile ilişkilidir.

Öte yandan, denetimsiz makine öğrenimi, tıbbi bilgi içeren bir derin öğrenme modeliyle ilgilenir. Veri setinde neye ve nasıl ilerleyeceğine dair net bir yönü olmadan işleyebileceği sinir ağı, yapılandırılmış verilerin otomatik olarak algılanmasını dener ve temel özellik çıkarımını gerçekleştirir. Bu modellerle Hepatit B için on altı çalışma yapıldığından bahsedilmiştir. Bayes ağlar ve NB modelleri astım sorunlarının teşhisini, KOAH sorunlarının teşhisi için ise SVM tahmin modelleri etkili olduğu vurgulanmıştır.

Sonuç olarak makine öğrenimi tıbbi bakımdaki yeni gelişmeler olarak kurulmakta ve elektronik verilere erişimi genişletmekte, bu karar için yeni kapılar açmaktadır. Bu modeller ile hasta bakım kalitesinin artırılması ve tıbbi maliyetlerin azaltılması hedeflenmiştir. Yapılan bu araştırmada genel olarak en iyi yaklaşımı belirlemek için standart bir yöntem olmadığını gösterilmektedir.

**[10]** **Hastane bilgi sistemlerinde veri madenciliği**

Keywords: Hastane bilgi sistemleri, Veri madenciliği, Bilgi keşfi

Hastane bilgi sistemleri hastalara ait demografik bilgiler, hastalık ve tedavi durumları, yapılan tetkikler, faturalama ve idari işlere ait bilgileri içerirler. Sağlık ve tıp, çağımızın en önemli bilimsel araştırma alanları olduğu için bu alandaki bilgi sistemleri de araştırmalar için en büyük veri kaynaklarıdır. Son otuz yılda dünyada sağlık bilgi sistemlerinde büyük gelişmeler yaşanmıştır. Daha sonraki yıllarda hastaların gelecekteki sağlık durumları ve maliyet tahminleri gibi konuları araştırmak için sinir ağları kullanılmaya başlanmıştır. Birçok insan, kalp hastalıkları, diyabet ve astım hastalıkları gibi kronik hastalıklarla yaşamak zorundadır. Veri madenciliği yöntemleri kullanılarak bu sistemlerdeki gizli ve önemli bilgiler keşfedilmelidir.

Veri Madenciliği (Data Mining) ve Bilgi Keşfi Temel olarak beş aşamadan oluşur; Veri seçimi, Önişleme, Dönüştürme, Veri Madenciliği, Yorumlama/Değerlendirme. Genellikle veri madenciliği modelleri tahmin edici ve tanımlayıcı olmak üzere ikiye ayrılırlar. Tahmin edici modeller bilinen verilerden yararlanarak, bilinmeyen bir değeri tahmin etmeye çalışırlar. Tanımlayıcı algoritmalar ise verilerdeki gizli ortak özellikleri ve ilişkileri araştırırlar. Geleneksel veri madenciliği algoritmaları çoğunlukla tek bir tablo veya düz bir dosya üzerinde çalışırlar.

İlişkisel kural analizi kullanılarak yapılan veri madenciliği çalışmasında, yaşın, cinsiyetin, hastanelerin özelliklerinin, kronik ve akıl hastalıklarının sürekli doktor ziyaret etme davranışında etkili oldukları ortaya çıkartılmıştır. Sağlık uygulamaları ve tedaviler büyük oranda maliyet gerektirirler. Tıp alanındaki makalelerden tedavi ve tanı ile ilgili yeni yaklaşımlar kavramlar arasındaki gizli ilişkiler ortaya çıkartılabilir. Elde edilen önemli bilgiler hem araştırmalara büyük destek sağlar hem de sağlık kurumlarının başarısını artırır.

Hacettepe Üniversitesi Hastaneleri ülkemizin en büyük üniversite hastanelerinden birisi olup birçok alanda öncü ve yenilikçi bir sağlık kurumu olarak bilgi sistemlerinin kullanımı açısından da diğer sağlık kurumlarına örnek teşkil etmiştir. Hastane genel bilgi sistemiyle tümleşik olarak çalışan Radyoloji Bilgi Sistemi, hastalar, tetkikler ve iş akışlarına ait verilerden oluşan büyük bir ilişkisel veri tabanına sahiptir. Radyoloji bölümünde kaynakların daha doğru planlanması, gelecek planları yapılabilmesi ve tıbbi açıdan yapılan çalışmalara katkıda bulunabilmesi için bu verilerden çıkartılacak değerli bilgilere ihtiyaç vardır. Bu nedenle radyoloji veri tabanı üzerinde ilişkisel veri madenciliği algoritmaları kullanılarak veri madenciliği çalışmaları yapılacak ve değerli bilgiler keşfedilmeye çalışılacaktır.

Sonuç olarak sağlık ve tıp, günümüzün en çok bilgi ihtiyacı olan araştırma alanlarıdır bu veriler ile önemli ölçüde gelişmeler ortaya çıkmıştır. Bu gelişmeler daha çok ve çeşitli verinin saklanabilmesini sağlamış ve beraberinde bilgi keşfi ihtiyacını ortaya çıkarmıştır. Veri Madenciliği, sağlık ve tıp alanındaki büyük veri tabanlarından değerli bilgileri ortaya çıkartarak hem tıp açısından hem de hizmet kalitesinin artırılması açısından büyük katkılar sağlar.

**[11] Veri madenciliğinin tıp ve sağlık hizmetlerinde kullanımı ve uygulamaları**

Keywords: Veri madenciliği yöntemleri, Veri madenciliği, Veri ambarı, İş zekası, Sağlık veri

Bu makalenin amacı, sağlıkta Veri Madenciliğinin kullanımı konusunda bir altyapı oluşturmak ve sağlık profesyonellerine sağlık sektöründe Veri Madenciliğinin kullanımı ile ilgili örnekler sunarak karar verme süreçleri açısından yeni bir bakış açısı kazandırmaktır. Sağlık sistemi politikalarının ve yönetsel kararlarının temeli veri ve veriden elde edilmiş bilgidir. Sağlık verileri hastaneler, diğer sağlık kurumları, sigorta şirketleri ve ilgili kamu kurumları başta olmak üzere birçok kuruluş tarafından toplanmaktadır. Günümüzde dijital verilerin hacmindeki artış beraberinde yeni sorun alanları da yaratmıştır. Bunların başlıcaları; çok miktarda, çok boyutlu ve karmaşık verileri işlemek için yöntem ya da sistemler geliştirmek; yeni türdeki verileri işlemek için yöntem ya da sistemler geliştirmek; dağılmış verileri işlemek için yöntem, protokol ya da altyapı geliştirmek; verilerin kullanımı ve güvenliği ile ilgili modeller geliştirmek olarak sıralanabilir.

