



SOSYAL VE BEŞERİ BİLİMLERE MULTİDİSİPLİNER BAKIŞ

Editörler

Prof. Dr. Ayhan Aytaç

Prof. Dr. Giray Saynur Derman

Prof. Dr. Mustafa Talas



SOSYAL VE BEŐERİ
BİLİMLERE
MULTİDİSİPLİNER BAKIŐ

EDİTÖRLER

PROF. DR. Ayhan AYTAÇ

PROF. DR. Giray Saynur DERMAN

PROF. DR. Mustafa TALAS

SOSYAL VE BEŞERİ BİLİMLERE MULTİDİSİPLİNER BAKIŞ

EDİTÖRLER

PROF. DR. Ayhan AYTAÇ

PROF. DR. Giray Saynur DERMAN

PROF. DR. Mustafa TALAS

Güven Plus Grup A.Ş. Yayınları: Aralık 41 / 2019

Yayıncı Sertifika No: 36934

E-ISBN: 978-605-7594-30-3

Güven Plus Grup A.Ş. Yayınları

Bu kitabının/derginin/kongrenin her türlü yayın hakkı GÜVEN PLUS GRUP DANIŞMANLIK A.Ş. YAYINLARI'na aittir. Yayınevinin yazılı izni olmadan, kitabın/derginin tümünün veya bir kısmının elektronik, mekanik ya da fotokopi yoluyla basımı, yayını, çoğaltımı ve dağıtımını yapılamaz. *Kitapta yer alan her bölüm/makale sorumluluğu, görseller, grafikler, direkt alıntılar ve etik/kurum iznine yönelik sorumluluk ilgili yazarlara aittir. Oluşabilecek Herhangi Hukuki bir olumsuzlukta Yayınevi başta olmak üzere kitabın hazırlanmasına destek sağlayan kurumlara, kitabın düzenlenmesi ve tasarımından sorumlular kurum(lar) ve kitap/dergi editörler/hakemler hiçbir konuda "maddi ve manevi" bir yükümlülük ve hukuki sorumluluğu kabul etmez, hukuki yükümlülük altına alınmaz. Her türlü hukuki yükümlülük ve sorumluluk "maddi ve manevi" yönden ilgili bölüm yazar(lar)ına aittir. Bu yöndeki haklarımızı maddi ve manevi yönden GÜVEN GRUP DANIŞMANLIK "YAYINCILIK" A.Ş. olarak saklı tutarız. Herhangi bir hukuki sorunda/durumda İSTANBUL mahkemeleri yetkilidir.* Güven Plus Grup Danışmanlık bünyesinde hazırlanan ve yayınlan bu eser ISO: 10002:2014-14001:2004-9001:2008-18001:2007 belgelerine sahiptir. Bu eser TPE "Türk Patent Enstitüsü" tarafından "Güven Plus Grup A.Ş.2016/73232" ve "2015/03940" nolu tescil numarası ile markalı bir eserdir. Bu bilimsel/akademik kitap/dergi ulusal ve uluslararası nitelikte olup, akademik teşvik kriterlerini karşılamaktadır. Çok bölümlü/yazarlı olan bu kitap/dergi E-ISBN'li olup Kültür Bakanlığı Millî Kütüphaneler tarafından ve 18 Farklı Dünya Ülkesiyle Anlaşmalı olan Millî Kütüphanenin E Erişim sistemi tarafından da taranmaktadır. Bu kitap/dergi maddi bir değer ile alınıp satılmaz. Kitap/dergi bölüm/makale yazarlarından, destekleyenlerden, kitap/dergiye emeği geçenlerden Güven Plus Grup A.Ş. Yayıncılık herhangi bir maddi bir gelir elde etmemiş ve talepte bulunmamıştır. Kitap/Dergide yer alan bölüm/makalelerden alıntı yapmak ve ilgili bölüm/makaleye atıf yapılmak koşulu ile kaynak gösterilmek üzere bilimsel ya da ilgili araştırmacılar tarafından kullanılabilir.

Metin ve Dil Editörü

Doç. Dr. Gülsemin HAZER

Dr. Öğr. Üyesi Gökşen ARAS

Doç. Dr. Ali Serdar YÜCEL

Kapak Tasarımı

Öğr. Gör. Ozan KARABAŞ

Sayfa Düzeni

Burhan MADEN

Baskı-Cilt

GÜVEN PLUS GRUP DANIŞMANLIK A.Ş. YAYINLARI®

Kayaşehir Mah. Başakşehir Emlak Konutları, Evliya Çelebi Cad. 1/A D Blok K4 D29 Başakşehir İstanbul Tel: +902128014061- 62 Fax:+902128014063 Mobile:+9053331447861

KONGRE/KİTAP/DERGİ İMTİYAZ SAHİBİ

GÜVEN PLUS GRUP DANIŞMANLIK A.Ş. YAYINLARI®

Kayaşehir Mah. Başakşehir Emlak Konutları, Evliya Çelebi Cad. 1/A D Blok K4 D29 Başakşehir İstanbul Tel: +902128014061-62-63 - +9053331447861 info@guvenplus.com.tr, www.guvenplus.com.tr

İNTERNET ARAMA VERİLERİ VE TOPLULUK REGRESYONUyla SATIŞI TAHMİNLEMESİ

İbrahim TOPAL

*Dr. Milli Savunma Üniversitesi, Deniz Astsubay Meslek Yüksekokulu,
Bilgisayar Teknolojileri Bölüm Başkanlığı, İstanbul / Türkiye*

Öz: İnternet, kullanımı küresel olarak hızla yayılan önemli bir bilgi kaynağı ve iletişim aracıdır. Gelişen teknoloji düşen maliyetler çok fazla insanı internete erişilebilir hale getirmiştir. Mobil teknolojiler tüketicilerin her an her yerden erişmesine ve uzun zamanlarını internette geçirmesine imkan tanımıştır. Tüketiciler özellikle ürün veya hizmet satın alma davranışında bulunurken bilgi alma, değerlendirme aşamalarında sıklıkla internete başvurumaktadırlar. Özellikle beğenmeli ürünlerde, fiyatının kolayda ürünlere göre nispeten fazla olması ve elde edilen faydanın uzun sürmesi gibi nedenlerle tüketicilerce bilgi arayışı ve karşılaştırma daha fazla olmaktadır. Beğenmeli ürünler kategorisinde yer alan otomobilin, teknik özelliklerinin de fazla olması nedeniyle tüketicilerce internet aramalarına konu olması muhtemeldir. Aynı zamanda otomotiv tüketiciler kadar ekonomiye ve diğer sektörler de etkisi olan önemli bir sektördür. Otomotiv sektörünün satışlarına yönelik yapılacak başarılı tahminlemenin birçok alana etkisi olabilecektir. Tahminlemenin internet arama verileri ile başarılı şekilde gerçekleşmesi kişisel bilgi olmaması, kolay erişilebilir, tutarlı olması ve tüm sektörü kapsayabilmesi yönlerinden avantaj sağlamaktadır. Yazında çeşitli sektörlerde ve ülkelerde internet arama verileri ile kullanılarak yapılmış çalışmalar olmakla birlikte olumlu/olumsuz sonuçlar alındığı görülmüştür. Bu çalışma, Türkiye’de Ocak 2010- Kasım 2019 tarih aralığını kapsamaktadır. Satışların %75’ini

