



## Veri Madenciliği Yöntemleri ile Türkiye’de Fertlerin E-Ticaret Kullanımını Etkileyen Faktörlerin Analizi

Deniz TANIR\*

Sahib RAMAZANOV\*\*

### Öz

Günümüzde e-ticaret kullanımının giderek yaygınlaşması ile işletmeler açısından internet üzerinden satış giderek daha fazla önemli olmaktadır. E-ticaretin kullanımındaki bu artış tüketicilerin e-ticaret kullanımını etkileyen faktörlerin neler olabileceği sorusunu da beraberinde getirmektedir. Bu çalışmada Türkiye’de fertlerin e-ticaret kullanım sıklıklarını etkileyen özellikler veri madenciliği yöntemlerinden karar ağaçları ve Destek Vektör Makineleri (DVM) algoritmaları ile analiz edilmiş ve yorumlanmıştır. Ayrıca e-ticaret kullanımını etkileyen değişkenlerden yararlanarak fertlerin e-ticaret kullanım sıklıklarını tahminleyici sınıflandırma modeli tasarlanmıştır. Çalışmanın bulgularına göre CHAID, C&R Tree ve QUEST algoritmalarında e-ticaret kullanımını etkileyen en önemli değişken son üç ay içinde özel kullanım amacıyla internet üzerinden mal veya hizmet satın alım sayısı olarak saptanmıştır. CHAID ve QUEST algoritmalarında ikinci sırada e-ticaret kullanımını etkileyen en önemli değişken yaş olurken C&R Tree algoritmasında son üç ay içinde özel amaçla (mobil uygulamalar dahil) E-posta gönderme/alma değişkeni olmuştur. E-ticaret kullanım sıklığı tahmini için en iyi sınıflandırma sonucunu %86,19 doğruluk oranı ile CHAID algoritması vermiştir.

**Anahtar Kelimeler:** Veri madenciliği, e-ticaret, karar ağaçları.

**Makale Türü:** Araştırma Makalesi

## Analysis of Factors Affecting Individuals' E-Commerce Use in Turkey Using Data Mining Methods

### Abstract

Nowadays, with the widespread use of e-commerce, online sales are becoming more and more important for businesses. This increase in the use of e-commerce brings with it the question of what factors may affect the use of e-commerce by consumers. In this study, the features that affect the e-commerce usage frequency of individuals in Turkey are analyzed with decision trees and Support Vector Machine (SVM) algorithms and interpreted. In addition, a classification model has been designed to predict e-commerce usage frequency of individuals by using variables that affect e-commerce usage. According to the results of the study, the most important variable which affects the e-commerce usage frequency in CHAID, C&R Tree and QUEST algorithms has been determined as the number of purchases of goods or services over the internet for private use in the last three months. While variable of age has been the second most important variable in CHAID and QUEST algorithms, the variable of sending/receiving e-mails for special purposes (including mobile applications) in the last three months has been the second most important variable in the C&R Tree. For e-commerce usage frequency estimation, the CHAID algorithm has given the best classification result with an accuracy rate of 86.19%.

**Keywords:** Data Mining, e-commerce, decision trees.

**Article Type:** Research Article

\* Dr. Öğr. Üyesi, Kafkas Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Yönetim Bilişim Sistemleri Bölümü, [tanirdeniz35@gmail.com](mailto:tanirdeniz35@gmail.com), ORCID iD: 0000-0001-6593-6625

\*\* Bilim Uzmanı, Kafkas Üniversitesi, Sosyal Bilimler Enstitüsü, İşletme Anabilim Dalı, [sahibramazanli@gmail.com](mailto:sahibramazanli@gmail.com), ORCID iD: 0000-0003-2582-3188

## 1. GİRİŞ

E-ticaret, son yıllarda hızla gelişen bir sektör olarak kabul edilmektedir. İnternetin yaygınlaşması ve artan mobil cihaz kullanımı, e-ticaret işletmelerinin satışlarını arttırmakta ve müşteri hizmetlerini geliştirmektedir. Ancak, e-ticaret kullanımını etkileyen birçok faktör bulunmaktadır. Bu makalede, e-ticaret kullanımını etkileyen faktörler incelenecektir. Öncelikle, e-ticaretin genel bir tanımı yapılacak ve daha sonra e-ticaret kullanımını etkileyen faktörler olarak teknolojik, ekonomik ve sosyal faktörler incelenecektir. Bu faktörler arasındaki etkileşimler ve e-ticaret işletmelerinin bu faktörleri nasıl yönetebilecekleri de ele alınacaktır.

E-ticaret, klasik ticaret yöntemlerinden farklı olarak ürünlerin ve hizmetlerin satışının tamamen veya büyük ölçüde internet üzerinden gerçekleştirildiği bir ticaret modelidir. Dünya Ticaret Örgütü tarafından yapılan tanıma göre e-ticaret mal ve hizmetlerin üretim, reklam, satış ve dağıtımlarının telekomünikasyon ağları üzerinden yapılmasıdır (Yaman, Geçgil ve Yavuz, 2018: 2). Bir diğer tanıma göre e-ticaret, tüketici ürünlerinin internet üzerinden satın alınması ve satılmasıyla ilişkilidir (Thanasankit, 2003: 150). E-ticaret, internet kullanarak satıcıların ve tüketicilerin ihtiyaçlarını karşılayan modern bir işletme olarak da tanımlanabilir (Goel, 2007: 1).

İçerisinde bulunduğumuz teknoloji çağındaki hızlı gelişimler toplumların teknoloji kullanımını yaygınlaştırmakla birlikte dünyadaki e-ticaret kullanımının da artmasında sebep olmuştur (Akçi ve Annaç Göç, 2015: 414). Türkiye’de de e-ticaret kullanım hacmi yıldan yıla artış göstermeye devam etmektedir. Türkiye Cumhuriyeti Ticaret Bakanlığının açıkladığı verilere göre her yıl ortalama %35 büyüyen e-ticaret hacmi 2020 yılında %66 ve 2021 yılında %69 artış göstermiştir. E-ticaret hacminin genel ticarete oranı 2019 yılında yüzde 9,8 iken, 2020 yılında 15,7 ve 2021 yılında bu oran %17,7 ye ulaşmıştır (Türkiye Cumhuriyeti Ticaret Bakanlığı, 2022). E-ticaret kullanımındaki bu artış işletmeler için e-ticareti daha da önemli kılmakta ve bu durum beraberinde e-ticaret sektörünü daha fazla rekabetçi hale getirmektedir. E-ticaret kullanımını etkileyen faktörlerin analizi işletmelerin e-ticaret alanında müşteri memnuniyeti sağlayabilmesi, pazarlama stratejileri oluşturabilmesi dolayısıyla da diğer işletmeler ile olan rekabet gücünü arttırabilmesi için önemlidir. Bu analiz için istatistiksel veya ekonometrik analiz yöntemleri kullanılabileceği gibi veri madenciliği algoritmalarından da yararlanılabilir.

Veri madenciliği, verilerdeki önceden bilinmeyen örüntüleri tespit ederek veri kümelerindeki gizli bilgilerin ortaya çıkarılmasında ve böylece daha etkili karar vermede yardımcı olur. Veri madenciliği büyük veri kümelerinden yararlı bilgileri ortaya çıkarmak için kullanılır. Anlamsız devasa veri kümelerinde gizli örüntüleri bulmak için veri analizi, tahmin ve sınıflandırma için farklı yaklaşımlar kullanılmaktadır (Adekitan ve Salau, 2019). Bu faydalı bilgiler, son teknoloji veri madenciliği algoritmaları tarafından çıkarılır. Bir veri madenciliği algoritması, faydalı bilgileri araştırmak ve veri kümelerindeki örüntüleri bularak sınıflandırma ve tahmin modelleri oluşturmak için izlenen bir dizi adımdır. Veri madenciliği yapmak için Karar Ağacı, K-En Yakın Komşular, Rastgele Orman, Destek Vektör Makinesi (DVM), Yapay Sinir Ağı ve Naive Bayes gibi birçok veri madenciliği algoritması kullanılmaktadır. Veri madenciliğinin yoğun ve başarılı bir biçimde kullanıldığı başlıca alanlar; pazarlama, finans (bankacılık, sigortacılık, borsa), perakendecilik, sağlık, telekomünikasyon, endüstri ve mühendislik, eğitim, tıp, biyoloji, genetik, kamu, istihbarat ve güvenlik biçiminde sıralanabilir.

## 2. LİTERATÜR TARAMASI

Çalışmada e-ticaret kullanımını etkileyen faktörler karar ağaçları ve destek vektör makineleri algoritmaları ile analiz edilmiştir. İlgili yöntem ve konuya ilişkin çalışmalar kronolojik sırayla aşağıda listelenmiştir.

Tanır, D. & Ramazanov, S. (2023). Veri Madenciliği Yöntemleri ile Türkiye’de Fertlerin E-Ticaret Kullanımını Etkileyen Faktörlerin Analizi. *KMÜ Sosyal ve Ekonomik Araştırmalar Dergisi*, 25(44), 46-65.

Emel ve Taşkın (2005), çalışmasında ele alınan bir perakendeci işletme için işletmenin müşterilere göre kişiselleştirilmiş satış hareketlerini içeren veri tabanından yararlanarak bir satış hacmi analizi yapılmıştır. Bunun için veri madenciliğinin bir sınıflama algoritması olan karar ağacı yöntemi uygulanmıştır. Yapılan çalışmada C&R Tree karar ağacı algoritması ile müşteriler harcama tutarına göre sınıflandırılmıştır.

Kayrı ve Boysan (2007), çalışmasında CHAID analizi kullanılarak üniversite öğrencilerinin stresli durumlarda kullandıkları baş etme stratejileri, demografik özellikleri ve psikolojik belirtilerin düzeyi arasındaki ilişkiler değerlendirilmiştir.

Otukei ve Blaschke (2010), bu çalışmada sırasıyla 1986 ve 2001'den Landsat TM ve ETM+ veri kümelerinin analizi için bir veri madenciliği tekniği olarak DT'lerin potansiyeli değerlendirilmiştir. Sonuçlar, DVM ve MLC sonuçlarıyla karşılaştırılmıştır. Her durumda, %85'in üzerinde tatmin edici doğruluk seviyelerine ulaşılmıştır. Genel olarak, DT'ler MLC ve DVM modellerinden daha iyi performans göstermiştir.

Türen, Gökmen ve Tokmak (2011), çalışmasında e-ticaret işlem hacmini etkilediği düşünülen internet kullanıcı sayısı, kişi başı gayri safi milli hasıla, enflasyon, e-ticaret ile ilgili yasal mevzuatlar ve ekonomik kriz gibi faktörler bağımsız değişken olarak ele alınmış ve e-ticaret işlem hacmi üzerindeki etkileri çoklu regresyon modeli kurularak analiz edilmiştir. Yapılan analizler sonucunda bağımsız değişkenlerin bağımlı değişken olarak tanımlanan e-ticaret işlem hacmini beklentiler yönünde etkilediği görülmüştür.