Veri Madenciliğinden bahsetmeden önce aktarılması gereken temel bir kavram veri tabanlarında bilgi keşfidir. Veri tabanlarında bilgi Keşfi, verinin nasıl depolanıp erişileceğinden, algoritmaların devasa veri setlerine nasıl ölçeklenebileceğine ve hala etkin olarak çalışmalarına, sonuçların nasıl yorumlanabileceği ve görselleştirilebileceğine ve bütün insan-makine interaksiyonunun kullanışlı olarak nasıl modellenip, desteklenebileceğine olmak üzere veriden bilginin keşfinin tüm süreçleri üzerine odaklanır. Veri Tabanlarında bilgi keşfi süreci interaktif ve yinelemeli, kullanıcı tarafından kararların verilmesini gerektiren adımlardan oluşmaktadır. Tanım olarak veri ambarı, pek çok farklı kaynaktan ve genellikle de farklı yapıda verinin depolandığı ve hepsinin de aynı birleşik çatı altında kullanılmasının ümit edildiği yapılardır. Veri ambarcı lığı, veri kümelerine veri tabanlarında bilgi keşfi aşaması için iki önemli yoldan yardımcı olur; Veri temizleme, Veri erişimidir.

Veri tabanlarında bilgi keşfi, sıklıkla, büyük hacimde veri koleksiyonundan faydalı bilgiyi keşfetmeyi hedefleyen, Veri Madenciliği olarak anılmaktadır. Veri tabanları günümüzde terabaytlarla ifade edilmektedir. Öğrenme problemleri kabaca denetimli ve denetimsiz olarak ikiye ayrılabilir. Denetimli öğrenmede hedef, girdi ölçülerinin sayısını temel alarak çıktı ölçüsünün değerini tahmin etmektir; denetimsiz öğrenmede ise çıktı ölçüsü yoktur ve hedef girdi ölçüleri kümesi arasındaki birliktelik ve örüntüleri betimlemektir. İş zekâsı terimi, işte karar vermeyi destekleyen ve bilgi teknolojilerini temel alan bütün süreçler, teknikler ve araçlar için genel anlamda kullanılan bir ifadedir. Veri madenciliği iş zekâsının yeni ve önemli bir bileşenidir.

Denetimli (Supervised) Veri Madenciliği yöntemleri:

* En yakın k komşuluk
* K-ortalamalar kümeleme
* Regresyon modelleri,
* Kural çıkarımı
* Karar ağaçları
* Sinir ağları

Denetimsiz (Unsupervised) Veri Madenciliği yöntemleri:

* Aşamalı kümeleme
* Kendi kendini düzenleyen olarak sınıflandırılabilir.

Regresyon analizi yaygın kullanılan bir modelleme tekniğidir. Bağımsız değişken olarak anılan tahmin edici değişkenlerin; bağımlı değişken denilen tahmin edilecek değişken değerini belirleyecek ağırlıklandırmaları içerecek bir bağımsız değişkenler birleşimidir.

K-En yakın komşuluk algoritması özellikle büyük veri tabanlarında kullanılan bir sınıflandırma tekniğidir. K-Ortalamalar kümeleme analizi tekniği segmentasyon, gruplama ve sınıflandırma yöntemidir. Karar ağaçları denildiğinde, sinir ağları ile birlikte ilk akla gelen yöntemlerden olan karar ağaçları, yeni jenerasyon veri madenciliği yöntemlerindendir. Sinir ağları insan beyninin hesaplama mantığı baz alınarak oluşturulmuş (yapay) sinir ağları, karar ağaçları gibi yeni jenerasyon veri madenciliği yöntemlerindendir.

Sağlık hizmetlerinin en hızlı, en doğru, en yüksek kalitede ve ihtiyaca cevap verecek şekilde sunulabilmesi için sağlık profesyonellerinin en doğru ve güncel bilgiye ulaşması ve bu bilgiyi karar destek sistemlerinden faydalanarak kullanması gerekmektedir. Veri Madenciliğinin kullanılması sağlık profesyonellerinin en optimal kararları almasına yardımcı olacaktır. Bu makaleden anladığımız veri madenciliği, sağlık profesyonellerinin en doğru ve güncel bilgiye ulaşmasını, en objektif ve optimum çözümleri kullanmasını sağlayacak bir karar destek aracıdır.

**[12]** **Hastalık teşhisi için makine öğrenimi tekniklerinin uygulanması**

Keywords: Data mining, Machine learning, Classification, Decision tree, Prediction

Veri madenciliği ve makine öğrenimi teknikleri sağlık alanında yaygın kullanım alanı olmuştur. Bu çalışmanın amacı, hastalıkları teşhis etmek için otomatik bir yöntem geliştirmektir. Bulanık mantık temelli rastgele orman yaklaşımı ve hastanın tıbbi kayıtlarının kapsamlı bir şekilde incelenmesi hastalığı teşhis etmek için kullanılır. Karar ağaçları genellikle sonuçları tahmin etmek için makine öğreniminde kullanılır. Bulanık veri kümeleri, tıbbi gerçekleri ve uzmanları tanımlamak için mükemmel bir seçimdir. Önerilen bu sistemde, rasgele orman algoritması ve bulanık karar ağaçları kullanarak hastalıkları teşhis eden bir uzman sistem sunulmaktadır. Bulanık karar ağaçları, teşhis sisteminin doğruluğunu arttırır.