oluşturan en fazla satışa sahip binek otomobil ve hafif ticari araç satışı yapan 10 markanın verileri dikkat alınmıştır. İnternet arama verileri Google Trends satış verileri Otomotiv Distribütörleri Derneği raporlarından alınmıştır. Analizlerde destek vektör regresyonu, karar ağaçları regresyonu, yapay sinir ağları ile birlikte çıkan sonuçlara dayalı olarak topluluk regresyonu kullanılmıştır. Topluluk regresyonu ile analiz sonuçlarında korelasyon %1-%23, ortalama mutlak yüzde hata değeri ise %2-%41 arasında geliştirilmiştir. Sonuç olarak çalışma, %90 korelasyon ile çok yüksek ilişkili ve %20,84 ortalama mutlak yüzde hata ile modellenmiştir. Çalışma ile internet arama verileri ile otomotiv sektörüne yönelik başarılı tahminleme yapılmasının mümkün olduğu görülmüştür. Çalışmanın yazına yaptığı katkının yanında satış ve pazarlama sorumlularınca kullanılabilir bir yöntem sunduğu değerlendirilmektedir.

GİRİŞ

İnternet, bilginin paylaşılması maksadıyla bilgi sistem cihazlarının birbiriyle bağlanması ve veri aktarımı yapmasına imkân tanıyan küresel bir yapıdır. İnsanlık tarihiyle karşılaştırıldığında yeni sayılabilecek bir teknoloji olan internet, diğer iletişim aygıtlarına benzer olarak zamanla kalitesi yükselmekte, fiyatlar düşmekte ve yaygınlaşmaktadır (Odlyzko, 2001). Günümüzde dünya nüfusunun yaklaşık %60'ını oluşturan 4,5 milyar kişi interneti kullanmakta her geçen gün yaklaşık 640.000 yeni kullanıcı internete ilk kez erişmektedir (Internetworldstats, 2019; Roser, Hannah ve Ortiz-Ospina, 2016). İnternetin yaygınlaşmasında, bilimsel, askeri maksatlarla yapıyı projeler alt yapıya geliştirerek, yazılım alanındaki gelişmeler ise kullanıcılar tarafından benimsenmesini sağlayarak olumlu katkı sağlamıştır.

İnternet altyapısının geliştiği ve kişisel bilgisayarların yayınlamaya başladığı dönemde World Wide Web (Dünya Çapında Ağ) ve HTML yazılım dili Tim Berners-Lee tarafından 1989 yılında icat edilmiştir. Web teknolojisi, internetin kullanıcılar için daha anlamlı ve anlaşılır olmasını sağlamıştır. Kullanıcılar, gazete okuma, araştırma yapma, alışveriş yapma gibi günlük faaliyetlerini yavaş yavaş internet üzerinden yapmaya başlamışlardır. İçerik sayısındaki artışla birlikte 1993 yılında ilk internet arama motorlarından olan Archie ortaya çıkmıştır (Wikipedia, 2019).

Bu dönemde yeterli teknik bilgiye sahip kişiler tarafından oluşturulan web siteleriyle kullanıcılara içerik sunulmuştur. Web 1.0 olarak bilinen bu dönemde kullanıcıların internetteki temel davranışları bilgi arama ve okuma olmuştur (Naik ve Shivalingaiah, 2008). Yazılım teknolojisindeki gelişme ve kullanıcıların web'i kullanım biçimlerindeki değişiklik Web 2.0 teknolojisinin çıkmasını sağlamıştır. Web 2.0 üzerinde yükselen sosyal ağlar, bloglar, forumlar, mikrobloglar, medya paylaşım siteleri, değerlendirme siteleri gibi birçok araç sosyal medyayı meydana getirmiştir. Öncesinde yalnızca okunabilir olan web, okunabilir ve yazılabilir web olarak değişmiştir (Getting, 2007). İçerik oluşturmak ve internet üzerinde herkesin görebileceği şekilde yayınlamak önemli oranda kolaylaşmıştır. Mobil teknolojiler ve mobil internetin de yaygınlaşması ile birlikte tüketiciler her an her yerden zahmetsizce internete ulaşabilmekte ve içerik üretebilmektedir. İnternetin ve yazılım teknolojisinin gelişerek yaygınlaşması tüketicilerin hayatlarında da değişime neden olmuştur.

Günümüz tüketicileri, yakın arkadaş ve çevresiyle yüz yüze ürün ve hizmetler hakkında fikir, görüş, tavsiyelerin karşılıklı olarak aktarıldığı ticari kaygı olmayan konuşmalar, diğer bir adıyla ağızdan ağıza iletişim, internete taşımaktadır. Tüketici satın alma sürecinin tamamında internet teknolojilerinden faydalanmaktadır. Örneğin tüketici satın alacağı telefonun teknik özelliklerine markanın resmi web sitesinden bakmakta, sonrasında kullanıcı görüşlerini alabilmek için forumlara girmekte, medya paylaşım veya blog sayfalarından uzman görüşlerini alabilmektedir. Tüketici değerlendirme ve karar sürecini tamandıktan sonra yine internet üzerinden sipariş verebilmekte ve kendi deneyimlerini çeşitli kanallar üzerinden paylaşabilmektedir. Tüketicilerin içeriği okumanın yanında içeriği üretebiliyor olması interneti dev bir bilgi yığını haline getirmektedir.

İnternet, birçok farklı konuyla ilgili ve sayısı her geçen gün hızla artan bilgiyi içerisinde barındırmaktadır. Sadece sosyal ağ sitesi Instagram'a günde 60 milyon fotoğraf yüklenmektedir (Boachie, 2018). Tüketicilerin içerik üretiminin anlık olarak arttığı ortamda aradıkları her bilginin site adreslerini bilmeleri mümkün değildir. Bunun yanında, her türlü ürün ve hizmet için aynı oranda arama yapılması da olası görülmemektedir.

Ürünler, kolayda, beğenmeli, özellikli ve aranmayan olarak sınıflandırılabilir. Sık satın alınan, çoğu yerde sıklıkla bulunabilen ve çok çaba harcanmayan ürünler kolayda sınıfındadır. Tüketici bu ürünler hakkında çok fazla karşılaştırma gereği duymaz. Kolayda ürünlere ampul, sakız, kibrit örnek gösterilebilir. Beğenmeli ürünlerde ise fiyat, kalite, renk, moda uygunluk karşılaştırma yapıldıktan sonra satın alınır. Tüketicinin ürün hakkında bilgisi azdır. Birim değerleri yüksektir ve kolayda ürünlere göre sık alınmamaktadır. Beğenmeli ürünlere beyaz eşya, otomobil örnek olabilir. Özelliği olan ürünlerde, belirli bir grup tüketicinin ürünü satın almak için özel gayret gösterdiği, yüksek fiyatlı ve her yerde satılmayan ürünlerdir. Pahalı kürk, lüks otomobiller özellikli ürünler grubundadır. Aranmayan ürünler ise tüketicilerin almak için özel çaba sarf etmediği örneğin mezar taşı gibi ürünlerdir (Zengin, 2014). Bu ürünler türleri arasında en fazla karşılaştırma ve bilgi arayışı beğenmeli ürünler arasında olmaktadır.