Coşkun Tuncel ve Inandi (2011), çalışmasında öğrencilerin öğrenmeye yönelik olumlu veya olumsuz tutumlarını etkileyen faktörleri belirlemek ve elde edilen sonuçlara göre önlem almak için CHAID analizi kullanılmıştır.

Saimurugan, Sugumaran ve Sakthivel (2011), çalışmasında dönme mekanik sisteminin iyi ve hatalı koşulları altında titreşim sinyallerinden çıkarılan istatistiksel özellikleri kullanarak hataları sınıflandırmak için dört çekirdek işlevli destek vektör makinesinin (DVM) c-DVC ve nu-DVC modellerinin kullanımı sunulmuştur. Öne çıkan özellikleri seçmek için karar ağacı algoritması kullanılmıştır. Bu özellikler, DVM'nin c-DVC ve nu-DVC modelinin eğitimi ve test edilmesi için girdi olarak verilmiş ve hata sınıflandırma doğrulukları karşılaştırılmıştır.

Kağnicioğlu ve Moğol (2014), çalışmasında beş yıldızlı bir otelin verilerine göre CHAID algoritması ile oda temizleme işleminde en sık kullanılan malzemeler ve aralarındaki ilişkiler belirlenmeye çalışılmıştır.

Çelik, Atalay ve Bayer (2014), çalışmasında depremin önceden tahmini için veri madenciliği yöntemlerinden yapay sinir ağları ve destek vektör makinaları kullanılmıştır ve sınıflandırma yapılmıştır. Yapay Sinir Ağları ile %83 doğruluk oranı bulunurken destek vektör makinaları ile %91 doğruluk oranı tespit edilmiştir.

Köse, Güraksın ve Deperlioğlu (2015), çalışmasında diyabet tespiti için GOA tabanlı bir destek vektör makinası kullanılmıştır. Burada destek vektör makinasının eğitimi için RBF çekirdek fonksiyonunun parametresinin belirlenmesinde GOA'dan yararlanılmıştır. Çalışmada Pima Yerlileri diyabet veri seti kullanılmıştır. Sonuç olarak GOA-DVM yöntemi sınıflandırmada başarılı olmuştur.

Bauboniené ve Guleviçiüté (2015), çalışmasında e-ticaret kullanımını etkileyen faktörler analiz edilmiştir. Bunun için güvenlik, hızlı teslimat, karşılaştırılabilir fiyat, uygunluk, daha ucuz fiyatlar ve daha geniş seçenek yelpazesi gibi faktörlerin analizi yapılmıştır. Araştırmada özellikle e-ticaret kullanımında yaş, cinsiyet veya meslek gibi faktörlerin etkileri araştırılmıştır. Elde edilen sonuçlara göre

kolaylık, basitlik ve daha iyi fiyat faktörleri tüketicilerin e-ticaret kullanımını etkilediğini göstermiştir. Cinsiyete göre bakıldığında erkeklerin daha düşük fiyatlı ürünleri tercih ettiği sonucuna varılmıştır. 25-35 yaş grubundaki katılımcıların zaman darlığı ve geniş ürün yelpazesi gibi nedenlerle daha çok internette alışveriş yaptığı sonucuna varılmıştır.

Basarir-Ozel ve Mardikyan (2017), çalışmasında e-ticaret kullanıcılarının davranışlarını etkileyen faktörler araştırılmıştır. Araştırma modeli olarak Teknoloji Kabul Modelinin güvene dayalı bir uzantısı kullanılmıştır. Faktörlere ilişkin veriler 186 katılımcıdan toplanmış ve model geçerliliğini ve hipotezleri test etmek için yapısal eşitlik modellemesi yöntemi kullanılmıştır. Elde edilen sonuçlara göre kullanıcıların kullanışlılık ve güven algılarının Türkiye’de e-ticaretin benimsenmesinde önemli faktörler olduğunu ortaya koymaktadır. Araştırmaya göre Türk kullanıcıların ürün ve hizmetlerin güvenilirliğini önemsendiği sonucu ortaya çıkmıştır. Ayrıca zaman ve para tasarrufu açısından kullanışlı olması ve birçok ürün ve hizmetin kolay ulaşılabilir olması benimsemeyi olumlu yönde etkilediği görülmüştür.

Alkan ve Güney (2019), çalışmasında Türkiye’deki TRA1 bölgesinde 2014-2018 yılları arasında ait TÜİK Hanehalkı Bilişim Teknolojileri Kullanım Araştırması verilerinden yararlanılarak e-ticaret kullanımını etkileyen faktörler araştırılmıştır. Bunun için veriler üzerinde binary lojistik regresyon analizi yapılmıştır. Elde edilen sonuçlara göre anket yılı, yaş, iş durumu, aylık gelir ve hanehalkı büyüklüğü değişkenlerinin e-ticaret kullanımında etkili değişkenler olduğu tespit edilmiştir.

Yılmaz ve Bayram (2020), çalışmasında COVID-19 pandemi döneminde e-ticaret verileri incelenmiş olup en çok ihracatın hazır giyim kategorisinde yapıldığı tespit edilmiştir. Seyahat ve turizm kategorisi ise salgın döneminde e-ticaret hacmi en çok azalan sektör olmuştur.

Khraisat (2020), bu çalışmada C5 karar ağacı sınıflandırıcısı ve Tek Sınıf Destek Vektör Makinesi birleştirilerek bir Hibrit IDS (HIDS) önerilmiştir (OC-DVM). SIDS’in güçlü yönleri ve anormallik tabanlı bir saldırı tespit sistemi HIDS’de (AIDS) birleştirilmiştir. SIDS geliştirmek için C5.0 Karar Ağacı Sınıflandırıcısı, AIDS (DVM) geliştirmek için tek sınıflı Destek Vektör Makinesi kullanılmıştır. Yüksek algılama doğruluğu ve düşük yanlış alarm oranları ile bu çerçeve hem iyi bilinen izinsiz girişleri hem de sıfır gün tehditlerini tanımaya çalışmıştır. Ağ Güvenliği Laboratuvarı-Bilgi Keşfi (NSL-KDD) ve Avustralya Savunma Kuvvetleri Akademisi’ne (ADFA) ait veri tabanları üzerinde önerilen HIDS’i değerlendirmek için kullanılmıştır. Araştırmaya göre, HIDS tespit oranı ve düşük yanlış alarm oranları açısından SIDS ve AIDS’den daha iyi performans göstermiştir.

Özcan ve Turna (2021), çalışmasında bireylerin internet üzerinden yapmış olduğu alışverişlerde bireyleri etkileyen özelliklerin analizi yapılmıştır. Analizde cinsiyet değişkenine bağlı olarak C5.0 algoritması kullanılmış olup tüketici yorumları en önemli değişken olarak tespit edilmiştir. Ayrıca yaş değişkenine ve aylık gelir durumuna bağlı olarak C&R Tree Algoritması ücretsiz kargo değişkeninin en önemli değişken olduğu sonucuna ulaşılmıştır.

### 3. AMAÇ VE ÖNEM

E-ticaret üzerinden faaliyetlerini sürdüren işletmeler için kullanıcıların iyi analiz edilmesi önemlidir. Özellikle reklam vererek pazarlama stratejisi yürüten işletmelerin hedef kitlenin özelliklerini en iyi şekilde belirlemesi en az maliyetle en iyi reklam stratejisini oluşturmalarını sağlayacaktır (Bauboniené ve Guleviçiüté, 2015: 75). Literatür incelendiğinde e-ticaret ile veri madenciliği yöntemleri konusunda birçok çalışma bulunmaktadır. E-ticaret ile veri madenciliği yöntemlerinin birlikte kullanıldığı çalışmalar az olmakla birlikte e-ticaret kullanımını etkileyen faktörlerin analizi birkaç çalışmada veri madenciliği yöntemleri dışındaki yöntemlerle yapılmıştır. Bu çalışmalar incelendiğinde e-ticaret kullanımını etkileyen faktörler çıkarılmış fakat bu faktörlerin ayrıntısına inilmemiştir. Örneğin yaş faktörünün e-ticaret kullanımını etkilediği elde edilmiş fakat hangi yaş aralığının e-ticareti daha

fazla kullandığı gibi ayrıntılara yer verilmemiştir. Bu nedenle bu çalışmada karar ağaçları tercih edilmiş olup böylece ayırıcı kriterlerin ayrıntılı analizi yapılmıştır.

Sayılı ve Büyükköroğlu (2012) çalışmasında e-ticaret işlemlerine başta kişisel bilgilerin çalınması ihtimali olmak üzere güvenilirlik sorunu olduğu sonucuna varmıştır. Yapılan çalışmada katılımcıların %65,36’lık bölümünün e-ticaret işlemlerine güvenmediği görülmüştür. Güvenmeme nedenleri arasında kişisel verilerin çalınması ihtimali en önemli neden olmuştur. Literatürdeki birçok çalışmada müşterileri elde tutma ve sadık kılmanın en önemli anahtarının müşterinin güvenini kazanmak olduğu vurgulanmıştır (Beg, Rafiq ve Siala, 2018; Guo, Ling ve Liu, 2012; Aslam vd, 2020). Yapılan bu çalışmada da e-ticaret kullanmama nedenleri arasında güven unsurunun etkili olup olmadığı araştırılmıştır.

Bu çalışmanın amacı Türkiye’de fertlerin e-ticaret kullanım sıklıklarını etkileyen özelliklerin analiz edilmesidir. Böylece elde edilen sonuçlar e-ticareti kullanan işletmelerin hedef kitlelerini belirlemelerine yardımcı olabilecektir. Bunun için veri madenciliği yönetimlerinden karar ağaçları ve destek vektör makineleri algoritmalarından yararlanılmış ve elde edilen sonuçlar yorumlanmıştır. Çalışmanın ana konusu e-ticaret kullanımını etkileyen faktörlerin analizi olmakla birlikte ayrıca e-ticaret kullanım sıklığı tahmini için sınıflandırma modeli kurulmuştur. Çalışmadan elde edilen bulgular diğer çalışmaların bulguları ile karşılaştırılmıştır.

## **4. YÖNTEM**

### **4.1. Veri Seti Açıklaması**

Çalışmada kullanılan veri seti Türkiye İstatistik Kurumu (TÜİK) tarafından yapılan hanehalkı bilişim teknolojileri kullanımı 2021 yılına ait B grubu mikro verilerden oluşmaktadır. TÜİK tarafından hazırlanan hane halkı bilişim teknolojileri kullanımı araştırma anketine 30.531 kişi katılmıştır. Bu verilerden bazıları veri ön işleme safhasında gereksiz olduğundan dolayı silinmiş olup 24.329 veri üzerinden analizler yapılmıştır.