Sağlık hizmetlerinde veri madenciliği, hataları azaltmak için tasarlanmıştır. Makalede bahsedilen kümeleme kavramı, nesneleri kategorilere ayırmak için kullanılan bir veri madenciliği tekniğidir. Sınıflandırma yaklaşımı ile bir veri kümesinde en yüksek doğruluk oranı ve en düşük hata oranı bu veri seti için en iyi sınıflandırma tekniği olarak seçilmiştir. Deneysel bulgular, bir ön işleme yaklaşımı olan özellik seçiminin sınıflandırma doğruluğunu önemli ölçüde iyileştirdiğini göstermiştir. Özellik seçim yöntemlerinden herhangi birinin benimsenmesinin sınıflandırıcı doğruluğunu iyileştirdiğini belirlemişlerdir. Aynı özellik seçim yaklaşımının aynı alandaki farklı veri kümeleri için en yüksek doğruluğu sağlayıp sağlamayacağını belirlemek için üç farklı meme kanseri veri kümesi üzerinde çeşitli testler gerçekleştirilmiştir. Bulgulara göre belirli bir özellik seçimi tüm meme kanseri veri kümeleri için en yüksek doğrulukta sonuçlanmayabilir. Sonuç olarak her ek veri kümesi değerlendirilir, aynı etki alanına bağlı daha önceden kanıtlanmış olanı değerlendirmek yerine sınıflandırıcı performansını iyileştirmek için en uygun olanı keşfetmek için birden fazla özellik seçim yaklaşımıyla deneme yapılmalıdır.

Makine öğreniminde karar ağaçları sıklıkla sonuçları tahmin etmek için kullanılır. Bulanık veri kümeleri, tıbbi gerçekleri ve uzman görüşlerini tanımlamanın harika bir yoludur. Bulanık karar ağaçları, temel karar ağaçları oluşturmak için bulanık girdi kullanır. Önerilen bu sistemde rastgele bir orman algoritması ve bulanık karar ağaçları kullanarak hastalığı teşhis eden bir uzman sistem sunulmaktadır. Bulanık karar ağaçları, teşhis sisteminin doğruluğunu arttırır. Önerilen teknik mevcut çözümlere göre hastalık tahmininde daha etkili olduğu gösterilir. Önerilen bulanık rastgele orman algoritması, kalp hastalığı sınıflandırması ve tahmini için %91,7’lik bir doğruluk oranı göstermiştir.

**[13] Veri analizinde istatistik mi veri madenciliğimi?**

Keywords: Veri madenciliği, İstatistik, Veri analizi

Veri analizinde yıllardır kullanılan istatistik, son yıllarda veri madenciliği ile yan yana yer almaktadır. Veri analizi çalışmalarında kullanılan istatistik ve veri madenciliği yöntemleri birçok farklı araştırma alanında kendilerine yer bulabilmektedir. Bu çalışmanın amacı, her iki alanın incelenerek aralarındaki farklılıkların ve benzerliklerin ortaya konmasıdır. Özellikle sağlık alanında veri madenciliğinin çok yaygın olarak bilinmemesi ve sağlık çalışanlarının genellikle istatistiksel yöntemlere bağlı kalıyor olması çalışmanın çıkış noktasını oluşturmaktadır. İstatistiksel yöntemler birçok alanda kullanılsa da veri analizinde temelini istatistiksel yöntemlerden alan veri madenciliği kavramı ortaya çıkarak gerek yapısal gerek yapısal olmayan farklı tipte ve büyük boyuttaki verinin analiz edilebilmesine olanak sağlamaktadır. Değişkenler arasındaki ilişkileri inceleyen, veri setini çeşitli hesaplamalarla analiz edip bulguları özetleyen, örneklemden çıkartılan sonuçların veri kütlesi için genelleştirmeye yarayan bir bilim dalı olarak istatistik veri madenciliğinden çok daha eski yıllarda ortaya çıkmış ve veri analiz etmede yüzyıllar boyunca kullanılmıştır. Geleceğe yönelik tahminde bulunulması gibi çeşitli analizler için istatistikten biliminden yararlanılır. Araştırmaya konu olan her iki makalede de hızla artan veri boyutlarından anlamlı bilgi elde edebilmek için, bilgisayar temelli yeni yöntemlere ihtiyaç duyulduğundan, veri tabanlarında bilgi keşfi ve veri madenciliği konularının araştırmacıların dikkatini çektiğinden bahsedilmektedir.

Veri madenciliği ile ilgili çalışmaların yıllar içerisinde arttığını göstermek adına ayrıca Yükseköğretim Kurulu Başkanlığı’na ait Tez Merkezi web sitesinde de bir tarama yapılmıştır. “Veri madenciliği” anahtar kelimeleri iki farklı filtre ile aranmıştır. İlk aramada veri madenciliği tez adları içerisinde, ikinci taramada ise aynı kavram herhangi bir kriter belirtmeden tüm alanlar içerisinde aranmıştır. Verinin hızlı artışı insanların bu veri yığınlarını yorumlayabilmelerini zorlaştırmaktadır. Bu yüzden büyük hacimli verinin analizi ya da yorumlanması için araştırmacılar veri madenciliği alanına yönelmektedir.

Özetlemek gerekirse farklı kaynaklarda yapılan farklı tanımlar aslında birbirine çok yakın noktalara parmak basmaktadırlar. Veri madenciliği de istatistik gibi çeşitli analizler yoluyla veriden öğrenmeye ve bilgi elde etmeye yarayan bir alandır. Bunun yanında veri madenciliği ile istatistiksel analizlerde olduğu gibi geleceğe dönük tahminler yapmak da mümkündür. Bu çalışmanın oluşturulması için literatür taraması yapılmıştır. 1996 ve 2015 yılları arasında yayımlanan konferans bildirisi, makale ve kitap olmak üzere, Türkçe, İngilizce ve Almanca dillerinde yazılmış toplam 24 kaynak incelenmiştir. Veri madenciliği yöntemlerinin birçoğu temelinde istatistiksel yöntemlere dayanır. Kullanılan analizler incelendiğinde; kümeleme, diskriminant, regresyon, korelasyon analizlerinin her iki alanda da kullanılan ortak yöntemler olduğu görülmektedir. Temellerini istatistikten alan veri madenciliği yöntemleri analizlerde istatistiksel yöntemlerin sınırlılıklarını ortadan kaldırmakta ve günümüzün veri bolluğunda her alandan araştırmacılara daha geniş kapsamlı ve esnek analizler yapma imkânı sunmaktadır.

