

**Citation:** Oralhan, B. (2020), Veri Madenciliği Yaklaşımı İle Telekomünikasyon Sektöründe Arıza Giderme Analizi, BMIJ, (2020), 8(1): 1026-1043 doi: <http://dx.doi.org/10.15295/bmij.v8i1.1220>

## VERİ MADENCİLİĞİ YAKLAŞIMI İLE TELEKOMÜNİKASYON SEKTÖRÜNDE ARIZA GİDERME ANALİZİ

Burcu ORALHAN<sup>1</sup>

Received Date (Başvuru Tarihi): 23/08/2019

Accepted Date (Kabul Tarihi): 21/10/2019

Published Date (Yayın Tarihi): 25/03/2020

### ÖZ

Telekomünikasyon operatörleri tarafından sunulan hizmetlerde problem veya arıza ile karşılaşılmasının yanı sıra meydana gelecek bu tür olumsuzlukların giderilememesi, müşteri güvenini azaltmakta ve sonuç olarak gelir kaybına yol açmaktadır. Veri madenciliği, telekomünikasyon endüstrisinde mevcut verilere ilişkin analizler sayesinde geliştirilmiş bilgiyi sağlayabilmektedir. Bu çalışmada, Türkiye’de telekomünikasyon sektöründe yer alan öncü bir firmanın mobil ağlarında sorun giderme süreçlerine veri madenciliği süreci uygulanmıştır. Bu kapsamda Mart-Mayıs 2019 tarih aralığında elde edilen 4032 veri ön işleme süreci sonucunda 3748 örneklem olarak analize dahil edilmiştir. Arıza kaydı sürecinde kayıt altına alınan Arıza Merkezi, İş Emri, Ekip Numarası, Hizmet Türü, Hizmet Süresi, Şikâyet Türü ve Sonuç verilerinden oluşan 7 farklı değişken incelenmiştir. Elde edilen sonuçlara göre J48, PART ve Multilayer Perceptron sınıflandırıcısının veri kümesinde daha iyi performans gösterdiği görülmüştür. Süreçlerin etkin bir şekilde kontrol etmesini sağlamada arıza giderme analizlerinde yol gösterici bir çalışma olması bakımından önem arz etmektedir.

**Anahtar Kelimeler:** Veri Madenciliği, Arıza Analizi, Telekomünikasyon Hizmetleri

**JEL Kodları:** L96, C80, C44

## TROUBLESHOOTING ANALYSIS IN TELECOMMUNICATION SECTOR USING DATA MINING APPROACH

### ABSTRACT

Experiencing problems without troubleshooting in the services that is offered by telecommunication operators causes the decreasing customer loyalty and loss of income. Data mining provides improved information through analysis of available data in the telecommunications industry. In this study, the data mining was applied on the troubleshooting process of broadband network in one of a leading telecommunication company in Turkey. In this scope, 4032 data that obtained from the company during the March-May 2019 period were used. 3748 samples were included the analysis after the pre-processing step. Seven different variables including Trouble Center, Work Order, Team Number, Service Type, Duration of Service, Complaint Type and Result data were recorded during the trouble recording process. According to the result, J48, PART and Multilayer Perceptron classifiers were performed better than others in the data set. The current research is important in terms of being a guiding work in ensuring effective control of processes in troubleshooting analysis.

**Keywords:** Data Mining, Troubleshooting Analysis, Telecommunication Services

**JEL Codes:** L96, C80, C44

<sup>1</sup> Dr. Öğretim Üyesi, Nuh Naci Yazgan Üniversitesi, [boralhan@nny.edu.tr](mailto:boralhan@nny.edu.tr)

<https://orcid.org/0000-0001-8905-0140>

## 1. GİRİŞ

Günümüzde telekomünikasyon şirketleri, müşteri profillerini genişletmek için en fazla çaba gösteren şirketler arasındadır. Sabit ağ hizmetleri ve mobil hizmetleri içeren telekomünikasyon hizmetleri pazarı, 2017 yılında yaklaşık 1,4 trilyon ABD doları değerine sahip olmuştur ve 2020 yılına kadar yaklaşık 1,46 trilyon ABD doları büyüyeceği tahmin edilmektedir (Telekomünikasyon Hizmetleri, 2018). Son yıllarda gelişmiş yeni ağların ve akıllı telefonların kullanılması, mobil geniş bant trafiğinin yaygın olarak benimsenmesi ve milyonlarca mobil uygulama nedeniyle telekomünikasyon sektöründe patlayıcı bir büyüme yaşanmıştır (Singh vd., 2016).

İşletmeler sundukları hizmetlerle mevcut müşterilerini korumakla birlikte rekabetçi pazar şartlarında yeni müşterileri kazanabilmeleri büyük ölçüde işletmelerin altyapılarının sağlam kurulmuş olmasına ve işleyişine bağlıdır. Süreç içinde herhangi bir hizmetin bozulması veya arızalanmasının yanında meydana gelecek bu tür olumsuzlukların giderilememesi, müşteri güvenini azaltmakta ve sonuç olarak gelir kaybına yol açmaktadır (Csikor & Pezaros, 2017). Bu küresel büyümenin sağlanmasında, hizmetin sunulmasının yanı sıra destek süreçlerin ve ekiplerin de müşteri ihtiyaçlarının karşılanması konusunda uzmanlaştırılmasına ve sürekli geliştirilmesine dikkat edilmelidir.

Telekomünikasyon sektöründe süreçler, müşterilerin bağlantı ağlarına yönelik sorunları gidermek için birçok karmaşık operasyona sahiptir (Saafein & Shaykhian, 2014). İletişim tekniklerinin ve ağlarının hızla gelişmesiyle, insanlar yoğun olarak kablosuz ağlara güvenmekte, daha uygun ve kaliteli hizmet almayı ummaktadırlar (Chen vd., 2017). Bununla birlikte, telekomünikasyon ağları daha karmaşık hale geldikçe, telekomünikasyon hizmetlerinin ve ağ altyapısının etkin yönetimi, telekomünikasyon operatörlerinin karşılaştığı en büyük zorluklarından biri haline gelmiştir (Network, 2009; Melero vd., 2003; Mishra, 2004; Pirinen, 2014). Günümüzde mevcut süreçlerde ağ bakım ve yönetim sistemleri otomasyondan uzaktır ancak bu ağların karmaşık yapısından ve sürekli gelişim göstermesinden dolayı manuel olarak kontrol etmek gün geçtikçe zorlaşmaktadır (Aliu vd., 2012). Bu ağlarda sorunların giderilmesi özel çaba ve uzmanlık gerektirmektedir (Calyam, P., 2014). Belirtilen sistemlerin aktif kullanımı, geliştirilmesi ve sistemler üzerinde meydana gelebilecek problemlerin giderilmesi için bilgi ve deneyime sahip çok sayıda kıdemli uzmana ihtiyaç duyulmaktadır. Bu alanda sınırlı sayıda uzmanların olması sebebiyle, gerçek ihtiyaçların karşılanması zorlaşmakta, ağ yönetimi süreçlerinde görevlerin tamamlanmasında

başarısızlıklar yaşanabilmekte ya da gecikmelerle karşılaşılabilir (Chen vd., 2017).

Telekomünikasyon süreçlerinde herhangi bir sorun ortaya çıktığında, olası olumsuz etkileri azaltmak için doğru ve zamanında teşhis ve problemin çözümü için büyük önem arz etmektedir. Sürekli artan ağ/veri trafiği talepleri (veri merkezlerinde işleme ve depolama) sonucunda farklı katmanlarda (örneğin yük dengeleyiciler, önbellekler, hızlandırıcılar) sorunlar ve arızalar oluşabilmektedir (Csikor & Pezaros, 2017). Bu durum bireylerin ve makinenin bir ekip olarak çalışabileceği, makine gücü ve insan uzmanlığının faydalarını birleştirecek bir yaklaşıma yönelmesine sebep olmaktadır (Steinhauer vd., 2017). Geçmişte, operatörler iş gücünü artırarak hızlı teknolojik değişimlerle başa çıkmayı başardılar. Ancak, mali baskılar nedeniyle bu artık geçerli bir strateji olmaktan çıkmıştır. Operasyonel maliyetlerin azaltılması için mümkün olan önemli seçeneklerden biri otomasyon seviyesinin artırılmasıdır (Barco vd., 2005).