TÜİK’in yapmış olduğu Hanehalkı Bilişim Teknolojileri Kullanım Araştırması veri seti 153 sütun 30531 satırdan oluşmaktadır. Sütunlar soruları temsil ederken satırlar katılımcıları temsil etmektedir. Sorular hane bilgileri, fert bilgileri, cep telefonu kullanımı, internet kullanımı, E-devlet kullanımı, E-ticaret, E-Beceriler, Kişisel Verilerin Gizliliği ve Güvenliği olmak üzere 8 kategoriye ayrılmıştır. Çalışmada E-ticaret kullanımını etkileyen faktörlerin analizi yapıldığından E-devlet kategorisine ait sorular çalışma ile ilişkisiz görüldüğünden çıkarılmıştır.

Ele alınan kategorilerden hane bilgileri kategorisinde hanede bilgisayar, cep telefonu, akıllı saat, Tv gibi elektronik cihazlardan hangilerinin bulunduğu, hanede internet erişiminin olup olmadığı, hangi tür internet bağlantı türlerinin kullanıldığı, hanenin toplam aylık geliri ne olduğu gibi sorular yer almaktadır. Fert bilgileri kategorisinde kullanıcılara ferdin yaşının, cinsiyetinin, eğitim durumunun, okuryazarlık durumunun, çalışma durumunun ne olduğu gibi sosyodemografik sorular yer almaktadır.

Cep telefonu kullanımı kategorisinde son üç ay içinde cep telefonu kullanıp kullanmadığı sorusu yer almaktadır.

İnternet kullanımı kategorisinde interneti en son ne zaman kullandığı, ortalama hangi sıklıkla internet kullandığı, interneti hangi cihazlar ile kullandığı, özel amaçla hangi faaliyetler için internet (mobil uygulamalar dahil) kullandığı, internet üzerinden hangi eğitim faaliyetlerini gerçekleştirdiği gibi sorular yer almaktadır.

E-Ticaret kategorisinde internet üzerinden özel kullanım amacıyla en son ne zaman mal/ hizmet satın aldığı ya da sipariş verdiği, web sitesi veya mobil uygulama üzerinden hangi mallardan satın aldığı, internet üzerinden hangi tür satıcılardan mal satın aldığı, web sitesi veya mobil uygulamalar kullanarak özel şahıslardan herhangi bir mal satın alıp almadığı, web sitesi veya mobil uygulamalar üzerinden özel şahıstan hanehalkı hizmeti (Armut.com, bakıcıburada.com) alıp almadığı, özel kullanım amacıyla bir web sitesi veya uygulama üzerinden herhangi bir ulaşım hizmeti satın alıp almadığı özel kullanım amacıyla bir web sitesi veya uygulama üzerinden rezervasyon yaptırıp yaptırmadığı veya kiralama yöntemiyle konaklama hizmeti alıp almadığı, özel kullanım amacıyla internet üzerinden kaç defa mal veya hizmet satın aldığı, internet üzerinden yaptığı satın alımlar için (Hisse senedi ve finansal araçlar hariç) tahmini olarak ne kadar harcadığı, web sitesi veya mobil uygulama üzerinden yaptığı satın alma işlemlerinde hangi sorunlarla karşılaştığı, web sitesi veya mobil uygulama üzerinden özel amaçla hangi finansal işlemleri gerçekleştirdiği, web sitesi veya mobil uygulama üzerinden herhangi bir şey satın almama nedenlerin neler olduğu gibi sorular yer almaktadır.

e-Beceriler kategorisinde dosya ve klasörleri cihazlar arasında kopyalama yapıp yapmadığı cihazlara yazılım ve uygulamaları indirme ve yükleme işlemlerini yapıp yapmadığı, yazılım, uygulama veya cihazların ayarlarını değiştirip değiştirmediği kelime işlemci yazılımlarını kullanıp kullanmadığı, sunum oluşturup oluşturmadığı, MS Excel gibi tablolama yazılımları kullanıp kullanmadığı, fotoğraf video veya ses dosyalarını düzenleyip düzenlemediği bir programlama dilinde kod yazıp yazmadığı, internet haber sitelerinde veya sosyal medyada (Facebook, Instagram, YouTube, Twitter) doğru olmadığını veya şüpheli olduğunu düşündüğü bilgi veya içerik (videolar, resimler) görüp görmediği internet haber siteleri ya da sosyal medyada karşılaştığı bilgi veya içeriğin doğruluğunu kontrol edip etmediği internette bulunan bilgi veya içeriğin doğruluğunu hangi faaliyetler ile kontrol ettiği internette bulunan bilgi veya içeriğin doğruluğunu kontrol etmeme nedenlerinin neler olduğu gibi sorular yer almaktadır.

Kişisel Verilerin Gizliliği ve Güvenliği kategorisinde internet üzerinden kişisel verilere erişimi yönetmek amacıyla hangi yöntemleri uyguladığı, web sitelerindeki çerezlerin, internet üzerindeki davranış eğilimlerini izleyerek profil oluşturduğunu ve bu verileri reklam amacıyla kullanabildiklerini bilip bilmediği, internete bağlandığı cihazlarda çerezleri önlemek veya sınırlandırmak amacı ile internet tarayıcısındaki ayarları değiştirip değiştirmediği, internete bağlandığı cihazlarda etkinliklerin izlenmesini sınırlayan bir yazılım kullanıp kullanmadığı gibi sorular yer almaktadır.

Katılımcılara internet kullanımı kategorisinde yöneltilen “İnterneti en son ne zaman kullandınız?” sorusuna hiç kullanmadı veya bir yıldan çok cevabını veren katılımcılara ait 6202 kayıt silinmiştir. Çünkü e-Ticaret kategorisinde yer alan sorular son 12 ay içinde internet kullananlar için sorulmuştur. “Son üç ay içinde web sitesi veya mobil uygulama üzerinden aşağıdaki mallardan hangilerini satın aldınız?” sorusu son üç ay içerisinde internet üzerinden özel kullanım amacıyla mal/hizmet satın alan ya da sipariş veren kullanıcılara yönetildiğinden üç aydan fazla süredir e-Ticaret kullanmayanların bu sorudaki yanıtları “boş” yerine “hayır” yani son üç ay içinde web sitesi veya mobil uygulama üzerinden herhangi bir mal almadım olarak değiştirilmiştir. Bunlar haricinde verilere önışleme aşamasında başka işlemler uygulanmamıştır.

Çalışmada E-Ticaret kategorisinde internet üzerinden özel kullanım amacıyla en son ne zaman mal/ hizmet satın aldığı ya da sipariş verdiği sorusuna verilen cevabı etkileyen faktörler tahmin edilmeye çalışılmıştır. Kullanıcılar bu soruya “Son üç ay içinde”, “Üç ay ile bir yıl arasında”, “Bir yıldan çok” ve “Hiç kullanmadı” olarak cevap vermişlerdir. Analizde bu değişken SON\_KULLANIM\_ETICARET olarak adlandırılmıştır.

Analizlerin yapılması amacı ile kullanılan program IBM SPSS Modeler 18.0’dır. IBM SPSS Modeler, IBM tarafından oluşturulmuş bir veri madenciliği ve metin analitiği yazılım uygulamasıdır. Kullanıcıların programlama yapmadan istatistiksel ve veri madenciliği algoritmalarından yararlanmalarını sağlayan görsel bir arayüze sahiptir.

#### 4.2. Tasarlanan Model

Analizde tahmin edilecek değişken olarak “İnternet üzerinden özel kullanım amacıyla en son ne zaman mal/ hizmet satın aldınız ya da sipariş verdiniz?” sorusu seçilmiştir. SON\_KULLANIM\_ETICARET olarak adlandırılan bu değişkeni en iyi tahmin edecek algoritmaları bulmak için IBM SPSS Modeler’in otomatik sınıflandırıcı düğümü kullanılmıştır.

En iyi sınıflandırıcıları tespit eden düğüm en iyi sonuç veren algoritmaları Şekil 1’deki gibi CHAID, C&R Tree, QUEST ve DVM olarak bildirmiştir. Bu algoritmaları kullanarak Şekil 2’deki gibi SPSS Modeler’da sınıflandırıcı model tasarlanmıştır.

Veriden öğrenme işlemi gerçekleştirilirken ezberlemeyi önlemek için veriyi eğitim ve test seti olarak ikiye bölmek önemlidir. Model önce eğitim seti üzerinde eğitildi daha sonra modelin doğruluğunu ölçmek için test setindeki veriler kullanıldı. Ampirik çalışmalar, verilerin %20-30’unu test için ve kalan %70-80’ini eğitim için kullanırsak en iyi sonuçların elde edildiğini göstermektedir (Nguyen vd, 2021:2). Bu çalışmada, verilerin %70’ini eğitim için ve %30’unu test için ayırdık. Modelin başarısını ölçmek ve yöntemleri karşılaştırmak için Accuracy (Doğruluk), Recall(Duyarlılık), Precision (Kesinlik) ve F-Score (F-Measure) metriklerinden yararlanılmıştır. (1-4) formüllerinde kullanılan metriklerin hesaplama formülleri verilmiştir. Burada belirtilen TP (True Positive) e-ticaret kullananları doğru tahmin etme durumunu, TN (True Negative) e-ticaret kullanmayanları doğru tahmin etme durumunu, FP(False Positive) e-ticaret kullananları yanlış tahmin etme durumunu ve FN(False Negative) e-ticaret kullanmayanları yanlış tahmin etme durumunu göstermektedir.

$$\text{Accuracy} = \frac{(TP + TN)}{TP + FP + TN + FN} \quad (1)$$

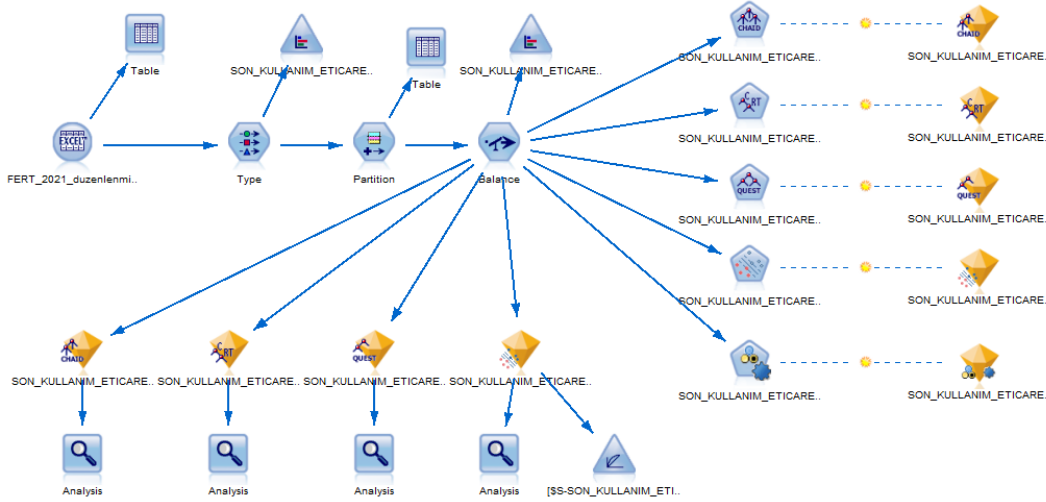
$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$F - \text{Measure} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Use?	Graph	Model	Build Time (mins)	Overall Accuracy (%)	No. Fields Used
<input checked="" type="checkbox"/>		CHAID 1	< 1	86.435	14
<input checked="" type="checkbox"/>		C&R Tree 1	< 1	86.345	20
<input checked="" type="checkbox"/>		Quest 1	< 1	81.047	25
<input checked="" type="checkbox"/>		SVM 1	< 1	79.563	110

Şekil 1. Auto Classifier İşlemcisi Sonuçları



Şekil 2. Kurulan Modelin Ekran Görüntüsü

### 4.3. Karar Ağacı Model Algoritmaları

Karar ağacı, sınıflandırma ve tahmin için güçlü teknikler sağlar. Karar ağacı modeli oluşturmak için çeşitli algoritmalar vardır (Elsalamony ve Elsayad, 2013: 392-393; Elder ve Miner, 2009: 278). Adından da anlaşılacağı gibi, bu model, tahmin doğruluğunu geliştirmek amacıyla bir ağaç yapısı oluşturmak için veri örneklerini yinelemeli olarak dallara ayırır. Her ağaç düğümü ya bir yaprak düğümdür ya da karar düğümüdür. Tüm karar düğümleri, veri özniteliklerinin bazı işlevlerinin değerlerini test eden bölmelere sahiptir. Karar düğümünün her bir dalı, testin farklı bir sonucuna karşılık gelir.