**[14] İlaç endüstrisinde insan sağlığı açısından tehlikeli maddelerin risk değerlendirmesi**

Keywords: İlaç endüstrisi, İnsan sağlığı, İlaç endüstirisi

Uçucu tehlikeli maddelerin endüstri alanlarında kullanılmaya başlanması; iç ortam hava kalitesini bozan ortam emisyonlarının oluşumunu beraberinde getirmektedir. Bu çalışmada iç ortam hava kalitesinde, ilaç endüstrisinde kullanılan uçucu tehlikeli maddelerin insan sağlığı açısından risk değerlendirilmesi için uygulamalı bir model kullanılıp, çıkan risklerin nasıl yönetilmesi gerektiği ve nasıl yönetildiği vurgulanmış, öneriler getirilmiştir.

Çevre yönetim sistemi içerisinde yapılan kalitatif risk değerlendirmelerinde tehlikeli maddelerin maruziyet risklerinin önemi yeteri kadar vurgulanmamaktadır. Risk yönetimindeki bu eksiklik üzerinden yola çıkılarak tehlikeli maddelerin insan sağlığı açısından risk değerlendirmesi için yapılmış olan model kullanılarak ortaya çıkan risklerin nasıl yönetilmesinin gerekliliği amaçlanmıştır. Bu çalışma kapsamında çevre yönetim sisteminin genel özelliklerine değinilerek, risk yönetimi altında insan sağlığı açısından risk değerlendirmesinin nasıl olması gerektiği vurgulanmıştır.

Risk değerlendirmesi yapmak için odaklanılan alanda bulunan maddelerin tanımlanması ve insan sağlığı değerlendirmesi ile ilişkili alan verilerinin derlenmesi ve analizini içeren adımdır. Derlenen verilerin değerlendirilmesi veri toplanmasından bağımsız tartışılamaz. Veri değerlendirmesi içinde yapılması gerekenler 7 kısımda özetlenebilir.

1. Saha araştırmalarından toplanan verilerin birleştirilmesi
2. Analitik metotların değerlendirilmesi
3. Miktar ve tespit limitlerinin değerlendirilmesi
4. Tanımlanmış ve kodlanmış verilerin değerlendirmesi
5. Şahitlerin değerlendirilmesi
6. Arka plan ile alan verilerinin kıyaslanması
7. Risk değerlendirmesinde kullanılacak bir kimyasal veri ve bilgi seti geliştirilmesi

İlaç endüstrisi dört alt kategoride sınıflandırılmaktadır. Bu alt kategoriler oluşturulurken üretim prosesi, kullanılan hammadde, oluşan ürün, atık su karakterizasyonu ve arıtılabilirliği göz önüne alınmaktadır. Bu alt kategoriler:

1. Fermantasyon
2. Biyolojik ve tabii ekstraksiyon
3. Kimyasal sentez
4. İlaç Araştırma ve Geliştirme

Çevre yönetim sistematiği içinde insan sağlığı risklerinin yönetimi tehlikeli maddelere maruz kalma ile değerlendirilerek bu risklere göre güvenli bir iç ortam havası sağlayan uygulamalara ulaşılması amaçlanmıştır.

**[15] Sağlık hizmetlerinde büyük veri**

Keywords: Büyük veri, Sağlık hizmetleri

Yüksek boyutlarda üretilen bu verilerin klasik yöntemlerle depolanamaması ve anlamlı hale getirilememesi durumu “büyük veri” kavramını ortaya çıkarmıştır. Veri hacmi küçük olan, depolanması, yönetilmesi ve analizi zor olmayan geleneksel verilerin yerini, bilgi ve iletişim teknolojilerindeki hızlı gelişme ile her an elde edilebilen, hacminin büyüklüğü nedeniyle geleneksel veri işleme sistemleri ile depolanamayan, yönetilemeyen ve analiz edilemeyen ‘Big Data’ kavramı almış bulunmaktadır. Büyük Veriler kısaca kolaylıkla geleneksel araçlar tarafından işlenemeyen veriler olarak tanımlanabilir. Büyük Veri aslen veri analitiği ve veriden anlam bulma ile ilgilidir. 3. Instagram’da günde 95 milyon fotoğraf ve video paylaşımı yapıldığı ve Twitter aracılığı ile günlük yaklaşık 500 milyon mesaj gönderildiği belirtilmektedir. Bu verilerle büyük veri oluşmaktadır. Günümüzde çoklu kaynaklardan üretilen bu devasa veri miktarı nedeniyle artık işletmeler terabayta varan verilerle değil petabyt (1024 tb) olarak isimlendirilen verilerle uğraşmak zorundadır. İşletmeler yönüyle değerlendirildiğinde Google, İnternet'teki trilyonlarca sayfadan elde edilen bilgileri kullanır ve kullanıcıların ihtiyaçlarını birkaç saniye içinde eşleştiren arama sonuçları üretmek için programlar ve formüller geliştirir.

Amazon, müşterilerine ürünleri doğru şekilde önermek için veri koleksiyonlarını kullanmaktadır. Amazon bu verilerin hızla analizi yoluyla müşterilerin önceki siparişlere ve diğer etkenlere dayanarak satın alması beklenen ürünlerin nakliye edilmesine dayanan bir patent geliştirmiştir. Sürekli büyüyüp, çeşitlenen “bilgi çöplüğü” olarak ifade edilen veri yığınları içinden anlamlı bilgilerin elde edilip doğru hedeflere sunulabilmesi bir değerdir. Ancak bu değerin elde edilmesi için bu verilerin işlenmesi ve anlamlandırılması çok kolay değildir.