Operasyondaki mevcut kablosuz ağlar genellikle büyük miktarlarda veri üretirler. Bu oluşturulan veriler arasında, ağ yönetiminde kablosuz ağların durumlarını analiz etmek için genellikle alarm verileri (ALM), çağrı geçmişi kayıtları (CHR) ve anahtar performans göstergeleri (KPI) genellikle kullanılır (Jakobson & Weissman, 1993; Barco vd., 2005; Sundsøy vd., 2016; Steinhauer vd., 2017; Singh vd., 2016). Ancak, oluşturulan (mevcut) verilerin çoğunu oluşturan günlük veriler genellikle dikkate alınmaz. Ne yazık ki, bu günlük veriler bol miktarda bilgiyi içermektedir (Chen vd., 2017). Sürekli artarak işlenen ve depolanan veriler büyük veri olarak literatürde yer almaktadır. Büyük veri analizi, uzman bilgisi kullanarak süreçle ilgili bilinenleri keşifsel veri analizi ile modellemek yerine (Tukey, 1962) süreçle ilgili verilerin istatistiksel karar verme yöntemlerini örneğin kümeleme gibi analizler kullanarak süreçle ilgili farklı bakış açıları elde edilmesini sağlamaktadır (Steinhauer vd., 2017). Büyük veri birçok kuruluş için büyük bir rekabet avantajı haline gelmiştir. Analitik platformlar tarafından sağlanan büyük verilerin analizi, organizasyonların süreçlerini iyileştirmek ve geliştirmek için önemli bir adımdır (Skračić & Bodrušić, 2017). Telekomünikasyon sistemlerinde yer alan bakım ve sorun giderme amacıyla kayıt altına alınan büyük verilerin analizi müşterilerin korunması için önemli rol oynamaktadır.

Tüm endüstriyel şirketlere nüfuz eden telekomünikasyon şirketleri, telefon, televizyon ve internet ağlarının operasyonları ve kullanıcıları hakkında muazzam miktarda bilgiye sahiptir. Bilgi ve iletişim sistemlerinin sürekli gelişmesiyle geçmiş, yeni veya iyileştirilen hizmetlerine yönelik büyük miktarda veriyi de depolarında saklamak zorundadırlar.

Kaydedilen bu büyük veriler analiz edilerek, hizmet kalitesinin artırılması, iletişim türünün belirlenmesi, suistimal faaliyetlerin tespit edilmesi, tarifelerin belirlenmesi gibi birçok konuda faydalı olabilmektedir. Bu konuda büyük verilerin analizinde veri madenciliği birçok sektörde olduğu gibi telekomünikasyon endüstrisinde de kaynak kullanımının daha iyi hale getirilebilmesinde yardımcı olacaktır

## **2. MATERYAL VE YÖNTEM**

Günümüz teknolojisinde telekomünikasyon sistemleri açısından da önemli değişiklikler olmuştur. Metal kablolar yerini fiber optik kablolarla bırakmış, eski telefon servisi (POTS- Plain old telephone service), IPTV, VoIP veya internet erişimi gibi yeni iletişim servisleri ile yer değiştirmiştir. Bakır tel tabanlı teknolojiler yeni tip hizmetlere ayak uyduramadığı için Dijital abone hattı DSL (Digital Subscriber Line) teknolojilerinin evrimi geliştirilmiştir. Asimetrik iletişim ise ADSL (Asimetrik Sayısal Abone Hattı) teknolojisi ile gerçekleştirilebilmektedir (Recommendation, 1999). Veri aktarım hızı (başka bir deyişle bit hızı) yükleme ve indirme işleminde farklılık gösterebilmektedir. İnternet'in ortalama kullanımına uygun olarak, yükleme bit hızı indirme işleminden daha düşüktür. İnternet bağlantısı, İnternet Protokolü Televizyonu (IPTV- Internet Protocol Television) ve İnternet Üzerinden Ses Protokolü (VoIP- Voice over Internet Protocol) hizmetleri özellikle konut müşterileri ve ofis kullanıcıları tarafından talep edilmektedir (Lilik, 2015). İki uç nokta arasındaki iletişimde, çok sayıda telekomünikasyon teknolojisi ve ağı kullanılmaktadır. Merkez ve ağ yapısı arasında optik ve mikrodalga bağlantılarında kullanılan iletim teknolojileri güvenilirdir. Bu sistemler aynı zamanda çok sayıda müşteriye iletişim hizmeti sunduğundan dolayı planlama, kurulum ve bakım sırasında teknik güvenlik büyük bir rol oynamaktadır (Lilik, 2015). Belirtilen tüm süreçlerde birçok veri kaydedilmekte, işlenmekte ve depolanmaktadır. Büyük veri analizi, telekomünikasyon endüstrisinde mevcut verilere ilişkin geliştirilmiş bilgi sağlamaktadır. Bu çalışmada, Türkiye'de telekomünikasyon sektöründe yer alan öncü bir firmanın mobil ağlarında sorun giderme verimliliğini ve değerini artırmak için veri analitiğinin kullanılabileceği uygulamalı olarak gösterilmiştir.

Telekomünikasyon şirketleri tarafından sunulan ve uygulanan teknolojilerin çeşitlendirilmesi müşteriler açısından memnuniyet vericidir. Ancak bu bağlantılar son derece güvenilir ve yönetilebilir olmasına rağmen erişim ağlarında yaşanan problemler sistemleri verimsizleştirmekte ve müşteri kayıplarına yol açabilmektedir. Bu sebeple sunulan hizmetler ve güvenlikleri sürekli olarak izlenmeli ve ortaya çıkabilecek arızalar derhal gözlemlenerek

onarılmalı ve hatalar bertaraf edilmelidir (Lilik, 2015). Sorun giderme işleminde ilk adım, performans göstergelerine ve alarmlara bağlı olarak kötü performans gösteren hücrelerin tanımlanması olan arıza tespittir. İkinci adım, tanı, neden tanımlamadır, yani sorunların nedenini bulmak için otomatik muhakeme mekanizmalarıdır. Son olarak, problem çözme, problemleri çözmek için eylemlerin yürütülmesidir (Barco vd., 2005).

Her operatör, örneğin hattan düşen aramalar, erişim hataları, tıkanıklık vb. gibi farklı performans göstergelerini temel alan problemlili hücreleri tanımlamak için farklı bir yöntem kullanır (Laiho vd., 2002). Problemlili alan tespit edildikten sonra bir mantıksal veya fiziksel bileşenin arızalı davranışına yönelik problemin teşhisi yapılmalıdır. Sorun giderme modelinin ana unsurları;

- Nedenleri: Neden olabilecek olası hataları
- Semptomlar: Sebeplerin belirtileridir.
- Koşullar: Üzerinde etkisi olabilecek faktörlerdir.
- Bağlantılar: Önceki unsurlar arasındaki ilişkilerdir (Barco vd., 2005).