Bir karar ağacı oluşturmak için genel algoritma aşağıdaki gibidir:

- Tüm eğitim alt kümesi boş bir ağaçla başlayın.
- Geçerli düğümdeki tüm eğitim örnekleri aynı c sınıfı etiketine sahipse, düğüm c etiketli bir yaprak düğüm olur.
- Eğitim örneklerini farklı sınıflara ayırmada en önemli olan ayırma niteliklerini seçin. Bu öznitelik bir karar düğümü olur.



- Her farklı s değeri için bir dal oluşturulur ve örnekler buna göre bölünür.

#### 4.4. CHAID

CHAID verileri homojen gruplara bölmek için güçlü bir tekniktir. Tekniğin amacı grup içi varyasyon minimum gruplar arası varyasyon maksimum olacak şekilde her adımda en iyi homojen gruplamaya dayanır. Bunun için  $\chi^2$ (ki-kare) testi ve Bonferroni p değerlerinden yararlanır (Doğan ve Özdamar, 2003: 393). Açıklayıcı değişkenler birbirleri ile karşılaştırılıp, en küçük Bonferroni p değerine sahip olan açıklayıcı değişkenin kategorilerine göre, veriler alt gruplara ayrılır. Böylece bağımlı değişkeni açıklayabilecek homojen alt gruplar oluşturulur ve oluşturulan bu alt gruplardan yola çıkarak kurallar oluşturulabilir. Algoritma yalnızca nominal veya sıralı değişken özniteliklerini kabul eder (Kass G. V., 1980: 120). CHAID karar ağacı modelleme algoritması aşağıdaki gibidir (Biggs vd., 1991:51).

Bonferroni yaklaşımı, her bir grubun ortalama vektörleri ile genel ortalama vektörleri arasındaki farkın sıfır olup olmadığını kontrol eder. Eğer fark sıfırsa gruplar arasında anlamlı fark olmadığı sıfır değilse anlamlı fark olduğu ve grupların birbirinden farklı olduğu şeklinde yorum yapılır. CHAID algoritmasının adımları aşağıdaki gibidir:

Adım 1: Veri setindeki bağımsız değişkenlerin kategorileri ile çapraz tablo oluşturulur. Bunların içerisinde önemi en az olan yani p değeri en büyük olan kategori çiftleri bulunur.

Adım 2: p değeri en büyük olan kategori çifti için, bu p değeri önceden belirlenmiş alfa düzeyi ile kıyaslanır. Eğer p değeri alfadan büyük ise bu kategori çifti birleştirilir ve 1. adıma geri dönülür aksi halde 3. adıma geçilir

Adım 3: Bonferroni düzeltmesi kullanılarak bağımlı ve bağımsız değişken kategorileri için düzeltilmiş p değeri hesaplanır.

Adım 4: En küçük düzeltilmiş p değerine sahip değişkenler arasından en önemli bağımsız değişken seçilir.

Adım 5: Ağaç durdurma kuralları gerçekleşene kadar bu işlem devam eder.

##### 4.4.1. Sınıflandırma ve Regresyon (C&R) Ağacı

C&R Ağacı algoritması bağımlı değişkenin kategorik olduğu durumlarda sınıflandırma, sürekli olduğu durumlarda tahminleme modeli kurar. Bu yöntemde amaç modelin doğruluğunun en yüksek değere sahip olmasıdır. Bunun için modelin doğruluğu bağımlı değişkenin kategorik olduğu durumlarda doğru tahmin edilen verilerin oranı sürekli olması durumunda ise ortalama hata kareler değeri ile hesaplanır.

C&R ağaç modelleri, varyansı en aza indiren algoritmaları kullanarak bağımsız öznitelik bölme kriterlerine dayalı olarak giderek daha homojen alt kümeler bulmak için verileri özyinelemeli olarak bölümlere ayırır. Bağımlı veriler, üst düğümlerden türetilen bir dizi azalan sol ve sağ alt düğümlere bölünür (Breiman vd., 1998: 297).

##### 4.4.2. QUEST

Uygulamada kullanılan bir diğer veri madenciliği algoritması ikili bölünmüş bir karar ağacı algoritması olan QUEST’dir. Wei-Yin Loh ve Yu-Shan Shih tarafından 1997 yılında geliştirilmiştir (Loh ve Shih, 1997: 817-819; Ture vd., 2009). QUEST algoritması diğer karar ağacı algoritmalarından daha hızlı çalışırken bu durum tahmin performansını etkilemez (Loh ve Shih, 1997: 815-840).

QUEST algoritması ağacın dallanması sırasındaki önyargılı seçimin daha genel hale getirilmesi ve hesaplama maliyetinin düşürülmesi amacıyla geliştirilmiştir. QUEST algoritması, sürekli ve sıralı ve nominal değişkenler için kullanılabilir. Sürekli ve sıralı değişkenler için Varyans Analizi (ANOVA) testi yapılırken, kategorik değişkenler için Pearson  $\chi^2$  testinden yararlanır.

#### 4.4.3. Destek Vektör Makinesi (DVM)

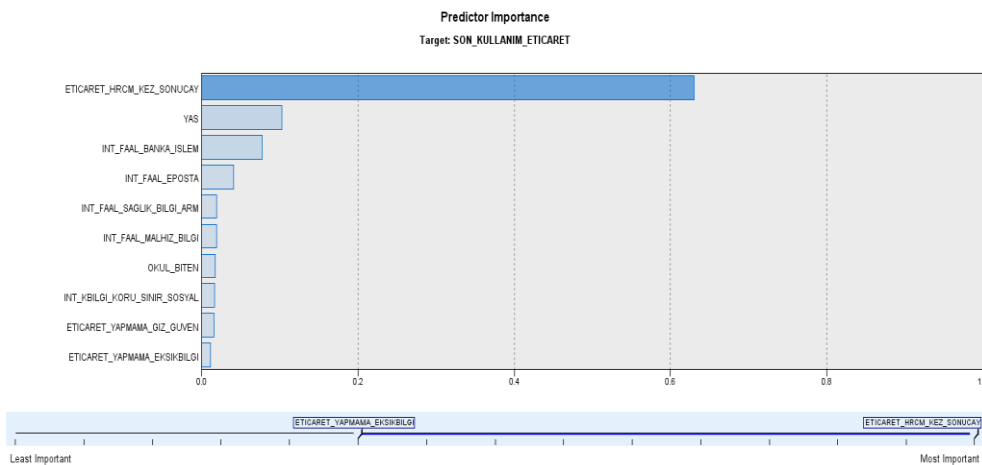
DVM temel olarak ikili sınıflandırmalarda kullanılan bir öğrenme tekniğidir. DVM'deki temel amaç, çok boyutlu bir veriyi bir hiper düzlem yardımıyla iki ayrı sınıfa ayırmaktır.

Gerçek hayat verilerinde, veriler genellikle doğrusal bir düzlemle ayrılmaz. Bu durumda veriler, çekirdek fonksiyonları kullanılarak verilerin doğrusal olarak ayrılabilmesi için yüksek boyutlu bir uzaya taşınır. Verileri yüksek boyutlu uzaya taşımak, genellikle hesaplama hatalarına ve fazla uyum sorunlarına neden olur. DVM yönteminde yüksek boyutlu uzayda elde edilen veriler doğrudan işlenmez, sadece o uzayda ortaya çıkan iç çarpım formüllerine ihtiyaç duyulur. Böylece hem hesaplama hem de çoklu uyum sorunları ortadan kalkar. Ayrıca, DVM'nin diğer öğrenme yöntemlerine göre avantajı, VC boyutunun açıkça hesaplanabilmesidir. VC boyutu, bir sistemin bilinmeyen veriler üzerinde başarı olasılığıdır. Örneğin yapay sinir ağları için böyle bir kriter yoktur.

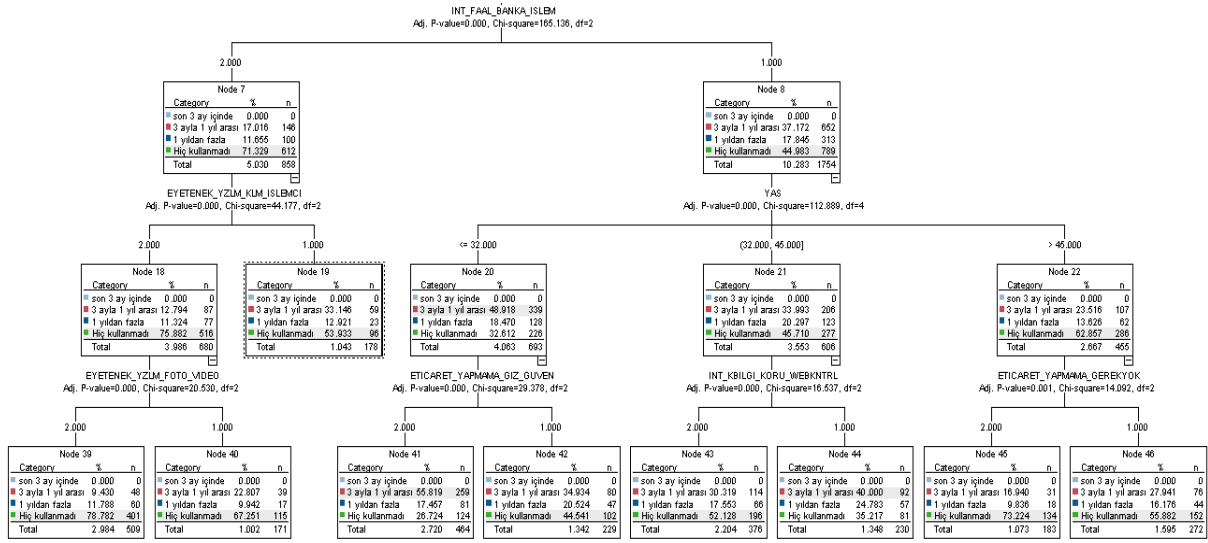
Genel olarak, DVM teoride sağlam temellere dayanan, sezgisel olarak güçlü ve pratikte başarılı olan bir öğrenme tekniğidir. DVM, regresyon çalışmaları için de kullanılmaktadır (Campbell ve Ying, 2011: 27). DVM'yi etkin bir şekilde kullanmak, DVM'nin nasıl çalıştığının iyi anlaşılmasını gerektirir. Bir DVM'yi eğitirken, uygulayıcı birçok konuda doğru kararlar vermelidir. Örneğin, ilk etapta verilerin nasıl işlenmesi gerektiği, hangi kernel fonksiyonunun kullanılması gerektiği ve DVM'de hangi parametrelerin kullanılması gerektiği sorularının cevabı, uygulayıcının kararına bağlıdır. Doğru yapılmayan seçimler algoritmanın performansının düşük olmasına neden olur.