Büyük Veri kavramının oluşumunda beş bileşen vardır. Çeşitlilik(Variety), Hız (Velocity), Veri Büyüklüğü (Volume), Doğrulama (Verification) ve Değer (Value) olarak adlandırılan bu kavramlar, İngilizce ifadelerinin baş harflerinden oluşarak kısaca 5V olarak ifade edilmektedir. Sağlık hizmetlerinde üretilen muazzam miktarlardaki verinin başarıyla entegrasyonu, sağlık sisteminde yer alan hastalar başta olmak üzere tüm paydaşlar için büyük gelişmeler sağlayabilir. Büyük Verinin sunduğu imkanlar sayesinde sağlık sektöründe yer alan tıbbi sigorta ve ilaç endüstrileri, reçete edilen ilaçların özellikleri hakkında bilgi sahibi olabilirler.

Tıp eğitimi ve öğretiminde Büyük Veri iki biçimde değerlendirilmektedir. Birincisi, yakın gelecekte eğitim müfredatında bir ihtiyaç olacağı öngörüsü ile yer alması gerekliliği diğeri de Büyük Verinin bir araç olarak eğitim ve öğretimde kullanılmasıdır. Büyük Veri ile geliştirilen programlar, hastaların bakım ve kişisel sağlıklarının teşviki ve geliştirilmesinde günlük kalori tüketimi, glikoz seviyesi gibi kişisel sağlık verilerini depolamak, almak ve yönetmek için benzersiz bir yol sağlar. Böylece kişisel sağlık hizmetinin geliştirilmesini hızlandırır.

Sağlık hizmetlerinde üretilen muazzam miktarlardaki verinin başarıyla entegrasyonu, sağlık sisteminde yer alan hastalar başta olmak üzere tüm paydaşlar için büyük fırsatlar sunabilir. Sağlıkta büyük verilerin kullanımının sunduğu fırsatlar ülkeleri, bunun önemini farkına varmaya ve sağlık bilgilerinin depolanabileceği çok geniş kapsamlı veri merkezleri kurmaya yöneltmektedir.

**[16]** **Makine öğrenimi ile eğitim verilerinin modellenmesindeki yöntemler**

Keywords: Machine learning, Deep learning, Artifical intelligence, Artifical neaural networks

Bu çalışmada amaç lise öğrencilerinde akademik başarıyı etkileyen demografik, sosyoekonomik, tutum, sosyal aktivite, motivasyon, sağlık ve spor, akademik başarı kategorilerinde yer alan anket soruları yardımı ile akademik başarının çalışmanın büyük çoğunluğunda hedef degişken olarak yer alması ve bu faktörlerin akademik başarı hedef degişkenini etkileme derecesinin tespit edilip hangi makine ögrenmesi modellerinin bu gücü anlamlı bir şekilde yorumlayabildiği değerlendirilmesi amaçlanarak, bu çalışmanın sonucunda akademik başarıyı etkileyen faktörlerin ve etkileme derecelerinin belirlenerek eğitim sistemine katkı getirmesi amaçlanmıştır. Akademik başarının artırılmasına yönelik ¸çalışmalar her gecen gün artmakla birlikte teknolojinin gelişmesi ile bilgisayar bilimleri, akademik başarıyı etkileyen faktörlerin değerlendirilmesinde büyük katkılar sağlamaktadır. Makine ögrenmesi algoritmaları kullanılarak eğitim verilerinin modellen dirilmesi ve veri madenciliği ve Yapay zekanın birleşimiyle verilerin sınıflandırma, tahmin ve kümeleme çalışmaları yapılmaktadır.

Yöntem olarak kaggle’dan edinilen veri ilk önce kullanılabilir olacak şekilde hazır hale getirilmiştir. Makine ögrenme algoritmaları ile Denetimli ögrenme modellerinden olan Sınıflandırma, Regresyon, Kolektif ögrenme modelleri uygulanmış ve başarılı sonuçlar elde edilmiştir. Derin ögrenme modeli olan Yapay Sinir Ağları modelleri veri setine uygulanmıştır.

Regresyon modelleri kendi içerisinde, sınıflandırma modelleri kendi içerisinde değerlendirilerek en iyi performansla çalışan modeller değerlendirildiğinde, Regresyon modelleri içerisinde Multilineer Regresyon, Lasso Regresyon ,Ridge Regresyon modellerinin eğitim ve test sonuçları (her üçünün de aynı ) sırası ile 0.87 ve 0.77 dir. Ortalama kareler hata katsayısı değerleri incelendiğinde içerisinde Multilineer Regresyon 6.40, Ridge Regresyon 6.41, Lasso Regresyon 6.39 ortalama kareler hata katsayısına sahiptir. Regresyon modellerinde değerlendirme yapıldığında diğerlerinden acık ara fark olmamak üzere skorlar değerlerine bakılarak en iyi performansla çalışan sınıflandırma modeli Lasso Regresyon olmuştur.

Bu çalışmanın sonunda, Türkiye’de farklı okul türleri, farklı sınıf düzeyleri, farklı bölgelerden oluşan geniş bir örneklemle ögrenmeyi etkileyen faktörler farklı kategorilerde ve geniş bir şekilde yer almak şartı ile ögrenmeyi etkileyen faktörlerin başarılı algoritmalar ve modeller ile toplanan veri setine uygulanması geliştirdiğimiz algoritmaları uygulayarak ülkemizde eğitime katkı sağlamaktır.

**[17]** **İlaç etkileşimlerini makine öğrenmesiyle tespit eden tahmin modeli**

Keywords: İlaç etkileşimi, Derin öğrenme, Yapay zeka

İlaç etkileşimleri birden fazla ilaç kullanan insanlarda görülen ve insanların kullandığı ilaçların etkileşimde bulunarak birbirlerinin etkisini olumlu veya olumsuz yönde etkilemesi durumudur. Bu görülen etkiler genellikle bir ilacın diğer ilaç etkisini artırması, diğer ilaç etkisini azaltması veya diğer ilacın etki ediş biçimini değiştirmesi şeklinde olabilmektedir. u hem insan sağlığı açısından hem de hastanelerde oluşan ekstra sağlık masrafları açısından büyük maliyetlere neden olabilmektedir. İlaçların birbirleriyle olan etkileşimlerinin bir kısmı tespit edilmiştir. Ancak henüz test edilmemiş ve tespit edilememiş olan birçok etkileşim bulunmaktadır.