Telekomünikasyon sektöründe birçok önemli farklı konunun ele alındığı istatistiksel karar vermeye yönelik büyük veri analizlerini kullanan çalışmalar mevcuttur (Skračić & Bodrušić, 2017; Singh vd., 2016). Örneğin, aboneleri hareketlerine göre sınıflandırması (Furletti vd., 2013), müşteri karlılığının operatör karlılığı üzerinde doğrudan bir etkiye sahip olup olmadığının incelenmesi (Li vd., 2016), WhatsApp veya Viber gibi over-the-top (OTT) uygulamalarını kullanırken müşteri deneyiminin tahmin edilmesi (Diaz-Aviles vd., 2015), nesnelerin internetinde sensor verilerinin analiz edilmesi (Din vd., 2015), ağ yapısının veri transfer hızına etkisi (Tomkos vd., 2015), mobil ağ performansında sorun giderme (Singh vd., 2015), xDSL üzerinden IPTV için sorun giderme testi için mevcut hat testi çözümlerinin verimliliğini artırma (Škaljo vd., 2018) vb. Sürekli gelişen ve değişen hizmetleri için telekomünikasyon şirketleri yeni ve verimli yöntemler aramaya devam etmektedir (Lilik vd., 2015). Bu gibi yeni uygulamalarla var olan verilerin yeniden kullanılarak analiz edilmesi, yakın gelecekte tüm işletmeler için önemli bir gelir kaynağı olacağı düşünülmektedir (Cao vd., 2014).

Veri analizi, veri toplama, veri analizi ve rapor oluşturma sürecini kapsamaktadır. Veri madenciliği iş akışı araçları da veri analizi sürecini yönetmek için kullanılır. Veri madenciliği sürecinde yer alan veri analizi bir dizi makine öğrenme algoritmasını içermektedir. Sistemde

yer alan fonksiyonel ve fonksiyonel olmayan faktörler ve bu faktörlerin kombinasyonlarının etkisinin anlaşılması, iş akışı araçlarının ve makine öğrenimi algoritmalarının seçilmesi için bir kılavuz oluşturur (Singh vd., 2016). Veri madenciliği sadece bazı araçların ve tekniklerin uygulanması olmayıp, verilerin toplanması, verilerin ayıklanması, modellerin oluşturulması, modellerin test edilmesi ve uygulanması gibi aşamaları içeren bir süreç olarak ifade edilebilmektedir (Savaş, Topaloğlu ve Yılmaz, 2012).

Veri madenciliği, veri tabanındaki bilgi keşfi süreçlerindedir (Rokach & Maimon, 2008). Veri madenciliği, mevcut karmaşık yapıdaki verileri, istatistiksel metotlarla yararlı ve anlaşılabilir hale getiren, özetleyen ve gizli ilişkileri ortaya koyan bir yöntemidir (Larose, 2005). Veri madenciliği için verilerin doğrudan kullanımı uygun değildir. Bu sebeple, araştırmacılar veri etkinliğini inceler ve bazı ön işlemleri uygulamaktadırlar. Bu çalışmada veri örnekleme, ön işleme, model oluşturma ve model değerlendirme aşamaları yer almaktadır. Veri ön işleme, eksik değerler, yanlış kelimeler, yanlış matematiksel semboller, çoğaltılan bilgiler ve bu gibi hatalardan oluşan ilgisiz bilgileri ortadan kaldırmaktadır. Veri örnekleme ile belirlenen telekomünikasyon şirketindeki iş tanımına göre, gerekli bilgileri içeren bir müşteri grubunu rastgele seçilmektedir. Araştırmada ön işleme tekniklerinden bazıları kullanılmıştır. Veri temizleme adımı, yanlış bilgi içeren veya yanlış tanımlanan hücrelerdeki verileri temizlenmektedir. Ham veriler genellikle, tipik büyüklükleri ve çoklu ve heterojen kaynaklardan muhtemel kökenleri nedeniyle genellikle gürültülü, tutarsız ve eksiktir. Acker'e göre, veri ön işleme toplam analiz çalışmasının %80'ine kadar olabilir ve bir kez birleştirildiğinde, temizlendikten ve dönüştürüldükten sonra tüketilen veriyi analiz ederek %20'ye kadar çıkabilir (Acker vd., 2013).

Veri ön işleme adımıyla veri kalitesinin boyutlarında iyileştirme sağlanmaktadır. Ön işleme tamamlanan veriler veri madenciliği analizlerine uygun hale getirilmiş olmaktadır. Veri madenciliğinde yapay zekâ, istatistik, makine öğrenmesi gibi alanlarda geliştirilmiş teknikler ve yöntemler uygulanmaktadır (Rygielski vd., 2002). Farklı amaçlarda farklı alanlarda kullanılabilen veri madenciliğinde sıklıkla verilerin tanımlanması, sınıflanması, kümelenmesi, ilişkilendirilmesi, tahmin edilmesi gibi uygulamalar kullanılmaktadır (Larose, 2005). Gözlemlerin ortak özelliklere sahip kümelere atanması yaygın bir biçimde uygulanmaktadır. Sınıflandırma ve kümeleme uygulamalarına yönelik literatürde müşteri profil analizi, ayrılan müşteri analizi gibi alanlarda yapılmış örnekler mevcuttur. Veri madenciliğinde sıklıkla kullanılan sınıflandırma uygulaması ikiye ayrılır. Eğitilen öğrenmede



sınıf sayısı ve örnek grubu önceden belirlenir. Bir öğrenme kümesi model olarak alınır. Bu model kullanılarak sınıfı belirlenmeyen diğer verilerin sınıfı belirlenmeye çalışılır (Akçetin ve Çelik, 2014). Veri madenciliği yöntemlerinden en popüler olanlar: Sınıflama ve regresyon modelleri, tahmin edici, kümeleme; birliktelik kuralları ve ardışık zamanlı örüntü modelleri ise tanımlayıcı modellerdir (Dunham, 2003; Aynekin, 2006; Albayrak ve Yılmaz, 2009; Bozkır, 2009). Karar ağaçları, Yapay Sinir Ağları ve Regresyon analizleridir. Çalışmada Weka 3.8V açık erişim programı kullanılarak veriler analiz edilmiştir. Program Bayes, MISC, Fonksiyonlar, Kurallar, Karar Ağacı, Tembel vb. gibi sınıflandırıcı algoritmaların bir kısmını içermektedir. Tablo 1’de veri madenciliğinde fonksiyon, teknik ve uygulamaları yer almaktadır.

**Tablo 1.** Veri Madenciliği İşlevselliği, Teknikleri Ve Örnek Uygulamaları

<i>İŞLEVSELLİK</i>	<i>TEKNİK</i>	<i>UYGULAMA</i>	<i>İŞLEVSELLİK</i>	<i>TEKNİK</i>	<i>UYGULAMA</i>
<b>İlişki</b>	Küme Teorisi	Çapraz Satışlar	<b>Tahmin</b>	Regresyon	Ayrılma Tahmini Suistimal Tahmini
	İstatistik				
	Bayes Sınıflaması				
<b>Kestirim</b>	Sinir Ağları	Döviz Kuru Tahmini	<b>Bölümleme</b>	Sinir Ağları	Pazar Bölümlendirme
	İstatistik				
	Zaman Serileri				
<b>Sınıflandırma</b>	Karar Ağacı	Kredi Zimmeti Pazar Bölümlendirme			
	Bulanık Sistemler				
	Sinir Ağı				
	<b>Genetik Algoritma</b>				

**Kaynak:** Hung, Yen ve Wang, 2006

Sınıflandırmayı oluşturan sistemler veri madenciliğinde yaygın olarak kullanılan araçlardan biridir. Bu tür sistemler, her biri az sayıda sınıftan birine ait olan ve sabit bir nitelik kümesi değerleri ile tanımlanan vakalardan oluşur. Yeni bir vakanın ait olduğu sınıfı doğru bir şekilde tahmin edebilen bir sınıflandırıcı elde edilmiş olur. Araştırmada kullanılan sınıflama algoritmaları;

### **2.1. Naive Bayes:**

En hızlı istatistik sınıflandırıcı algoritmasından birisi olup veri örneğinde yer alan tüm özelliklerin ayrı ayrı olasılığı üzerinde çalışmaktadır ve daha sonra bunları doğru bir şekilde sınıflandırmaktadır. Sınıf üyeliği olasılıklarını, yani belirli bir sınıfa ait olup olmadığına ilişkin pay hakkında olasılık tahmin etmek için kullanılmaktadır (Theodorescu, 1968). Bayes sınıflandırması Bayes teoremine dayanmaktadır (Hilden, 1984). Özet olarak, Naive Bayes şartlı bir olasılık modelidir: sınıflandırılacak bir problem örneği verildiğinde, bazı  $n$  özelliklerini (bağımsız değişkenler) temsil eden bir  $X = (x_1, x_2, \dots, x_n)$  vektörüyle temsil edilen

bir problem örneği verildiğinde, bu örnek olasılıklara verir (Friedman vd., 1997). Olası sonuç veya sınıfların her biri için  $p(C_k | x_1, \dots, x_n)$ .