Beğenmeli ürünler grubunda olan otomobil, çok fazla teknik özelliği içerisinde barındıran kişilerin ve ailelerinin fiziksel ihtiyaçlarının yanında duygusal ihtiyaçlarını da tatmin etmesi beklenen maliyetli bir üründür. Otomobil konusunda, tüketiciler ihtiyacın ortaya çıkmasından sonraki araştırma, karşılaştırma süreçlerinde çok sayıda miktarda bilgiye gereksinim duymaktadır. Çözüm olarak ise internet ve internet arama motorları görülebilmektedir.

İnternette arama motorları, çoğu zaman tüketicilerin ürün veya hizmetler hakkında bilgi arayışında ilk nokta olmaktadır. Başlarda sadece FTP dosyalarının aratılabildiği Archie ile başlayan arama motorları arasına 1997 yılına kadar geçen sürede W3Catalog, Yahoo! Search, AltaVista, Go.com, Yandex gibi bazıları halen kullanılan birçok site katılmıştır. 1998 yılında ve sonrasında ise Google, Baidu, Live Search, Ask.com, Bing gibi siteler internette arama işlemini yapmaya başlamıştır (Wikipedia, 2019). Büyük bir bilgi yığınında tüketicilerin istediği sonuçlara ulaşabilmesi için kullanılan arama motorları, internet trafiğinin %93'üne ulaşılmasında aracı olmaktadır (Blueopps, 2018). Arama motorları, internet kullanımının önemli bir parçası haline gelmiştir.

Arama motorlarında arasında ise en bilinen Google'dır. Google, sade yapısı, sonuçların hızla ekrana getirmesi ve "... mı demek istediniz?"

İNTERNET ARAMA VERİLERİ VE TOPLULUK REGRESYONUyla SATIŞI TAHMİNLEMESİ şeklinde öneri sunmasıyla 2003 yılında pazarın %55'ine sahip olmuştur. 2005 yılında, otomatik tamamlamayla başlayan pazar payındaki yükselişini sesli arama gibi özelliklerle ekleyerek arttırmıştır (Tekula, 2014). Tüm arama motorlarının dâhil olduğu arama motoru pazarında Google %93'le en büyük paya sahiptir (Sullivan, 2016). Site de saniye de 70.000 yılda ise yaklaşık 2 trilyon arama yapılmaktadır (Prater, 2019). Google, bu aramalar hakkında net bilgi vermese de ölçeklendirilmiş bir değerler tablosu oluşturulmasına imkân tanımaktadır. Google Trends belirtilen kelimenin yer bilgisi ve zaman aralığı belirtilerek 0-100 arasında değerler alarak gösterimini yapmaktadır.

Google Trends verilerinin özellikle beğenmeli ürün grubunda olması nedeniyle hakkında fazla sayıda arama yapılması muhtemel otomobil sektöründe kullanılması mümkün görülmektedir. Otomobilin tüketiciler için önemini yanında küresel endüstri açısından önemi de fazladır. Çelik, demir, tekstil, kauçuk ve plastik gibi birçok sanayinin ürünleri otomobil üretiminde kullanılmaktadır. Dünya cam üretiminin dörtte biri, akaryakıt ve kauçuğun ise yarısı otomobil sektörüne gitmektedir (Saberı, 2018). Otomobil, hem tüketici hem de diğer sektörler için önemli bir üründür.

Otomobil sektörünün yaşadığı gelişmeler büyük bir ekonomiyi etkileyebilmektedir. Bu nedenle, otomobilin satış rakamlarının gerçeğe en yakın şekilde tespit edilmesi avantaj sağlayabilecektir. Talep tahmininde yöntemler temel olarak ikiye ayrılmaktadır. Kalitatif (nitel) ve kantitatif (nicel) yöntemlerdir. Kalitatif yöntemler daha çok istatistikî yöntemler kullanılmadan tecrübeye dayalı yapılmaktadır. Kantitatif yöntemlerde ise sayısal veriler kullanılarak istatistiksel yöntemlerle tahmin edilmesi söz konusudur (Yavuz, 2019). Yapay zeka ve derin öğrenme yöntemleri talep tahmininde kullanılabilir kantitatif yöntemlerdendir. Bu yöntemlerin, 1980'lerden itibaren işletme araştırmalarında kullanıldığı görülmektedir (Tkáč ve Verner, 2016; Wong, Bodnovich ve Selvi, 1996)

İşletmeler tahminlemelerinde kantitatif analizleri kullanabilmesi için büyük miktarda düzenli veriye ihtiyaç olmaktadır. Zamanla birlikte işletmelerin mevcut tüketicilerinin verilerini kişisel bilgi olduğu nedeniyle kullanılması sorun doğurabilmektedir (Park ve Huh, 2019). Tüketicilere yönelik ve kişisel bilgi sayılmayan, düzenli ve uzun zamanı

kapsayan bilgiye Google Trends Sitesi üzerinden kısa sürede ulaşılması mümkündür.

Çalışmada, internetin hayatına girmesiyle bilgi arama ve karşılaştırma için sıklıkla arama motorlarını kullanan tüketicinin kişisel olmayan verileri kullanılarak beğenmeli ürün grubunda olan otomobile yönelik sektör bazında yapay sinir ağları, destek vektör regresyonu ve karar ağaçları regresyonu algoritmaları ve bunların topluluk halinde kullanımıyla tahminleme yapılması amaçlanmaktadır.

YÖNTEM

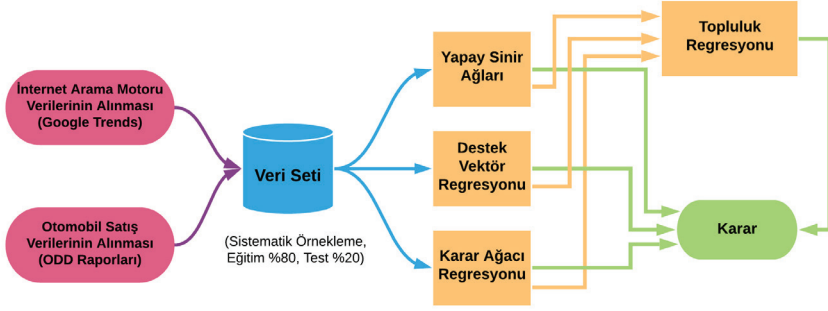
Araştırma alanı olarak %83'ün üzerinde internet erişimine sahip Türkiye seçilmiştir (Internetworldstats, 2019).

Zaman dilimi olarak Türkiye'de internet erişiminin %40 olduğu 2010 yılı (WorldBank, 2019) başlangıç, olası en yakın tarih olması nedeniyle Kasım 2019 ise bitiş tarihi olarak belirlenmiştir.