Bu tezde bilinmeyen ilaç etkileşimlerinin tespit edilebilmesi için çeşitli yapay zekâ ve makine öğrenmesi algoritmaları kullanılacaktır. Bu çalışmanın çıktısı birden fazla ilaç kullanmanın gerektiği durumlarda hem ilacı veren hekimi hem de ilacı kullanan kişiyi uyarmak için kullanılabilir. Böylece istenmeyen ilaç reaksiyonlarının önüne geçilebilecektir.

İlaç etkileşiminin tahmin edilmesinde temel olarak üç yöntem türünden bahsetmek mümkündür. Bunlar, ilaçların benzerliklerini esas alarak geliştirilen yöntemler, network ağları kullanılarak geliştirilen yöntemler ve makine öğrenmesi algoritmaları kullanan yöntemlerdir. Bu tezi hazırlamaktaki amacımız ilaç etkileşimlerini tahmin etmek üzere var olan makine öğrenmesi yöntemlerini kıyaslayarak en verimli modeli ortaya çıkarmaktır.

Makalede kullandığımız veriler Drugbank veri tabanından alınmıştır. DrugBank, ilaçlar ve ilaç hedefleri hakkında bilgi içeren kapsamlı, erişimi ücretsiz bir veri tabanıdır. DrugBank, ilaç endüstrisi, tıbbi kimyagerler, eczacılar, doktorlar, öğrenciler ve genel halk tarafından yaygın olarak kullanılmaktadır. Drugbank veri tabanı dosyası yapısal olarak oldukça karmaşık bir yapıdadır. Çalışmada kullanılacak olan verileri her seferinde bu dosyadan ayıklayarak kullanmak hem yazılımsal olarak maliyet getireceği ve karmaşıklığı artıracağı için hem de çalışmaları yapacağımız makina sistemlerini fazladan kullanacağı için bu veriler bir ilişkisel veri tabanına aktarılmıştır. Drugbank veri tabanı dosyası yapısal olarak oldukça karmaşık bir yapıdadır. Çalışmada kullanılacak olan verileri her seferinde bu dosyadan ayıklayarak kullanmak hem yazılımsal olarak maliyet getireceği ve karmaşıklığı artıracağı için hem de çalışmaları yapacağımız makina sistemlerini fazladan kullanacağı için bu veriler bir ilişkisel veri tabanına aktarılmıştır. Bu verilerde her bir ilaç diğer ilaçlar ile eşleştirildiğinde toplamda 210,925 adet ilaç çifti elde edilmiştir.

Bu tezde geçen bir diğer terim ATK, ilaç kullanımını izlemek ve kullanım kalitesini artırmak amacıyla araştırmalar yapmak için geliştirilmiş bir sınıflandırma sistemidir. İlaçları etken maddelerine göre sınıflandırmaya yarar. Bu sınıflandırma yapılırken ilacın etki ettiği yer ve ilaçların terapötik, kimyasal ve farmakolojik özellikleri dikkate alınır. İlaç enzimlerinin ve ilaç hedeflerinin makine öğrenmesi modeline girmesini sağlamak amacıyla veriler ATK kodlarına benzer şekilde bir ön işleme tabi tutulmuştur. İlaçların ATK kodu, moleküler yapısı, hedef proteinleri ve enzim bilgilerini kullanarak yaptığımız denemelerde en verimli yöntemi birinci veri seti üzerinde derin öğrenme sinir ağları kullanarak ilaçların farmakokinetik ve farmakodinamik etkileşimlerini tahmin ederken tespit edilmiştir. Modellerin eğitilmesi sonucu en verimli çalışan modelimizde giren örneklem kümesine göre %73 ile %98 arasında değişen ve ortalaması %88 olan başarı oranını elde edildi görülmüştür.

**[18] Sağlık hizmetlerinde büyük veriye ilişkin makine öğrenimi yaklaşımı**

Keywords: Big data, Machine learning

Sağlık hizmetlerinde büyük veri işleme, klasik veri işleme yöntemleriyle çıkarılamayacak kadar geniş veya karmaşık klinik verilerinin üretilmesi, toplanması, analiz edilmesi ve tutulması anlamına gelir. Büyük veri orijinal olarak, bir hastanın bakımıyla ilgili bilgileri içeren sağlık hizmeti sağlayıcıları tarafından çeşitli veri üretim süreçlerinden gelen verilerin hacmini, hızını ve çeşitliliğini gösterir. Büyük veri uygulamaları, çeşitli yeni bilgiler için yeni fırsatlar sunar ve sağlık hizmetlerinin kalitesini iyileştirmek için yenilikçi yöntemler yaratır.

Sağlık hizmetlerinde derin öğrenme ilaçların keşfedilmesine ve geliştirmesine yardımcı olur. Teknoloji hastanın tıbbi geçmişini analiz eder ve onlar için en iyi tedaviyi sağlar. Birden fazla makine öğrenme algoritması kullanılmış elde edilen sonuçlara göre lojistik regresyon ve rastgele ormanın daha doğru sınıflandırıcılar olduğu tespit edilmiştir.

Günümüzde ve ilaç endüstrisinde sağlık bakımı ile ilgili temel sınırlamalardan biri, hastalığın biyolojisinin anlaşılmasıdır. Büyük veri; DNA’dan, proteinlerden, metabolitlerden hücrelere, dokulara, organlara kadar bir hastalığı neyin oluşturduğuna ilişkin birden fazla ölçek etrafına giderek daha fazla bilgi toplama safhasında devreye giriyor. Bu yazıda çeşitli makine öğrenimi tekniklerini kullanarak büyük verilerin uygulamalarını, işlenmesini ve ele alınması tartışılmıştır. Ayrıca makine öğrenme modellerinin performanslarını değerlendirmek için kullanılan ölçütler büyük verilere dayanmaktadır. Sağlık hizmetlerinde büyük veri analitiği ve uygulamaları belirsiz bir gelişim aşamasındadır, ancak platformlardaki ve araçlardaki hızlı gelişmeler bunların gelişen sürecini hızlandırabilir.