Eğer özelliklerin sayısı yani  $n$  büyükse veya bir özellik çok sayıda değer alabilirse, o zaman böyle bir modeli olasılık tablolarına dayandırmak olanaksızdır. Bu nedenle modeli daha izlenebilir hale getirmek için yeniden düzenlemeye ihtiyaç duyulmaktadır. Bayes teoremi koşullu olasılık olarak tanımlanabilir.

$$p(C_k|X) = \frac{p(C_k)p(X|C_k)}{p(X)} \quad (1)$$

## 2.2. J48:

Bilgi tabanlı bir ölçü olan kazanç oranı, J48 sınıflandırıcı algoritması tarafından varsayılan bir bölme kriteridir, sınıflandırma için basit bir C4.5 karar ağacıdır (Korting,2006). Denetimli sınıflandırma yöntemidir. ID3 algoritmasının bir uzantısıdır. Bu sınıflandırıcıda Divide and Conquer yaklaşımı verileri sınıflandırmak için kullanılmaktadır. Amacı verileri eğitim örneğinde bulunan özellik değerine göre aralıklara bölmektir (Korting, 2006).

J48 yapı olarak tek değişkenli bir karar ağacıdır. Bu yaklaşım menzil tabanlı ve tek değişkenli olduğu için çok değişkenli yaklaşımdan daha iyi performans göstermemektedir. Bu karar ağacı değerlerin tahmin edilmesinde faydalı bir yöntemdir. Doğru sınıflandırılmış örneğin J48 doğruluğu, tek değişkenli olan diğer algoritmalarından çok daha fazladır. Akış şeması benzeri bir düğüm ve dalların olduğu bir ağaç yapısı ile oluşmaktadır. Düğümler, sınıf etiketlerini veya sınıf dağılımını temsil etmektedir.

Karar ağacı üretimi iki aşamadan oluşur. Ağaç yapısında tüm eğitim örnekleri kökte toplanmıştır. Seçili özneliğe bağlı olarak olarak tekrarlı bölümlenme ile kademeler alt dallara doğru ilerlemektedir. Karar ağacının kullanımında bilinmeyen bir veri örneğinin sınıflandırılması karar ağacına göre test edilmektedir (Apté, C., & Weiss, S., 1997; Fayyad, 1994).

## 2.3. Multilayer Perceptron:

Çok Katmanlı Perceptron, Perceptron'u temel alan doğrusal olmayan bir sınıflandırıcıdır. Günümüz dünyasında en popüler ağ mimarisidir. Her birim girdilerinin etkili ağırlıklı toplamını gerçekleştirmekte ve çıktılarını üretmek için bu etkinleştirme seviyesini bir aktarma işlevinden geçirmektedir. Birimler, katmanlı ileri beslemeli bir topolojide düzenlenmektedir. Ağ, ağırlıklar ve eşikler ile basit bir giriş-çıkış modeline sahiptir. Bu tür



ağlar, fonksiyon karmaşıklığını belirleyen, katman sayısı ve her katmandaki birim sayısı ile neredeyse karmaşık olan işlevleri modelleyebilmektedir.

Çok Katmanlı Perceptron'daki önemli konular, gizli katman sayısının ve bu katmanlardaki birim sayısının tasarım özellikleridir. Çok Katmanlı Bir Algılayıcı (MLP), giriş ve çıkış katmanı arasında bir veya daha fazla katman bulunan bir geri yayılım sinir ağıdır.

#### **2.4. JRip:**

JRip (RIPPER), temel ve en popüler algoritmalarından birisidir. Sınıflar artan boyutta incelenir ve sınıf için başlangıçtaki bir kurallar dizisi artan hata kullanılarak üretilir. JRip (RIPPER), belirli bir yargılamaya ait tüm örnekleri, eğitim verilerindeki bir sınıf olarak ele almakta ve bir kurallar dizisi bularak ilerlemektedir. Bu sınıfın tüm üyelerini kapsamaktadır. Bundan sonra bir sonraki sınıfa ilerlemekte ve aynısını yapmakta, tüm sınıflar kapsanana kadar bunu tekrar etmektedir (Rajput vd., 2011). JRip, IREP'in optimize edilmiş bir versiyonudur (Cohen, 1995). William W. Cohen tarafından tanıtıldı. Tekrarlanan artımlı JRip ile hata azaltma üretmektedir. (Shahzad vd., 2013)

#### **2.5. Random Forest:**

Breiman tarafından önerilen rastgele orman sınıflandırıcısı, birçok bireysel sınıflandırma ağacından (bu çalışmada ağaç sayısı 10'dur) oluşmaktadır; burada her ağaç, sınıflandırma çıktısı için belirli bir ağırlık verilen bir sınıflayıcıdır (Weka'da tüm ağaçlara aynı ağırlık verilir). Tüm ağaçlardan sınıflandırma çıktıları, tüm ağaç sınıflandırma çıktılarının modunu (en fazla oy alan çıktı) seçerek yapılan genel sınıflandırma çıktısını belirlemek için kullanılmaktadır. (Gislason vd., 2006)

Rastgele orman (RF), veriyi ikili bölme kurallarını kullanan karar ağaçlarına dayalı bir yöntemdir. Sınıflandırma problemlerinde, verileri bölmek için kullanılan ana kurallar Gini endeksi, sapma ve çekme kuralıdır (Breiman, 2001, MathWorks, 2017). Bu kurallar arasında, düğüm kirliliğini ölçen Gini endeksi en yaygın kullanılanıdır. (De Santana, F.B. 2019)

#### **2.6. OneR:**

OneR algoritması, eğitim verilerinin her bir niteliği için tek bir kural oluşturmakta ve ardından kuralı en düşük hata oranıyla almaktadır. Öznitelik için bir kural oluşturmak üzere, her öznitelik değeri için en fazla yinelenen sınıf oluşturulmalıdır. En çok tekrarlayan sınıf, bu öznitelik değeri için en sık görünen sınıftır. Bu kural, özniteliğin dayandığı en çok tekrarlanan

sınıfa bağılı bir öznitelik değerleri kümesidir. (Devasena vd., 2011).

Kuraldaki öznitelik değerinin bağlanması ile uyuşmayan egzersiz verisi örneklerinin sayısı hata oranını vermektedir. OneR kuralı en düşük hata oranıyla seçer. İki veya daha fazla kural aynı hata oranına sahipse, kural rastgele seçilmektedir (Devasena vd., 2011).

### **2.7. PART:**

Bu bir PART karar listesi oluşturmak için bir sınıftır. Ayır, böl yaklaşımını kullanmakta, her yinelemede kısmi bir C4.5 karar ağacı oluşturmakta ve "en iyi" yaprağı bir kural haline getirmektedir (Devasena vd., 2011).

Bu araştırmada Türkiye'nin en büyük telekomünikasyon şirketinden elde edilen arıza kaydı verilerine yönelik veri madenciliğinin sınıflandırma algoritmaları uygulanmıştır.