Binek otomobil ve hafif ticari araçlar, Türkiye'de benzer maksatlarla kullanılabilir olması ve Otomotiv Distribütörleri Derneği raporlarında da tek başlık altında yer alması nedeniyle otomotiv sektörünü temsil etmektedir.

Araştırma Modeli

Araştırmanın modeli, verilerin farklı kaynaklardan toplanması ve veri setinin oluşturulmasıyla başlamaktadır. Veri seti eğitim ve test olarak ayrıldıktan sonra, Destek Vektör Regresyonu, Karar Ağaçları Regresyonu ve Yapay Sinir Ağları uygulanmıştır. Çıkan sonuçlar topluluk regresyonu tekrar değerlendirilmiş ve karar aşamasına gelinmiştir. Çalışmanın modeli Şekil 1.'de görülmektedir.



Şekil 1. Araştırmanın Modeli

Araştırma Grubu

Çalışmada Türkiye’de binek otomobil ve hafif ticari araç satışı yapan 10 marka değerlendirmeye alınmıştır. Bu firmaların Kasım 2019 itibariyle satış rakamları toplam satışın %75’ini oluşturmaktadır. Bu markalar satış rakamlarına göre; (1) Renault, (2) Fiat, (3) Volkswagen, (4) Ford, (5) Toyota, (6) Hyundai, (7) Peugeot, (8) Dacia, (9) Nissan, (10) Mercedes-Benz olarak sıralanmaktadır.

Verilerin Toplanması

İnternet arama motoru verileri, Google Trends sitesi (<https://trends.google.com/trends/>) üzerinde marka adı, tarih ve yer bilgisi girilerek elde edilmiştir. Örneğin sitede ülke bölümünden “Türkiye” seçilmiş, tarih aralığı olarak 01 Ocak 2010 ile 30 Kasım 2019 seçilmiş ve “Renault” yazılarak arama yapılmıştır. İşlem 10 marka için tekrarlanmıştır. İşlemler sonunda veriler analizde kullanılabilmesi amacıyla düzenlenmiştir.

Otomobil satış verileri, Otomotiv Distribütörleri Derneği internet sayfasında Raporlar bölümünde Pazar- Otomobil ve Hafif Ticari Araç sayfasındaki bilgilerden derlenmiştir.

Otomobil satışlarının en az aylık bazda belirleniyor olması nedeniyle veriler buna uygun olarak hazırlanmıştır. Veri örneği Tablo 1.’de görülmektedir.

Tablo 1. Örnek Veriler

Ay	Re-nault	FIAT	Volk-swagen	Ford	Toy-ota	Hyun-dai	Peu-geot	Da-cia	Nis-san	Merce-des-Benz	Satış Adedi
1	54	42	35	54	36	61	55	32	41	50	20.095
2	53	45	35	55	40	56	53	33	44	25	31.172
3	55	49	37	57	37	71	59	80	51	49	51.769
4	53	43	35	55	38	63	60	84	50	34	54.946
...
8	68	64	85	87	59	87	78	67	76	52	26.246
9	81	75	89	86	72	98	77	73	72	53	41.992
10	76	75	95	88	71	100	86	70	76	53	49.075
11	77	73	85	93	64	96	100	73	85	64	58.176

Verilerin Analizi

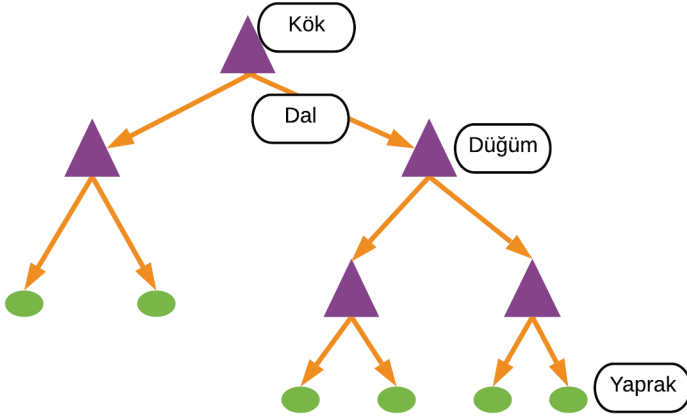
Kullanılan yöntemler, Karar Ağaçları Regresyonu, Destek Vektör Regresyonu ve Yapay Sinir Ağlarıdır. Bu çalışmada kullanılan yöntemler alt başlıklarda açıklanmaktadır. Analizler yapılırken veriler sistematik örnekleme metoduyla %80 eğitim ve %20 test olarak ayrılmıştır. Böylece 96 aylık veri eğitim 23 aylık veri ile test için kullanılmıştır.

Karar Ağaçları Regresyonu

Karar Ağaçları (KA), problemin veri setlerine bağlı olarak bir ağaç yapısına benzer şekilde sınıflandırma ve regresyon modeli oluşturulmasını sağlamaktadır. KA, karmaşık karar verme süreçlerini daha basit kararlar bütününe dönüştürmektedir. Bu nedenle medikal tanılardan, radar sistemlerine kadar birçok alanda yaygın olarak kullanılmaktadır (Safavian ve Landgrebe, 1991).

KA'da, ortaya koydukları bilginin gösterimi kolay olmakla birlikte açıklamasını da ortaya koymaktadır (Emel ve Taşkın, 2005). KA ile sınıflandırma veya regresyon yapılabilir. Öznitelikler kategorik verilerden oluşuyorsa sınıflandırma ağacı sürekli verilerden oluşuyorsa regresyon ağacı olarak tanımlanmaktadır (Gokceoglu, Nefeslioglu, Sezer, Bozkir ve Duman, 2010).

KA regresyonu düğüm, dal ve yapraklardan oluşur. Öznitelikler düğümler tarafından gösterilmektedir. Genel olarak görüntüsü ters çevrilmiş ağaç şeklindedir. En üst kısımda kök bulunurken en altta yapraklar yer almaktadır. Kök, düğüm ve yapraklar arasındaki bağlantıyı sağlayan kısımlara ise dal adı verilmektedir. Kökten başlayarak yapraklara doğru verilere ilişkin sorular sorularak ağacının büyümesi devam etmektedir (Kavzoğlu, Şahin ve Çölkesen, 2012). KA yapısı Şekil 2.'de görülmektedir.



Şekil 2. Karar Ağacı

KA'da ID3, C4.5 ve CART (Classification and Regression Tree [Sınıflandırma ve Regresyon Ağacı]) gibi algoritmalar kullanılmaktadır. CART algoritması, regresyon ağacı oluştururken iyi performans sergilemektedir. Regresyon probleminin çözümünde karesel ortalama hatayı en aza indirecek bölünmeleri hesaplayarak ağacın büyümesini ve dalanmasını sağlamaktadır. Yapraklardaki tahminler düğüm için ağırlıklı ortalamalara bağlıdır. CART algoritmasında Twoing kuralı uygulanır ve tüm öznitelikler değerlendirilerek yapılan eşleşmelerden sonra düğümler iki dala ayrılır (Rokach ve Maimon, 2015).