## 5. BULGULAR

CHAID karar ağacı algoritmasında Şekil 3'teki değişkenler en iyi önem derecesine sahiptir. Şekil 4'te karar ağacının bir kısmı yer almaktadır. CHAID algoritmasının eğitim ve test verileri üzerinde sınıflandırma başarıları karşılaştırmalı olarak Tablo 1'de gösterilmiştir.



Şekil 3. CHAID Yöntemi Değişken Önem Dereceleri

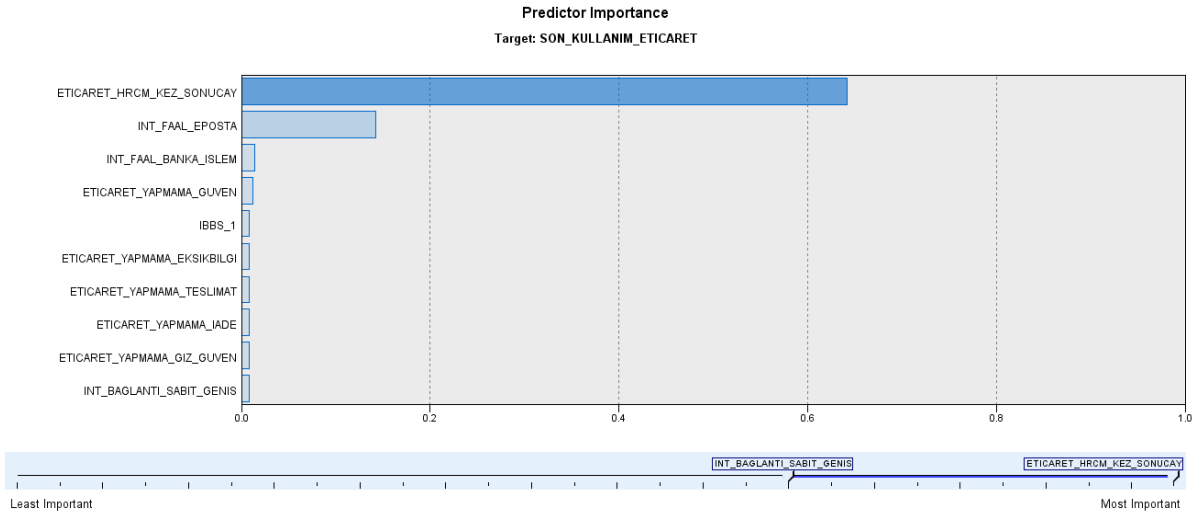


Şekil 4. CHAID Karar Ağacının Bir Kısmı

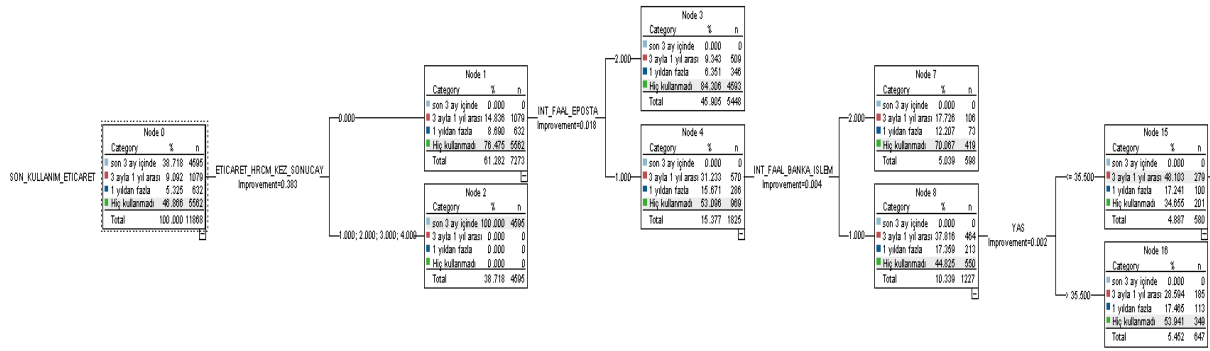
Tablo 1. CHAID Sınıflandırma Algoritmasının Eğitim ve Test Verileri Üzerindeki Başarılarının Karşılaştırması

	TP	TN	FP	FN	Accuracy	Recall	Precision	F-measure
Eğitim	7097	7811	205	1945	87,40%	78,49%	97,19%	0,8685
Test	3022	3289	80	879	86,81%	77,47%	97,42%	0,8631

C&R Tree karar ağacı algoritmasında Şekil 5’teki değişkenler en iyi önem derecesine sahiptir. Şekil 6’da karar ağacının tamamı yer almaktadır. C&R Tree algoritmasının eğitim ve test verileri üzerinde sınıflandırma başarıları karşılaştırmalı olarak Tablo 2’de gösterilmiştir.



Şekil 5. C&R Tree Yöntemi Değişken Önem Dereceleri

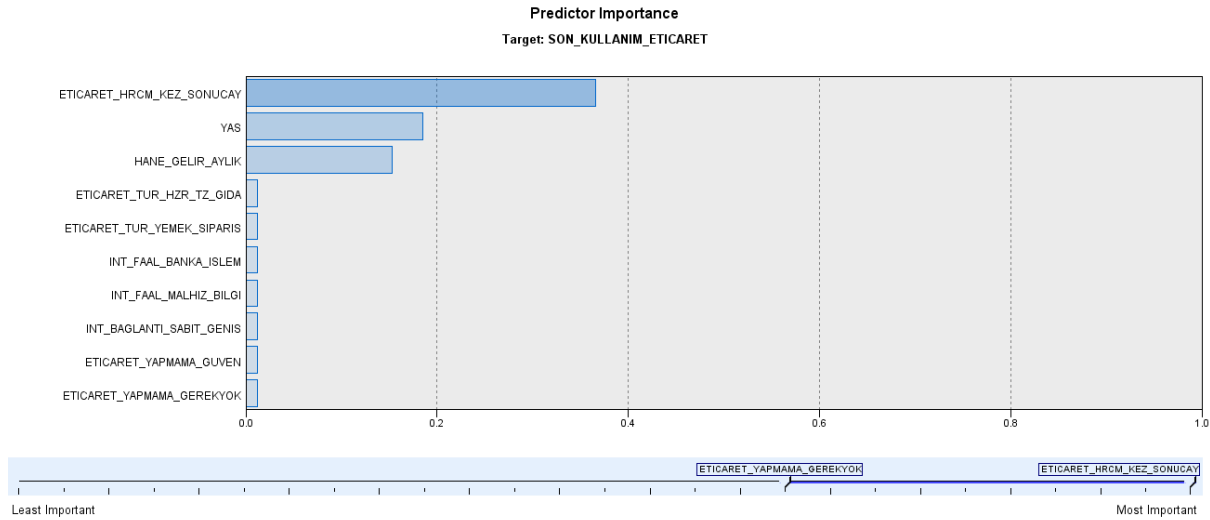


Şekil 6. C&R Tree Karar Ağacı Dallanması

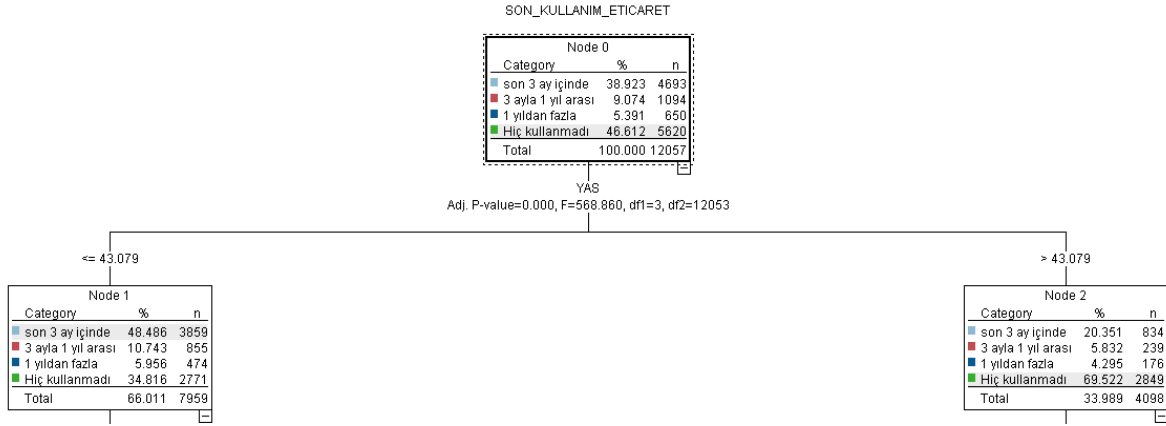
Tablo 2. C&R Tree Sınıflandırma Algoritmasının Eğitim ve Test Verileri Üzerindeki Başarılarının Karşılaştırması

	TP	TN	FP	FN	Accuracy	Recall	Precision	F-measure
Eğitim	6970	7881	135	2072	87,06%	77,08%	98,10%	0,8633
Test	2957	3312	57	944	86,23%	75,80%	98,11%	0,8552

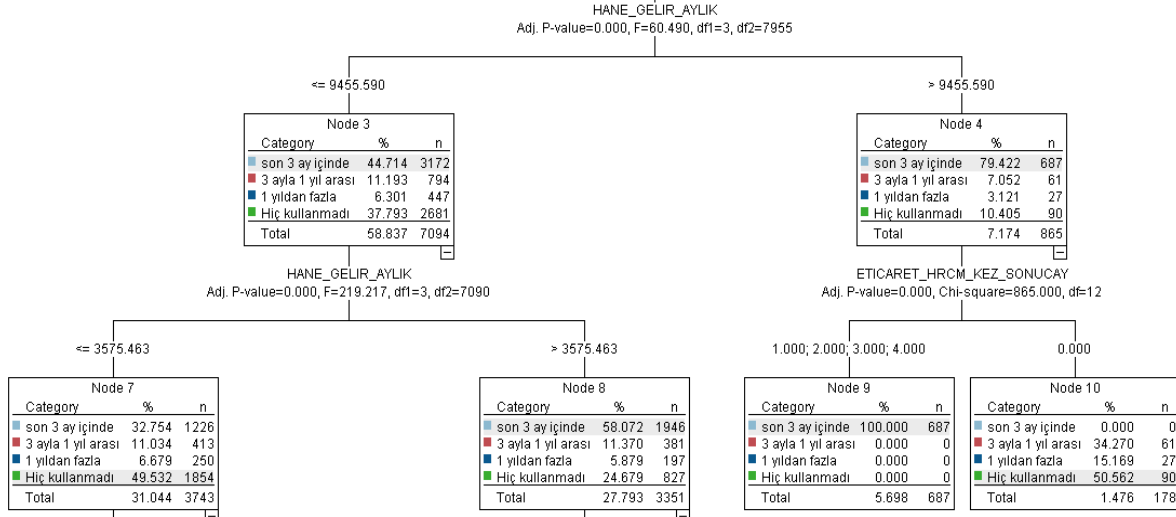
QUEST karar ağacı algoritmasında Şekil 7’deki değişkenler en iyi önem derecesine sahiptir. Şekil 8 ve Şekil 9’da karar ağacının bir kısmı yer almaktadır. QUEST algoritmasının eğitim ve test verileri üzerinde sınıflandırma başarıları karşılaştırmalı olarak Tablo 3’te gösterilmiştir.