### **3. VERİ SETİ**

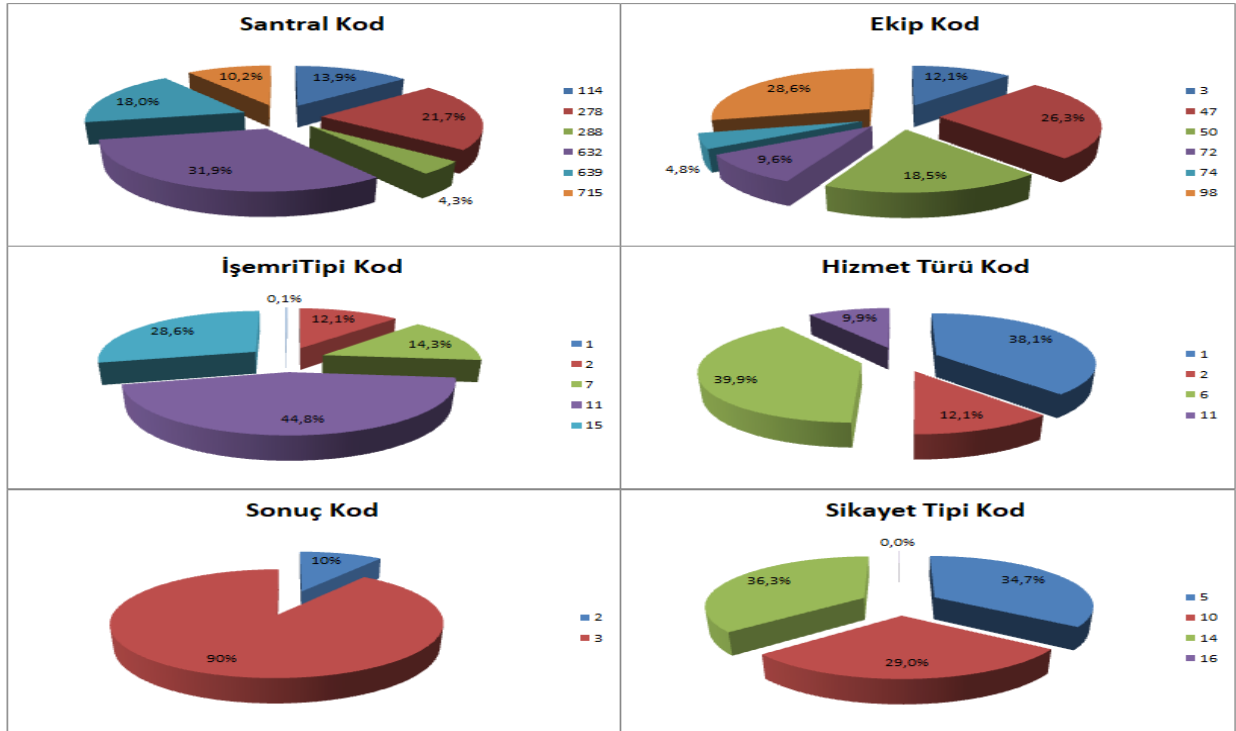
FFF Veri madenciliği için verilerin doğrudan kullanımı uygun değildir. Bu sebeple, araştırmacılar veri etkinliğini inceler ve bazı ön işlemleri uygulamaktadırlar. Çalışmamızda veri örnekleme, ön işleme, model oluşturma ve model değerlendirme aşamaları yer almaktadır.

Çalışmada kullanılan veri seti, Türkiye'nin en büyük telekomünikasyon şirketlerinden birinden Mart-Mayıs 2019 tarih aralığında elde edilen gerçek zamanlı verilerden oluşmaktadır. Bu çalışmada, veri madenciliği yöntemleri ile telekomünikasyon sisteminde izlenen 7 farklı faktörün arıza sorunu giderme analizi üzerindeki etkisi araştırılmıştır. İşletme yetkilileriyle yapılan beyin fırtınası ve tutulan kayıtların incelenmesi sonucunda en fazla arıza kayıtlarına yönelik üzerinde durulan değişkenlerin araştırma kapsamında kullanılması planlanmıştır. Bu kapsamda veri kaynağı değişkenleri Arıza Merkezi, İş Emri, Ekip Numarası, Hizmet Türü, Sonuç Türü, Hizmet Süresi, Şikâyet Türü'nden oluşmaktadır. 4032 örnekten oluşan işlenmemiş veriler, ön işlemlere tabi tutularak 3748 örneklem olarak düzenlenmiştir. Çalışmada kullanılan değişkenler ve yapılan kodlamalar aşağıdaki Tablo 2'de yer almaktadır.

**Tablo 2.** Araştırmada Kullanılan Değişkenler Ve Kodları

Kodu	Arıza Merkezi	Kodu	Ekip No.
114	AYDINLIKEVLER-121	3	1
278	ESKİŞEHİR BAĞLARI-121	47	5
288	FEVZİÇAKMAK SNT-122	50	7
632	ŞEKER -121	72	10
639	TALAS SNT-122	74	11
715	YENİMAHALLE -121	98	12
Kodu	İşemri Tipi Kod	Kodu	Sikayet Tipi Kod
1	CPE Arıza İslah	5	ERİŞİM HIZI DÜŞÜK/ERİŞİLEMİYOR
2	FDH Arıza İslah	9	HATALI PORT ARIZASI
7	Lokal Arıza İslah	10	IP ALAMIYORUM/INTERNETE GİRİLMİYOR
11	Repartitör Arıza İslah	14	MODEMLER BULUSMUYOR (ISIK YOK)
15	XDSL Operasyon Arıza İslah	16	SINYAL YOK
Kodu	Sonuç	Kodu	Hizmet Türü Kod
2	ÇÖZÜLEMEDİ	1	ADSL
3	GİDERİLDİ	2	FIBER INTERNET
Kodu	Hizmet Süresi (dk)	Kodu	Hizmet Türü Kod
-	Nümerik veri	11	VDSL2

Araştırmada kullanılan faktörler ve alt faktörlerin dağılımı Şekil 1'de sunulmuştur.



**Şekil 1.** Araştırmada Kullanılan Verilerin Dağılımı

Araştırmada yer alan arızanın çözülüp çözülemediğine ilişkin Sonuç Kod'un diğer değişkenlerle ilişkisinin tespit edilmesi amacıyla korelasyon analizi uygulanmıştır. Korelasyon analizi sonuçları Tablo.3'de yer almaktadır.

**Tablo 3.** Korelasyon Analizi Sonuçları

Değişken	Santral	Ekip	İşemri Tipi	Hizmet Türü	Isemri Tamamlanma Süresi (Dk)	Sikayet Tipi	
Sonuç Kod	r	,086**	-,142**	,433**	-,320	-,458**	-,266**
	P	,000	,000	,000	,044	,000	,000
	N	3748	3748	3748	3748	3748	3748