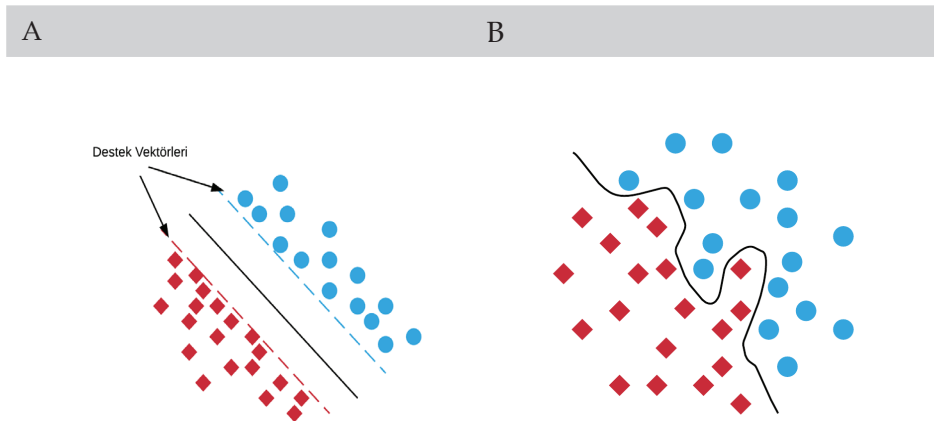
Çalışmada tahmin edilmeye çalışılan satış miktarları sürekli değişkendir. Satış tahminine yönelik KA regresyon hesaplamasında CART algoritması kullanılmıştır.

Destek Vektör Regresyonu

Destek vektör makineleri en iyi öğrenme algoritmaları arasında gösterilmektedir. Destek vektör ağı hatalardan ayırlamayan eğitim verisinin olduğu durumlar göz alınarak geliştirilmiştir. Giriş vektörlerin yüksek boyutlu özellik uzayında dağınık olduğu durumda doğrusal bir karar yüzeyinin oluşturulmasını sağlamaktadır. Öğrenen makinede bu karar yüzeyinin kendine has özellikleri ile yüksek genelleme mümkün olmaktadır (Cortes ve Vapnik, 1995).

Destek vektör makineleri algoritmaları ayırlamaz görünen hataların en az hata ile ayrılmasını sağlamaktadır. Şekil 3.'te doğrusal ve doğrusal olmayan çizgilerle ayrımı yapılacak veriler görülmektedir. Doğrusal olmayan çizgi ile ayrımın yapılabilmesi doğrusala göre daha zordur ve kullanılan hattın kıvrılması gereklidir. Destek vektör makineleri, nesnelere arasında ayırım yapmak için oluşturulan çizgilerin farklılaştırılmasına imkân tanıyan hiper düzlem sınıflandırıcılardandır (Mandhala, Sujatha ve Devi, 2014). Destek vektör makineleri sınıflandırmanın yanında regresyon maksadıyla da kullanılmaktadır (Shevade, Keerthi, Bhattacharyya ve Murthy, 2000).

Destek vektör regresyonunun en sık kullanılan çekirdekler Radial Basis Function ve Gaussian'dır. Bu çalışmada Gaussian çekirdeği daha iyi verdiği performansa sahip olduğu için tercih edilmiştir.



Şekil 3. Doğrusal (A) ve Doğrusal Olmayan (B) Verilerde Ayırım

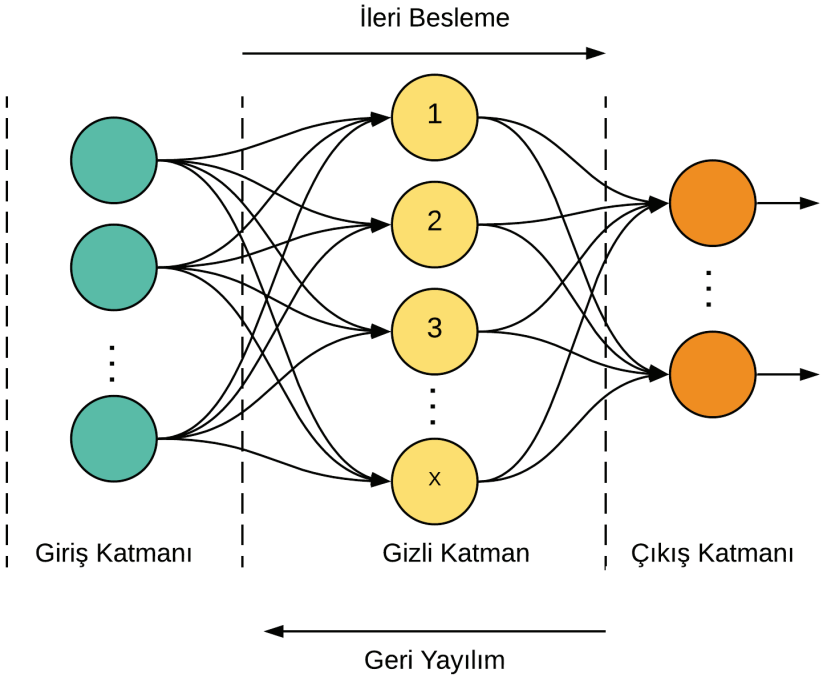
Yapay Sinir Ağları

Beyindeki sinir ağlarına benzer olarak geliştirilen yapay sinir ağları, nöronların ağırlıklı bağlantılar ile bağlı, belleği olan elemanların oluşturduğu paralel ve dağıtılmış bilgi işleme yapılarıdır. Bu yapı nöron, bağlantı ve öğrenme algoritmasından oluşmaktadır (Çuhadar ve Kayacan, 2005). Çok sayıda nörondan oluşan ağlarda nöronlar giriş sinyali alır ve çıkış sinyali üretir. Ağ, ağırlıklandırılmış sinapsis dizisinden meydana gelmektedir (Zhang, 2018: 1). Yapay sinir ağlarının ileri, geri, kademeli olmak üzere üç çeşidi vardır. Kademeli, önceki katmandan en az bir geri beslendiği; geri, en az bir hücrenin bir sonraki katman tarafından beslendiği; ileri beslemeli ağ yalnızca önceki katmanlardan beslendiği ağdır (Cerit ve diğerleri, 2017).

İleri beslemeli ağda giriş, gizli ve çıkış katmanları bulunmaktadır. Her katmanda çeşitli sayıda nöron olmakla birlikte aynı katmandaki nöronlar arasında ilişki bulunmamaktadır. Giriş katmanı, verilerin sisteme dahil olduğu katmandır. Belirlenen öznitelik sayısı kadar nöron giriş katmanında bulunur. Giriş katmanındaki veriler gizli katmana iletilir. Gizli katmanda ileri seviye istatistiksel yöntemlerle çıkarımlarda bulunulur ve oluşturulan sinyal çıkış katmanına iletilir. Çıkış katmanındaki nöronlar tarafından oluşturulan sinyal tüm ağın cevabını niteliğindedir (Zhang, 2018: 3). İleri beslemeli yapay sinir ağı Şekil 4.'te görülmektedir. Yapılan araştırmanın özelliklerine göre giriş ve çıkış katmanındaki nöron sayıları değişebilmektedir. Gizli katmandaki nöron sayısı deneme yoluyla belirlenebilmektedir.