**[19] Kişisel sağlık verilerinin dijitalleşmesi ve büyük veri**

Keywords: Kişisel veri, Kişisel sağlık verisi, Elektronik sağlık kaydı, Büyük veri

Yapılan çalışmanın amacı, kişisel sağlık verilerinin dijitalleşmesi sonucu oluşan büyük veri kavramının önemini ortaya koymaktır. Bu anlamda; kişisel sağlık verileri, elektronik sağlık kayıtları ve büyük veri kavramları ile ilgili literatür taraması yapılmıştır. Ayrıca büyük veri ile kişisel sağlık verileri arasındaki ilişki ele alınmıştır. Yapılan literatür taramasında, bilgi ve iletişim teknolojilerinin gelişmesi sonucu, kişisel sağlık verilerini işleme, depolama ve aktarma aşamalarında değişiklik ve elektronik sağlık kayıtlarının oluşturulduğu tespit edilmiştir. Diğer bir yandan, büyük veri kavramının gelişiminde; siber saldırılar, kötü niyetli yazılım şirketleri veya kullanıcı olarak bireylerin sağlık bilgi sistemlerine karşı olası güvensizlik kaygısına yönelik olumsuz tutum ve davranışlar gibi, önemli zorlukları oluşturduğu değerlendirilmiştir.

Yin vd. (2022) tarafından yapılan bir çalışmada elektronik sağlık kayıtları ile bilgi sistemlerinin hastalara tavsiye ettiği ilaç ile fiziki kağıtlar kullanılarak hekimlerin verdiği ilaçların etkisi karşılaştırılmıştır. Bu doğrultuda bilgi sistemlerinin verdiği ilaçların etkililiği daha yüksek çıkmıştır. Aradaki farkın sebepleri araştırıldığında ise insan hatası, ayıklama ve dönüştürme hataları ve yanlış teşhis hatasından kaynaklandığı tespit edilmiştir. Elektronik sağlık kayıtları dijital dönüşümün sağlık hizmetlerinde ortaya çıkardığı en önemli değişimlerden biri olarak kabul edilmektedir. Teknolojinin yaşamakta olduğu hızlı gelişim sürecine paralel olarak sağlıkta kullanılan diğer dijital araçlar ile elektronik sağlık kayıtlarının işlem hacminin artması sağlıkta büyük veri kavramının yer almasına yol açmaktadır.

Büyük veri kavramının önemli bir özelliği verinin bulut vb. dijital teknolojiler aracılığıyla paylaşılabilir nitelikle olmasıdır ve böylelikle farklı sistemlerle bütünleşmiş bir şekilde birlikte çalışabilir niteliğe sahip olması olmaktadır. Bu bakımdan büyük veri olarak üretilen ve saklanan elektronik sağlık kayıtları kurumlar arasında paylaşılabilmektedir. Elektronik sağlık kayıtları ve büyük veri için bir diğer önemli zorluk ise verilerin doğruluğunun sağlanması noktasında ortaya çıkmaktadır. Veriler toplanırken belirlenen yapıya uygun toplanmaması, eksik veya taraflı olması, güvenilirlik konusu ile amacın gerçekleştirilmesini tehlikeye atacaktır.

Bu çalışmada öncelikle kişisel veri kavramından hareket ederek özel nitelikli veri olarak kabul edilen sağlık verilerinden bahsedilmiştir. Sağlık verilerinin uğradığı dijitalleşme süreci değerlendirilerek ortaya çıkan elektronik sağlık kayıtları detaylı olarak incelenmeye çalışılmıştır. Bu veriler gerek tıbbi geçmişin incelenmesinde gerek tedavi sürecinin izlenmesinde gerekse de toplum sağlığına dair politikaların geliştirilmesinde kullanılabilmektedir. Ancak, büyük veri kavramı ve teknolojilerinin ortaya çıkışı ve yapılan tartışmalar bu anlayışın bir değişimle karşı karşıya kaldığını göstermektedir. Nitekim yukarıda ifade edildiği üzere modern sağlık hizmetleri sistemleri büyük veri teknolojilerinin kullanımı yoluyla toplum veya birey sağlığına dair tahminlerde bulunma kabiliyetini kavramıştır. Bu bakımdan olası sağlık sorunlarının tespiti ve önlenmesi veya sağlığın geliştirilmesi gibi faaliyetlerde büyük veri ile elde edilen bilgi daha iyi sağlık hizmetinin üretilmesi amacıyla karar verme mekanizmasına destek olacaktır.

**[20]** **Türkiye’de sağlık alanında veri madenciliği kullanım alanları**

Keywords: Veri madenciliği, Sağlık, Büyük veri

Gelişen teknoloji ve artan veri tabanı nedeniyle insanlar verileri anlamlı hale getirmek istediler. Bu nedenle ortaya çıkan veri madenciliği kavramı birçok kişi tarafından farklı tanımlanabilmektedir. İnsanoğlu her gün verilerle beraber onları yorumlayarak işine yarayacak hale getirmeye çalıştı. Başlangıç aşaması ENIAC (ilk sayısal bilgisayar) olarak düşünülmektedir.

Bu çalışmanın amacı, sağlık sektöründe veri madenciliği üzerine yapılan son yıllardaki çalışmaları inceleyerek bir araya getirilmesidir.

Veri madenciliği için izlenmesi gereken bazı adımlar vardır. Bu adımları incelediğimizde karşımıza çıkan durumları şu şekilde özetleyebiliriz. Bunlar:

* Problemin tanımlanması
* Verilerin hazırlanması
* Modelin kurulması ve değerlendirilmesi
* Modelin izlenmesi

Sağlık sektörü hem big dataların olduğu hem de sürekli değişen verilerin olduğu bir alandır. Fayez’in 2018 yılındaki göre Koroner kalp hastalığı ölümle sonuçlandığı için çok fazla dikkat çekilen bir konuydu. Veri madenciliğinin sınıflandırma tekniğini ve Phyton programlama dilini kullanarak koroner kalp hastalığı teşhisinde maliyet ve zaman olarak daha iyi bir sistem tasarlayan Fayez çalışmasında yüksek oranda doğruluk elde etmiş. Çalışmasını farklı koroner kalp hastalığı verilerinde uygulayarak Random forest algoritmasında %99, Cleveland’da %94, SVM algoritmasında %58 sonuç elde etmiştir. Bir diğer önemli çalışma Altun’un 2018 yılında yaptığı 2009 – 2017 yılları arasında beyin tümörü teşhisi konulan ve koyulmayan hastaların Mr verileri kullanılmıştır. Çalışmaya göre MR Spektroskopide ölçülen kolin metabolit değerini çok önemlidir. Hazırlanan veri seti WEKA programı yardımı ile çözülmüştür. Sınıflandırmanın beş yöntemi denenmiş olup bunlardan KNN Algoritması %95.69 oranı ile en yüksek başarıyı elde etmiştir.