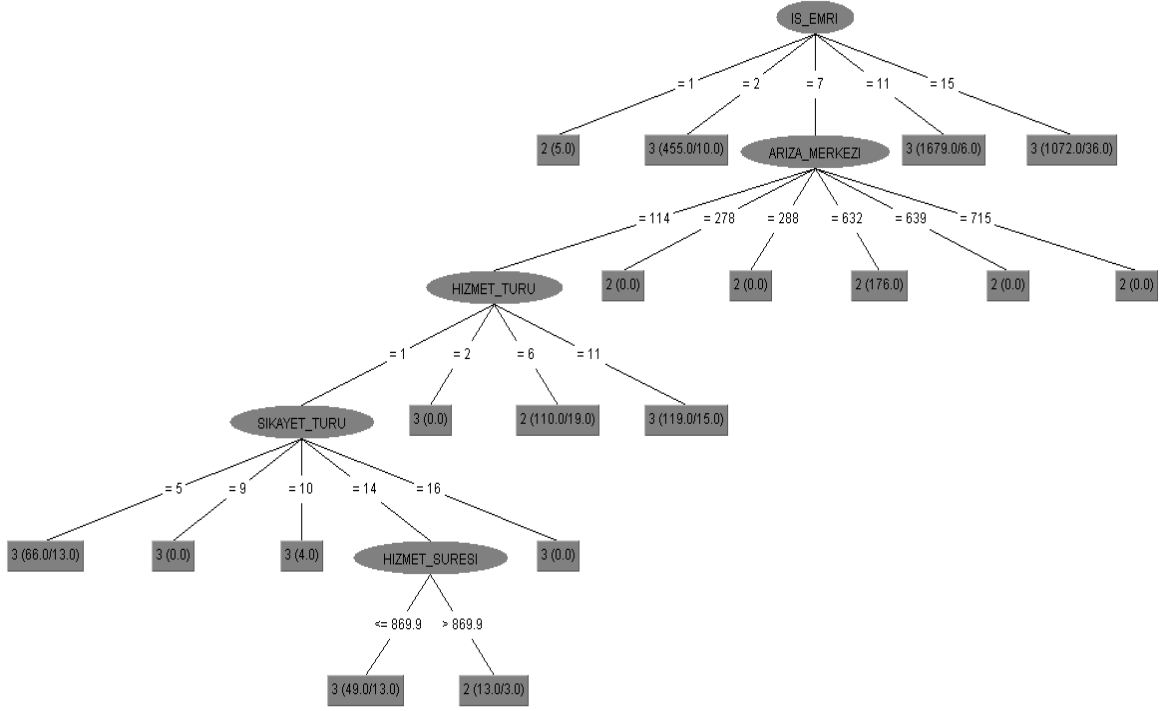
Yapılan korelasyon analizine göre araştırma kapsamında belirlenen değişkenlerle arızanın sonuç durumu arasında istatistiksel olarak anlamlı ilişkiler mevcuttur ( $p < 0,05$ ). Ancak ilişkilerin düzeyi orta ve düşük seviyededir. Sonuç olarak belirlenen değişkenlerin istatistiksel olarak anlamlı olmasından dolayı araştırma kapsamında kullanılabileceği anlaşılmıştır.

#### 4. ANALİZ

Bu çalışmada literatürde en fazla kullanılan sınıflandırma algoritmaları uygulanmıştır. Kullanım kolaylığı, açık erişimle ulaşılabilmemesinden dolayı WEKA 3.8V (Hall vd., 2009) programı araştırma analizlerinde kullanılmıştır. Veri setlerinin eğitim ve test setleri olarak ayrılmasında ise, 10-kat çapraz doğrulama (Kohavi, 1995) yöntemi uygulanmıştır. Farklı sınıflandırma algoritmaları kullanılarak, kurulan modellerin performansları karşılaştırılmıştır. Doğru sınıflandırılmış örnekler performansı en yüksek olan 7 algoritma sonuçları ve karşılaştırması Tablo 3'de yer almaktadır. Elde edilen sonuçlara göre J48, PART ve Multilayer Perceptron sınıflandırıcılarının veri kümesinde doğru sınıflandırılmış örnekler özelliğine göre daha iyi performans gösterdiği görülmüştür.

**Tablo 3.** Sınıflandırıcı Algoritmaların Performanslarının Karşılaştırılması

Sınıflayıcı	Modeli Oluşturma Süresi	Doğru Sınıflandırılmış Örnekler	Kappa İstatistiği	Ortalama Mutlak Hata	Ortalama Mutlak Hata Karesi	Relatif Mutlak Hata
J48	0,01	96,79%	0,8046	0,0554	0,169	30,79%
Naive Bayes	0,01	92,26%	0,6479	0,0774	0,2689	42,90%
JRIP	0,13	96,58%	0,7899	0,0626	0,1791	34,72%
PART	0,06	96,74%	0,7987	0,0546	0,1692	30,29%
Random Forest	0,17	95,09%	0,7221	0,0545	0,1929	30,23%
MultilayerPerceptron	13,45	96,42%	0,7857	0,0519	0,1710	28,79%
OneR	0,01	94,82%	0,6268	0,0518	0,2275	28,71%



**Şekil 2.** J.48 Karar Ağacı

Tablo 3’de sunulduğu üzere doğruluk performansı en yüksek olan sınıflandırma algoritması J48 olduğu için arıza çözümlenebilmesine yönelik bölümünde karar ağacı oluşturularak yapı modellenmeye çalışılmıştır. Karar ağacında her satır, bir düğümü, alt satırlar, ilgili satırın alt düğümlerini göstermektedir. Düğümlerde yer alan parantezin içindeki soldaki sayı veri kümesindeki vaka sayısını varsa ikinci sayı yanlış olarak sınıflandırılan vakaların sayısını ifade etmektedir.

Şekil 2’de görüldüğü üzere iş emri, arıza merkezi, hizmet türü, şikâyet türü, hizmet süresi olacak şekilde kademeler yukarıdan aşağıya doğru sıralanmaktadır. Algoritmasının sonuçlarına göre;

1. İlk dallanmada eğer iş emri kodu CPE Arıza Islah ise arızanın çözülemediği, FDH Arıza Islah, Repartitör Arıza Islah, XDSL Operasyon Arıza Islah arızaları ise çözülebildiği görülmüştür. Eğer iş emri kodu Lokal Arıza Islah ise bir alt dallanma oluşmaktadır.

2. İkinci dallanmada, eğer arıza merkezi 114 kodlu Aydınlık-121 ise bir alt dallanmaya geçildiği, diğer merkezlerinde karşılaşılan bir arıza ise arızanın çözülemediği görülmektedir.

3. Üçüncü dallanmada eğer hizmet türü FBER NET ve VDSL2 ise arızaların giderilemediği eğer hizmet türü NDSL ise arızanın giderilemediği görülmektedir. Şayet ADSL

arızası ise bir alt dallanma ile değerlendirilmelidir.

4. Dördüncü dallanmada eğer şikayet türü “erişim hızı düşük/erişilemiyor”, “hatalı port arızası”, “IP alamıyorum/internete girilmiyor”, “sinyal yok” arızalarından biri ise arızanın çözümlendiği, eğer “Modemler Buluşmuyor (Işık Yok)” ,ise bir alt dallanma ile değerlendirilmesi gerekmektedir.

5. Beşinci dallanmada eğer hizmet süresi 869dk’yı geçen arızaların çözülemediği, 869dak’dan az olan arızaların çözümlenebildiği görülmektedir.

## 5. SONUÇLAR VE TARTIŞMA

Telekomünikasyon sektörünün güçlü bir şekilde gelişmesiyle birlikte, servis sağlayıcılar süreçleri ve müşterileri hakkında daha fazla bilgiye ihtiyaç duymaktadır. Günümüzün zorlu rekabet şartlarında, mevcut müşterileri elde tutmak ve yeni müşteriler kazanmak, süreçlerin iyi yönetilmesine bağlıdır. Bu süreçlerin en önemlilerinden birisi de müşterilere satış sonrası sunulan hizmetlerin etkinliği ve kalitesidir. Telekomünikasyon operasyonlarında şikayetler ve operasyonel problemler büyük bir sorun olarak karşımıza çıkmaktadır. Sunulan hizmetlerin geliştirilmesi ve ortaya çıkan sorunların giderilebilmesinde sürekli işlenen ve depolanan büyük verinin analiz edilmesi sürecin sorunsuz işletilmesinde yol gösterici olmaktadır (Han & Kamber, 2006) . Muhtemel oluşabilecek arızaları tahmin etmek ve oluşmadan önlenmesi veya oluştuysa bir an önce giderilebilmesi için veri madenciliği faydalı bir araç olarak uygulanabilmektedir. Arıza sisteminin takibini manuel işleyen sistemler yerine otomasyon sistemler üzerinden takip etmek süreç verimliliği açısından faydalı olacaktır. Otomasyon, radyo ağının çok farklı alanlarında (Barco vd., 2005), frekans planlamasında (Wille vd., 2002), parametre optimizasyonunda (Toril vd., 2002), sorun giderme (Barco vd., 2002) vb. birçok alanda uygulandığı görülmüştür.