Geri yayılım yöntemleri olarak Levenberg–Marquardt, BFGS Quasi-Newton, Conjugate Scaled Conjugate Gradient, Gradient with Powell/Beale Restarts, Conjugate Gradient with Powell/Beale Restarts, Polak-Ribière Conjugate Gradient, Resilient Backpropagation, One Step Secant, Variable Learning Rate Gradient Descent, Fletcher-Powell Conjugate Gradient, Gradient Descent with Momentum, Gradient Descent, Bayesian Regularization, kullanılmaktadır.

Bu çalışma da daha iyi performansa sahip olması nedeniyle Bayesian Regülasyon yöntemi tercih edilmiştir. Yapay sinir ağının yapısı Tablo 2.'de verilmiştir.



Şekil 4. İleri Beslemeli Yapay Sinir Ağı

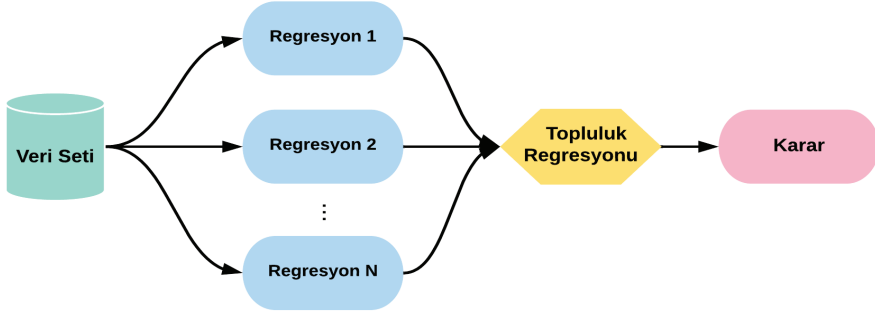
Tablo 2. Yapay Sinir Ağı Yapısı

Katman	Öznitelik	Açıklama	Nöron
Giriş	Ay Bilgisi	Yılın kaçınıcı ayı olduğu	I1
	Renault	“Renault” internet arama motoru verisi	I2
	FIAT	“FIAT” internet arama motoru verisi	I3
	Volkswagen	“Volkswagen” internet arama motoru verisi	I4
	Ford	“Ford” internet arama motoru verisi	I5
	Toyota	“Toyota” internet arama motoru verisi	I6
	Hyundai	“Hyundai” internet arama motoru verisi	I7
	Peugeot	“Peugeot” internet arama motoru verisi	I8
	Dacia	“Dacia” internet arama motoru verisi	I9
	Nissan	“Nissan” internet arama motoru verisi	I10
	Mercedes-Benz	“Mercedes-Benz” internet arama motoru verisi	I11
Çıkış	Satış Bilgisi	Otomotiv Distribütörleri sitesinden alınan aylık satış verisi	O1

Topluluk Regresyonu

Topluluk regresyonunda, birkaç yöntemle yapılan tahminlemeler bir arada kullanılarak daha başarılı sonuç elde edilmek için yapılmaktadır. Topluluk regresyonun, ortalama, ağırlıklı ortalama, durumsal ortalama, bagging, boosting ve stacking gibi yöntemleri bulunmaktadır (Emra, 2017).

Bu çalışma sıklıkla kullanımı olan ortalama topluluk yöntemi tercih edilmiştir. Bu yöntemde N sayıda regresyon ile sistem kurulur ve bunların tahminlemeleri alınır. Çıkan sonuçların ortalaması alınarak performans değerleri yeniden hesaplanır. Performans değerlerine göre karar verilir. Topluluk regresyonu Şekil 5.’te görülmektedir.



Şekil 5. Topluluk Regresyonu

Karar

Araştırmanın son kısmında analizler neticesinde elde edilen sonuçların performans değerlendirmesi yapılmıştır. Performans değerlendirmesi için Denklem 1’de görülen Pearson korelasyon (R), Denklem 2’de görülen Ortalama Hataların Karesi (MSE), Denklem 3’te görülen Ortalama Mutlak Yüzde Hata (MAPE), Denklem 4’te görülen Ortalama Mutlak Sapma (MAD) ve Denklem 5’te görülen Kare Ortalamalarının Karekökü (RMSE), Denklem 6’da görülen Ortalama Yüzde Hata Kareleri Toplamı Kökü (RMPSE) değerlerine bakılmıştır.

Formüllerde “G” ile gerçek değer, “T” ile tahmin edilen değer, “n” ile veri setinin büyüklüğü, “ μ_G ” ile gerçek değerlerin ortalaması, “ σ_G ” gerçek değerlerin standart dağılımı, “ μ_T ” ile tahmin edilen değerlerin ortalaması, “ σ_T ” tahmin edilen değerlerin standart dağılımı gösterilmektedir.

$$\rho(G, T) = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{G_i - \mu_G}{\sigma_G} \right) \left(\frac{T_i - \mu_T}{\sigma_T} \right) \quad (1)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (G_i - T_i)^2 \quad (2)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|G_i - T_i|}{G_i} \times 100 \quad (3)$$

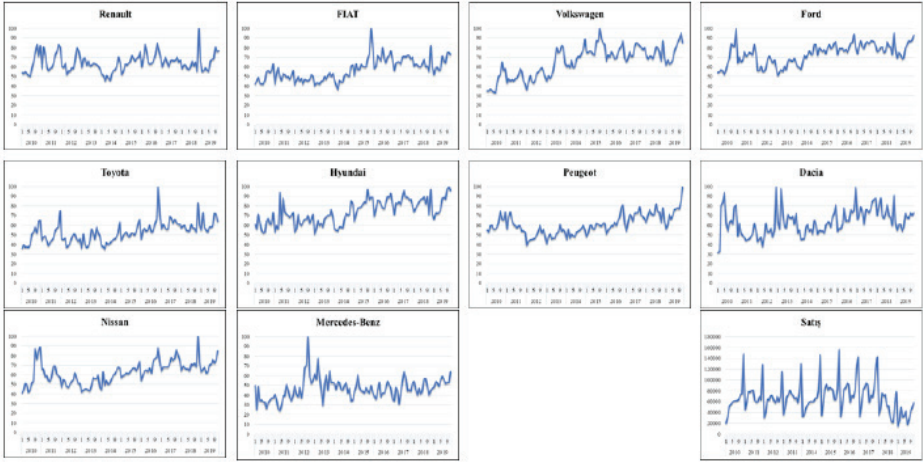
$$MAD = \frac{1}{n} \sum_{i=1}^n |G_i - T_i| \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (G_i - T_i)^2} \quad (5)$$

$$RMSPE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{|G_i - T_i|}{G_i} \times 100 \right)^2} \quad (6)$$

BULGULAR

Türkiye’de en fazla binek otomobil ve hafif araç satışına sahip 10 markanın internet arama motoru ve toplam binek otomobil ve hafif araç satış verileri Şekil 6.’da görülmektedir.