Şekil 7. QUEST Yöntemi Değişken Önem Dereceleri



Şekil 8. QUEST Karar Ağacı Birinci Dallanması



Şekil 9. QUEST Karar Ağacı İkinci ve Üçüncü Dallanması

Tablo 3. QUEST Sınıflandırma Algoritmasının Eğitim ve Test Verileri Üzerindeki Başarılarının Karşılaştırması

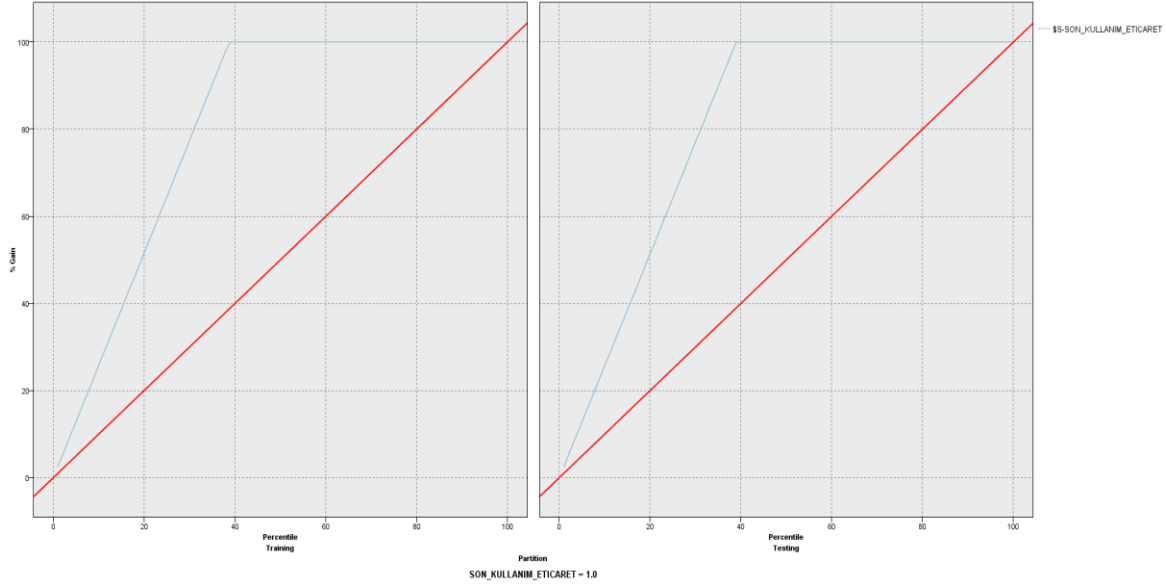
	TP	TN	FP	FN	Accuracy	Recall	Precision	F-measure
Eğitim	5675	8016	0	3367	80,26%	62,76%	100,00%	0,7712
Test	2407	3369	0	1494	79,45%	61,70%	100,00%	0,7632

Destek Vektör Makinası (DVM) algoritmasının eğitim ve test verileri üzerinde sınıflandırma başarıları karşılaştırmalı olarak Tablo 4’te gösterilmiştir. DVM algoritmasına ait ROC eğrisi Şekil 9’daki gibidir.

Tablo 4. DVM Sınıflandırma Algoritmasının Eğitim ve Test Verileri Üzerindeki Başarılarının Karşılaştırması

	TP	TN	FP	FN	Accuracy	Recall	Precision	F-measure
Eğitim	7711	5916	2100	1331	79,89%	85,28%	78,60%	0,8180
Test	3172	2546	823	729	78,65%	81,31%	79,40%	0,8034

DVM algoritmasına ait ROC eğrisi Şekil 10’daki gibidir.



Şekil 10. DVM Algoritmasına Ait ROC Eğrisi

## 6. TARTIŞMA VE SONUÇ

Bu çalışmada, Türkiye’de fertlerin e-ticaret kullanım sıklıklarının etkileyen özellikler veri madenciliği yönetimlerinden karar ağaçları ve destek vektör makineleri algoritmaları ile analiz edilmiş ve e-ticaret kullanım sıklıklarının tahmin edebilmek için bir sınıflandırma modeli tasarlanmıştır.

Elde edilen bulgulara göre kullanmış olduğumuz sınıflandırma algoritmalarının hepsi %70 üzerinde doğruluğa sahip olup veriler üzerinde başarılı bir sınıflandırma performansı elde edilmiştir. Bu algoritmalar içinde en iyi sınıflandırma başarısı gösteren algoritma CHAID olup test verileri üzerinde %86,81’lik doğrulukta sınıflandırma başarısı göstermiştir. Diğer algoritmaların test verileri üzerindeki başarıları sırasıyla C&R Tree 86,23, QUEST 79,45 ve DVM 78,65 olarak ölçülmüştür.

Çalışmanın bulgularına göre CHAID, C&R Tree algoritmalarında en önemli değişken son üç ay içinde özel kullanım amacıyla internet üzerinden mal veya hizmet satın alım sayısı olarak saptanmıştır. QUEST algoritması için en önemli değişken yaş olmuştur. CHAID için ikinci en önemli değişken son üç ay içinde özel amaçla (mobil uygulamalar dahil) e-posta gönderme/alma, QUEST algoritması için hanenin aylık toplam net geliri olurken, C&R Tree algoritmasında son üç ay içinde özel amaçla (mobil uygulamalar dahil) e-posta gönderme/alma değişkeni olmuştur. CHAID algoritması için üçüncü en önemli değişken internet bankacılığı (web sitesi veya mobil bankacılık uygulamaları) kullanma, QUEST algoritması için son üç ay içinde özel kullanım amacıyla internet üzerinden mal veya hizmet satın alım sayısı olurken, C&R Tree algoritmasında internet bankacılığı (web sitesi veya mobil bankacılık uygulamaları) kullanma değişkeni olmuştur.

Bu çalışmada 2021 yılına ait TÜİK Hanehalkı Bilişim Teknolojileri Kullanım Araştırması verilerinden yararlanılmış olup Alkan ve Güney (2019) çalışmasında aynı verinin 2014-2018 arasındaki verilerinin e-ticaret kullanımını etkileyen faktörlerini araştırmıştır. Kullanılan yöntemler farklı olmakla birlikte yaş ve aylık gelir her iki çalışmada da e-ticaret kullanımında etkili değişkenler olduğu tespit edilmiştir. Bu da çalışmaların birbirleri arasında tutarlı olduğunu göstermektedir.

Elde edilen sonuçlar yorumlanacak olursa son üç ay içinde özel kullanım amacıyla internet üzerinden mal veya hizmet satın alım sayısının en önemli değişken olması aşikar bir sonuçtur. Çünkü

son üç ay içerisinde mal veya hizmet alımı yapıldıysa bu kullanıcıların tamamı son üç ay içerisinde e-ticareti kullanmıştır yapılmadıysa bu kullanıcıların son üç ay içerisinde e-ticareti kullanmama ihtimali yüksektir. Bu da iyi bir ayırıcı nitelik olmasına sebep olmaktadır. QUEST için en önemli değişken olan yaş incelendiğinde yaşı 16-43 arasında olanların %48,48’i son üç ay içerisinde, %10,74’ü 3 ayla 1 yıl arasında, %5,95’i 1 yıldan fazla süre önce e-ticareti kullanmış olup %34,81’i ise e-ticareti hiç kullanmamıştır. Yaşı 43-74 arasında olanların %20,35’i son üç ay içerisinde, %5,83’ü 3 ayla 1 yıl arasında, %4,29’u 1 yıldan fazla süre önce e-ticareti kullanmış olup %69,52’si ise e-ticareti hiç kullanmamıştır. Bu durum 43 yaş altında olanların 43 yaş üstünde olanlara göre oransal olarak daha fazla e-ticaret kullandığını göstermektedir. QUEST için ikinci en önemli değişken olan hanenin aylık toplam net geliri incelendiğinde hanenin toplam geliri 9.455 TL altında olanların %44,71’i son üç ay içerisinde, %11,19’u 3 ayla 1 yıl arasında, %6,30’u 1 yıldan fazla süre önce e-ticareti kullanmış olup %37,79’u ise e-ticareti hiç kullanmamıştır. Toplam geliri 9.455 TL üstünde olanların %79,42’si son üç ay içerisinde, %7,05’i 3 ayla 1 yıl arasında, %3,12’si 1 yıldan fazla süre önce e-ticareti kullanmış olup %10,40’ı ise e-ticareti hiç kullanmamıştır. Bu durum hanenin aylık toplam net gelirinin 9.455 TL üstünde veya altında olması durumunu ayırıcı kriter yaparken geliri 9.455 TL üstünde olanların oransal olarak daha fazla e-ticareti kullandığını göstermektedir. QUEST dışındaki diğer algoritmalarda dallanma son üç ay içinde özel kullanım amacıyla internet üzerinden mal veya hizmet satın alım sayısının 0 olması durumundan devam ettiği için yorumlanabilir anlamlı sonuçlar çıkmamıştır. Bu algoritmalar sadece sınıflandırma başarılarıyla ön plana çıkmıştır.

Çalışmada literatürdeki diğer çalışmaların bulgularında bulunan güven sorununun e-ticaret kullanımını etkileyip etkilemediği konusuna bakıldığında CHAID algoritması ağacının 5. seviyesinde son 3 ay içinde ödeme güvenliği ve gizlilik endişesi sebebiyle e-ticaret kullanmayanların %44,5’inin e-ticareti hayatında hiç kullanmadığı görülmüştür. Elde edilen bu sonuç literatürdeki diğer çalışmalar ile paralellik göstermektedir.