Telekomünikasyon sektöründe müşteri ilişkileri yönetiminde müşteri kaybı en kritik konulardan birisidir. Sektörde rekabet edebilmek için, hizmet sağlayıcıların, süreçlerde ortaya çıkabilecek problemleri ve olasılıklarını tahmin edebilmesi ve mevcut müşterileri elde tutmak, müşteri sadakatini sağlamak için proaktif bir yaklaşım benimsemesi gerekir. Bu araştırmada, telekomünikasyon sektöründe kayıt altına alınan arıza kayıtlarını WEKA veri madenciliği algoritmalarından farklı teknikleri kullanılarak sınıflandırmak amaçlanmıştır.

Yapılan çalışmada, veri madenciliği teknikleri kullanılarak Türkiye’nin önde gelen Telekomünikasyon firmasının arıza kaydı analizlerine yönelik kayıtları inceleyerek, arıza



sürecindeki sonuçları ortaya koyan modeller geliştirilerek; arızaların çözüme ulaşım, ulaşmamasına yönelik karar verme süreçlerine önerilerde bulunulmuştur. Verilerin temizlenmesi ve gerekli formata dönüştürülmesini içeren ön işleme sonrasında; WEKA açık erişim programının sınıflandırma algoritmaları kullanılarak veriler modellenmiştir. Sınıflama tekniklerinden NaiveBayes, OneR, Multilayer Perceptron sinir ağları, JRIP ve PART kuralları, Random Forest ve J48 karar ağaçları olmak üzere 7 farklı algoritma kullanılmıştır. Algoritmaların performansları karşılaştırılarak sınıflandırma için en başarılı algoritmanın J.48 karar ağacı olduğu saptanmıştır. J.48 karar ağacı ile arızanın giderilip, giderilemeyeceğine yönelik bir karar ağacı modellenmiştir. Yapılan çalışma Türkiye’de Telekomünikasyon sektöründe arıza sorun giderme konusunda veri madenciliği yaklaşımını ilk defa uygulayan çalışma olması özelliğiyle önem arz etmektedir. Veri madenciliği konusu süreçlerin etkin yönetilmesinde verilerdeki gizli bilgileri çeşitli algoritmalar kullanarak ortaya çıkardığı için yol gösterici bir araç olarak tercih edilmektedir.

Arıza veya sorun giderme konusunda yapılacak analizler sayesinde süreç içindeki bilgilerin aktif olarak kullanılabilmesi için gerek analizlerin yapılabilmesinde gerekse arızanın giderilebilmesinde özel yetiştirilmiş uzman personele olan ihtiyaç vardır. Bu süreçler otomasyona bağlanarak işletmeler bünyesinde anahtar performans göstergeleriyle (KPI) desteklenerek akışın sonuç doğruluk ve zamanlama performansının iyileştirilmesi sağlanabilecektir. KPI ölçümü mobil operatörler ve mobil ağ altyapısı satıcıları da olmak üzere telekomünikasyon işletmelerinde, sistem/ağ darboğazlarını tespit etmede, arıza/sorun gidermede ve anormalliklerin tanımlanması gibi süreçlerde de kullanılabilir (Singh vd., 2016). Araştırmada sunulan veri madenciliği uygulamalı çalışması, yöneticilerin veya akademisyenlerin gerek sektör gerek konu içeriği değiştirilerek yapacakları veri madenciliği çalışmasında yol gösterici olacaktır. İlerideki çalışmalarda, uzman personel istihdamı veya otomasyon süreçlerinin kullanılması sonucunda arıza yönetim süreçlerindeki değişimin değerlendirilebilmesi sağlanabilecektir. Farklı değişkenler ilave edilerek sürecin karar ağacındaki yapıya nasıl etki ettiği incelenebilecek, yapılan çalışmanın etkinliği ölçülebilecektir. Müşteri ilişkileri yönetiminde satış sonrası hizmetler ve operasyonel süreçler maliyet, zaman ve memnuniyeti direk etkileyen arıza yönetimi konusuna gerekli önem verilmelidir.

## KAYNAKÇA

- Acker, O., Blockus, A., & Pötscher, F. (2013). Benefiting from big data: A new approach for the telecom industry. *Strategy&, Analysis Report*.
- Aliu, O. G., Imran, A., Imran, M. A., & Evans, B. (2012). A survey of self organisation in future cellular networks. *IEEE Communications Surveys & Tutorials*, 15(1), 336-361.
- Akçetin, E., & Çelik, U. (2014). İstenmeyen elektronik posta (spam) tespitinde karar ağacı algoritmalarının performans kıyaslaması. *Journal of Internet Applications & Management/İnternet Uygulamaları ve Yönetimi Dergisi*, 5(2).
- Albayrak, A. S., & YILMAZ, Ö. G. Ş. K. (2009). Veri madenciliği: Karar ağacı algoritmaları ve İMKB verileri üzerine bir uygulama. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 14(1), 31-52.
- Apté, C., & Weiss, S. (1997). Data mining with decision trees and decision rules. *Future generation computer systems*, 13(2-3), 197-210.
- Aynekin, G. (2006). *İnternet içerik madenciliğinde yapay sinir ağları ve bir uygulama* (Master's thesis, Uludağ Üniversitesi).
- Barco, R., Guerrero, R., Hylander, G., Nielsen, L., Partanen, M., & Patel, S. (2002, September). Automated troubleshooting of mobile networks using bayesian networks. In *Proc. IASTED International Conference on Communication Systems and Networks (CSN02), Malaga, Spain* (p. 105110).
- Barco, R., Wille, V., & Díez, L. (2005). System for automated diagnosis in cellular networks based on performance indicators. *European Transactions on Telecommunications*, 16(5), 399-409.
- Bozkır, A. S. (2009). OLAP ve veri madenciliği teknolojilerinden yararlanılarak web tabanlı bir karar destek sisteminin gerçekleştirilmesi. *Yayımlanmış Yüksek Lisans Tezi. Hacettepe Üniversitesi, Ankara*.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Calyam, P., Dovrolis, C., Jörgenson, L., Kettimuthu, R., Tierney, B., & Zurawski, J. (2013). Monitoring and troubleshooting multi-domain networks using measurement federations [guest editorial]. *IEEE Communications Magazine*, 51(11), 53-54.
- Cao, H., Dong, W. S., Liu, L. S., Ma, C. Y., Qian, W. H., Shi, J. W., ... & Cohen, D. (2014). SoLoMo analytics for telco Big Data monetization. *IBM Journal of Research and Development*, 58(5/6), 9-1.
- Chen, R., Ji, W., Duan, S., Ling, Q., & Li, F. (2017, August). A novel method to analyze logs generated by wireless telecommunication systems. In *2017 2nd International Conference on Advanced Robotics and Mechatronics (ICARM)* (pp. 235-238). IEEE.
- Cohen, W. W. (1995). Fast effective rule induction. In *Machine learning proceedings 1995* (pp. 115-123). Morgan Kaufmann.
- Csikor, L., & Pezaros, D. P. (2017, December). End-Host driven troubleshooting architecture for software-defined networking. In *GLOBECOM 2017-2017 IEEE Global Communications Conference* (pp. 1-7). IEEE.
- de Santana, F. B., Neto, W. B., & Poppi, R. J. (2019). Random forest as one-class classifier and infrared spectroscopy for food adulteration detection. *Food chemistry*, 293, 323-332.
- Devasena, C. L., Sumathi, T., Gomathi, V. V., & Hemalatha, M. (2011). Effectiveness evaluation of rule based classifiers for the classification of iris data set. *Bonfring International Journal of Man Machine Interface*, 1(Special Issue Inaugural Special Issue), 05-09.
- Diaz-Aviles, E., Pinelli, F., Lynch, K., Nabi, Z., Gkoufas, Y., Bouillet, E., ... & Salzwedel, J. (2015, October). Towards real-time customer experience prediction for telecommunication operators. In *2015 IEEE International Conference on Big Data (Big Data)* (pp. 1063-1072). IEEE.
- Din, S., Ghayvat, H., Paul, A., Ahmad, A., Rathore, M. M., & Shafi, I. (2015, December). An architecture to analyze big data in the internet of things. In *2015 9th International Conference on Sensing Technology (ICST)* (pp. 677-682). IEEE.
- Dunham, M. H. (2006). *Data mining: Introductory and advanced topics*. Pearson Education India.
- Fayyad, U. M. (1994, October). Branching on attribute values in decision tree generation. In *AAAI* (pp. 601-606).

- Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian network classifiers. *Machine learning*, 29(2-3), 131-163.
- Furletti, B., Gabrielli, L., Renso, C., & Rinzivillo, S. (2013, October). Analysis of GSM calls data for understanding user mobility behavior. In *2013 IEEE International Conference on Big Data* (pp. 550-555). IEEE.
- Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recognition Letters*, 27(4), 294-300.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *SIGKDD Explor. Newsl.*, 11(1), 10-18. doi: 10.1145/1656274.1656278.
- Han, J., Kamber, M., & Pei, J. (2006). *Data mining: concepts and techniques*. 2001. San Francisco: Morgan Kauffman.
- Hilden, J. (1984). Statistical diagnosis based on conditional independence does not require it. *Computers in biology and medicine*, 14(4), 429-435.
- Jakobson, G., & Weissman, M. (1993). Alarm correlation. *IEEE network*, 7(6), 52-59.
- Kohavi, R. (1995, August). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai* (Vol. 14, No. 2, pp. 1137-1145).
- Korting, T. S. (2006). C4. 5 algorithm and multivariate decision trees, image processing division. *National Institute for Space Research-INPE, SP, Brazil*.
- Laiho, J., Kylvaja, M., & Hoglund, A. (2002, May). Utilization of advanced analysis methods in UMTS networks. In *Vehicular Technology Conference. IEEE 55th Vehicular Technology Conference. VTC Spring 2002 (Cat. No. 02CH37367)* (Vol. 2, pp. 726-730). IEEE.
- Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data: an introduction to data mining* (Vol. 4). John Wiley & Sons.
- Li, H., Yang, D., Yang, L., & Lin, X. (2016, October). Supervised massive data analysis for telecommunication customer churn prediction. In *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom)(BDCloud-SocialCom-SustainCom)* (pp. 163-169). IEEE.
- Lilik, F., Simonyi, P., & Kóczy, L. T. (2013). Computational intelligence in performance evaluation and fault prognosis in telecommunication access networks. *Czasopismo Techniczne*.
- MathWorks, Inc. (2017). MATLAB 2017b.
- Melero, J., Halonen, T., & Romero, J. (2003). *GSM, GPRS and edge performance: evolution towards 3G/UMTS*. The British Library, TJ International.
- Mishra, A. R. (2004). *Fundamentals of cellular network planning and optimisation*. Evolution to. Wiley-Interscience.
- Network, S. O. (2009). *NEC's proposals for next-generation radio network management*. NEC Corporation Technical white paper.
- Pirinen, P. (2014, November). A brief overview of 5G research activities. In *1st International Conference on 5G for Ubiquitous Connectivity* (pp. 17-22). IEEE.
- Rajput, A., Aharwal, RP, Dubey, M., Saxena, SP ve Raghuvanshi, M. (2011). E-yönetişim verileri için J48 ve JRIP kuralları. *Uluslararası Bilgisayar Bilimi ve Güvenliği Dergisi (IJCSS)* , 5 (2), 201.
- Rokach, L.& M. Oded (2008), *Data Mining With Decision Trees*, World Scientific, New Jersey.
- Rygielski, C., Wang, J.-C. & Yen, D. C.. (2002). Data mining techniques for customer relationship management. *Technology in Society*, 24 (4), 483-502.
- Saafein, O., & Shaykhian, G. A. (2014). Factors affecting virtual team performance in telecommunication support environment. *Telematics and Informatics*, 31(3), 459-462.
- Savaş, S., Topaloğlu N., & Yılmaz, M. (2012). Veri madenciliği ve türkiye'deki uygulama örnekleri. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, sayı 21, s. 1-23.
- Shahzad, W., Asad, S., & Khan, M. A. (2013). Feature subset selection using association rule mining and JRip

classifier. *International Journal of Physical Sciences*, 8(18), 885-896.

Singh, S., Liu, Y., Ding, W., & Li, Z. (2016, June). Evaluation of data mining tools for Telecommunication Monitoring Data using design of experiment. *IEEE International Congress on Big Data (BigData Congress)* (pp. 283-290). IEEE.

Singh, S., Liu, Y., Ding, W., & Li, Z. (2016). Empirical evaluation of big data analytics using design of experiment: case studies on telecommunication data. *Services Transactions on Big Data*, 3(2).

Škaljo, N., Begović, A., Turajić, E., & Behlilović, N. (2018). On ability of troubleshooting by observing some physical layer parameters of xdsl transceivers. *International Journal of Electronics and Telecommunications*, 64.

Skračić, K., & Bodrušić, I. (2017, May). A Big Data solution for troubleshooting mobile network performance problems. *40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (pp. 472-477). IEEE.

Steinhauer, H. J., Helldin, T., Karlsson, A., & Mathiason, G. (2017, November). Topic modeling for situation understanding in telecommunication networks. *27th International Telecommunication Networks and Applications Conference (ITNAC)* (pp. 1-6). IEEE.

Sundsøy, P., Bjelland, J., Reme, B. A., Iqbal, A. M., & Jahani, E. (2016, January). Deep learning applied to mobile phone data for individual income classification. *International Conference on Artificial Intelligence: Technologies and Applications*. Atlantis Press.

Theodorescu, R. (1968). Good, I. J.: The Estimation of Probabilities. An essay on modern Bayesian Methods. Research Monograph No. 30, The M. I. T. Press Cambridge 1965. 109 S., 7 Tab., 115 Lit. *Biometrische Zeitschrift*, 10(1), 87-87.

Tomkos, I., Kachris, C., Khodashenas, P. S., & Soldatos, J. K. (2015, July). Optical networking solutions and technologies in the big data era. *17th International Conference on Transparent Optical Networks (ICTON)* (pp. 1-1). IEEE.

Toril, M., Pedraza, S., Ferrer, R., & Wille, V. (2002, May). Optimization of signal level thresholds in mobile networks. In *Vehicular Technology Conference. IEEE 55th Vehicular Technology Conference. VTC Spring 2002*. (Vol. 4, pp. 1655-1659). IEEE.

Telekomünikasyon hizmetleri - İstatistikler ve Gerçekler, 2018. <https://www.statista.com/topics/2665/telecommunication-services/>, erişim tarihi: 11.12.2018

Tukey, J. W. (1962). The future of data analysis. *The annals of mathematical statistics*, 33(1), 1-67.

Wille, V., Kuurne, A., Burden, S., Dunn, G., & Barco, R. (2002, September). Simulations and trial results for mobile measurement based frequency planning in GERAN networks. In *Proceedings IEEE 56th Vehicular Technology Conference* (Vol. 1, pp. 625-628). IEEE.