Yapılan analizlerde gerçek veri ile tahmin edilen değer arasındaki ilişkiye bakılmıştır. Korelasyon değerleri destek vektör regresyonunda, eğitim 0,85 test 0,73; karar ağacı regresyonunda eğitim 0,92 test 0,89; yapay sinir ağlarında eğitimin 0,90 test 0,88 olarak belirlenmiştir. Her üç analiz yönteminde de test değerleri için %70'in üzerinde korelasyon olduğu gözlenmiştir.

Yapay sinir ağlarında en yüksek performansa sahip Bayesian Re-gülasyon geri yayılım yöntemiyle elde edilen sonuçlar görülmektedir. Ağda gizli katmanın 31 nörondan oluşması en iyi sonucu vermiştir. Destek vektör regresyonu, karar ağacı regresyonu ve yapay sinir ağlarına ait performans değerleri Tablo 3.'tedir.

Tablo 3. Destek Vektör Regresyonu, Karar Ağacı Regresyonu ve Yapay Sinir Ağlarının Performans Değerleri

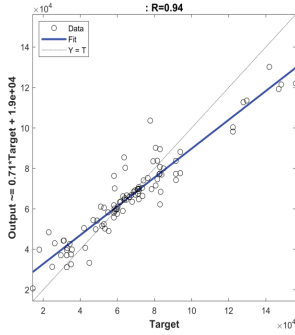
	Destek Vektör Regresyonu		Karar Ağacı Regresyonu		Yapay Sinir Ağları	
	Eğitim	Test	Eğitim	Test	Eğitim	Test
Korelasyon	0,85	0,73	0,92	0,89	0,90	0,88
MSE	314.946.112,67	386.782.705,52	122.979.906,09	162.730.878,19	147.204.539,59	242.065.222,13
MAPE	18,31	29,43	13,33	21,34	15,40	22,66
MAD	10.878,30	14.107,63	7.425,38	9.055,18	8.053,41	10.923,18
RMSE	17.746,72	19.666,79	11.089,63	12.756,60	12.132,79	15.558,45
RMPSE	32,40	53,82	20,97	44,72	28,91	43,87

Topluluk regresyonunda eğitim için korelasyon değeri 0,94; test için ise 0,90 olarak tespit edilmiştir. Beklendiği üzere topluluk regresyonu diğer analiz yönteminden daha iyi sonuç vermiştir. Topluluk yöntemine dair detaylı performans değerleri Tablo 4’te, korelasyon değerlerinin grafik olarak gösterimi Şekil 7.’dedir.

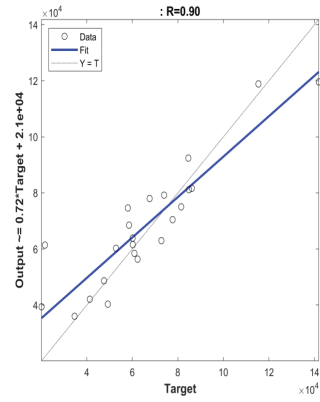
Tablo 4. Topluluk Regresyonu Performans Değerleri

	Eğitim Verileri	Test Verileri
Korelasyon	0,94	0,90
MSE	115.574.462,03	150.382.343,41
MAPE	13,51	20,84
MAD	7.519,97	8.681,94
RMSE	10.750,56	12.263,05
RMPSE	23,30	44,77

Eğitim



Test



Şekil 7. Topluluk Regresyonu Eğitim ve Test Korelasyon Grafiği

TARTIŞMA ve SONUÇ

İnternet tüketici satın alma davranışlarının her adımında etkisi olan bir iletişim aracıdır. Tüketicilerin bilgi alma ve değerlendirme maksadıyla interneti kullanmak istemeleri internet arama motorunun kullanımını arttırmıştır.

İnternet arama motorunun, satın almadaki yeri nedeniyle çeşitli ülke ve sektörlerde arama motoru verileri kullanılarak araştırmalar yapılmıştır. Amerika Birleşik Devletleri, Kanada ve Birleşik Krallık'ta otomotiv, seyahat, emlak ve perakende sektöründe araştırma yapılmış ve otomotiv sektörüne yönelik yapılan tahminde %18 geliştirici olduğu görülmüştür (Choi ve Varian, 2012). Perakende sektöründe yapılan tahminlerde internet arama verilerinin önemli katkılarının olduğu görülmüştür (Schmidt ve Vosen, 2009). Buna karşın internet arama motoru verileriyle yapılan bir başka araştırma da ise beklenen sonuçlara ulaşılamamıştır. Portekiz, İtalya ve Fransa'yı kapsayan internet arama verileri ile işsizlik ve otomotiv satışı tahminine yönelik yapılan çalışma da Google Trends'ten alınan verilerin hipotezlere az destek sağladığı görülmüştür (Barreira, Godinho ve Melo, 2013). İnternet arama verileriyle yapılan çalışmalarda araştırmanın yapıldığı ülke ile birlikte kullanılan yöntemde önemli olduğu görülmektedir.

Çalışmada internet arama verilerinin marka bazlı olarak toplam satışın %75'lik kısmını içeren veriler alınmıştır. Test verileri için korelasyon değerleri destek vektör regresyonunda 0.73 ile yüksek, karar ağaçları regresyonu 0,89 ile çok yüksek ve yapay sinir ağları ile yapılan analizlerde ise 0,88 ile çok yüksek ilişkili olarak bulunmuştur. Topluluk regresyonu ile yapılan analiz neticesinde ise test verileri için korelasyon 0,90 ile çok yüksek ilişki çıkmıştır. Yapılan ilk analizler topluluk regresyonuyla %1 ile %23 arasında geliştirilmiştir. Ortalama mutlak yüzde hata, destek vektör regresyonu %29,43, karar ağacı regresyonu %21,34 ve yapay sinir ağları %22,66 ve topluluk regresyonunda ise %20,84 olarak bulunmuştur. Ortalama mutlak yüzde hata değerleri topluluk regresyonuyla %2 ile %41 arasında gelişmiştir. Diğer hata değerlerinde de benzer gelişmeler görülmektedir.

Çalışma sonucunda gerçek ile tahmin edilen veriler arasında çok yüksek ilişki bulunmuştur ve %20,84 ile modellenmiştir. Topluluk regresyonu analizleri geliştirmekle birlikte en fazla fark destek vektör regresyonu ile arasında olduğu görülmüştür.

Otomotiv sektörüne yönelik internet arama verileriyle yapılan çalışmalarda ülke ve yöntemle ilgili olarak olumlu veya olumsuz sonuçlar elde edilebilmektedir. Türkiye'de yapılan bu çalışmada kullanılan yöntem başarılı olmuştur.

Çalışma, internet arama motoru verileri ile satış tahminlerinin geliştirilmesi mümkün olduğu ve satış pazarlama birimlerindeki yetkililerce dikkate alınması gereken bir veri kaynağı olduğu göstermiştir.

ÖNERİLER

Bu çalışmada, sadece internet arama verileri kullanılarak otomotiv sektöründe tahmin yapılmıştır. Gelecek çalışmalarda, veri kaynağı çeşitlendirilerek örneğin ekonomik verilerde kullanılarak tahminlemedeki hata değeri düşürülebilir. Aynı zamanda sektörün çeşitlendirilmesi de mümkündür. Çalışması muhtemel sektörler arasında bankacılık veya turizm olduğu değerlendirilmektedir. Örneğin kredi kullandırma oranları, erken rezervasyon internet aramalarının çok yapılabileceği alanlar olarak görülmektedir.