Bu derlemede veri madenciliği hakkındaki bilgiler tek çatı altında toplanmaya çalışılmış, ülkemiz sağlık sektöründe veri madenciliği uygulamalarına ve kullanım alanlarına göz atılmıştır. Derlemedeki örnek uygulamalar ülkemizde yapılan çalışmalardan seçilmiş olup hastalık riski üzerine yapılan çalışmaların daha fazla olduğu gözlemlenmiştir. Veri madenciliğinin sağlık sektöründeki çalışmaları hastalık riski ve ilaç dozu ile sınırlı olmayıp sağlık sektöründe çalışanların beklentileri derleme içerisinde verilmiştir. Sağlık alanında ve diğer alanlarda bu büyük verilerin hastalıkları teşhis ve tespit etmede çok büyük kolaylıklar sağladığı görülmektedir.

**KAYNAKLAR**

[1] *” Büyük veride sağlık bilişimi ve analitiği”,* <https://www.sciencedirect.com/science/article/pii/S0957417420302128?ref=pdf_download&fr=RR-2&rr=761634ee695450be>

[2] *“Hastalar ve sağlık hizmetleri ile deneyime dayalı ortak tasarımı daha güvenli ilaç kullanımı için teoriye dayalı müdahaleler geliştirme”,* <https://www.sciencedirect.com/science/article/pii/S1551741121002060?ref=pdf_download&fr=RR-2&rr=761636682db450be>

[3] *“Sağlık hizmeti analizine dayalı makine öğrenimi algoritmalarının değerlendirilmesi”,* <https://www.sciencedirect.com/science/article/pii/S0965997822001867?ref=pdf_download&fr=RR-2&rr=76163047391150be>

[4] *“Sağlık Analitiğinde hassas ilacın karar vermeyi ne ölçüde etkilediği”,* <https://www.sciencedirect.com/science/article/pii/S1877050921024923?ref=pdf_download&fr=RR-2&rr=761637eb087550be>

[5] *“Makine öğrenme modelleri kullanarak verimli hastalık teşhisi “,* <https://downloads.hindawi.com/journals/jhe/2021/9983652.pdf>

[6] *“Veri madenciliğinde CART ve Lojistik regresyon analizinin yeri: İlaç provizyon sistemi verileri üzerine araştırma”,*   
 <https://dergipark.org.tr/en/download/article-file/152170>

[7] *“Denetimli makine öğrenimi sınıflandırıcıları ile sağlık hizmetleri kararlarını tahmin etme”,*  
<https://www.researchgate.net/profile/Muhammad-Amin-40/publication/329515048_Performance_Evaluation_of_Supervised_Machine_Learning_Classifiers_for_Predicting_Healthcare_Operational_Decisions/links/5c0c277aa6fdcc494fe4a3a6/Performance-Evaluation-of-Supervised-Machine-Learning-Classifiers-for-Predicting-Healthcare-Operational-Decisions.pdf>

[8] *” Büyük veri çağında işletmelerde veri bilimi”,* <https://www.researchgate.net/profile/Selin-Goekalp/publication/331481196_Buyuk_Veri_Caginda_Isletmelerde_Veri_Bilimi/links/5c7c0cb0a6fdcc4715ac8d47/Bueyuek-Veri-Caginda-Isletmelerde-Veri-Bilimi.pdf>

[9] *“Kronik hastalık tanısında makine öğrenimi öngörü modellerinin uygulamaları”,* <https://www.mdpi.com/2075-4426/10/2/21>

[10] *“Hastane bilgi sistemlerinde veri madenciliği”,* <https://ab.org.tr/ab08/kitap/Bildiriler/Yildirim_Uludag_Gorur_AB08.pdf>

[11] *“Veri madenciliğinin tıp ve sağlık hizmetlerinde kullanımı ve uygulamaları”,* <https://dergipark.org.tr/en/download/article-file/75259>

[12] *“Hastalık teşhisi için makine öğrenimi tekniklerinin uygulanması”,* <https://www.sciencedirect.com/science/article/pii/S2214785321072679?ref=pdf_download&fr=RR-2&rr=76164caf6cd7514d>

[13] *“Veri analizinde istatistik mi veri madenciliğimi?”,* <https://dergipark.org.tr/en/download/article-file/297864>

[14] *“İlaç endüstrisinde insan sağlığı açısından tehlikeli maddelerin risk değerlendirmesi”,* <https://polen.itu.edu.tr:8443/server/api/core/bitstreams/4f32870f-01de-42d0-8633-241fd96e136a/content>

[15] *“Sağlık hizmetlerinde büyük veri”,*   
<https://dergipark.org.tr/en/download/article-file/463040>

[16] *“Makine öğrenimi ile eğitim verilerinin modellenmesindeki yöntemler”,* <https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp>

[17] *“İlaç etkileşimlerini makine öğrenmesiyle tespit eden tahmin modeli”,* <http://earsiv.medeniyet.edu.tr:8080/xmlui/bitstream/handle/123456789/467/AhmetSener_Tez_2022.pdf?sequence=1&isAllowed=y>

[18] *“Sağlık hizmetlerinde büyük veriye ilişkin makine öğrenimi yaklaşımı”,* <http://www.aimspress.com/aimspress-data/bdia/2020/1/PDF/bigdia-05-005.pdf>

[19] *“Kişisel sağlık verilerinin dijitalleşmesi ve büyük veri”,* <https://dergipark.org.tr/en/download/article-file/2540120>

[20] *“Türkiye’de sağlık alanında veri madenciliği kullanım alanları”,* <https://dergipark.org.tr/en/download/article-file/688182>