Gelecekteki çalışmalarda e-ticaret kullanım sıklığını etkileyen özelliklerin araştırılmasında rastgele orman, XGboost, Yapay sinir ağları (YSA) ve farklı kümeleme teknikleri (iki aşamalı, Kohonen-SOM) gibi makine öğrenmesi algoritmaları kullanılabilir.

### **Etik Beyan**

Çalışma verileri TÜİK tarafından temin edildiğinden Etik Kurul onayını gerektiren bir çalışma değildir.

### **Katkı Oranı Beyanı**

Yazarlar makalenin her aşamasında katkı sağlamış olup, makalenin son halini onaylamışlardır.

### **Çatışma Beyanı**

Yapılan bu çalışma herhangi bir çıkar çatışmasına yol açmamıştır.

### **KAYNAKÇA**

- Adekitan, A. I. ve Salau, O. (2019). The Impact of Engineering Students’ Performance in the First Three Years on Their Graduation Result Using Educational Data Mining. *Heliyon*, 5(2), 1250.
- Akçi, Y. ve Annaç Göç, S. (2015). Tüketicilerin E-Ticaret Algılarının İncelenmesi (Gaziantep ve Adıyaman Örneği). *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 1(13).
- Alkan, Ö. ve Güney, E. (2019). TRA1 Bölgesinde E-Ticaret Kullanımını Etkileyen Faktörlerin Araştırılması. *Hoca Ahmet Yesevi 2. Uluslararası Bilimsel Araştırmalar Kongresi 6-8 Aralık 2019 ERZURUM*.

- Tanır, D. & Ramazanov, S. (2023). Veri Madenciliği Yöntemleri ile Türkiye’de Fertlerin E-Ticaret Kullanımını Etkileyen Faktörlerin Analizi. *KMÜ Sosyal ve Ekonomik Araştırmalar Dergisi*, 25(44), 46-65.
- Aslam, W., Hussain, A., Farhat, K. ve Arif, I. (2020). Underlying Factors Influencing Consumers’ Trust and Loyalty in E-Commerce. *Business Perspectives and Research*, 8(2), 186-204.
- Basarir-Ozel, B., ve Mardikyan, S. (2017). Factors Affecting E-commerce Adoption: A Case of Turkey. *The International Journal of Management Science and Information Technology (IJMSIT)*, 23, 1-11.
- Baubonienė, Ž. ve Gulevičiūtė, G. (2015). E-Commerce Factors Influencing Consumers ‘Online Shopping Decision. *Social Technologies*, 5(1), 62-73.
- Beg, S., Rafiq, M. ve Siala, H. (2018). The Effect of Customer Reciprocity on E-Loyalty through Relationship Quality: An International Perspective. *The Business & Management Review*, 9(3), 20–20.
- Biggs, D. Ville, D. B. ve Suen, E. (1991). A Method of Choosing Multi-Way Partitions for Classification and Decision Trees. *Journal of Applied Statistics*, 18(1), 49-62.
- Breiman, L. Friedman, J. H. Olshen, R. A. ve Stone, C. J. (1998) *Classification and Regression Trees*, Florida: CRC Press.
- Burges, J. C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2(2), 121-167.
- Campbell, C. ve Ying, Y. (2011). *Learning with Support Vector Machines*, Williston: Morgan Claypool Publishers.
- Cortes, C. ve Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20(2), 273-297.
- Çelik, E. Atalay, M. ve Bayer, H. (2014). Yapay Sinir Ağları ve Destek Vektör Makineleri ile Deprem Tahmininde Sismik Darbelerin Kullanılması. *2014 IEEE 22nd Signal Processing and Communications Applications Conference (SIU 2014)*.
- Doğan, N. ve Özdamar, K. (2003). CHAID Analizi ve Aile Planlaması ile İlgili Bir Uygulama. *Türkiye Klinikleri Tıp Bilimleri Dergisi*, 23(5), 392-397.
- Elder, R. J. ve Miner, G. (2009). *Handbook of Statistical Analysis and Data Mining Applications*. Burlington: Academic Press.
- Emel, G. G. ve Taşkın, Ç. (2005). Veri Madenciliğinde Karar Ağaçları ve Bir Satış Analizi Uygulaması. *Sosyal Bilimler Dergisi*, 6(2).
- Floares, A. A. (2012). Birlutiu. Decision Tree Models for Developing Molecular Classifiers for Cancer Diagnosis. *WCCI 2012 IEEE World Congress on Computational Intelligence June*, 10-15.
- Friedrichs, F. ve Igel, C. (2005). Trends in Neurocomputing. *The 12th European Symposium on Artificial Neural Networks*, (64), 107-117.
- Frohlich, H. ve Zell, A. (2005). Efficient Parameter Selection for Support Vector Machines. *IEEE International Joint Conference on Neural Networks*, (3), 1431-1436.
- Goel, R. (2007). E-commerce. New Delhi: New Age International.
- Guo, X., Ling, K. C. ve Liu, M. (2012). Evaluating Factors Influencing Consumer Satisfaction Towards Online Shopping in China. *Asian Social Science*, 8(13), 40
- Hany, A. ve Alaa, M. (2013). Bank Direct Marketing Based on Neural Network and C5.0 Models. *International Journal of Engineering and Advanced Technology (IJEAT)*, 2(6), 2249-8958.



- Tanır, D. & Ramazanov, S. (2023). Veri Madenciliği Yöntemleri ile Türkiye’de Fertlerin E-Ticaret Kullanımını Etkileyen Faktörlerin Analizi. *KMÜ Sosyal ve Ekonomik Araştırmalar Dergisi*, 25(44), 46-65.
- Haykin, S. Networks, N. (1999). *A Comprehensive Foundation*. New Jersey: Prentice-Hall.
- Kagnicioglu, C. H. ve Mogol, M. (2014). Implementation of CHAID Algorithm: A Hotel Case. *International Journal of Research in Business and Social Science*, 3(4), 42-51.
- Kass, G. V. (1980). An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics*, 29(2), 119-127.
- Kayri, M. ve Boysan, M. (2007). Araştırmalarda CHAID Analizinin Kullanımı ve Baş Etme Stratejileri ile İlgili Bir Uygulama. *Journal of Ankara University Faculty of Educational Sciences*, 40(2), 133-149.
- Khraisat, A. Gondal, I. Vamplew, P. Kamruzzaman, J. ve Alazab, A. (2020). Hybrid Intrusion Detection System Based on the Stacking Ensemble of C5 Decision Tree Classifier and One Class Support Vector Machine. *Electronics*, 9(173), 10.
- Köse, U. Güraksın, G. E. ve Deperlioğlu, Ö. (2015). Diabetes Determination via Vortex Optimization Algorithm Based Support Vector Machines. *In 2015 Medical Technologies National Conference (TIPTEKNO)*, 1-4.
- Lim, T. S. Loh, W. Y. ve Shih, Y. S. (2000). A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-Three Old and New Classification Algorithms. *Machine Learning Journal*, (40), 203-228.
- Loh, W. Y. ve Shih, Y. S. (1997). Split Selection Methods for Classification Trees. *Statistica sinica*, 7(4), 815-840.
- Muthusamy, S. M. Ramachandran, K. I. Sugumaran, V. ve Sakthivel, N. (2011). Multi Component Fault Diagnosis of Rotational Mechanical System Based on Decision Tree and Support Vector Machine. *Expert Syst*, 38(4), 3819-3826.
- Nguyen, Q. H., Ly, H. B., Ho, L. S., Al-Ansari, N., Le, H. V., Tran, V. Q., ... ve Pham, B. T. (2021). Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil. *Mathematical Problems in Engineering*, 2021.
- Otukei, J. R. ve Blaschke, T. (2010). Land Cover Change Assessment Using Decision Trees, Support Vector Machines and Maximum Likelihood Classification Algorithms. *International Journal of Applied Earth Observation and Geoinformation*, 12, 27-31.
- Pham, H. (2006). *Springer Handbook of Engineering Statistics*. London: Springer.
- Qabbaah, H. Sammour, G. Vanhoof, K. Voronin, A. Ziatdinov, Y. Aslanyan, L. ve Honchar, A. (2019). Decision Tree Analysis to Improve E-Mail Marketing Campaigns. *International Journal of Information Theories and Applications*, 26(4), 3-36.
- Saimurugan, M., Ramachandran, K. I., Sugumaran, V. ve Sakthivel, N. R. (2011). Multi Component Fault Diagnosis of Rotational Mechanical System Based on Decision Tree and Support Vector Machine. *Expert Systems with Applications*, 38(4), 3819-3826.
- Sayılı, M. ve Büyükköroğlu, A. (2012). E-Ticaret Yoluyla Gıda Maddeleri Satın Almaya Yönelik Tüketicilerin Tutumunu Etkileyen Faktörlerin Analizi. *Journal of Agricultural Sciences*, 18 (3), 246-255 . DOI: 10.1501/Tarimbil\_0000001212
- Silahtaroglu, G. (2013). *Veri Madenciliği Kavram ve Algoritmaları*. İstanbul: Papatya Yayınevi.

- Tanır, D. & Ramazanov, S. (2023). Veri Madenciliği Yöntemleri ile Türkiye’de Fertlerin E-Ticaret Kullanımını Etkileyen Faktörlerin Analizi. *KMÜ Sosyal ve Ekonomik Araştırmalar Dergisi*, 25(44), 46-65.
- Smola, A.J. (1996). *Regression Estimation with Support Vector Learning Machines*. (Master’s Thesis). Technische Universität München, Germany.
- Thanasankit, T. (Ed.). (2003). *E-commerce and Cultural Values*. Hershey: Idea Group Publishing.
- Tokatlı, M. F. ve Kurt, I. (2009). Using Kaplan–Meier Analysis Together with Decision Tree Methods (C&RT, CHAID, QUEST, C4.5 And ID3) in Determining Recurrence-Free Survival of Breast Cancer Patients. *Expert Systems with Applications*, 36(2), 2017-2026.
- Türen, U., Gökmen, Y. ve Tokmak, İ. (2011). Türkiye’de E-Ticaret İşlem Hacmini Etkileyen Faktörler Üzerine Bir Araştırma: Bir Model Önerisi. *Savunma Bilimleri Dergisi*, 10(1), 49-71.
- Türkiye Cumhuriyeti Ticaret Bakanlığı (2022), “2021 Yılı E-Ticaret Verileri Açıklandı”, [https://www.eticaret.gov.tr/haberler/10040/detay#:~:text=2021%20y%C4%B1%C4%B1nda%20%C3%BClkemizde%20e%2Dticaret.%17%2C7%20olarak%20ger%C3%A7ekle%C5%9Fti,\(08.04.2022\).](https://www.eticaret.gov.tr/haberler/10040/detay#:~:text=2021%20y%C4%B1%C4%B1nda%20%C3%BClkemizde%20e%2Dticaret.%17%2C7%20olarak%20ger%C3%A7ekle%C5%9Fti,(08.04.2022).)
- Vapnik, V. N. (1998). *Statistical Learning Theory*. New York: John Wiley & Sons.
- Yaman, B., Geçgil, G. ve Yavuz, G. (2018). Elektronik Ticaret ve Mobil Ticaret Üzerine Bir İnceleme: Meta-Analiz Çalışması. *R&S-Research Studies Anatolia Journal*, 1(2), 142-153.
- Yılmaz, Ö. ve Bayram, O. (2020). COVID-19 Pandemi Döneminde Türkiye’de E-Ticaret ve E-İhracat. *Kayseri Üniversitesi Sosyal Bilimler Dergisi*, 2(2), 37-54.