KAYNAKÇA

Barreira, N., Godinho, P. ve Melo, P. (2013). Nowcasting unemployment rate and new car sales in south-western Europe with Google Trends. *NETNOMICS: Economic Research and Electronic Networking*, 14(3), 129-165. doi:10.1007/s11066-013-9082-8

Blueopps. (2018). SEO DRIVES ORGANIC SEARCH ENGINE RANKINGS. blueopps.com. 16 Aralık 2019 tarihinde <https://www.blueopps.com/blog/2018/2/8/seo-drives-organic-search-engine-rankings> adresinden erişildi.

Boachie, P. (2018). User-Generated Content Brings Authenticity to Brands - Adweek. Adweek. <https://www.adweek.com/digital/user-generated-content-brings-authenticity-to-brands/> adresinden erişildi.

Cerit, I., Yildirim, A., Ucar, M. K., Demirkol, A., Cosansu: ve Demirkol, O. (2017). Estimation of antioxidant activity of foods using artificial neural networks. *Journal of Food and Nutrition Research*, 56(2), 138-148.

Choi, H. ve Varian, H. (2012). Predicting the Present with Google Trends. *Economic Record*, 88(SUPPL.1), 2-9. doi:10.1111/j.1475-4932.2012.00809.x

Cortes, C. ve Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20, 273-297. doi:10.1017/CBO9781107415324.004

Çuhadar, M. ve Kayacan, C. (2005). Yapay Sinir Ağları Kullanılarak Konaklama İşletmelerinde Doluluk Oranı Tahmini : Türkiye ' deki Konaklama İşletmeleri Üzerine Bir Deneme, 24-30.

Emel, G. G. ve Taşkın, Ç. (2005). Veri Madenciliğinde Karar Ağaçları ve Bir Satış Analizi Uygulaması. *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, 6(2), 221-239. doi:10.17494/ogusbd.06798

Emra, A. H. (2017). Ensemble Modeling for beginners. ahmethamzaemra.github.io/. 23 Aralık 2019 tarihinde <https://ahmethamzaemra.github.io/2017/12/14/Ensemble-Modeling.html> adresinden erişildi.

Getting, B. (2007). Basic Definitions: Web 1.0, Web 2.0, Web 3.0. <http://www.practicalecommerce.com/Basic-Definitions-Web-1-0-Web-2-0-Web-3-0> adresinden erişildi.

Gokceoglu, C., Nefeslioglu, H. A., Sezer, E., Bozkir, A. S. ve Duman, T. Y. (2010). Assessment of landslide susceptibility by decision trees in the metropolitan area of Istanbul, Turkey. *Mathematical Problems in Engineering*, 2010. doi:10.1155/2010/901095

Internetworldstats. (2019). Internet World Stats Usage and Population Statistics. 6 Aralık 2019 tarihinde <http://www.internetworldstats.com/stats.htm> adresinden erişildi.

Kavzoğlu, T., Şahin, E. K. ve Çölkesen, İ. (2012). Heyelan Duyarlılığının İncelenmesinde Regresyon Ağaçlarının Kullanımı: Trabzon Örneği, (April 2015), 21-33.

Mandhala, V. N., Sujatha, V. ve Devi, B. R. (2014). Scene classification using support vector machines. 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies içinde (ss. 1807-1810). IEEE. doi:10.1109/ICACCCT.2014.7019421

Naik, U. ve Shivalingaiah, D. (2008). Comparative Study of Web 1.0, Web 2.0 and Web 3.0. 6th International CALIBER, 499-507. doi:10.4018/978-1-61520-611-7.ch121

Odlyzko, A. (2001). Internet pricing and the history of communications. *Computer Networks*, 36(5-6), 493-517. doi:10.1016/S1389-1286(01)00188-8

Park, ve Huh: (2019). A Social Network-Based Inference Model for Validating Customer Profile Data, 36(4), 1217-1237.

Prater, M. (2019). 25 Google Search Statistics to Bookmark ASAP. [blog.hubspot.com](https://blog.hubspot.com/marketing/google-search-statistics). 16 Aralık 2019 tarihinde <https://blog.hubspot.com/marketing/google-search-statistics> adresinden erişildi.

Rokach, L. ve Maimon, O. (2015). *Data Mining With Decision Trees Theory And Applications 2nd Edition*. World Scientific Publishing. Singapore: World Scientific Publishing.

Roser, M., Hannah, R. ve Ortiz-Ospina, E. (2016). Internet. ourworldindata.org/. <https://ourworldindata.org/internet> adresinden erişildi.

Saberi, B. (2018). The Role Of The Automobile Industry In The Economy Of Developed Countries. *International Robotics & Automation Journal*, 4(3), 179-180. doi:10.15406/iratj.2018.04.00119

Safavian, R. S. ve Landgrebe, D. (1991). A Survey of Decision Tree Classifier Methodology. *IEEE Transactions on systems, man, and cybernetics*, 21(3), 660–674.

Schmidt, T. ve Vosen: (2009). Forecasting Private Consumption. *Economic Papers*, 155, 23.

Shevade: K., Keerthi: S., Bhattacharyya, C. ve Murthy, K. R. K. (2000). Improvements to the SMO algorithm for SVM regression. *IEEE Transactions on Neural Networks*, 11(5), 1188–1193. doi:10.1109/72.870050

Sullivan, D. (2016). Google now handles at least 2 trillion searches per year. *Search Engine Land*. 28 Temmuz 2019 tarihinde <https://searchengineland.com/google-now-handles-2-999-trillion-searches-per-year-250247> adresinden erişildi.

Tekula. (2014). Your Approach to Organic Search is Obsolete: How to Evolve in 2014. *distilled.net*. 16 Aralık 2019 tarihinde <https://www.distilled.net/blog/your-seo-approach-is-obsolete-evolve-in-2014/> adresinden erişildi.

Tkáč, M. ve Verner, R. (2016). Artificial neural networks in business: Two decades of research. *Applied Soft Computing Journal*, 38, 788–804. doi:10.1016/j.asoc.2015.09.040

Wikipedia. (2019). Web Search Engine Architecture. 16 Aralık 2019 tarihinde https://en.wikipedia.org/wiki/Web_search_engine adresinden erişildi.

Wong, B. K., Bodnovich, T. A. ve Selvi, Y. (1996). Neural network applications in business: A review and analysis of the literature (1988-95), 19, 301–320. doi:10.1016/S0167-9236(96)00070-X

WorldBank. (2019). Internet User. <https://data.worldbank.org>. <https://data.worldbank.org/indicator/IT.NET.USER.P2> adresinden erişildi.

Yavuz, A. (2019). Talep Tahminleri. 17 Aralık 2019 tarihinde <http://www.mku.edu.tr/files/339-fe1f52d7-578f-4b4f-b078-bbcbae358b91.pdf> adresinden erişildi.