---

#### Genişletilmiş Öz

---

#### Veri Madenciliği Yöntemleri ile Türkiye’de Fertlerin E-Ticaret Kullanımını Etkileyen Faktörlerin Analizi

---

İçerisinde bulunduğumuz teknoloji çağındaki hızlı gelişimler toplumların teknoloji kullanımını yaygınlaştırmakla birlikte dünyadaki e-ticaret kullanımının da artmasında sebep olmuştur. Türkiye’de de e-ticaret kullanım hacmi yıldan yıla artış göstermeye devam etmektedir. Türkiye Cumhuriyeti Ticaret Bakanlığının açıkladığı verilere göre her yıl ortalama %35 büyüyen e-ticaret hacmi 2020 yılında %66 ve 2021 yılında %69 artış göstermiştir. E-ticaret hacminin genel ticarete oranı 2019 yılında yüzde 9,8 iken, 2020 yılında 15,7 ve 2021 yılında bu oran %17,7 ye ulaşmıştır. E-ticaret kullanımındaki bu artış işletmeler için e-ticareti daha da önemli kılmakta ve bu durum beraberinde e-ticaret sektörünü daha fazla rekabetçi hale getirmektedir. E-ticaret kullanımını etkileyen faktörlerin analizi işletmelerin e-ticaret alanında müşteri memnuniyeti sağlayabilmesi, pazarlama stratejileri oluşturabilmesi dolayısıyla da diğer işletmeler ile olan rekabet gücünü arttırabilmesi için önemlidir. Bu analiz için istatistiksel veya ekonometrik analiz yöntemleri kullanılabilir. Sosyal faktörler de e-ticaret kullanımını etkileyen önemli bir rol oynamaktadır. İnternet kullanımının yaygınlaşmasıyla birlikte tüketicilerin alışveriş alışkanlıkları değişmektedir. Online alışveriş, zaman tasarrufu sağlama, geniş ürün seçeneklerine erişim, kolay karşılaştırma yapma gibi avantajlar sunar. Ayrıca, sosyal medyanın etkisiyle tüketiciler arasında yapılan yorumlar ve değerlendirmeler, e-ticaret işletmelerinin itibarını ve güvenilirliğini belirlemede önemli bir faktör haline gelmiştir.

Veri madenciliği, verilerdeki önceden bilinmeyen örüntüleri tespit ederek veri kümelerindeki gizli bilgilerin ortaya çıkarılmasında ve böylece daha etkili karar vermede yardımcı olur. Veri madenciliği büyük veri kümelerinden yararlı bilgileri ortaya çıkarmak için kullanılır. Anlamsız devasa veri kümelerinde gizli örüntüleri bulmak için veri analizi, tahmin ve sınıflandırma için farklı yaklaşımlar kullanılmaktadır. Bu faydalı bilgiler, son teknoloji veri madenciliği algoritmaları tarafından çıkarılır. Bir veri madenciliği algoritması, faydalı bilgileri araştırmak ve veri kümelerindeki örüntüleri bularak sınıflandırma ve tahmin modelleri oluşturmak için izlenen bir dizi adımdır. Veri madenciliği yapmak için Karar Ağacı, K-En Yakın Komşular, Rastgele Orman, Destek Vektör Makinesi (DVM), Yapay Sinir Ağı ve Naive Bayes gibi birçok veri madenciliği algoritması kullanılmaktadır.

Bu makalede, e-ticaret kullanımını etkileyen faktörler incelenmektedir. Bu çalışmanın amacı Türkiye’de fertlerin e-ticaret kullanım sıklıklarını etkileyen özelliklerin analiz edilmesidir. Böylece elde edilen sonuçlar e-ticareti kullanan işletmelerin hedef kitlelerini belirlemelerine yardımcı olabilecektir. Bunun için veri madenciliği yönetimlerinden karar ağaçları ve destek vektör makineleri algoritmalarından yararlanılmış ve elde

---

edilen sonuçlar yorumlanmıştır. Çalışmanın ana konusu e-ticaret kullanımını etkileyen faktörlerin analizi olmakla birlikte ayrıca e-ticaret kullanım sıklığı tahmini için sınıflandırma modeli kurulmuştur.

Çalışmada kullanılan veri seti Türkiye İstatistik Kurumu (TÜİK) tarafından yapılan hanehalkı bilişim teknolojileri kullanımı 2021 yılına ait B grubu mikro verilerden oluşmaktadır. TÜİK tarafından hazırlanan hane halkı bilişim teknolojileri kullanımı araştırma anketine 30.531 kişi katılmıştır. Bu verilerden bazıları veri ön işleme safhasında gereksiz olduğundan dolayı silinmiş olup 24.329 veri üzerinden analizler yapılmıştır. Analizde tahmin edilecek değişken olarak “İnternet üzerinden özel kullanım amacıyla en son ne zaman mal/hizmet satın aldınız ya da sipariş verdiniz?” sorusu seçilmiştir. SON\_KULLANIM\_ETICARET olarak adlandırılan bu değişkeni en iyi tahmin edecek algoritmaları bulmak için IBM SPSS Modeler’ın otomatik sınıflandırıcı düğümü kullanılmıştır. En iyi sınıflandırıcıları tespit eden düğüm en iyi sonuç veren algoritmaları CHAID, C&R Tree, QUEST ve DVM olarak bildirmiştir.

Çalışmadan elde edilen bulgular diğer çalışmaların bulguları ile karşılaştırılmıştır. Elde edilen bulgulara göre kullanmış olduğumuz sınıflandırma algoritmalarının hepsi %70 üzerinde doğruluğa sahip olup veriler üzerinde başarılı bir sınıflandırma performansı elde edilmiştir. Bu algoritmalar içinde en iyi sınıflandırma başarısı gösteren algoritma CHAID olup test verileri üzerinde %86,81’lik doğrulukta sınıflandırma başarısı göstermiştir. Diğer algoritmaların test verileri üzerindeki başarıları sırasıyla C&R Tree 86,23, QUEST 79,45 ve DVM 78,65 olarak ölçülmüştür.

Çalışmada literatürdeki diğer çalışmaların bulgularında bulunan güven sorununun e-ticaret kullanımını etkileyip etkilemediği konusuna bakıldığında CHAID algoritması ağacının 5. seviyesinde son 3 ay içinde ödeme güvenliği ve gizlilik endişesi sebebiyle e-ticaret kullanmayanların %44,5’inin e-ticareti hayatında hiç kullanmadığı görülmüştür. Elde edilen bu sonuç literatürdeki diğer çalışmalar ile paralellik göstermektedir.

#### Extended Abstract

##### **Analysis of Factors Affecting Individuals' E-Commerce Use in Turkey Using Data Mining Methods**

The rapid advancements in technology in the current age have contributed to the widespread use of technology and the increasing prevalence of e-commerce worldwide. The volume of e-commerce usage in Turkey continues to increase year by year. According to the data announced by the Ministry of Commerce of the Republic of Turkey, the e-commerce volume, which has grown by an average of 35% every year, has increased by 66% in 2020 and 69% in 2021. While the ratio of e-commerce volume to general trade is 9.8 percent in 2019, it has reached 15.7 percent in 2020 and 17.7% in 2021. This increase in the use of e-commerce makes e-commerce more important for businesses and this makes the e-commerce sector more competitive. Analysis of the factors affecting the use of e-commerce is important for businesses to provide customer satisfaction in the field of e-commerce, to create marketing strategies, and thus to increase their competitiveness with other businesses. For this analysis, statistical or econometric analysis methods can be used, as well as data mining algorithms. Social factors also play an important role in influencing the use of e-commerce. With the widespread use of the Internet, consumers' shopping habits are changing. Online shopping offers advantages such as saving time, access to a wide range of products, and easy comparison. In addition, the comments and evaluations have been made among consumers under the influence of social media have become an important factor in determining the reputation and reliability of e-commerce businesses.

Data mining helps uncover hidden information in datasets by detecting previously unknown patterns in data, thus helping to make more effective decisions. Data mining is used to extract useful information from large data sets. Different approaches are used for data analysis, prediction and classification to find hidden patterns in meaningless huge datasets. This useful information is extracted by state-of-the-art data mining algorithms. A data mining algorithm is a series of steps followed to search for useful information and find patterns in datasets to create classification and prediction models. Many data mining algorithms such as Decision Tree, K-Nearest Neighbors, Random Forest, Support Vector Machine (DVM), Artificial Neural Network and Naive Bayes are used for data mining.

This article examines the factors affecting the use of e-commerce. The aim of this study is to analyze the features that affect the frequency of individuals' e-commerce use in Turkey. Thus, the results obtained will help businesses using e-commerce to determine their target audience. For this, decision trees and support vector machine algorithms from data mining managements have been used and the results have been interpreted. The main subject of the study is the analysis of the factors affecting the use of e-commerce, and also a classification model has been established for e-commerce usage frequency estimation.

The data set used in the study consists of B-group micro data from the Household Use of Information and Communication Technologies survey has been conducted by the Turkish Statistical Institute (TUIK) for the year 2021. A total of 30,531 individuals have participated in the survey. Some of the data has been deleted because of unnecessary during the data preprocessing stage, resulting in analyses being conducted on 24,329 records. The variable selected as the target variable for prediction in the analysis is "When is the last time you

purchase or order goods/services for personal use over the Internet?" The automatic classifier node of IBM SPSS Modeler has been used to find the algorithms that best predict this variable. The node, which identifies the best classifiers, has reported CHAID, C&R Tree, QUEST, and DVM as the algorithms that yield the best results.

The findings obtained from the study were compared with the findings of other studies. According to the results, all the classification algorithms used in the study have an accuracy of over 70% and have achieved successful classification performance on the data. Among these algorithms, the CHAID algorithm has shown the highest classification success, with an accuracy of 86.81% on the test data. The success rates of the other algorithms on the test data were measured as follows: C&R Tree 86.23%, QUEST 79.45%, and DVM 78.65%.

The study examined whether the trust issue, as found in the findings of other studies in the literature, affects e-commerce usage. When looking at the CHAID algorithm tree at the 5th level, it was observed that 44.5% of those who did not use e-commerce due to payment security and privacy concerns in the past three months have never used e-commerce in their lives. This result is consistent with other studies in the literature.

---